FINANCIAL MATHEMATICS TEAM CHALLENGE

A collection of the four reports from the 2024
Financial Mathematics Team Challenge.

AIFMRM

AFRICAN INSTITUTE OF FINANCIAL MARKETS AND RISK MANAGEMENT

Preamble

One of the key aims of the FMTC is for South African postgraduate students in Financial and Insurance Mathematics to have the opportunity to focus on a topical, industry-relevant research project, while simultaneously developing links with international students and academics in the field. An allied objective is to bring a variety of international researchers to South Africa to give them a glimpse of the dynamic environment that is developing at UCT in the African Institute of Financial Markets and Risk Management. The primary goal, however, is for students to learn to work in diverse teams and to be exposed to a healthy dose of fair competition.

The Ninth Financial Mathematics Team Challenge was held from the 24th of June to the 4th of July 2024. The challenge brought together four teams of Masters and PhD students from Mexico, China and South Africa to pursue intensive research in Financial Mathematics. Each team worked on a distinct research problem over the twelve days. Professional and academic experts from the USA, South Africa, and the UK mentored the teams; fostering teamwork and providing guidance. As they have in the past, the students applied themselves with remarkable commitment and energy.

This year's research included topical projects on (a) Pricing Index Based Insurance for Maize Crops in South Africa, (b) Generative Modelling of Arbitrage-Free Markets via Neural SDEs, (c) Monte Carlo Arithmetic Asian Options Pricing with Variance Reduction Techniques, and (d) Climate Risk Analysis of the South African Interest Rate Swap Market. These were either proposed directly by our academic/industry partners or chosen from areas of current relevance to the finance and insurance industry. In order to prepare the teams, guidance and preliminary reading was given to them a month before the meeting in Cape Town. During the final two days of the challenge, the teams presented their conclusions and solutions in extended seminar talks. The team whose research findings were adjudged to be the best was awarded a floating trophy. Each team wrote a report containing a critical analysis of their research problem and the results that they obtained. This volume contains these four reports and will be available to future FMTC participants. It may also be of use and inspiration to Masters and PhD students in Financial and Insurance Mathematics.

FMTC X will take place in June/July 2025, and is already being organised!

David Taylor, *University of Cape Town*
Andrea Macrina, *University College London & University of Cape Town*

Contents

---

[1]Winning team of the ninth Financial Mathematics Team Challenge

Pricing Index Based Insurance for Maize Crops in South Africa

TEAM 1

SIPHOKAZI HLALUKANA, University of Cape Town
CHRISTOPHER DAVIS , University of Cape Town
FORTUNE KHOZA, University of Cape Town
TJAART VAN DER WALT, Stellenbosch University

Supervisors:
CORINA CONSTANTINESCU, Liverpool University
KHOUZEIMA MOUTANABBIR, University of Cape Town

African Institute of Financial Markets and Risk Management

Contents

1 Introduction

In many developing countries, especially those in Africa and some parts of Asia, agriculture is a critical sector, contributing significantly to GDP, employment, and exports (Prokopchuk et al., 2020). However, the sector faces numerous challenges, such as low-income levels, low capital-labour ratios, and high susceptibility to natural risks (Jain, 2004). Consequently, agricultural production in these developing countries is largely characterized by low productivity, poor infrastructure, and high risk due to the heavy dependency on nature. Farmers face risks such as extreme weather events, uncertain crop yields, pests, and diseases (Mookerjee et al., 2014). These risks are further exacerbated by limited financial resources and the lack of modern agricultural practices.

A large number of farmers in these developing countries are often not insured due to scepticism surrounding insurance. This scepticism arises from the long payout periods and high premiums associated with insurance (Barnett et al., 2008). As a result, farmers usually engage in informal risk management strategies instead of taking on insurance. To lower their exposure, farmers are unlikely to invest in technology that could increase yields and profits for fear of the technology being damaged in the event of a catastrophe (Barnett et al., 2008). Some farmers might choose a more drought-resistant cultivar with lower profits to reduce the effects of drought on income. This is because these cultivars are often significantly more expensive than the more weather-prone cultivars (Govindaraj et al., 2018). In the event of a disaster, farmers are often forced to sell assets at massive discounts to cope with the loss of income due to the catastrophe.

Most of the risks that these farmers face can be mitigated using agricultural insurance, thus providing financial security to the farmers and encouraging the adoption of improved agricultural technologies. Moreover, agricultural insurance can help reduce the financial burden on the government by streamlining relief efforts during times of natural disasters. However, there are some issues that arise when implementing agricultural insurance in developing countries. These issues include the lack of reliable data on crop yields and losses, diverse agricultural practices, and poverty among farmers (Jain, 2004). Additionally, the high administration costs and lack of reinsurance support further complicate the implementation (Mookerjee et al., 2014).

The most common type of agricultural insurance is crop insurance. When designing a crop insurance product, one needs to carefully consider a few elements. These elements include deciding on the types of risk that the insurance product will cover; the product could cover all risks, in which case it would be known as yield insurance, or cover specific risks (Jain, 2004). Farmers would prefer yield

4

insurance, but it is often more costly and can be difficult due to challenges in identifying losses from uninsured events. The insurance product can also be operated by the public sector, private sector, or through a combination. Moreover, one also needs to decide on the approach to assessing losses; there are two approaches: the Individual or Area approach (Jain, 2004). In the case of the individual approach, the loss assessment is made separately for each farmer, while in the case of the area approach, indemnity is determined for a group of farmers on the basis of the average loss experienced by a specified homogeneous area. Additionally, the insurance product can be voluntary or compulsory. Compulsory schemes have the advantage of reducing adverse selection and administration costs but may cause dissatisfaction among low-risk farmers (Jain, 2004).

Crop insurance can be indemnified in the traditional way or using an index, in which case it would be known as crop index insurance. Traditional indemnity-based insurance in agriculture is usually very difficult to implement due to the aforementioned reasons and also due to the risk of adverse selection and moral hazard (Mookerjee et al., 2014). There are two categories of index insurance: Weather Index Insurance and Area-Yield Index Insurance. Under Weather Index Insurance, the insurance pays out based on specific weather events, such as rainfall deficits, which are measured using objective indices. The main advantage is the ease of measurement, but it may not always correlate perfectly with actual crop losses. On the other hand, under Area-Yield Index Insurance, the insurance estimates crop loss by measuring the actual crop output on a sample of farms within an area. Although more closely related to actual losses, it is costlier to implement and administer.

The main benefits of index insurance relative to traditional insurance are that under index insurance, moral hazard is reduced since the payouts are based on a predetermined index that is not influenced by the insured. Additionally, there is transparency in claim handling and payouts, and administrative costs are reduced, which in turn lowers the premium (Prokopchuk et al., 2020). The main drawback of index insurance is basis risk. Basis risk arises when the index measurements do not match an individual's actual losses (Hess and Syroka, 2005). The two main sources of basis risk in index insurance are product design basis risk and geographical basis risk (Mookerjee et al., 2014). Product design basis risk stems from poorly designed products, and geographical basis risk stems from geographical elements. The main goal of index insurance is to minimize basis risk. Minimizing basis risk is essential to attract more customers. According to Mookerjee et al. (2014), basis risk can be minimized by using accurate weather data collection, selecting appropriate indices, and ensuring the insurance coverage closely matches the actual risks faced by farmers. There is, however, often a trade-off between ensuring that the target audience can easily understand how the index insurance product works and en-

suring that the correlation between the index and actual losses is high enough.

The primary objective of this paper is to design a weather index insurance product that is both affordable and has a low level of basis risk. Rainfall and temperature are used as indices since maize is mainly affected by these two factors. The paper carefully constructs products based on each of these indices and, thereafter, a product based on both indices. Furthermore, the paper will employ different methods of pricing the products, i.e., calculating the premium and comparing the prices of the product based solely on the rainfall index, the product based solely on the temperature index, and the product based on both indices. This will help determine which product is more appropriate. To the best of our knowledge, this is the first paper that designs and prices multi-index-based insurance.

It is worth noting that most index insurance products are based on a single index. Although prior literature has priced using Black Scholes, it has not compared its prices with the actuarial approach, and the accuracy and appropriateness of using Black Scholes have not been verified. We will thus compare our actuarial approach to the Black Scholes approach.

Additionally, instead of making use of one threshold, the paper will also explore a product that has multiple thresholds.

2 Literature Review

Traditional insurance has low participation amongst Ghanaian farmers due to four primary factors. First, the majority of farmers in low-income regions work small lots, which only have the capacity for small profit margins and, due to their small size, can only induce low liability limits. Traditional insurance is thus economically unfeasible (Barnett et al., 2008). Second, farming regions in low-income areas comprise many smaller, individually owned farm-lots. This increases the volume of unique damage assessments that need to be carried out, driving up the cost of insurance. Particularly so in an underdeveloped insurance market that would not have the infrastructure necessary to ease the damage assessment process. Third, objective damage assessment can be costly, in terms of the administrative process and the hiring of personnel to carry out the assessment itself. Lastly, traditional insurance rarely, if ever, covers correlated risks (Okine, 2014).

Index-based insurance addresses these factors as it has an automatic and objective payout structure that avoids assessment and administrative fees, potentially lowering costs to economic feasibility. Furthermore, indices are chosen based on high correlation with what has been insured so that the farmer may choose to use the insurance payout to address correlated weather damages if they are so inclined. Examples of successful implementations of index-based insurance include the case of Malawi, where the weather reports of local meteorological stations are used to determine the occurrence of severe drought in an area, and the case of Northern Kenya, where independent satellites are used to measure the amount of vegetation that is available for livestock to consume (Okine, 2014).

2.1 The Case of Ghana

Okine (2014) prices index-based insurance for maize crops. The growth of maize depends on rainfall, thus, Okine (2014) first established the relationship between rainfall and annual maize production in Ghana. Having found the highest correlation with maize yields to be in March, they used the cumulative rainfall over the month of March. That is, if meteorological stations report that the insured region experienced cumulative rainfall in March that fell below an amount specified in the insurance contract, the farmer would receive a lump sum amount. Due to the lump sum payout being based on whether or not some numerical value falls below a certain value at a specified date, they interpreted the index-based insurance as a cash-or-nothing put option. Okine (2014) applied standard Black-Scholes-Merton theory to simplify pricing, citing the reason as the rainfall data following a log-normal distribution.

2.2 The Case of India

One developing country that has a somewhat rich history of parametric crop insurance is India. India's experience with crop insurance spans several decades, evolving from experimental schemes to comprehensive national programs. From 1972 to 1979, India went through an experimental phase; during this phase, schemes focused on specific crops in certain districts but were found to be economically unviable (Jain, 2004). In 1979, the Pilot Crop Insurance Scheme (PCIS) was implemented in 13 states and covered 600,000 farmers. It used the area approach and covered various crops for loanee farmers. Thereafter, in 1985, the Comprehensive Crop Insurance Scheme (CCIS) was launched. This scheme was linked to short-term crop credit and used a homogenous area approach. It was optional for state governments and was criticized for financial non-viability, limited crop coverage, and deficiencies in yield assessment (Jain (2004)).

As a result, the National Agricultural Insurance Scheme (NAIS) was introduced between 1999 and 2000 to address these criticisms. The NAIS expanded coverage to include non-loanee farmers and additional crops. It provided for greater risk coverage and aimed for financial viability through a rationalised premium structure. Despite the improvements, the NAIS faced issues such as high claims ratios and the need for better yield assessment methods (Jain, 2004). Another scheme that was implemented is the Farm Income Insurance Scheme (FIIS). This scheme aimed to protect farmers' incomes by integrating production and market risks, providing income protection through a guaranteed minimum income based on yield and price movements.

2.3 The Case of Malawi

In Malawi, the World Bank and MicroEnsure launched a weather index insurance product to protect farmers against rainfall deficits. The farmers were encouraged to take bank loans to access funding for high-quality seeds and fertilizer. For each farmer, the insurance was then linked to the bank loan. Initially, the product design divided the crop growing season into three periods. If the rainfall was below a certain trigger level in any of these periods, a payout would be made depending on the rainfall level.

To improve the product design, the insurer held discussions with the farmers to better understand their practices (Mookerjee et al., 2014). It was found that dividing the crop growing season into three phases resulted in some issues of basis risk (Mookerjee et al., 2014). The solution was to divide the crop growing season into nine phases of 10 days each, which was found to reduce the basis risk (Mookerjee et al., 2014). Overall, the project highlighted the importance of linking insurance to agricultural loans and refining products based on local experiences.

2.4 The Case of the Philippines

In 2009, MicroEnsure launched a weather index insurance product that covered 446 smallholder rice farmers in the Philippines against high wind speeds. The project involved partnering with rural banks and MFIs that acted as delivery channels for the insurance product. Just as in the case of Malawi, the insurance product was linked to a loan from either the rural bank or MFI in order to obtain funding for the farm inputs. To assess the basis risk of the product, the product design was tested against crop loss data across the whole country and it was found that the payouts were well correlated with the crop losses (Mookerjee et al., 2014). However, before the product was launched, the level of basis risk was not known, so it was combined with a government multi-peril product that was subsidised by the insurance broker. It was later found that typhoons with low wind speed and high rainfall were not picked up well by the product, and it was suggested that the insurer should use some form of satellite rainfall estimation technique. However, satellite rainfall estimators are known to not be particularly accurate for high rainfall events (Mookerjee et al., 2014).

3 Problems with Traditional insurance

In traditional crop insurance, the indemnities are tied to the actual yield losses. This structure gives rise to two main problems: adverse selection and moral hazard (Horton (2021)). First, adverse selection refers to a situation in which there is asymmetrical information between the buyer and the seller of the insurance, leading to the insurer potentially facing higher-than-expected claims. It occurs when parties who are at a higher risk are more likely to buy the insurance, while those at lower risk are less likely to do so. This can result in the insurer's pool of policies being skewed towards higher risk, which in turn leads to higher expected payouts (Horton (2021)). Second, moral hazard occurs when the existence of insurance coverage encourages the insured party to increase their risk exposure based on the expectation that the insurer will pay for any resulting losses. To mitigate these issues, insurers often incur higher administrative costs (Hess and Syroka (2005)).

Additionally, several challenges make traditional crop insurance problematic, especially for microinsurance customers such as farmers in developing countries. These challenges include a relative lack of data on loss or damage, the expenses involved in underwriting individual risks, and the time taken to verify and pay claims. Consequently, finding a reinsurer for such products becomes difficult (Mookerjee et al. (2014)). Traditional multiple-peril crop insurance also faces large correlated risks, necessitating the additional cost of reinsurance Hess and Syroka (2005)). For farmers in developing countries where crop insurance is not well established, particularly in most African countries, these additional costs can be significant. Thus, traditional multiple-peril crop insurance is not a viable solution for managing agricultural risk (Hess and Syroka (2005)).

4 Parametric Insurance

4.1 How Parametric Insurance Works

Parametric insurance, also known as index-based insurance, provides coverage where payouts to the insured party are triggered and calculated by a predetermined index. This index is used as a threshold, and the payout amount corresponds to the value of the index Mookerjee et al. (2014). Therefore, under parametric insurance, there is no need for individual loss assessments. The key components of parametric insurance are:

4.1.1 Index Definition:

The first step is to define an index that will be used to decide whether a payout is made and, if so, what the payout amount should be. The index needs to be reliable

and objective and must be closely correlated with the risk being insured. Common indices include weather data, such as rainfall, temperature, or other measurable factors.

4.1.2 Trigger Events:

Specific thresholds need to be established for the index. When the index reaches or exceeds these thresholds, it triggers a payout. For the trigger to be viable, it needs to be fortuitous, independently and reliably monitored, and capable of being modelled (Tadesse et al. (2015)). Common triggers in crop insurance within agricultural insurance include deficit or excess rainfall, temperature, etc.

4.1.3 Payout Mechanism:

The payout structure is predefined based on the index level, i.e., if the parameter or index threshold is reached or exceeded, the insured party will receive the pre-agreed payout regardless of the actual losses sustained.

4.2 Advantages of Parametric Insurance

Using parametric insurance instead of traditional indemnity insurance has many benefits. One of these benefits is that in parametric insurance, adverse selection and moral hazard are significantly reduced altogether (Mookerjee et al. (2014)). This is mainly because the payout is based on a predetermined index which is not influenced by the insured party, so it should not matter to the insurer who buys the insurance or what actions they take on their farm after buying the insurance. Additionally, since the payouts are based on an index whose value can be easily determined, it implies that the payouts are easily calculated as well, which means the administrative costs are significantly reduced (Mookerjee et al. (2014)). This, in turn, leads to the insurer charging lower premiums. Furthermore, since the claims are connected to an objective and independent source of information, the chances of fraudulent claims are significantly reduced. Altogether, these benefits significantly increase the chances of obtaining reinsurance (Tadesse et al. (2015)).

4.3 Disadvantages of Parametric Insurance

One of the main drawbacks of parametric insurance is basis risk. Basis risk arises when the index measurements do not match the insured party's actual losses due to an imperfect correlation between the index and the insured party's output (Hess and Syroka (2005)). Depending on the correlation between the actual losses incurred and the index, basis risk may be significant, which may then result in a

large disparity between losses and payouts (Mookerjee et al. (2014)). For parametric insurance products that are based on a weather index, which is often the case, setting up and maintaining weather stations and the required infrastructure for measuring the indices may be challenging and not financially viable (Mookerjee et al. (2014)). Additionally, it is often difficult to construct a suitable index and product design that will have low levels of basis risk (Tadesse et al. (2015)). Moreover, it is often the case that parametric insurance products with low levels of basis risk will be expensive and more complex and, hence, not easily understood by the target market.

5 Catastrophe Cover

5.1 Catastrophe Insurance

Catastrophe insurance is insurance taken by businesses and the general public to protect against low-probability natural disasters. In the event of a catastrophe, the insurer would have to pay its clients for the losses they have suffered because of the disaster. The size of the payout would be determined by the actual losses suffered and the payout is meant to indemnify the client. This means that the insurer would have to assess the individual losses of each client and make the payment based on the realised losses. In addition to the costs of making the payouts, the process of assessing losses is usually costly for the insurer (Barnett et al., 2008). Moreover, the process is lengthy and clients often have to wait long periods to receive their payout. To transfer some of this risk, insurers may take on reinsurance, which may also take long to indemnify the insurer.

5.2 Catastrophe Bonds

As an alternative to reinsurance, insurance companies can issue catastrophe bonds. Catastrophe bonds (CAT) bonds are instruments generally issued by insurers as a form of reinsurance for the insurer. When a predefined catastrophe occurs, such as an earthquake, tsunami, hurricane, etc, the bond pays the issuer to help with dealing with losses arising from catastrophe insurance claims. When these bonds are issued, investors buy them as a bet against natural disasters until the bond matures. In return, the investor receives coupon payments typically higher than corporate bonds. If the specified trigger catastrophe does not happen until the bond matures, then the investors receive the principal amount plus the final coupon. However, if the specified catastrophe occurs, the insurer receives a payment to indemnify it from the losses arising from the disaster. This means that the investors get only a fraction or none of their principal payments back. In the case of CAT bonds as opposed to reinsurance, the insurer receives the payment quicker.

6 Maize Production in South Africa

Maize is the most extensively grown field crop in South Africa and is considered one of the main staples in the country. Both white (about 60%) and yellow (about 40%) maize are produced in South Africa, with white maize primarily for human consumption and yellow maize used for animal feed. The maize industry is an important part of the country's economy as an employer as well as an earner of foreign currency due to its multiplier effect. This is because maize can also be used as a raw material for a range of manufactured products such as textiles, medicine and paper. Of all the maize produced in South Africa, more than two thirds is consumed by the local market, this consumption is split between humans (37.2%) and the animal feed industry (39.2%), with the rest being used for seed and industrial uses (23.6%)(Department of Agriculture and Fisheries, 2021).

Maize production is widely spread across South Africa and is mostly produced under dryland conditions, with only 10% being produced under irrigated conditions, indicating that weather conditions play a vital role in production. The Free State is the province with the largest share of maize production with approximately 45% of total production taking place in the province. The main districts where maize production takes place in the Free State are Lejweleputswa, Motheo and Thabo Mofutsanyana. We selected eight weather stations in different regions across these districts to ensure that we captured the average weather conditions for the entire area. These were Bloemfontein, Bothaville, Bultfontein, Ficksburg, Harrismith, Heilbron, Senekal and Viljoenskroon (Department of Agriculture and Fisheries, 2021).

6.1 Climate Requirements

6.1.1 Temperature

The ideal temperature for maize cultivation is between 19°C and 25°C. Extreme heat, temperatures exceeding 32°C are detrimental to crop yield. Lower temperatures are not as critical to the yield, but it will slow the growth of the crop(Department of Agriculture, Land Reform and Rural Development, 2022).

6.1.2 Rainfall

Maize needs between 450 and 600 mm of water per season to reach its full potential. Since the majority of production occurs under dry land conditions, this water must come mainly from rainfall. (du Plessis, 2003)

6.2 Growth Cycle

The optimal planting time for maize depends on the area in which it will be grown. Broadly this can be divided into three different regions. For the eastern region with

cooler climates, planting takes place from the start of October to the first week of November. For the central region, planting should start from the last two weeks in October to mid-November. Finally, for the western region with a drier climate, planting should take place from the last two weeks in November to mid-December. After planting maize needs around 5 months to grow, dependent on external factors, before it is ready to harvest. For the purposes of this paper, we will assume that planting takes place at the end of November(Department of Agriculture, Land Reform and Rural Development, 2022).

7 Product Structuring

We consider five different product types: single-index rainfall insurance, single-index temperature insurance, single-index rainfall insurance with multiple thresholds, single-index rainfall insurance with multiple thresholds, and a double-index temperature-rain insurance product. For the single-index temperature insurance and double-index product, we make the assumption that the insurance is issued on the first day of November and extends to the last day of February (time T_H), ignoring the existence of leap years. For the single-index rainfall insurance product we assume that the insurance is issued on the first day of December and extends to the last day of December (time T_R).

The temperature index we use in our pricing is the daily maximum temperature, averaged over the time period from the first of November to the last day of February of the following year. The rainfall index we use in our pricing is the cumulative rainfall over the month of December. These indices were chosen because, as discussed in Section 6, maize is sensitive to abnormally high temperatures over its entire growth period, and requires rainfall shortly after planting, and so we consider these indices to be a reasonable predictor for whether or not maize grew within sufficient climate conditions. Further, we average the indices over the eight regions discussed in Section 6.

The insurance products follow a cash-or-nothing payoff structure. Suppose $R(t)$ represents the observed accumulated rainfall over the December period, then for our single-index rainfall insurance product, we offer at the end of the insurance period, T_R, a payoff

$$V(R(T_R), T_R) = \begin{cases} P & \text{if } R(T_R) < K_R \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

That is, if the total rainfall received over December, $R(T_R)$, is insufficient relative to some level, K_R, agreed to beforehand, the client receives a lump-sum amount P.

The single-index temperature insurance product follows a similar structure. Suppose $H(t)$ represents the observed average temperature over the November to February period, then here we offer at the end of the insurance period, T_H, a payoff

$$V(H(T_H), T_H) = \begin{cases} P & \text{if } H(T_H) > K_H \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

That is, if the maximum daily observed temperature over the November-February period has an average value, $H(T_H)$, that is excessive relative to some level, K_H, agreed to beforehand, the client receives a lump-sum amount P.

For single-index rainfall insurance with multiple thresholds, suppose we have a descending sequence of rainfall thresholds $K_{R,1} > ... > K_{R,n}$ and an increasing sequence of payouts $P_1 < ... < P_n$, we offer a payout at the end of the December period of

$$V(R(T_R), T_R) = \begin{cases} P_i & \text{if } K_{R,i} > R(T_R) > K_{R,i+1} \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

That is, if the cumulative rainfall over December, $R(T_R)$, is realised between (and only between) the threshold levels i and $i + 1$, the client receives payout P_i.

For single-index temperature insurance with multiple thresholds, suppose we have an increasing sequence of temperature thresholds $K_{H,1} < ... < K_{H,n}$ and an increasing sequence of payouts $P_1 < ... < P_n$, we offer a payout at the end of the December period of

$$V(H(T_H), T_H) = \begin{cases} P_i & \text{if } K_{H,i} < H(T_H) < K_{H,i+1} \\ 0 & \text{otherwise,} \end{cases} \qquad (4)$$

That is, if the maximum daily observed temperature over the November-February period has an average value, $H(T_H)$, that is realised between (and only between) the threshold levels i and $i + 1$, the client receives payout P_i.

The double-index follows a multivariate cash-or-nothing payoff structure, where we offer at the end of the insurance period, T_H, a payoff

$$V(R(T_H), H(T_H), T_H) = \begin{cases} P & \text{if } R(T_H) < K_R \text{ and } H(T_H) > K_H \\ 0 & \text{otherwise,} \end{cases} \qquad (5)$$

That is, if the rainfall received over December, $R(T_R) = R(T_H)$, and the maximum daily observed temperature over the November-February period has an average value, $H(T_H)$, that has crossed their predefined trigger levels, the client receives a lump-sum amount P.

7.1 On the risk-free rate

Due to the binary payoff structure one may be attempt to model the insurance premium using the language of financial derivatives, however this may have a critical consequence. As the fair price of any instrument is the discounted expectation of the future payoff (Baškot and Stanić, 2020), one would require a risk-free interest rate to base their pricing on. However, the underlying "asset", the weather, is not tradeable. One would require additional assumptions. One may attempt to take the approach used by Jewson and Zervos (Jewson and Zervos, 2003) and assume the existence of a highly liquid, standardly traded, linear weather forward derivative. In such a market, one may proceed with risk-neutral valuation and, if the risk-free rate is say r, use a discount factor of e^{-rT}, and offer the insurance at a price at the present time of

$$\text{Price} = e^{-rT}\mathbb{E}[V(T)] \tag{6}$$

However, such a market does not exist. Furthermore, (Cao and Wei, 2004) propose a utility-based valuation approach that allows the study of the market price of risk of weather derivatives, and they show that risk-neutral valuation provides a severe misestimation of weather option prices. However, the assumption that discounting by the risk-free rate is acceptable remains common in literature, such as in (Rui Zhou and Pai, 2019) where they priced temperature derivatives. So, here forth, we assume that e^{-rT} as our discount factor is a sufficient approximation.

8 Pricing Parametric Insurance

As shown in Section 7, the payoff of index insurance is triggered when the threshold K_T is exceeded; otherwise the insurer pays nothing. The price of the insurance can be calculated as the discounted expectation of the payoff. In the case of insurance indexed by rainfall, the premium can be calculated as

$$\text{Premium} = \mathbb{E}[Pe^{-rT_R}\mathbb{I}_{R_T < K_R}] \tag{7}$$

and would be given by

$$\text{Premium} = \mathbb{E}[Pe^{-rT_H}\mathbb{I}_{H_T > K_H}] \tag{8}$$

for insurance indexed by temperature.
Thus the premium is

$$\text{Premium} = Pe^{-rT_H}\mathbb{P}[\text{Index} > K] \tag{9}$$

8.1 Comparison to Black Scholes

This payoff is similar to that of a cash-or-nothing option, where, in the case of a put option the payoff is given by

$$\text{Payoff} = \begin{cases} P & \text{if } S_T \leq K \\ 0 & \text{otherwise,} \end{cases}$$

where P is the payoff at maturity, S_T is the stock price at maturity, and K is the strike price. As seen in equations 10 and 8, for the insurance indexed by rain, the insurance pays out once the rain is below the trigger level, making it analogous to a put option. However, in the case of temperature, the insurance pays out when the temperature in the region exceeds the threshold, making it analogous to a call option. Since Black Scholes assumes that the stock price is lognormally distributed, various studies (see Okine (2014);) have used Black Scholes to price index-based insurance when the index follows a lognormal distribution.
The option price is given by

$$\text{Premium} = P e^{-r T_R} N(-d_2) \tag{10}$$

where $N(-d_2)$ is the cumulative normal distribution, r is the risk-free interest rate with d_2 given by

$$\frac{\ln(\frac{H_0}{K_H}) + \mu t}{\sigma \sqrt{t}}.$$

with $\mu = \frac{1}{n-1} \ln(\frac{H_n}{H_1})$, $\sigma = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (u_j - \bar{u})^2}$, with $u_j = \ln\left(\frac{H_j}{H_{j-1}}\right)$. We will thus compare our actuarial approach to the Black Scholes approach by pricing using both approaches.

9 Pricing

9.1 Data

Daily maximum temperature (in °C) and daily rainfall (in mm) data was obtained from the Open-Meteo Weather API (Zippenfenig (2023)) for the period from 1940 to 2024. Since we considered eight different regions in the Free State the data was averaged across these regions to obtain the average daily maximum temperature and the average daily rainfall. For the temperature index, this data was augmented into an index that contains the average maximum temperatures of each year from November to February the following year, and for the rainfall index, the data was augmented to obtain the total rainfall over the month of December for each year.

9.1.1 Temperature

Figure 1 shows the histogram for the values of the temperature index. Different distributions were fit to this data, and it was found that the log-normal distribution provided the best fit.
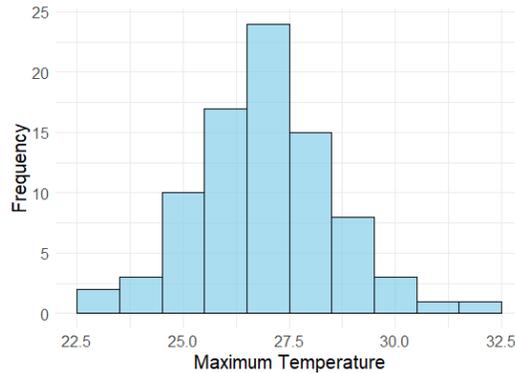
Figure 1: Histogram of Temperature Index Values.

Table 1: Goodness of Fit Test Results

Test	Statistic	P-value
Kolmogorov-Smirnov	0.0408	0.998
Anderson-Darling	0.1553	0.998

Table 1 displays the test statistics and p-values that where obtained for the Kolomogorov-Smirnov and Anderson-Darling tests, since the p-values for both are much larger

than 0.05 it indicates that the log-normal distribution is a good fit for the temperature index data.

The estimated parameters of the fitted log-normal distribution are given in table 2.

Table 2: Estimated Parameters for Log-Normal

Parameters	Estimate	Std. Error
Mean	3.2914	0.0065
Std. Dev.	0.0597	0.0046

9.1.2 Rainfall

Figure 1 shows the histogram for the values of the rainfall index. Different distributions were fit to this data, and it was found that the gamma distribution provided the best fit.
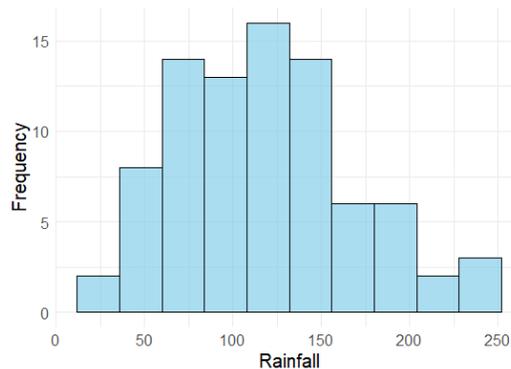
Figure 2: Histogram of Rainfall Index Values.

Table 3: Goodness of Fit Test Results

Test	Statistic	P-value
Kolmogorov-Smirnov	0.0655	0.841
Anderson-Darling	0.2893	0.946

Table 3 displays the test statistics and p-values that where obtained for the Kolomogorov-Smirnov and Anderson-Darling tests, once again the p-values for both are much

larger than 0.05 which indicates that the gamma distribution is a good fit for the rainfall index data.

The estimated parameters of the fitted gamma distribution are given in table 4.

<div align="center">

Table 4: Estimated Parameters for Gamma

Parameters	Estimate	Std. Error
Shape	4.8633	0.7236
Rate	0.0412	0.0065

</div>

9.2 Pricing single-index insurance

Using the parameters of the distribution of the historical index data, we simulate $n = 1000000$ samples of index values from the respective distributions. Based on the simulated index, we calculate the probability that the index exceeds the threshold.

9.3 Pricing Multi-Index Insurance

The multi-index insurance product considers both the temperature and rainfall indices to determine when a payout is triggered. Three different approaches were followed for pricing this product: the copula and multivariate approaches, which take the dependence between the two indices into account, and pricing on the assumption that the indices are not dependent.

9.3.1 Copula Approach

Copulas can be used to model the dependence between random variables. The joint probability returned by a copula is a function of the marginal probabilities. The fact that the marginal behaviour of random variables are modelled separately from their dependence is what makes the copula approach so attractive. Kole et al. (2007)

Let $X_1, \cdots, X_n$ be random variables with marginal continuous distribution functions $F_1, \cdots, F_n$, respectively, and joint distribution function H. Then $(X_1, \cdots, X_n)^T$ has a unique copula, denoted by $\mathbb{C}(\cdot)$. The standard copula representation for the distribution of $(X_1, \cdots, X_n)^T$ then becomes:

$$H(x_1, \cdots, x_n) = P\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = \mathbb{C}\left(F_1(x_1), \cdots, F_n(x_n)\right) \qquad (11)$$

The transformations used in the above representation, $X_i \mapsto F_i(X_i)$, are the probability-integral transformations (to uniformity) and form a standard tool in simulation

methodology.Embrechts et al. (2001)

The selection of an appropriate bivariate copula family to use for the temperature and rainfall index was done by fitting all possible bivariate copula families and selecting the best one based on the AIC and BIC selection criteria. Under both the AIC and BIC the rotated Gumbel copula 270 degrees was selected as the most appropriate. The rotated Gumbel copula 270 degrees is defined as:

$$\mathbb{C}(u,v) = exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right), \tag{12}$$

where θ is the Gumbel parameter given by $\hat{\theta} = \frac{1}{1-\tau}$ with τ the correlation between the variables (Rajini and Jayalakshmi, 2021).

Figure 3 shows the dependence structure between the simulated values for temperature and rainfall using the rotated Gumbel copula 270 degrees. This indicates that the fitted copula appropriately models the non-linear dependence between the two indices.
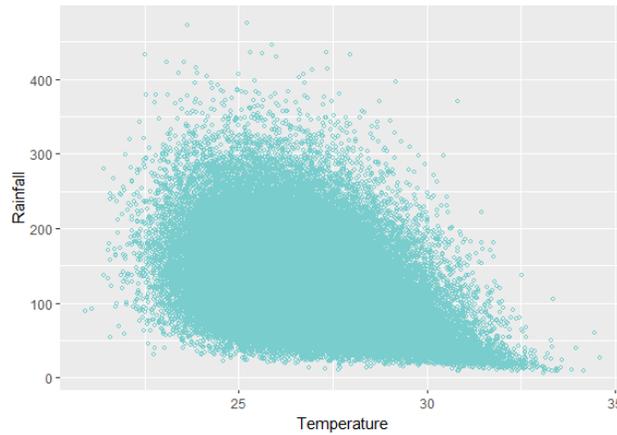
Figure 3: Dependence structure between simulated Temperature and Rainfall

9.3.2 Multivariate Normal Approach

As a point of comparison, we also model the joint index as a multivariate normal distribution using a method described by Li and Hammond (Li and Hammond, 1975). This modelling method is less than desirable than the copula approach as it does not capture the dependence between variables as well as the copula approach does, however it is perhaps more intuitive as it only involves straightforward transformations of probability distributions.

Suppose we wish to generate two random variables, X and Y, where X follows distribution F and Y follows distribution G, and further suppose that we wish for X and Y to have (Pearson) correlation ρ. Then, we may proceed by generating a multivariate normal variable $(Z_X, Z_Y)^t$ (here the superscript t denotes transposition) with mean $(0,0)^t$ and correlation matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. We then make use of the inverse transform method by generating the correlated uniform variates $N(Z_X)$ and $N(Z_Y)$, where N is the normal cumulative distribution function, and then generating the correlated target variates $F^{-1}(N(Z_X))$ and $G^{-1}(N(Z_Y))$.

9.3.3 Assumed Independence Approach

If $X_1, \cdots, X_n$ are independent random variables with marginal continuous distribution functions $F_1, \cdots, F_n$, respectively. Then the joint distribution function H of $(X_1, \cdots, X_n)^T$ is given by:

$$H(x_1, \cdots, x_n) = P\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = F_1(x_1) \cdot F_2(x_2) \cdots F_n(x_n). \quad (13)$$

Thus, we can simply multiply the marginal probabilities to obtain the joint probabilities.

10 Single-Index Insurance

10.1 Rain-based Insurance

Table 5 shows the premiums based on different trigger levels for rain insurance. Since maize requires high levels of rainfall, the insurance pays out when rainfall is below the trigger level.

Table 5: Premiums for Rain-Based Insurance

	$q_{0.3}$	$q_{0.2}$	$q_{0.1}$	$q_{0.05}$	$q_{0.025}$	$q_{0.01}$
Rainfall (mm)	80.01	70.43	54.20	39.94	28.47	17.49
Premium	29.87	19.91	9.96	4.98	2.49	0.98

As the threshold increases, the premium increases as well. This is because a higher rainfall threshold triggers a payout more frequently than a lower one, making insurance more expensive.

10.2 Temperature-based Insurance

Table 6 shows the premiums based on different trigger levels for rain insurance. Since maize is sensitive to extremely high temperatures, the insurance pays out when the temperature is above the threshold.

Table 6: Premiums for Temperature-Based Insurance

	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.975}$	$q_{0.99}$
Temperature (°C)	27.75	28.29	29.05	29.69	30.25	30.93
Premium	29.51	19.67	9.83	4.91	2.46	0.98

As the threshold decreases, the premium increases. This is because a lower temperature threshold triggers a payout more frequently than a higher one, making insurance more expensive.

10.3 A comparison with Black Scholes Prices

Since the temperature is lognormally distributed, we compute the price using the Black-Scholes approach as implemented in existing literature. Table 7 shows the prices from both approaches.

Table 7: Premiums for Temperature-Based Insurance

	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.975}$	$q_{0.99}$
Temperature (°C)	27.75	28.29	29.05	29.69	30.25	30.93
Black-Scholes	70.11	60.18	45.24	33.23	23.99	15.27
Actuarial	29.51	19.67	9.83	4.91	2.46	0.98

The Black-Scholes prices are more than double the actuarial approach prices, going as high as almost 16 times. This indicates that the Black-Scholes formula cannot be used directly to price index-based insurance.

10.4 Rain-based Insurance with multiple thresholds

As an alternative, farmers could opt for rain-based index insurance with multiple thresholds. The payoff for such a product is shown in Table 8. For every one percent increase in the strike quantile, the payoff is assumed to increase by an additional one percent from the original 100 payoff. Such a product will have a lower level of basis risk due to its payoff structure. In other words, if the farmer suffers a higher loss, they will receive a higher payout, as opposed to the case where there is only one threshold with one fixed payment. So, this is one attempt at minimising basis risk.

Table 8: Payoff structure for Rainfall-Index Insurance with multiple thresholds

Rainfall (mm)	**Payoff**
$x > 80.1$	0
$70.43 < x \leq 80.1$	100
$54.20 < x \leq 70.43$	110
$39.94 < x \leq 54.20$	120
$28.47 < x \leq 39.94$	125
$17.49 < x \leq 28.47$	127.5
$x \leq 17.49$	130

The price of such a product is then calculated as a combination of the prices of single rain-index insurance products, i.e. a combination of cash-or-nothing puts. The price for this product is calculated to be 33.19. As can be seen, the premium of the product with multiple thresholds is higher than all of the premiums for the products with one threshold in 5. This result demonstrates the trade-off between basis risk and the premium charged for the insurance, i.e., an insurer cannot reduce basis risk without having to charge more for the insurance.

10.5   Temperature-based Insurance with multiple thresholds

Table 9: Payoff structure for Temperature-Index Insurance with multiple thresholds

Temperature (°C)	Payoff
$x < 27.75$	0
$27.75 \leq x < 28.29$	100
$28.29 \leq x < 29.05$	110
$29.05 \leq x < 29.69$	120
$29.69 \leq x < 30.25$	125
$30.25 \leq x < 30.93$	127.5
$x \geq 30.93$	130

The price for this type of insurance is calculated as a combination of cash-or-nothing calls and is found to be 32.77. As can be seen, the premium of the product with multiple thresholds is higher than all of the premiums for the products with one threshold in 7.

11   Multi-Index Insurance

11.1   Copula Approach

Table 10: Premiums from Copula Approach

		Rain (mm)					
		$q_{0.3}$	$q_{0.2}$	$q_{0.1}$	$q_{0.05}$	$q_{0.025}$	$q_{0.01}$
	$q_{0.7}$	16.17	12.25	7.24	4.05	2.16	0.87
	$q_{0.8}$	12.34	9.82	6.27	3.67	2.02	0.84
Temperature (°C)	$q_{0.9}$	7.24	6.22	4.48	2.93	1.74	0.77
	$q_{0.95}$	4.04	3.64	2.91	2.16	1.41	0.69
	$q_{0.975}$	2.17	2.03	1.74	1.42	1.02	0.57
	$q_{0.99}$	0.92	0.88	0.80	0.69	0.56	0.38

Table 10 contains the insurance premiums obtained when using the copula approach to calculate the joint probabilities of exceeding the thresholds. These prices are deemed to be the most accurate since using copulas to model the joint probability distribution does a better job at capturing the dependency between the two indices. As expected the premium decreases as the threshold is moved further into the tails of the distributions of the respective indices since these thresholds are less likely to be breached.

11.2 Multivariate Approach

Table 11: Premiums from Multivariate Approach

		Rain (mm)					
		$q_{0.3}$	$q_{0.2}$	$q_{0.1}$	$q_{0.05}$	$q_{0.025}$	$q_{0.01}$
	$q_{0.7}$	15.11	11.08	6.26	3.43	1.85	0.80
	$q_{0.8}$	11.07	8.34	4.90	2.78	1.54	0.68
Temperature (°C)	$q_{0.9}$	6.26	4.90	3.05	1.81	1.04	0.48
	$q_{0.95}$	3.43	2.77	1.82	1.13	0.67	0.33
	$q_{0.975}$	1.84	1.53	1.05	0.68	0.42	0.21
	$q_{0.99}$	0.79	0.68	0.48	0.33	0.21	0.11

Table 11 presents the insurance premiums when modelling the joint probability distribution as a correlated multivariate distribution, and estimating the joint probability that the observed rainfall and average temperatures have crossed their payout threshold levels. As above, respective thresholds are set at the upper and lower 30%, 20%, 10%, 5%, 2.5%, and 1% quantiles of the marginal distributions. Immediately, we notice that the multivariate approach universally prices lower than the copula approach in table 10, perhaps due to the multivariate approach not being able to capture the dependence between rainfall and temperature as well as the copula approach does. This would lead to a misestimation of the probability that low rainfall and high temperature are observed simultaneously.

11.3 Assumed Independence Approach

Table 12: Premiums from Assumed Independence Approach

		Rain (mm)					
		$q_{0.3}$	$q_{0.2}$	$q_{0.1}$	$q_{0.05}$	$q_{0.025}$	$q_{0.01}$
	$q_{0.7}$	8.88	5.93	2.99	1.49	0.75	0.30
	$q_{0.8}$	5.90	3.95	1.99	0.99	0.50	0.20
Temperature (°C)	$q_{0.9}$	2.92	1.95	0.98	0.49	0.25	0.10
	$q_{0.95}$	1.47	0.98	0.49	0.25	0.12	0.05
	$q_{0.975}$	0.74	0.49	0.25	0.12	0.06	0.02
	$q_{0.99}$	0.29	0.20	0.10	0.05	0.02	0.01

Table 12 shows the premiums as calculated using the approach where we assume that the two indices are independent. It is observed that this approach is underpricing the insurance product across all combinations of the indices. This underpricing

is due to the fact that not taking the dependence into account underestimates the probability of high temperatures coinciding with low rainfall.

12 Government Relief vs Paying Premiums

In the event of a natural disaster, national governments, especially in developing countries, provide relief by allocating funds from their budgets that were, in some cases, meant for development. The International Monetary Fund (IMF, 2016) found that these costs can average around 1.8% of the GDP of the country.

In South Africa, in 2022, R1 Billion was allocated for emergency relief for the floods experienced in April in KwaZulu-Natal. From a GDP of R4.60 trillion, this was 0.1% of the GDP in the year Treasury (2016). By taking on index-based insurance, the government knows beforehand how much it needs to pay for the insurance and the premium is paid at predetermined intervals. Moreover, since the government only provides relief when the losses are severe, the government would only take the insurance for the more extreme weather conditions as this would be significantly cheaper.

13 Conclusion and Recommendations for Future Work

In this paper, we calculated prices for single-index-based insurance, double-index-based insurance, and single-index multiple-threshold index-based insurance. Single-index rainfall insurance is priced by fitting historical cumulative rainfall data to a gamma distribution and estimating the probability that rain falls below specified quantiles of its distribution. Single-index temperature insurance is priced by fitting historical average maximum temperature data to a log-normal distribution and estimating the probability that temperature exceeds specified quantiles of its distribution. Single-index multiple-threshold products are also considered, as the higher payoff for more extreme weather conditions may allow for minimization of basis risk in terms of the client being better compensated for more extreme losses. Double-index-based insurance is priced by modelling the distribution of the joint-index with a 270° rotated Gumbel copula and, as a point of comparison, we also model the joint distribution under the assumption that the dependence between the indices is linear, and the assumption that the indices are independent.

We found that allowing multiple thresholds produces a universally more expensive insurance product, and the combination of indices to create a double-index product produces a product universally cheaper than the single-index products. Modelling the joint distributions by assuming linear or no independence results in lower pricing than the prices generated from the copula approach, due to inadequate modelling of the dependence between rainfall resulting in misestimations of the probability that low rainfall coincides with high temperatures.

We find that since the premiums are very low for severe loss events, it might be cheaper for the government to buy the insurance than pay for relief when extreme conditions strike.

As the temperature data was found to be log-normally distributed, we attempted pricing the single-index temperature insurance under the assumptions of standard Black-Scholes-Merton and risk-neutral valuation theory and found that this assumption leads to significant over-pricing relative to the actuarial approach. This suggests that risk-neutral pricing is inappropriate for the pricing of weather-index insurance, in line with previously published literature (Cao and Wei, 2004). As potential future work, one might model the weather-index processes via the financial approach as appropriate stochastic differential equations and transform those to a risk-neutral measure to create a new risk-neutral valuation framework.

Although having a long history of data is desirable, because of climate change historical data might not be a good reflection of current and future climatic conditions and potential hazards. Thus future work should account for the effects of climate

change.

Bibliography

Department of Agriculture, F., Fisheries, 2021. A PROFILE OF THE SOUTH AFRICAN MAIZE MARKET VALUE CHAIN. Department of Agriculture, Land Reform and Rural Development Website. "Retrieved from: `https://webapps1.daff.gov.za/AmisAdmin/upload/Maize%20profile%202021.pdf`".

Barnett, B.J., Barrett, C.B., Skees, J.R., 2008. Poverty traps and index-based risk transfer products. World Development 36, 1766–1785.

Baškot, B., Stanić, S., 2020. Parametric crop insurance against floods: The case of bosnia and herzegovina. Economic Annals 65, 83–100.

Cao, M., Wei, J., 2004. Weather derivatives valuation and market price of weather risk. Journal of Futures Markets 24, 1065–1089. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/fut.20122`, doi:`https://doi.org/10.1002/fut.20122`, arXiv:`https://onlinelibrary.wiley.com/doi/pdf/10.1002/fut.20122`.

Department of Agriculture, Land Reform and Rural Development, 2022. Maize Production in South Africa. Retrieved from: `https://www.dalrrd.gov.za/phocadownloadpap/Brochures_and_Production_Guidelines/Maize%20brochure%20updated%20Nov%202022.pdf`.

Embrechts, P., Lindskog, F., McNeil, A., 2001. Modelling dependence with copulas. Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich 14, 1–50.

Govindaraj, M., Pattanashetti, S.K., Patne, N., Kanatti, A.A., Ciftci, Y., 2018. Breeding cultivars for heat stress tolerance in staple food crops. Next generation plant breeding 76480.

Hess, U., Syroka, J., 2005. Weather-based insurance in southern africa: The case of malawi .

Horton, J., 2021. Parametric insurance as an alternative to liability for compensating climate harms. carbon climate law rev 12: 285–296.

IMF, 2016. Small states' resilience to natural disasters and climate change - role for the IMF (IMF Policy Paper). Washington, DC: International Monetary Fund. Retrieved from: `https://www.imf.org/external/np/pp/eng/2016/110416.pdf`.

Jain, R., 2004. Challenges in implementing agriculture insurance and re-insurance in developing countries. The Journal 2004, 14–23.

Jewson, S., Zervos, M., 2003. The black-scholes equation for weather derivatives. SSRN .

Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. Journal of Banking & Finance 31, 2405–2423.

Li, S.T., Hammond, J.L., 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. IEEE Transactions on Systems, Man, and Cybernetics SMC-5, 557–561. doi:`10.1109/TSMC.1975.5408380`.

Mookerjee, A., Clarke, D., Grenham, D., Sharpe, J., Stein, D., 2014. Crop microinsurance: tackling poverty, one insurance policy at a time. Annals of Actuarial Science 8, 253–280.

Okine, A.N., 2014. Pricing of index insurance using Black-Scholes framework: A case study of Ghana. Master's thesis. Illinois State University.

du Plessis, J., 2003. Maize Production. Department of Agriculture Website. "Retrieved from: `https://www.arc.agric.za/arc-gci/fact%20sheets%20library/maize%20production.pdf`".

Prokopchuk, O., Prokopchuk, I., Mentel, G., Bilan, Y., 2020. Parametric insurance as innovative development factor of the agricultural sector of economy. AGRIS on-line Papers in Economics and Informatics 12, 69–86.

Rajini, A., Jayalakshmi, C., 2021. Modeling temperature and precipitation in hyderabad and medak using copula. Indian Journal of Science and Technology 14, 1274–1282. doi:`10.17485/IJST/v14i16.417`.

Rui Zhou, J.S.H.L., Pai, J., 2019. Pricing temperature derivatives with a filtered historical simulation approach. The European Journal of Finance 25, 1462–1484. URL: `https://doi.org/10.1080/1351847X.2019.1602068`, doi:`10.1080/1351847X.2019.1602068`, arXiv:`https://doi.org/10.1080/1351847X.2019.1602068`.

Tadesse, M.A., Shiferaw, B.A., Erenstein, O., 2015. Weather index insurance for managing drought risk in smallholder agriculture: lessons and policy implications for sub-saharan africa. Agricultural and Food Economics 3, 1–21.

Treasury, N., 2016. Funding Framework for the Response to the National Disaster. National Treasury Media Statement. Retrieved from: `https://www.imf.org/external/np/pp/eng/2016/110416.pdf`.

Zippenfenig, P., 2023. Open-meteo.com weather api. URL: `https://open-meteo.com/`, doi:`10.5281/zenodo.7970649`.

Generative Modelling of Arbitrage-Free Markets via Neural SDEs

TEAM 2

JUSTIN VAN DER MERWE, University of Cape Town
NICHOLAS BODEN, University of Cape Town
EMMANUEL ABIODUN, University of Cape Town
MARIO GONZÁLEZ, University College London

Supervisor:
JOHANNES WIESEL, Carnegie Mellon University

African Institute of Financial Markets and Risk Management

Contents

Chapter 1

Introduction

1 Background

The modelling of financial markets is a crucial yet challenging task faced by financial engineers, because, unlike scientific fields where most models are governed by physical laws, financial market dynamics are shaped by abstract forces (such as supply and demand), which are influenced by a host of unquantifiable factors like investor sentiment and global economic policies (Gierjatowicz et al., 2020).

Traditionally, the approach to market modelling has been to handcraft assumption-based models that were as parsimonious and computational undemanding as possible (Cuchiero et al., 2020). Such models, however, while valuable for their tractability, are limited in their ability to capture complex market dynamics and the parameters of such models tend to be highly sensitive to training data.

Fortunately, the adoption of neural networks allows for a more robust and data-driven approach to financial market modelling (Gierjatowicz et al., 2020). In some sense, the dynamics of a financial market are much like that of a neural network: a black-box system with a multitude of parameters all interlinked to form a complex structure that is governed by an unclear and non-intuitive pattern.

2 Problem Statement

The crux of financial market modelling is model calibration, which involves the fine-tuning of model parameters to fit to observed market prices. In times past, the calibration speed of a model has been a critical factor influencing its adoption in practice; however, this perspective was shifted by the pioneering work of Hernandez (2016) called *Model Calibration with Neural Networks*.

In their seminal paper, Hernandez (2016) describes the expedition of the model calibration process using feed-forward neural networks, and has prompted scholars

to rethink the notion of model calibration time as a limitation. Essentially, the bulk of neural network training can be performed offline while online calibration (which just involves simple matrix multiplication) proceeds swiftly.

Of all the papers published on the subject, Cuchiero et al. (2020) have identified three main ways that neural networks can be applied to the calibration of financial market models, namely:

1. Calibrating model parameters to market prices

2. Calibrating model parameters to model prices

3. Generative adversarial modelling using neural SDEs

The first two approaches involve the traditional approach of specifying a parametric model and using neural networks to fit the model parameters. The third approach, however, models the stochastic diffusion dynamics using neural networks, and is model-free as it allows the data to dictate the model (hence the term generative modelling).

Our work in this report investigates the use of generative modelling using neural SDEs, building upon the work published by Gierjatowicz et al. (2020). Our main objective is to use conditional expectation in the pricing of different options using a local volatility neural stochastic differential equation with the following form:

$$dX_t^\theta = b(t, X_t^\theta, \theta)\, dt + \sigma(t, X_t^\theta, \theta)\, dW_t \qquad (1.1)$$

where X_t^θ is a stochastic process, θ is a set of parameters and b and σ are neural networks.

The remainder of our report is structured as follows: Chapter 2 briefly reviews literature on neural networks and conditional expectation techniques. Chapter 3 discusses the mathematical framework underpinning model calibration using neural SDEs as well as the variance reduction (conditional expectation and hedging portfolio) strategies. Chapter 4 provides a breakdown of the methodology used for investigation. Chapter 5 reports on the results and key insights of our investigation and, finally, in Chapter 6 conclusions and further recommendations are presented.

Chapter 2

Literature Review

1   Neural Networks

(Artificial) neural networks are mathematical machine learning models, inspired by the neural structure of the brain, that enable computers to learn from vast datasets and perform human-like tasks like pattern recognition, decision-making and time series prediction (Da Silva et al., 2017). At their core, neural networks comprise a system of interconnected artificial neurons, whose structure is depicted in Figure 2.1.
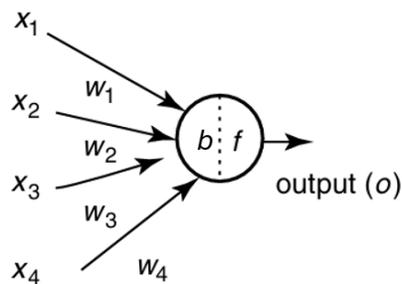
Figure 2.1: The basic structure of an artificial neuron (Abraham, 2005)

In reference to Figure 2.1, each artificial neuron is a mathematical model that takes a weighted (w_1, w_2, w_3, w_4) combination of its inputs (x_1, x_2, x_3, x_4), adds a bias term (b) and then transforms the combined result using a transfer function, f, to obtain an output (o) (Abraham, 2005).

A number of transfer functions, f, have been developed for artificial neuron modelling, for example, the sigmoid function, given by $f(\lambda) = (1 + e^{-\lambda})^{-1}$, transforms input values into a range between 0 and 1, and is particularly useful or applications

where outputs need to represent probabilities (Amiri and Derakhshandeh, 2011).

Finally, in consolidation of all the artificial neuron components described above, a generalised equation relating the inputs of a neuron to its output(s) is given by the equation below Krenker et al. (2011):

$$o = f\left[\sum_{i=1}^{n}(\mathbf{w}_i \cdot \mathbf{x}_i + b) \right] \tag{2.1}$$

1.1 Neural Network Architecture

The combination of neurons into a neural network can either be done such that information flows only in one direction from inputs to outputs, or it can be done such that the network utilises both forward connections and feedback loops between its neurons (Basheer and Hajmeer, 2000). These two neural network architecture types — the former being a feed-forward neural network (FNN) and the latter being a recurrent neural network (RNN) — are contrasted in Figure 2.2.

It is worth mentioning that the feed-forward architecture of a neural network should not be mistaken for forward propagation. The latter forms part of the training process of artificial networks, as will be outlined in Section 1.2, which has been informed by *Artifical Neural Networks*, a book published by Da Silva et al. (2017).
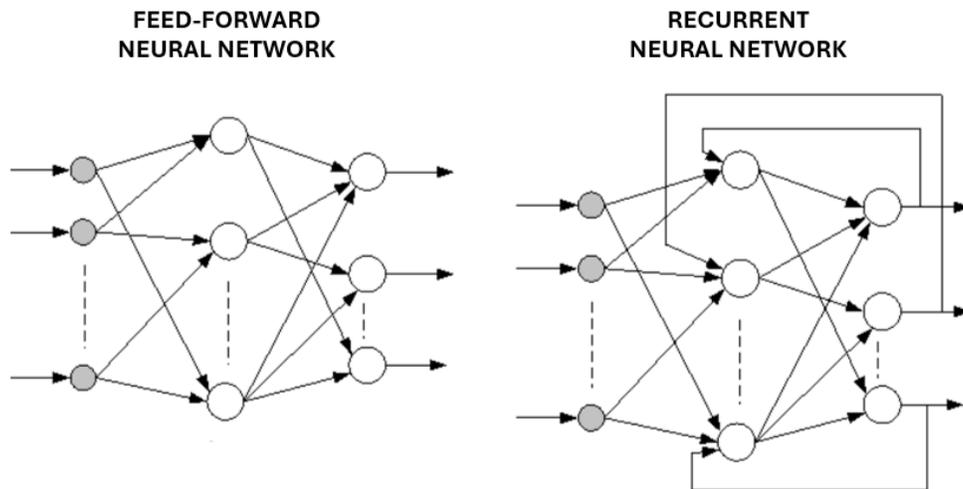
Figure 2.2: A comparison between the two main neural network architecture types

For the purposes of financial market modelling, the feed-forward artificial network architecture is preferred for two main reasons (1) it is a simple architecture which would make computation more stable and (2) given that one wants stock price

process to be adapted for option pricing, using a recurrent neural network would conflict with this major assumption, which underpins the model.

1.2 Neural Network Training Process

The training of a neural network is a recursive process that mirrors a feedback-loop. In the same way that one accumulates knowledge and refines their understanding over time, a neural network processes large volumes of data and adjusts model weights and biases to improve future predictions.

Initially, the neural network is initialized with a random set of weights and biases, akin to an uninformed mind. Through a process called forward propagation, it computes predictions based on input data. These predictions are then compared to the actual outcomes using a defined loss function, providing feedback on its performance. Backpropagation adjusts the network's weights and biases in response to this feedback, akin to learning from mistakes.

This iterative cycle of prediction, evaluation, and adjustment continues for multiple epochs until the network's performance converges to an optimal state. Ultimately, the neural network learns to generalize from the training data, enabling it to make accurate predictions on unseen data. Thus, a crucial activity in the training of a neural network is to create a substantial training set, enabling the neural network to generalize effectively from it Hernandez (2016).

2 Conditional Expectation Estimation

A well studied approach to conditional expectation estimation is Kernel smoothing. Kernel smoothing is a non-parametric technique that assign weights to neighboring data points based on a chosen kernel function, typically a symmetric and non-negative function such as the Gaussian or Epanechnikov kernel Tibshirani and Wasserman (2013). A review of this method will be discussed further in Section 3 of the Methodology.

Longstaff and Schwartz (2001) introduced a method for approximating American options prices using least squares to estimate the conditional expected payoff for continuation. Their approach, otherwise known as the Least Squares Monte Carlo (LSM) method, is simple to implement and applicable for conditional expectation estimation in this report.

Ólafsson (2023) reviews the LSM method for approximating conditional expectations and compares it to the Gaussian mixture model (GMM) which promises more efficient and accurate calculations for option pricing and hedging strategies.

Chapter 3

Mathematical Framework

1 Generative Market Modelling

The majority of the discussion in this section is derived from the study conducted by Gierjatowicz et al. (2020).

1.1 Neural Stochastic Differential Equations

Assuming constant interest rate $r \in \mathbb{R}$ and considering parameter space $\Theta = \Theta^b \times \Theta^\sigma \subseteq \mathbb{R}^p$ and parametric functions $b : \mathbb{R}^d \times \Theta^b \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \Theta^\sigma \to \mathbb{R}^{d \times n}$. Let $(W_t)_{t \in [0,T]}$ be an n-dimensional Brownian motion supported on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{Q})$ so that $\mathbb{Q}$ is the Wiener measure and $\Omega = C([0,T]; \mathbb{R}^n)$. We consider the following parametric SDE:

$$dX_t^\theta = b(t, X_t^\theta, \theta) \, dt + \sigma(t, X_t^\theta, \theta) \, dW_t \tag{3.1}$$

We split X_t^θ into traded and non-traded assets and focus only on the first ones (traded) defined as S, then the SDE for these assets can be defined as:

$$dS_t^\theta = rS_t^\theta \, dt + \sigma^S(t, S_t^\theta, \theta) \, dW_t \tag{3.2}$$

This process defined in a discounted value $(((e^{-rt}S_t)_{t \in [0,T]}))$ has the property of local martingale and thus the model is free of arbitrage (Shreve et al., 2004). If (b, σ) are defined to be neural networks, the SDE defined above is called **neural SDE** and we denote by $\mathcal{M}^{nsde}(\theta)$ the class of solutions. For simplicity, (Gierjatowicz et al., 2020) both b and σ have been modelled as feed-forward neural networks.

We denote the law X^θ on $C([0,T]; \mathbb{R}^d)$ by $\mathbb{Q}(\theta) := \mathcal{L}((X_t)_{t \in [0,T]})$. Given a loss function $l : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$, the search for a calibrated model can be written as

$$\theta^* \in \arg\min_{\theta \in \Theta} \sum_{i=1}^{M} l\left(\mathbb{E}_{\mathbb{Q}(\theta)}[\Phi_i], p(\Phi_i)\right), \tag{3.3}$$

where

$$\mathbb{E}_{\mathbb{Q}(\theta)}[\Phi] = \int_{C([0,T],\mathbb{R}^d)} \Phi(\omega)\mathcal{L}(X^\theta)(d\omega). \tag{3.4}$$

1.2   Model Calibration Optimization Problem

Let $\ell : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ be a convex loss function such that $\min_{x \in \mathbb{R}, y \in \mathbb{R}} l(x,y) = 0$. The aim is to solve the following optimisation problems:

- Find model parameters θ^* such that model prices match market prices:

$$\theta^* \in \arg\min_{\theta \in \Theta} \ell\left(\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i], p(\Phi_i)\right). \tag{3.5}$$

- Find model parameters $\theta^{l,*}$ and $\theta^{u,*}$ which provide robust arbitrage-free price bounds for an illiquid derivative, subject to available market data:

$$\theta^{l,*} \in \arg\min_{\theta \in \Theta} \mathbb{E}^{\mathbb{Q}(\theta)}[\Psi], \text{ subject to} \sum_{i=1}^{M} \ell\left(\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i], p(\Phi_i)\right) = 0. \tag{3.6}$$

$$\theta^{u,*} \in \arg\max_{\theta \in \Theta} \mathbb{E}^{\mathbb{Q}(\theta)}[\Psi], \text{ subject to} \sum_{i=1}^{M} \ell\left(\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i], p(\Phi_i)\right) = 0. \tag{3.7}$$

The no-arbitrage price of Ψ over the class of neural SDEs used is then in the range $[\mathbb{E}^{\mathbb{Q}(\theta^{l,*})}, \mathbb{E}^{\mathbb{Q}(\theta^{u,*})}]$

1.3   Learning hedging strategy as a control variable

A practical algorithm is to estimate $\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi]$ using a Monte Carlo estimator based on the idea that due the Law of Large Numbers and Central Limit Theorem $\mathbb{E}^{\mathbb{Q}^N(\theta)}[\Phi]$ converges to $\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi]$ and increasing N, there is a reduction in a confidence interval, respectively. This increase of N, increases the overall computational cost, so a better strategy is to find a control variate Φ^{cv} such that:

$$\mathbb{E}^{\mathbb{Q}^N(\theta)}[\Phi^{cv}] = \mathbb{E}[\Phi] \quad \text{and} \quad \mathbb{V}ar[\Phi^{cv}] < \mathbb{V}ar[\Phi]. \tag{3.8}$$

According to the Martingale Representation Theorem, there is a methodology for finding Monte Carlo estimators with the above stated properties (Cohen and Elliott, 2015).

If Φ is such that $\mathbb{E}^{\mathbb{Q}}[|\Phi|^2] < \infty$, then there exists a unique process $Z = (Z_t)_t$ adapted to the filtration $(\mathcal{F}_t)_{t \in [0,T]}$ with $\mathbb{E}^{\mathbb{Q}}[\int_t^T |Z_s|^2 ds] < \infty$ such that

$$\mathbb{E}[\Phi|\mathcal{F}_0] = \Phi - \int_0^T Z_s dW_s. \tag{3.9}$$

Define

$$\Phi^{cv} := \Phi - \int_0^T Z_s dW_s \tag{3.10}$$

and note that

$$\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi^{cv}|\mathcal{F}_0] = \mathbb{E}^{\mathbb{Q}(\theta)}[\Phi|\mathcal{F}_0] \quad \text{and} \quad \mathbb{V}ar^{\mathbb{Q}(\theta)}[\Phi^{cv}|\mathcal{F}_0] = 0 \tag{3.11}$$

The process defined above (Z) requires an approximation, which in this case will be a neural network: $\hbar : [O,T] \times C([0,T],\mathbb{R}^d) \times \mathbb{R}^p \to \mathbb{R}^d$ with parameters $\xi \in \mathbb{R}^{p'}$ with $p' \in \mathbb{N}$ and define the following learning task, in which θ is fixed:

$$\xi^* \in \arg\min_{\xi} \mathbb{V}ar \left[\Phi((X_t^\theta)_{t \in [0,T]}) - \int_0^T \hbar(s, (X_{s \wedge t}^\theta)_{t \in [0,T]}, \xi) dW_s | \mathcal{F}_0 \right]. \tag{3.12}$$

Following the same way, we can obtain new equations from (3.6) and (3.7) by changing (Ψ, Φ) to (Ψ^{cv}, Φ^{cv}) which have lower Monte Carlo variance. These new equations provide us a hedging strategy for trading in the underlying asset (S) to replicate the derivative payoff.

Recall that we focus only on traded assets for simplicity, so the control variate is adapted for the tradeable asset.

The idea of a hedging strategy is as follows: let $\bar{S}_t^\theta := e^{-rt} S_t^\theta$ and following (3.2)

$$d(\bar{S}_t^\theta) = e^{-rt} \sigma^S(t, X_t^\theta, \theta) dW_t \tag{3.13}$$

Therefore, a hedging strategy will be defined as:

$$h^S = e^{-rt} \bar{h}_t^S \sigma^S(t, X_t^\theta, \theta) dW_t \tag{3.14}$$

then it would be generating the same path (fluctuations) as the tradable asset (S). Later, a hedging strategy for an option is estimated.

Linking everything together, we have that, by the martingale representation theorem, we know that

$$\mathbb{E}[\Phi|\mathcal{F}_0] = \Phi - \int_0^T Z_s dW_s.$$

The integral term can be replaced by the stochastic integral of the hedging strategy with respect to the discounted stock price. In the case of Gierjatowicz et al. (2020), this hedging strategy is computed via a second neural network.

1.4 Time Discretisation

Throughout the implementation of the model ((3.6) and (3.7)), the stochastic integral (3.12) is approximated using a sum of increments over a partition π of $[0, T]$ with N steps. The approximation of the parameters of the neural SDE plays an important role as there is a possibility that the moments of the simulated paths via the classical Euler scheme could blow up. In order to avoid this issue, the *Tamed Euler scheme* is used:

$$X_{t_{k+1}}^{\pi,\theta} = X_{t_k}^{\pi,\theta} + \frac{b(t_k, X_{t_k}^{\pi,\theta}, \theta)}{1 + \left|b(t_k, X_{t_k}^{\pi,\theta}, \theta)\right|\sqrt{\Delta t_k}} \Delta t_k + \frac{\sigma(t_k, X_{t_k}^{\pi,\theta}, \theta)}{1 + \left|\sigma(t_k, X_{t_k}^{\pi,\theta}, \theta)\right|\sqrt{\Delta t_k}} \Delta W_{t_k}$$

(3.15)

where $\quad \Delta t_k = t_{k+1} - t_k \quad$ and $\quad \Delta W_{t_k} = W_{t_{k+1}} - W_{t_k}$.

1.5 Stochastic Gradient Descent

Define the following classical optimisation problem

$$\min_{\theta \in \Theta} h(\theta), \quad h(\theta) := \mathbb{E}[H(\theta)].$$

(3.16)

Under suitable conditions on H and the learning rates $(\eta_k)_{k=1}^{\infty}$ with $\eta_k > 0$ for all k, the classical gradient descent algorithm can be applied to this optimization problem. Since $\mathbb{E}[H(\theta_k)]$ cannot be computed explicitly, we use stochastic gradient descent. This results in the following update rule:

$$\theta_{k+1} = \theta_k - \eta_k \frac{1}{N} \sum_{i=1}^{N} \partial_\theta H^i(\theta_k),$$

(3.17)

where $(H^i(\theta))_{i=1}^{N_{batch}}$ are independent samples from the distribution of $H(\theta)$ and N is the size of the mini-batch. The choice of a good estimator (unbiased) for $\mathbb{E}[H(\theta)]$, the stochastic gradient descent converges to a minimum of h.

The main objective of this project is to minimize the loss

$$\sum_{i=1}^{M} \ell\left(\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i^{cv}], p(\Phi_i)\right).$$

(3.18)

To compute (3.18), we use stochastic gradient descent described above, where differentiation is justified under the conditions in (Glasserman, 2004) and we have $\partial_\theta \mathbb{E}^{\mathbb{Q}}[G(X^\theta)] = \mathbb{E}^{\mathbb{Q}}[\partial_\theta G(X^\theta)]$. In conclusion

$$\partial_\theta = \sum_{i=1}^{M} (\partial_x \ell)(\mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i^{cv}], p(\Phi_i)) \partial_\theta \mathbb{E}^{\mathbb{Q}(\theta)}[\Phi_i^{cv}]$$
$$= \sum_{i=1}^{M} (\partial_x \ell)(\mathbb{E}[\Phi_i^{cv}(X^\theta)], p(\Phi_i)) \mathbb{E}[\partial_\theta \Phi_i(X^\theta)]. \tag{3.19}$$

Eminently, a key result underpinning the use of feed-forward neural networks in stochastic modelling was developed by Hornik (1991). This theorem, which is stated below, postulates that any given continuous function can be approximated by one-layer neural networks.

Theorem 1.1. Universal Approximation Theorem

Let $\mathcal{N}_{N_I,N_O}^{\phi}$ denote the set of all feed-forward neural networks with input layer dimension N_I, output layer dimension N_O and activation function ϕ. Given that ϕ is bounded and non-constant then the following statements hold

1. *For any finite measure, μ on $(\mathbb{R}^{N_0}, \mathcal{B}(\mathbb{R}^{N_0}))$ and $1 \le p < \infty$, the set $\mathcal{N}_{N_I,N_O}^{\phi}$ is dense in $L^p(\mathbb{R}^{N_0}, \mathcal{B}(\mathbb{R}^{N_0}), \mu)$*

2. *If in addition $\phi \in C(\mathbb{R}, \mathbb{R})$ then $\mathcal{N}_{N_I,N_O}^{\phi}$ is dense in $C(\mathbb{R}^{N_0}, \mathbb{R})$ for the topology of uniform convergence on compact sets.*

2 Conditional Expectation and pricing of derivatives

There are two main methods for pricing derivatives:

1. Black-Scholes-Merton Partial Differential Equation (PDE)

2. Discounted expected payoff under Risk-Neutral measure via Stochastic Differential Equation (SDE)

By no-arbitrage, these two methods should provide the same result. Feynman-Kac theorem provides a link between the two methods.

2.1 Feynman-Kac

We assume that the process S_t is defined by:

$$dS_t^\theta = rS_t^\theta dt + \sigma^S(t, S_t^\theta, \theta)dW_t \tag{3.20}$$

and that $(S, t) \mapsto G(S, t)$ is a function that satisfies the following PDE:

$$\frac{\partial G(S_t, t)}{\partial t} + rS_t^\theta \frac{\partial G(S_t, t)}{\partial S_t} + \frac{1}{2}(\sigma^S(t, S_t^\theta, \theta))^2 \frac{\partial^2 G(S_t, t)}{\partial S_t^2} - rG(S_t, t) = 0 \qquad (3.21)$$

with the boundary condition $G(s, T) = \Phi(s, T)$. Then, the **solution** of this PDE is given by

$$G(s, t) = \mathbb{E}[e^{-r(T-t)} \Phi(S_T, T) \mid S_t = s] \qquad (3.22)$$

2.2 Hedging Strategy

We assume that $G(S_t, t)$ is a function of an option value and satisfies the Feynman-Kac theorem, where S_t follows the process (3.20). Therefore, by Itô's Lemma, the dynamics of the process $G(S_t, t)$ are given by

$$dG(S_t, t) = \left[\frac{\partial G(S_t, t)}{\partial t} + \frac{\partial G(S_t, t)}{\partial S}rS_t + \frac{1}{2}\frac{\partial^2 G(S_t, t)}{\partial S^2}\sigma^S(t, S_t, \theta)^2 \right] dt$$
$$+ \frac{\partial G(S_t, t)}{\partial S}\sigma^S(t, S_t, \theta) \, dW_t. \qquad (3.23)$$

The idea is to find a strategy (h) which hedges the uncertainty of the option.

$$\begin{aligned} h_t &= \gamma_t S_t + \phi_t B_t \\ dh_t &= \gamma_t dS_t + \phi_t r B_t dt \\ &= (\gamma_t S_t + \phi_t B_t)r dt + \gamma_t \sigma^S(t, S_t, \theta) \, dW_t \end{aligned} \qquad (3.24)$$

The fluctuations of both portfolios must be equal:

$$\begin{aligned} \frac{\partial G(S_t, t)}{\partial S}\sigma^S(t, S_t, \theta) \, dW_t &= \gamma_t \, \sigma^S(t, S_t, \theta) \, dW_t \\ \frac{\partial G(S_t, t)}{\partial S} &= \gamma_t \end{aligned} \qquad (3.25)$$

This result shows that the number of shares of the stock S_t must be equal to the delta of the option.

2.3 Conditional Expectation and Conditional Variance

As we have stated, Φ is a random variable which is associated with a Control Variate Φ^{cv} defined as

$$\Phi^{cv} := \Phi - \int_t^T Z_s dW_s. \qquad (3.26)$$

14

By the properties defined in (3.11), this implies that Φ^{cv} is an unbiased estimator of Φ but with a reduced variance.

$$\mathbb{V}\mathrm{ar}(\Phi^{cv}) = \mathbb{V}\mathrm{ar}\left(\Phi - \int_t^T Z_s dW_s\right). \tag{3.27}$$

Variance reduction is achieved through the orthogonal decomposition of Φ into a predictable part and a martingale

$$\Phi = \mathbb{E}(\Phi|\mathcal{F}_t) + \int_t^T Z_s dW_s. \tag{3.28}$$

Therefore,

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(\Phi^{cv}) &= \mathbb{V}\mathrm{ar}\left(\Phi - \int_t^T Z_s dW_s\right) \\
&= \mathbb{V}\mathrm{ar}\left(\mathbb{E}(\Phi|\mathcal{F}_t) + \int_t^T Z_s dW_s - \int_t^T Z_s dW_s\right) \\
&= \mathbb{V}\mathrm{ar}\left(\mathbb{E}(\Phi|\mathcal{F}_t)\right).
\end{aligned}
\tag{3.29}
$$

By (3.28) and because $\mathbb{V}\mathrm{ar}(\int_t^T Z_s dW_s) > 0$, it is clear that $\mathbb{V}\mathrm{ar}(\Phi^{cv}) \leqslant \mathbb{V}\mathrm{ar}(\Phi)$.

This result provides an insight into how the Monte Carlo variance is decreased through learning of conditional expectations, or equivalently Z. The orthogonal decomposition stated above allows us to understand how Φ can be split into a predictable part and a martingale part. By using a control variate (Φ^{cv}), we achieve a reduction in the variance of the estimator, thereby improving the efficiency of the Monte Carlo method. This variance reduction is based on the lower conditional variance of the predictable part compared to the total variance of Φ.

Chapter 4

Methodology

1 Crude Monte Carlo

The crude Monte Carlo method is a very popular numerical method of estimating variables. It is often used in financial scenarios to calculate expected gains and losses, as well as pricing various options. The method involves running various scenarios and using the average of the outcomes to provide an estimated sample. Since it is a simple method, with most methods built around it, it is used as a benchmark for all the different methods we discuss on the following pages.

Algorithm 1: Generative model calibration to European call option prices for one maturity

---

Input: $\pi = \{t_0, t_1, \cdots t_{N_{\text{steps}}}\}$ time grid for numerical simulations

Input: $(\Phi_i)_{i=1}^{N_{\text{prices}}}$ option payoffs

Input: $\mathfrak{p}(\Phi_j), j = 1, \cdots, N_{\text{prices}}$ market option prices

Initialisation: θ for neural SDE parameters $N_{\text{trn}} \in \mathbb{N}$

for $= 1 : N_{epochs}$ **do**

  1) Generate N_{rn} paths: $(x_{t_n}^{\pi,\theta,i})_{n=0}^{N_{\text{steps}}}$ for $i = 1, \cdots, N_{\text{rn}}$ using the Tamed Euler scheme in Equation 3.15

  2) Create copies of the generated paths, which no longer depend on θ.
  (i.e., create $(\tilde{x}_{t_n}^{\pi,i})_{n=0}^{N_{\text{steps}}}$) such that $\partial \tilde{x}_{t_n}^{\pi,i} = 0$)

  3) During every epoch: Use Adam (developed by [Kingma and Ba (2014)]) to update θ where:

$$\theta = \arg\hat{\min_\theta} \sum_{j=1}^{N_{\text{prices}}} \left(\mathbb{E}^{N_{\text{trn}}}[\Phi_j(X^{\pi,\theta})] - \mathfrak{p}(\Phi_j) \right)^2$$

end
return θ

---

2 Hedging strategy using a neural network

The model by Gierjatowicz et al. (2020) (which is refered to as Algorithm 1 in their paper) for calibration to market European option prices for one maturity as explained in Chapter 1 Section 1, is utilised in a two-step process. The model works in an iterative manner. As the first matter of business, stocks paths and payoffs must be simulated. This paper uses the tamed Euler method for this, as it avoids blow ups by the moments of the stipulated paths (equation 3.15). It controls the step size in a way that prevents the numerical solution from becoming unstable, especially in the presence of coefficients that grow super-linearly.

These paths are realised for all maturities and all relevant strikes, leading to the payoffs being calculated and the options priced. The algorithm works by alternating between optimising for the single factor NSDE and then focusing on optimising for the control variate. The Algorithm is then extended to a two-factor NSDE with both the diffusion process and the drift process having a stochastic nature.

For the optimising of the NSDE step, the algorithm computes estimated realisations of stock paths based off of the parameters discovered from the previous iteration. These realisations are then used to calculate conditional expectations passed into the neural network training. The predicted expectations are then compared to the true prices from the heston model.

The comparison of the prices is done through the use of Mean Squared Error (MSE). The MSE is therefore the loss function used in the model calibration and is then used to calculate new parameters for the NSDE. The updated parameters are then passed forward for the next iteration. This process repeats itself for 20 iterations and then the best loss value (i.e. the lowest RMSE value) is recorded and documented. The epoch is then ended.

The next step in the algorithm focuses on minimizing the variance of the control variate. Control variates are used in order to decrease variance and, therefore, obtain more accurate results. The neural networks loss function for this step is therefore to minimise the variance of the vanilla call option. This is done by developing new parameters for the control variate neural network. This process occurs in each iteration with the sum of the variance from the expected vanilla option price being stored as the loss function and, therefore, minimised. It is then documented for further analysis.

The program is able to switch between these two neural networks by freezing the step that is not being minimised.This results in the other neural network parameters (e.g., the NSDE when the diffusion neural network is being optimised) being fixed for each alternating epoch and, therefore, for the corresponding 20 iterations in that epoch.

The loop continues until there are realisations for every strike and maturity pair. The crude monte carlo estimates are then calculated and stored. The variance of the control variates calculated using the neural network control variate step is then added to the crude Monte Carlo simulations. This, then, due to it being approximated as a perfect hedge, reduces the variance and allows the NSDE to be minimised effectively.

The parameters are then inputted again and this recurs until a suitable error has been reached. Overall, the model is well-optimised and runs efficiently. The algorithm used can be seen below in Algorithm 2

Algorithm 2: Generative model calibration to European call option prices for one maturity

---

Input: $\pi = \{t_0, t_1, \cdots t_{N_{\text{steps}}}\}$ time grid for numerical simulations

Input: $(\Phi_i)_{i=1}^{N_{\text{prices}}}$ option payoffs

Input: $\mathfrak{p}(\Phi_j), j = 1, \cdots, N_{\text{prices}}$ market option prices

Initialisation: θ for neural SDE parameters $N_{\text{trn}} \in \mathbb{N}$

for $= 1 : N_{epochs}$ **do**

> **1)** Generate N_{rn} paths: $(x_{t_n}^{\pi,\theta,i})_{n=0}^{N_{\text{steps}}}$ for $i = 1, \cdots, N_{\text{rn}}$ using the Tamed Euler scheme in Equation 3.15
>
> **2)** Create copies of the generated paths, which no longer depend on θ. (i.e., create $(\tilde{x}_{t_n}^{\pi,i})_{n=0}^{N_{\text{steps}}}$) such that $\partial \tilde{x}_{t_n}^{\pi,i} = 0$)
>
> **3)** During every odd epoch: Freeze ξ, Use Adam (developed by [Kingma and Ba (2014)]) to update θ where:
>
> $$\theta = \arg\hat{\min}_{\theta} \sum_{j=1}^{N_{\text{prices}}} \left(\mathbb{E}^{N_{\text{trn}}} \left[\Phi_j(X^{\pi,\theta}) - \sum_{k=0}^{N_{\text{steps}}-1} \bar{\mathfrak{h}}(t_k, X_{t_k}^{\pi,\theta}, \xi_j) \Delta \tilde{\bar{S}}_{t_k}^{\pi,\theta} \right] - \mathfrak{p}(\Phi_j) \right)^2$$
>
> **4)** During every even epoch: Freeze θ, Use Adam to update ξ which is the vanilla call option sample variance given by:
>
> $$\xi = \arg\hat{\min}_{\xi} \sum_{j=1}^{N_{\text{prices}}} \mathbb{V}\text{ar}^{N_{\text{trn}}} \left[\Phi_j(X^{\pi,\theta}) - \sum_{k=0}^{N_{\text{steps}}-1} \bar{\mathfrak{h}}(t_k, X_{t_k}^{\pi,\theta}, \xi_j) \Delta \tilde{\bar{S}}_{t_k}^{\pi,\theta} \right]$$

end

return θ, ξ_j *for all prices*$(\phi_i)_{i=1}^{N_{prices}}$

---

3 Kernel Regression

Given a set of paired data $(\boldsymbol{X}, \boldsymbol{Y}) = ([X_1, X_2, \cdots, X_n]^T, [Y_1, Y_2, \cdots, Y_n]^T)$, it is often of interest to want to predict $\boldsymbol{Y}$ from $\boldsymbol{X}$. In particular, this report is concerned with the optimal and unbiased prediction of payoff functions $\boldsymbol{Y} = \Phi(S_T)$ from underlying asset price realisations $\boldsymbol{X} = S_t$, which can be performed by finding an estimate for the regression function, $m(x)$.

$$m(x) = \mathbb{E}(\boldsymbol{Y}|\boldsymbol{X} = x). \tag{4.1}$$

One of the main classes of approaches towards the non-parametric estimation of $m(x)$ is Kernel regression, chief of these being Nadaraya-Watson kernel regression, where the regression function estimate, $\hat{m}(x)$ is defined as follows:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i \, K\left(\frac{||x - X_i||}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{||x - X_i||}{h}\right)}. \tag{4.2}$$

Here h (the bandwidth) is a non-negative parameter and $K(x)$ is a symmetric and non-negative (smoothing kernel) function, with the following properties:

$$\int K(x) \, dx = 1, \quad \int x K(x) \, dx = 0 \quad \text{and} \quad \sigma_K^2 \equiv \int x^2 K(x) \, dx > 0 \tag{4.3}$$

Although several different smoothing kernel functions exist (e.g, the rectangular, triangular, Gaussian, and Epanechnikov kernels), it has been asserted by Tibshirani and Wasserman (2013) that the choice of $K(x)$ is of far less importance than, the choice of the bandwidth, h.

Given the specification of the kernel estimator, it suffers from the curse of dimensionality as the expression needs to be computed for each x. To see this, recall that we require the realisations of the payoff at each maturity. We then need to calculate the conditional expectation based on earlier times. It would clearly be inefficient to implement the conditional expectation approach element-wise for each realisation of S_t. Therefore it is necessary to address this through vectorisation of program code.

3.1 Bandwidth Selection

In essence, the selection of the bandwidth involves balancing estimator bias and variance: a smaller bandwidth provides more detailed but potentially more noisy estimates, while a larger bandwidth gives smoother, more stable estimates, but may overlook important local features in the data.

Bandwidth selection is an art, rather than a science; nevertheless, Tibshirani and Wasserman (2013) outlines a simple technique, called leave-one-out cross validation. If we define $R(h)$ as the risk of using a particular bandwidth, h, then an

estimate of this risk, $\hat{R}(h)$, can be estimated using the using leave-one-out cross-validation score

$$\hat{R}(h) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{m}_{(-i)}(X_i))^2 \tag{4.4}$$

where $\hat{m}_{(-i)}$ is the regression function estimate obtained by omitting the ith data pair (X_i, Y_i). Under this metric, the bandwidth, h, is chosen in order to minimize $\hat{R}(h)$.

3.2   Bias Corrected Estimator

There is a common issue with many kernel-based methods: boundary bias. This issue leads to an overestimation or underestimation of the true regression function because the kernel function has fewer neighboring points to average at the edges of the data support. (Cheruiyot, 2020)

There are several methods to mitigate the boundary bias, such as: local linear regression, variable bandwidth selection, extrapolation, etc.

For this purpose, we proceed throughout **Local Polynomial Regression**, which instead of using a simple average, it fits a local linear model reducing the bias because it better approximates the behavior of the regression function near the boundaries. (Jácome et al., 2005)

$$\hat{m}_{BC}(x) = \hat{m}(x) - \frac{\hat{m}''(x)h^2}{2} \tag{4.5}$$

Overall by implementing the Kernel regression above and calculating a new loss function for the neural network, one can implement the algorithm below.

Algorithm 3: Using conditional expectation through Kernel Regression in order to calibrate the model to calibrate to European call option prices for a single maturity

---

Input: $\pi = \{t_0, t-1, \cdots t_{N_{\text{steps}}}\}$ time grid for numerical simulations
Input: $(\Phi_i)_{i=1}^{N_{\text{prices}}}$ option payoffs
Input: $\mathfrak{p}(\Phi_j), j = 1, \cdots, N_{\text{prices}}$ market option prices
Initialisation: θ for neural SDE parameters $N_{\text{trn}} \in \mathbb{N}$
for $= 1 : N_{epochs}$ **do**

    1) Generate N_{rn} paths: (S_i^θ) for $i = 1, \cdots, N_{\text{rn}}$ using the Tamed Euler scheme in Equation 3.15

    2) Create copies of the generated paths, which no longer depend on θ. (i.e., create $(\tilde{S}_i^\theta)$ such that $\partial \tilde{S}_i^\theta = 0$

    3) Use Adam (developed by [Kingma and Ba (2014)]) to update θ where:

$$\theta = \arg \hat{\min_\theta} = \sum_{j=1}^{N_{\text{prices}}} (\mathbb{E}^{N_{\text{trn}}}[\Phi_j(S^\theta)|\mathcal{F}_j] - \mathfrak{p}(\Phi_j))^2$$

end
return θ

---

4 Hedging using conditional expectation

By the derivation from Feynman-Kac, the process h can be given as the derivative of the conditional expectation with respect to the asset price. As we have an estimator for the conditional expectation, one can calculate the hedging delta directly by applying the central difference method for the first derivative as opposed to approximating the hedge via a neural network. One would expect this method to perform similarly to the method of Gierjatowicz et al. (2020) as we are essentially using the same control variate. It has the additional advantage over the standard conditional expectation approach in that the bias will cancel out in the central difference calculation, however, it suffers from the same limitation as their approach in that the accuracy depends on the number of time steps. To implement this method, we consider the algorithm below.

Algorithm 4: Using conditional expectation as a hedging strategy for model calibration to European call option prices for a single maturity

---

Input: $\pi = \{t_0, t_1, \cdots t_{N_{\text{steps}}}\}$ time grid for numerical simulations
Input: $(\Phi_i)_{i=1}^{N_{\text{prices}}}$ option payoffs
Input: $\mathfrak{p}(\Phi_j), j = 1, \cdots, N_{\text{prices}}$ market option prices
Initialisation: θ for neural SDE parameters $N_{\text{trn}} \in \mathbb{N}$

for $= 1 : N_{epochs}$ **do**

    1) Generate N_{rn} paths: (S_i^θ) for $i = 1, \cdots, N_{\text{rn}}$ using the Tamed Euler scheme in Equation 3.15

    2) Create copies of the generated paths, which no longer depend on θ. (i.e., create create $(\tilde{S}_i^\theta)$ such that $\partial \tilde{S}_i^\theta = 0$

    3) During every epoch: Use Adam (developed by [Kingma and Ba (2014)]) to update θ where:

$$\theta = \arg \hat{\min_\theta} \sum_{j=1}^{N_{\text{prices}}} \left(\mathbb{E}^{N_{\text{trn}}} \left[\Phi_j \left(X^{\pi, \theta} \right) - \sum_{k=0}^{N_{\text{steps}}-1} \frac{\Delta \mathbb{E}^{N_{\text{trn}}} [\Phi_j(\tilde{S}_i^\theta) | \tilde{S}_i^\theta = s]}{\Delta \tilde{S}_i^\theta} \right] - \mathfrak{p}(\Phi_j) \right)^2$$

end

---

return θ

.

Chapter 5

Results and Discussion

1 Comparison of Main Methods

As shown in Chapter 3, the variance of the conditional expectation approach is equal to the approach by Gierjatowicz et al. (2020). The disadvantages of their approach are the time discretisation and the approximation of the second neural network. The disadvantages of the conditional expectation approach are that some estimators are biased as well as the fact that the first and second derivatives need to be calculated for the hedge and the bias correction respectively.

Table 5.1: Disadvantages of Main Methods

$\mathbb{Var}[\Phi^{cv}]$	$\mathbb{Var}[\mathbb{E}[\Phi|\mathcal{F}]]$
Time discretisation for hedging portfolio	Potentially biased estimators
Approximation of second neural network for hedging portfolio	Approximation of derivatives

2 Gaussian Kernel Estimation Bias

As mentioned in Chapter 3, the kernel estimator is a biased estimator of conditional expectation. To illustrate this, we perform a simulation using the conditional expectation method and 100 simulations.

Table 5.2: Parameters

S0	1
K	1
T	1
r	0.025
c	10 000
Batch Size	100
Number of Epochs	100
Bump size	0.0001
h	0.3
Number of Conditional Expectations	8

Figure 5.1: Bias Correction for Conditional Expectation
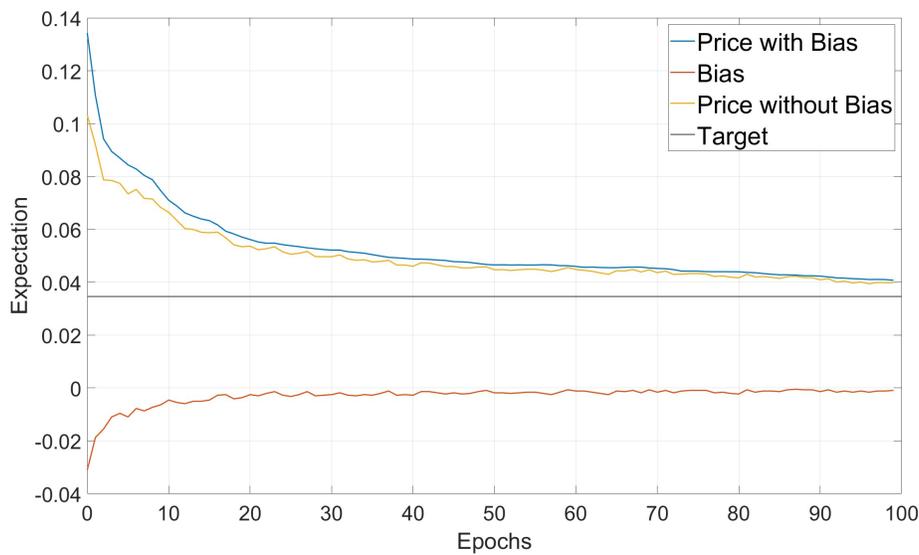
Using the bias correction stated in Chapter 3, we calculate the second derivative of the Nadaraya-Watson estimator using the central difference method with a bump size of 0.0001. As seen in Figure 5.1, there is a clear upward bias for the Gaussian kernel estimator. We observed this behaviour across all strikes. We observed that the bias correction was mostly negative, as expected. However, as the number of epochs is increased this effect becomes less pronounced. This method can therefore be seen more as a means of increasing the convergence rate.

3 Comparison of RMSE

Devroye and Lugosi (2001) note that one of the advantages of using a kernel esti-
mator is that it performs relatively well for a low number of simulations. There-
fore, for a fair comparison of the four methods stated in Chapter 4, we calculate the
RMSE for each method using 1000 simulations using the parameters in Table 5.3.

Table 5.3: Parameters

S0	1
K	[0.8-1.2]
T	1
r	0.025
c	10 000
Batch Size	1 000
Number of Epochs	50
Tolerance for RMSE	2e-05
Bump size	0.0001
h	0.3
Number of Conditional Expectations	8

Figure 5.2: RMSE for the Four Methods
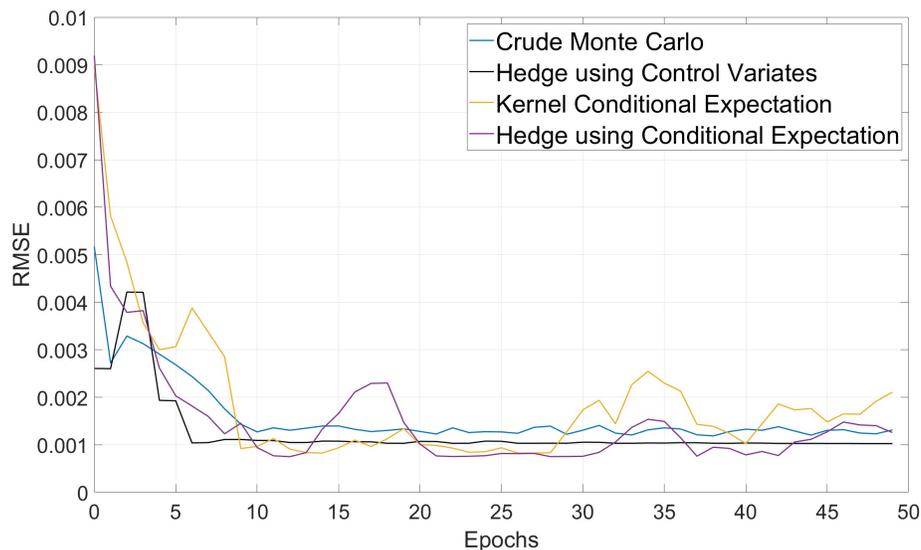
Table 5.4: Best RMSE

Crude Monte Carlo	0.0011868
Hedge using Control Variate	0.0010264
Kernel Conditional Expectation	0.0008264
Hedge using Conditional Expectation	0.0007501

Note that the approach by Gierjatowicz et al. (2020) uses a two step approach and hence we keep the RMSE the same for every second epoch as the second loss function of the method by Gierjatowicz et al. (2020) is not comparable.

In figure 5.2 and table 5.4, we observe that all three methods have a lower RMSE than the crude Monte Carlo method as expected. The results suggest that a more accurate calculation of the first and second derivatives of the conditional expectation would lead to better results for the same number of epochs. The estimate of the conditional expectation as the weighted average of the conditional expectations for certain time steps could also be improved by considering a finer grid.

Chapter 6

Conclusions and Recommendations

1   Conclusions

In this report, we have explored the application of neural stochastic differential equations (SDEs) for generative modeling. The aim of the project was to replace the second neural network present in the hedging strategy by Gierjatowicz et al. (2020) with a conditional expectation. This conclusion was made through using Feynman-Kac Theorem linking the value of an option with the expected discounted payoff, as well as the relationship between the Conditional Expectation and the Control Variate through the reduction variance.

The results suggest that each method has its advantages and disadvantages. Advantages of the neural network hedging strategy resulted in less computational stress and arrived at a result faster than that of the conditional expectation counterpart. This advantage however was short lived with the conditional expectation being able to perform a lower loss function value and therefore match the target data better, which can be seen in Figure 3.

After adjusting for the bias of the kernel estimator, the predicted price of the generative model more closely matched the target market prices. The bias was negative as the bias correction as negative in order to cancel out the upward bias of the kernel. Moreover, the bias of the kernel decreased as as the model became more rigorously trained (i.e. as the number of epochs increased).

2   Recommendations

Based on the promising results, we recommend that this approach be explored further. Specifically, one can consider a larger number of simulations for each

method, different kernel functions, different loss function parameters, different loss functions, different bump sizes for the derivative approximations, as well as other methods of computing conditional expectations so as Least-Squares Monte Carlo. Furthermore, one can apply the two conditional expectation methods to simulated prices based on other models as well as actual market data.

Bibliography

Abraham, A., 2005. Artificial neural networks. Handbook of Measuring System Design .

Amiri, M., Derakhshandeh, K., 2011. Applied artificial neural networks: from associative memories to biomedical applications, in: Artificial Neural Networks - Methodological Advances and Biomedical Applications. Intech.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods 43, 3–31.

Cheruiyot, L.R., 2020. Local linear regression estimator on the boundary correction in nonparametric regression estimation. Journal of Statistical Theory and Applications 19, 460–471.

Cohen, S.N., Elliott, R.J., 2015. Stochastic calculus and applications. volume 2. Springer.

Cuchiero, C., Khosrawi, W., Teichmann, J., 2020. A generative adversarial network approach to calibration of local stochastic volatility models. Risks 8, 101.

Da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F., da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F., 2017. Artificial Neural Networks. Springer.

Devroye, L., Lugosi, G., 2001. Combinatorial methods in density estimation. Springer Science & Business Media.

Gierjatowicz, P., Sabate-Vidales, M., Šiška, D., Szpruch, L., Žan Žurič, 2020. Robust pricing and hedging via neural sdes. arXiv:2007.04154.

Glasserman, P., 2004. Monte Carlo methods in financial engineering. volume 53. Springer.

Hernandez, A., 2016. Model calibration with neural networks doi:10.2139/ssrn.2812140.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural networks 4, 251–257.

Jácome, M.A., Gijbels, I., Cao, R., 2005. Comparison of the nadaraya-watson and local linear methods in presmoothed density estimation under censoring. Universidade de Vigo. Spain .

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization `arXiv:1412.6980`.

Krenker, A., Bešter, J., Kos, A., 2011. Introduction to the artificial neural networks, in: Artificial Neural Networks - Methodological Advances and Biomedical Applications. Intech.

Longstaff, F.A., Schwartz, E.S., 2001. Valuing american options by simulation: a simple least-squares approach. The review of financial studies 14, 113–147.

Ólafsson, V.., 2023. Gaussian mixture models for estimating conditional expectations in the context of option valuation. Ph.D. thesis.

Shreve, S.E., et al., 2004. Stochastic calculus for finance II: Continuous-time models. volume 11. Springer.

Tibshirani, R., Wasserman, L., 2013. Nonparametric regression. Statistical Machine Learning, Springer .

Monte Carlo Arithmetic Asian Options Pricing with Variance Reduction Techniques

TEAM 3

ERICA ZHANG, Stanford University
YINUO SHAN, University College London
ZOË ARENDSE, University of Cape Town
DAMIAN SPARKS, University of Cape Town

Supervisor:
MESIAS ALFEUS, Stellenbosch University

African Institute of Financial Markets and Risk Management

Contents

3

1 Introduction

Option pricing is a hot topic in Mathematical Finance due to its complexity and the variety of options available which can be broadly categorized into vanilla and exotic options. Vanilla options include standard European and American options which depend on the terminal value of the underlying asset. European options can only be exercised at maturity, whereas the long party in an American option contract has the option to exercise at any time at or before maturity. Exotic options are any options priced differently [1], one commonly traded example, especially in commodity markets, being the Asian option. Introduced by Ingersoll (1987), they were originally used in Tokyo to price average option contracts, see for instance Zhang (2003). Asian options consider the geometric or arithmetic average of the price of the underlying asset over the contract term. This means that the variance of these options is lower than that of the underlying and therefore the options are cheaper than equivalent vanilla options Mudzimbabwe et al. (2012).

A variety of established methods exist for option pricing, including closed-form solutions, semi-analytical methods, and Monte Carlo estimation. Arithmetic Asian options, due to it being non-Markovian and the arithmetic structure, do not have a closed-form solution. There can be various forms of Arithmetic Asian options. Specifically, we focus on the Arithmetic Asian option which writes calls on the arithmetic mean of a stock with payoff at a fixed maturity T and with a fixed strike price K:

$$f(A(S,T), K) := \max(0, A(S,T) - K), \tag{1}$$

where

$$A(t) = \frac{1}{T} \int_0^T S_t dt, \tag{2}$$

and the underlying stock process $(S_t)_{t \geq 0}$ follows a standard geometric Brownian motion (GBM) stochastic model.

Due to the path-dependency of such options (as seen in its payoff function 1), Arithmetic Asian options require Monte-Carlo based technique to price them since closed-form solutions are not available. In general, closed-form solutions are the simplest approach, however few formulas exist with the most popular being the Black-Scholes-Merton formula which assumes constant volatility. This however fails to capture the stochasticity of volatility. Stochastic volatility models, such as the Heston and stochastic alpha-beta-rho (SABR) models, solve this issue by allowing instantaneous volatility process to evolve stochastically (see Fukasawa and Gatheral (2021)). Recent empirical evidence on volatility versus time indicates that

---

[1]Exotic options are options that have a more complicated payoff structure, often dependant on the entire evolution of the stock price process rather than just the terminal stock price.

volatility dynamics are rougher than a Brownian semi-martingale, therefore there have been efforts to modify the aforementioned models to form rough models. El Euch and Rosenbaum (2019) note the that fractional Brownian motions with small Hurst parameters model log-volatility dynamics reasonably well. Gatheral et al. (2018) specifically show that rough models reproduce the empirical volatility correlation structure well, closely follow the entirety of the volatility surface and generate the typically observed explosion of the at-the-money skewed term structure for short maturities, especially in Equity markets. Analytical methods use the solution to partial differential equations with given boundary conditions to model prices, while probabilistic methods use numerical integration on risk-neutral conditional expectations (Yin, 2015). Fu et al. (2001) and others show that Monte Carlo methods are advantageous given that convergence is typically independent of the number state variables, making it a preferred choice for higher dimensional problems. With regard to rough processes with dynamics following a fractional Brownian motion, Monte Carlo analysis becomes computationally intensive as the non-Markovian nature makes this analysis complex and inefficient, therefore other methods should be developed when utilizing rough models.

Additionally, it is well known that Monte Carlo option pricing is computationally expensive and that it requires a hundred fold increase in sample size to reduce the standard deviation of the estimate by a factor of ten (Glasserman, 2004). Another way to reduce the standard deviation of the estimated option price is to reduce the standard deviation of the simulated discounted payoff. This is done through variance reduction techniques. The aim of these techniques is to make the option price estimates more accurate and efficient. Commonly implemented techniques include the use of antithetic variates, control variates and endpoint stratification. The goal of this paper however is to reassess the implementation of the diffusion operator integral (DOI) variance reduction technique proposed by Heath and Platen (2002). This paper is organised as follows. In Section 2, the notation and terminology used throughout the paper will be explained and standardised. In Section 3, the price dynamics of the four models will be discussed including a derivation of the processes under a risk neutral measure or equivalent martingale measure. This will be followed by a critical analysis of the variance reduction techniques implemented with special attention paid to the diffusion operator integral method which requires derivation of the Greeks for each model in order to be implemented in Section 4. Section 5 then provides an overview of the numerical methods that will be implemented in simulation including discretisation methods and Monte Carlo pricing algorithms. Finally, we conclude and discuss future avenues of research in Section 6. The Appendix 7 provides additional information regarding mathematical derivations and methodology required in order to implement the four models and relevant variance reduction techniques.

To the best of our knowledge, our work is the first to offer theoretical analysis as well as effective implementation of the Diffusion Operator Integral (DOI) variance

reduction techniques with an application of rough stochastic volatility models such as the rough Heston model and the rough SABR model. It is a highly non-trivial task to implement and successfully apply variance reduction techniques to rough models, due to their fractional Brownian motion components, which corrupt the nice Markovian structure in standard models. Our paper is also among the first to give a comprehensive survey and empirical comparison of the popular variance reduction methods to rough stochastic volatility models, especially with respect to the path-dependent Arithmetic Asian Options, which are known to be difficult to price due to their lack of a closed analytic form. Finally, to contribute to the wider computational finance community, we release our codes through Github to the general public.

2   Notation

Our theoretical analysis for the paper will be based in a continuous time framework. On a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a right continuous filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying $\mathcal{F}_0 = \{\emptyset, \Omega\}$, we consider a market model consisting of assets, on which the price of the Arithmetic Asian options is derived. We describe the price evolution dynamics of an asset in the market in terms of a stochastic process $(S_t)_{t \geq 0}$, which is assumed to be an Itô process adapted to the filtration $\mathcal{F}$.

Throughout the paper, we assume the existence of a risk-free asset and denote its fixed interest rate to be r. Furthermore, given a $(\mathcal{X}, \mathcal{A}, \mu)$-measurable function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{A}$ is a sigma-algebra on set $\mathcal{X}$ and μ is the Lebesgue measure, we use the simplified notation L^p, where $1 \leq p < \infty$, to denote $L^p(X, \mathcal{A}, \mu)$, i.e. the equivalent classes of measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that

$$\int \|f\|^p d\mu < \infty,$$

where $\| \cdot \|$ is the L^p-norm.

3   Price Dynamics Models

A good model for asset evolution dynamics should reproduce the stylized facts of modern electronic markets. The vanilla geometric Brownian model characterized by stochastic differential equation (SDE) under a physical or historical measure

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

assumes constant drifts (μ) and volatility (σ) and is widely criticized for not capturing the reality of market observed market phenomena, where volatility is stochastic. To address this short-coming, models such as Heston models and SABR as well as their rough variants are proposed.

3.1 Heston Model

A very popular stochastic volatility model for an asset price is the Heston model. This model assumes that the underlying stock price, S_t, follows a Black Scholes-type stochastic process, but with a stochastic variance v_t that follows Cox et al. (1985). The Heston model with parameter $\theta = \{\mu, \kappa, \xi, \eta, \rho, s_0, v_0\}$.

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dW_t, \;\; S_0 = s_0$$
$$dV_t = \kappa(\xi - V_t)dt + \eta \sqrt{V_t} dB_t, \;\; V_0 = v_0 \tag{3}$$
$$\langle dW_t, dB_t \rangle = \rho dt.$$

Here, μ is the drift under the physical measure $\mathbb{P}$, $\kappa > 0$ is the mean reversion speed for the variance, $\xi > 0$ is the mean reversion level for the variance, $\eta > 0$ is the volatility of the variance, $s_0 > 0$ is the initial asset price, $v_0 > 0$ is the initial level of the variance, and $\rho \in [-1, 1]$ is the correlation between the two standard Brownian motions W and B.

The Heston model described in System 3 is defined under the historical measure $\mathbb{P}$ and provides a reasonable dynamics for the volatility surface. It is a celebrated model because it reproduces several important features of low frequency price data, including leverage effect, time-varying volatility, and fat tails.

For free-arbitrage pricing, we rewrite process defined in System 3 under the risk-neutral measure $\mathbb{Q}$. In the Heston model, this is done by modifying each SDE in System 3 separately by an application of Girsanov's theorem.

First, the risk-neutral process for the stock price is:

$$dS_t = r S_t dt + \sqrt{V_t} S_t d\tilde{W}_t, \tag{4}$$

where

$$\tilde{W}_t = \left(W_t + \frac{\mu - r}{\sqrt{V_t}} t \right),$$

Here r is the riskless fixed interest rate. Notice that if we designate $\mu = r$ in System 3, the term $\frac{\mu - r}{\sqrt{V_t}}$ vanishes and $\tilde{W}_t = W_t$.

Using Itô's formula [2], the risk-neutral process of the log price therefore is

$$d \ln S_t = (r - 1/2 V_t)dt + \sqrt{V_t} d\tilde{W}_t. \tag{5}$$

Second, we introduce the volatility risk premium $\lambda(S, V, t) = \lambda v_t$ into the drift of dv_t in System 3, where λ is a constant. [3] The risk-neutral process for the variance is then obtained as follows:

$$dV_t = [\kappa(\xi - V_t) - \lambda V_t]dt + \eta \sqrt{V_t} d\tilde{B}_t, \tag{6}$$

---

[2] Derivation can be found in the Appendix 7.

[3] As explained in Heston (1993) and Breeden (1979), consumption model yields a premium proportional to the variance, so that $\lambda(S, V, t) = \lambda V_t$ for some constant λ.

where

$$\tilde{B}_t = (B_t + \frac{\lambda V_t}{\eta\sqrt{V_t}}t).$$

Here λ is a constant scaling parameter. Hence, we rewrite the risk-neutral Heston model as

$$
\begin{aligned}
dS_t &= rS_tdt + \sqrt{V_t}S_td\tilde{W}_t, \ \ S_0 = s_0 \\
dV_t &= \kappa^*(\xi^* - V_t)dt + \eta\sqrt{V_t}\tilde{B}_t, \ \ V_0 = v_0 \\
\langle d\tilde{W}_t, d\tilde{B}_t \rangle &= \rho dt,
\end{aligned}
\tag{7}
$$

where risk-neutral parameters $\kappa^* = \kappa + \lambda$ and $\xi^* = \kappa\xi/(\kappa + \lambda)$. Note that when $\lambda = 0$, we have $\kappa^* = \kappa$ and $\xi^* = \xi$ so that the parameters under the physical and risk-neutral measures are the same. That is:

$$\tilde{B}_t = B_t.$$

3.2   Rough Heston Model

In the Heston model, volatility follows a Brownian diffusion. But Gatheral et al. (2014) showed that in fact log-volatility time series behave essentially like a fractional Brownian motion, with Hurst parameter of order approximately 0.1, especially in high frequency data. The fBM is a generalization of the classical Brownian motion, but in the former, the increments are not independent.

Definition 3.1 (Fractional Brownian Motion (fBM))**.** The Fractional Brownian Motion (fBM) is a continuous-time centered self-similar Gaussian process indicated as $\{B_t^H; t \in \mathbb{R}\}$ on [0,T], which starts at zero and has expectation zero for all $t \in [0, T]$. It has stationary increments and covariance function:

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2}\{|t|^{2H} + |s|^{2H} - |t - s|^{2H}\},$$

where $H \in (0, 1)$ is called the Hurst parameter that describes the "roughness" of the motion. Classical Brownian motion is retrieved with $H = 1/2$.

The Rough Fractional Stochastic Volatility Model (RFSV) is proposed to address the roughness by modeling log-volatility dynamics as a fractional Brownian motion B_t^H with Hurst parameter $H < \frac{1}{2}$. RFSV models follow the following form:

$$
\begin{aligned}
dS_t &= \mu S_tdt + \sqrt{V_t}S_tdW_t \\
dV_t &= \lambda(V_t)dB_t^H \\
\langle dW_t, dB_t^H \rangle &= \rho dt.
\end{aligned}
\tag{8}
$$

8

Since the increments of fBM are no longer independent, the rough models are not Markovian and its variance process is no longer a semi-martingale. This makes it hard for to use Monte-carlo based techniques to simulate these models. Proposed by Euch and Rosenbaum (2017), rough Heston models aim to incorporate roughness into the classical Heston Models whilst also allowing tractable simulation of such process.

Euch and Rosenbaum (2017) modify the Heston model to its rough version with parameter $\theta = \{\mu, \kappa, \xi, \eta, \rho, \alpha, s_0, v_0\}$:

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dW_t, \quad S_0 = s_0$$
$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \kappa(\xi - V_s) ds$$
$$+ \frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} dB_s, \quad V_0 = v_0$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

(9)

where $\alpha \in (1/2, 1)$. The Hurst parameter in this case is $H = \alpha - 1/2$. Notice that when $\alpha = 1$, we recover the original vanilla Heston model in 3 We can rewrite system 9 as the following:

$$dS_t = \mu S_t dt + S_t \sqrt{V_t} dW_t, \quad S_0 = s_0$$
$$V_t = V_0 + \int_0^t K(t-s)[\kappa(\xi - V_s)ds + \eta\sqrt{V_s}dB_s],$$
$$V_0 = v_0$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

(10)

where the kernel function is

$$K(s) = \frac{1}{\Gamma(\alpha)} s^{\alpha-1}$$

(11)

with $\alpha \in (1/2, 1)$.

To see why system 10 incorporates similar effect of the fBM term in RFSV in 8, recall that a fractional Brownian motion B_t^H with Hurst parameter H can be built through the general Mandelbrot-van Ness representation given below:

$$B_t^H = \frac{1}{c_1(H)} \int_{\mathbb{R}} f_t(s) dW_s$$
$$f_t(s) = ((t-s)^+)^{H-\frac{1}{2}} - ((-s)^+)^{H-\frac{1}{2}}$$

(12)

and $c_1(H)$ is a scaling constant that is a function of H chosen appropriately to normalize the fractional Brownian motion. One example of Mandelbrot-van Ness

represented fractional Brownian motion is given by the following (Euch and Rosenbaum, 2017):

$$B_t^H = \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^{0} \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) dW_s$$
$$+ \frac{1}{\Gamma(H + \frac{1}{2})} \int_{0}^{t} (t-s)^{H-\frac{1}{2}} dW_s, \tag{13}$$

where the scaling constant

$$c_1(H) = \Gamma(H + \frac{1}{2}).$$

Restricting ourselves to horizon $[0, t]$, the fractional Brownian motion B_t^H simplifies to

$$B_t^H = \frac{1}{\Gamma(H + 1/2)} \int_0^t (t-s)^{H-1/2} dW_s$$
$$= \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} dW_s. \tag{14}$$

Note that the dependence of increments of fBM and the roughness it exhibits are driven by kernel $K(x)$ in equation 11. Roughness is introduced in a similar way in the rough Heston model as described in 9 by incorporating such kernel. In particular, as $\sqrt{V_t}$ is finite for all $t \geq 0$, we note that the component

$$\frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} dB_s$$

in System 9 is a $(\eta-)$scaled fractional Brownian motion B_t^H in equation 14, where we have $H = \alpha - 1/2$. In this way, roughness parameter $\alpha \in (1/2, 1)$ governs the smoothness of the volatility sample paths as the Hurst parameter $H \in [0, 1]$ governs the roughness of the path.

Roughness is introduced when $\alpha < 1$ (or equivalently, when the Hurst parameter $H < 1/2$ in fBM). And the stochastic process $(S_t)_{t \geq 0}$ described by System 9 is neither Markovian nor a semimartingale. However, by constructing a suitable sequence of the nearly unstabled Hawkes processes (which are semi-martingales) that converge to System 9, characteristic function of the rough Heston model can be determined in closed-form (Euch and Rosenbaum, 2017). Therefore, by defining a fractional BM in term of a standard BM instead of driving the volatility process by a fraction BM, one effectively circumvent the issue of non-Markovianity and non-semimartingularity that from which the naïve fractional BM models suffer.

To price assets without arbitrage under the rough Heston model, we similarly introduce the risk-neutral version, where we have assumed the same volatility premium function:

$$dS_t = rS_t dt + S_t \sqrt{V_t} d\tilde{W}_t, \quad S_0 = s_0$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \kappa^* (\xi^* - V_s) ds$$

$$+ \frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} d\tilde{B}_s, \quad V_0 = v_0$$

$$\langle d\tilde{W}_t, d\tilde{B}_t \rangle = \rho dt,$$

(15)

where

$$\tilde{W}_t = W_t + \frac{\mu - r}{\sqrt{V_t}}$$

$$\tilde{B}_t = B_t + \frac{\lambda V_t}{\eta \sqrt{V_t}}$$

for all $t \geq 0$ and $\kappa^* = \kappa + \lambda$ and $\xi^* = \kappa\xi/(\kappa + \lambda)$ as in 7.

3.3 SABR Model

The stochastic alpha-beta-rho (SABR) model is another popular stochastic volatility model. First proposed by Hagan et al. (2002), the celebrated SABR model is widely used for modeling the dynamics of forward prices. The original SABR model makes the arbitrage-free market assumption with zero risk-free rate, where the forward price do not exhibit a predictable trend (drift) when expressed in terms of a risk-neutral measure.

Following Yin (2015), we consider a slight generalization of the original SABR model by allowing the possibility of a deterministic trend in the forward price evolution. Specifically, we consider the SABR model describing the dynamics of a forward, F_t, by the following SDE:

$$dF_t = \mu F_t^\beta dt + V_t F_t^\beta dW_t, \quad F_0 = f_0$$

$$dV_t = \alpha V_t dB_t, \quad V_0 = v_0$$

$$\langle dW_t, dB_t \rangle = \rho dt,$$

(16)

where $\alpha \geq 0$ is the magnitude of volatility, $0 \leq \beta \leq 1$ is a parameter controlling the elasticity of variance, and $-1 \leq \rho \leq 1$ describes the correlation between the forward price and its volatility. Observe that 16 describes an asymptotically arbitrage-free Markovian process with no arbitrage dynamics of the implied volatility, which has zero drift. The conditions for which the model 16 is appropriate are given in Fukasawa and Gatheral (2021).

11

Let market fixed interest rate be r. The risk-neutral version of System 16 can be derived by adjusting the drift of the forward price process:

$$dF_t = rF_t^{\beta}dt + V_tF_t^{\beta}d\tilde{W}_t, \quad F_0 = f_0$$
$$dV_t = \alpha V_td\tilde{B}_t, \quad V_0 = v_0 \tag{17}$$
$$\langle d\tilde{W}_t, d\tilde{B}_t \rangle = \rho dt,$$

where the risk-neutral Brownian motions are

$$d\tilde{W}_t = W_t + \frac{\mu - r}{V_t}dt$$
$$d\tilde{B}_t = B_t. \tag{18}$$

For the sake of simplicity and considering that SABR model models forward prices, which in reality are calculated by discounting w.r.t. the given market interest rate, we will set $\mu = r$ to work directly with the SABR model under an equivalent martingale measure.

To recover the underlying stock price process, we recall that

$$F_t = S_t e^{rt}. \tag{19}$$

Then under the risk-neutral SABR model, the stock price dynamics is described by the following SDE:

$$dS_t = re^{-\beta rt}S_t^{\beta}dt + V_te^{-\beta rt}S_t^{\beta}dW_t, \quad S_0 = e^{-rt}f_0$$
$$dV_t = \alpha V_tdB_t, \quad V_0 = v_0 \tag{20}$$
$$\langle dW_t, dB_t \rangle = \rho dt.$$

A closed-form solution for the implied volatility smile exists for this model, however issues include inconsistency with market data and inability to reproduce term structures of the power-law type for typically observed at-the-money markets (Fukasawa and Gatheral, 2021).

3.4 Rough SABR Model

Fukasawa and Gatheral (2021) adapt the SABR model to be a rough volatility model where the volatility has a rougher path than a semi-martingale. Specifically, the rough SABR model derived in Fukasawa and Gatheral (2021) presents a generalized version that includes both the SABR model and the rough Bergomi model proposed by Bayer et al. (2016) as particular cases.

The proposed rough SABR model under an equivalent martingale measure $\mathbb{Q}$ is governed by the following dynamics:

$$dF_t = V_t\beta(F_t)dW_t, \quad F_0 = f_0$$
$$d\xi_t(s) = K(s - t)\xi_t(s)dB_t, \quad t < s, \quad \xi_0(0) = v_0^2 \tag{21}$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

where

$$V_t = \sqrt{\xi_t(t)}$$
$$K(t) = \eta\sqrt{2H}t^{H-1/2}, \;\; \eta > 0, \; H \in (0, 1/2] \tag{22}$$

and

$$\xi_t(s) = \mathbb{E}^{\mathbb{Q}}[\xi_s(s)|\mathcal{F}_t]$$
$$= \xi_0(s)\exp\{\eta\sqrt{2H}\int_0^t (s-u)^{H-1/2}dB_u \tag{23}$$
$$- \frac{1}{2}\eta^2(s^{2H} - (s-t)^{2H})\}, \;\; 0 \le t \le s.$$

Here, correlation coefficient $\rho \in [-1, 1]$, $\beta(\cdot)$ is a positive continuous function, $\{\xi_0(s)\}_{s\ge 0}$ is a family of $\mathcal{F}_0$ (filtration at time 0) measurable random variables, and the curve $s \mapsto \xi_0(s)$ is continuous.

To allow for more flexibility of possible drifts in the forward price process, one may similarly consider a slight general form of the rough Heston model in 21 by including a drift term in the forward process as the following:

$$dF_t = \mu\beta(F_t)dt + V_t\beta(F_t)dW_t, \;\; F_0 = f_0$$
$$d\xi_t(s) = K(s-t)\xi_t(s)dB_t, \;\; t < s, \; \xi_0(0) = v_0^2 \tag{24}$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

where $V_t, K(t), \xi_t(s)$ are defined in the same way as defined in 22-23. In particular, System 24 is risk-neutral when we set $\mu = r$. But for the sake of simplicity, we will work with model 21 .[4]

For clarity, we write out the corresponding underlying stock price process following the risk-neutral Rough SABR:

$$dS_t = r\beta(e^{-rt}S_t)dt + V_t\beta(e^{-rt}S_t)dW_t, \;\; S_0 = e^{-rt}f_0$$
$$d\xi_t(s) = K(s-t)\xi_t(s)dB_t, \;\; t < s, \; \xi_0(0) = v_0^2 \tag{25}$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

Notice that the generalization that the rough Heston model in 24 offers is achieved through the positive continuous function $\beta(\cdot)$ and through modeling volatility V_t as a function of the family of $\mathcal{F}_0-$measurable random variables $(\xi_t(s))_{s\ge 0}$. For instance, if we consider the below positive continuous function (since we assume by convention that $F_t \ge 0$ for all $t \ge 0$)

$$\beta(F_t) = F_t^\beta, \;\; 0 \le \beta \le 1$$

---

[4]The risk neutral version of Model 24 can be derived in a similar manner as the risk-neutral naïve SABR model. The specific form of the risk-neutral rough SABR model can be found in the Appendix 7.

and we require that

$$\xi_0(s) = \alpha_0^2 \exp\{\frac{1}{4}\eta^2 s\},$$

then when $H = 1/2$, we recover the naïve SABR model in 16, where $\eta = 2\alpha$ (α as defined in model 16). Derivation of this result can be found in the Appendix 7. In addition, when $\beta(F_t) = F_t$, we recover the rough Bergomi model introduced in Bayer et al. (2016).

The the family of $\mathcal{F}_0-$measurable random variables $(\xi_t(s))_{s \geq 0}$ has nice financial interpretations. The mapping $s \mapsto \xi_t(s)$ is the forward variance curve at time t:

$$\int_t^s \xi_t(u)ds = \int_t^s \mathbb{E}^{\mathbb{Q}}[\alpha_u^2|\mathcal{F}_t]du$$
$$= \mathbb{E}^{\mathbb{Q}}[d\langle \log S \rangle_u|\mathcal{F}_t],$$

and ξ can be determined from weighted variance swap rates:

$$\int_t^s \xi_t(u)ds = \mathbb{E}^{\mathbb{Q}}[\int_t^s \frac{S_u^2}{\beta(S_u)^2}d\langle \log S \rangle_u|\mathcal{F}_t]$$

On a higher level, as in the rough Heston model in 9, roughness is brought by the kernel $K(s-t)$. In the following, we will show in detail how and why roughness is introduced.

First, let us re-represent the volatility SDE in the SABR model 16. Since the volatility follows a log-normal martingale, we can apply the Ito's formula. Consider $f(V_t) = \ln(V_t)$. Then:

$$d(\ln(V_t)) = \frac{1}{V_t}dV_t + \frac{1}{2}(-\frac{1}{V_t^2}\langle dV \rangle_t)$$
$$= \frac{1}{V_t}(\alpha V_t dB_t) + \frac{1}{2}(-\frac{1}{V_t^2}(\alpha^2 V_t^2 dt)$$
$$= \alpha dB_t - \frac{1}{2}\alpha^2 dt,$$

where we have used that

$$\langle dV \rangle_t = \langle \alpha V_t dB_t \rangle = \alpha^2 V_t^2 \langle dB \rangle_t = \alpha^2 V_t^2 dt.$$

It follows that

$$\ln(V_t) = \int_0^t d(\ln(V_t)) = \int_0^t \alpha dB_t - \frac{1}{2}\alpha^2 dt$$
$$= \ln(V_0) + \alpha B_t - \frac{1}{2}\alpha^2 t.$$

14

Taking the exponential gives

$$V_t = v_0 \exp\{\alpha B_t - \frac{1}{2}\alpha^2 t\}. \tag{26}$$

Motivated by the fact that log-volatility behaves like a fractional Brownian Motion with Hurst exponent H of order 0.1 at any reasonable time scale (Gatheral et al., 2014), the rough SABR model in 21 essentially extends the class of naïve SABR models defined in 26 via the following:

$$V_t = V_0 \exp\{\gamma B_t^H - \frac{1}{2}\mathbb{V}\mathrm{ar}B_t^H\}, \tag{27}$$

for some constant factor γ ($\gamma = 1/2$ in our case), where

$$B_t^H = \eta\sqrt{2H} \int_0^t (s-u)^{H-1/2} dB_u \tag{28}$$

is a fractional Brownian motion with Hurst parameter H and can be obtained from the Mandelbrot-Van-Ness formula 12 by setting the scaling constant

$$\frac{1}{c_1(H)} = \eta\sqrt{2H}. \tag{29}$$

A more detailed derivation can be found in the Appendix 7.

This rough SABR model proposed by Fukasawa and Gatheral (2021) allows an explicit asymptotic arbitrage-free approximation of the implied volatility under various models, such as the rough Bergomi model and the celebrated Black-Scholes model. Furthermore, these models more closely fit the volatility surface and require less parameters than the standard SABR method. For instance, empirical evidence suggests that for a large range of time to expiry, the at-the-money (ATM) volatility skew is of the form

$$\psi(\tau) = C\tau^{-\alpha},$$

where C is a constant and $0 < \alpha < \frac{1}{2}$. The rough SABR model 21, through using fractional Brownian motion in log-volatility 27, is able to capture this form in the model as the following

$$\psi(\tau) = C\tau^{H-\frac{1}{2}},$$

whereas the naïve SABR model cannot.

In addition, the empirical results of Fukasawa and Gatheral (2021) show that Model 24 more accurately approximates the prices for higher values of H, and that the scaling of $y(k,\tau)$ offsets the normalized smile's dependence on maturity. It is suggested that the H can be used to parameterize the classical SABR model for time dependent parameters implied by the market.

4 Variance Reduction Techniques

Monte Carlo estimation results in high variance, requiring variance reduction techniques to make this method more computationally efficient. A variety of established techniques exist for variance reduction, including antithetic variates, control variates and endpoint stratification, and the diffusion operator integral (DOI) method.

4.1 Diffusion Operator Integral (DOI)

The diffusion operator integral (DOI) method, originally proposed by Heath and Platen (2002), is related to Itô integral representation methods and measure transformation methods for variance reduction. While this method is less commonly utilized in comparison to its counterparts, it is advantageous given its applicability to a broader range of valuation problems and its ability to improve most parabolic partial differential equations (PDEs). Heath and Platen find that application of this method to the Heston model results in a drastic reduction in variance of 23000 times that achieved by standard antithetic variate Monte Carlo estimation.
Heath and Platen propose that for the SDE

$$dX_t^d = a(t, X_t^d)dt + \sum_{j=1}^{m} b^j(t, X_t^d)dW_t^j, \tag{30}$$

where $\{X_t^d\}_{t \in [\tau, T], \tau \in [0, T]}$ is a general d-dimensional diffusion process with initial value $X_\tau^d = x^d \in \Gamma$. Here the drift coefficient $a : [0, T] \times \Gamma \to \Re^d :$ and diffusion coefficient $b^j : [0, T] \times \Gamma \to \Re^d, j \in \{1, 2, \ldots, m\}$, satiesfy appropriate conditions so that 30 has a unique strong solution and is Markovian, seeKloeden and Platen (2008).
Consider the payoff function

$$u(t, x) = E(h(T, X_T^d) | X_t^d) \tag{31}$$

Here h satisfies appropriate integrability conditions so that the process $\{M_t\}_{t \in [0, T]}$ with $M_t = E(h(T, X_T^d) | \mathcal{F}_t)$ is a square integrable $(\mathcal{F}_t, \mathbb{P})$-martingale. Using the martingale representation theorem, together with the Markov property for X, it can be inferred that there exists an m-dimensional $\xi = \{\xi_t = (\xi_t^1, \ldots, \xi_t^m)^\top, t \in [0, T]\}$ with $M_t = u(t, X_T^d) = u(0, x) + \sum_{j=1}^{m} \int_0^T \xi_s^j \, dW_s^j$
Our aim will be to find an unbiased variance-reduced estimator for $u(0, x)$ given an approximation function $\bar{u}$.
Applying Itô formula and by martingale property of M_t, refer to Heath and Platen

(2002), the approximation to the value of the option is as follows:

$$\bar{u}(t, X_t^d) = \bar{u}(0, x) + \int_0^\tau L^0 \bar{u}(t, X_\tau^d) dt +$$
$$\sum_{j=1}^m \int_0^t L^j \bar{u}(t, X_t^d) dW_\tau^j, \tag{32}$$

where the operators L^0 and L^j are

$$L^0 f(t, x) = \frac{\delta f}{\delta t}(t, x) + \sum_{i=1}^d a^i(t, x) \frac{\delta f}{\delta x^i}(t, x)$$
$$+ \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t, x) b^{k,j}(t, x) \frac{\delta^2 f}{\delta x^i \delta x^k}(t, x), \tag{33}$$

$$L^j f(t, x) = \sum_{i,k=1}^d b^{i,j}(t, x) \frac{\delta f}{\delta x^i}(t, x), \tag{34}$$

These operations apply to functions f which are sufficiently smooth and $(t, x) \in (0, T) \times \Gamma$. The unbiased estimator for the option price, $u(0, x)$, is the following random variable named the DOI estimator

$$\bar{Z}_\tau = \bar{u}(0, x) + \int_0^\tau L^0 \bar{u}(t, X_\tau^d) dt. \tag{35}$$

$\bar{Z}_\tau$ will have a small variance if $L^j u$ is close to $L^j \bar{u}$ which is the case if $\bar{u}$ is a good approximation of u. This technique can be applied to parabolic PDE approximation, however to approximate $\bar{u}$ the majority of applications of this method utilize a diffusion process approximation whose dynamics are assumed to be Markovian (Heath and Platen, 2002). A convenient chose for this diffusion process approximation is the Black-Scholes type model. For sufficiently smooth $\bar{u}$, $\bar{L}^0 \bar{u}(t, x) = 0$, so the value function can be adjusted as follows

$$u(0, x) = \bar{u}(0, x) + E\left(\int_0^\tau L^0 \bar{u}(t, X_\tau^d) dt\right),$$
$$= \bar{u}(0, x) + E\left(\int_0^\tau (L^0 - \bar{L}^0) \bar{u}(t, X_\tau^d) dt\right), \tag{36}$$

where the DOI estimator shown in 37 have a small variance if the approximation to the operator and L^0 are close.

$$\bar{Z}_\tau = \bar{u}(0, x) + \int_0^\tau (L^0 - \bar{L}^0) \bar{u}(t, X_\tau^{0,x}) dt. \tag{37}$$

Now we describe a DOI variance reduction technique which can be utilized to approximate the option price

$$c(0, x) = e^{-rT} u(0, x), \tag{38}$$

where

$$u(0, x) = \tilde{E}((\frac{1}{T} \int_0^t S_\tau d\tau - K)^+). \tag{39}$$

Recall the risk-neutral dynamics of the Heston model 7, we introduce I_t to represent path dependency of Asian option,

$$I_t = \int_0^t S_\tau d\tau, \tag{40}$$

then the option price can be written as

$$c(0, x) = e^{-rT} \tilde{E}((\frac{I_t}{T} - K)^+). \tag{41}$$

Consider the three-dimensional diffusion process $X_t = (S_t, V_t, I_t)$ and the two-dimensional Wiener process $Z_t = (\tilde{W}_t, \tilde{B}_t)$. For many practical applications, a systematic way of obtaining an approximation $\bar{u}$ satisfying 39 is to first find an approximation $\bar{X}_t$ to the diffusion process X_t. Then by utilizing the Greeks given by the Black-Scholes model, we derive the DOI estimator for the Asian option price.

In summary, the DOI variance reduction technique relies on finding an approximated underlying price dynamics process, denoted $\bar{S}_t$, which sets the volatility term of the stochastic volatility SDE of the original model to 0. This strategy follows the following three-step procedure:

- Derive the PDE to obtain equality condition.

- Derive expression for $\bar{\sigma}_t$.

- Derive the Greeks based on the above steps.

In the below, we develop explicit formulation for the Greeks in the DOI variance reduction method for each of our four models in pricing Arithmetic Asian options.

4.1.1 Deriving DOI Greeks for Heston and Rough Heston Models

Based on the Black-Scholes model, a convenient choice of diffusion process is as follows: let $\bar{X}_t = (\bar{S}_t, \bar{V}_t, \bar{I}_t)$ be the three-dimensional diffusion process which satisfies the SDE

$$\begin{aligned} d\bar{S}_t &= r\bar{S}_t dt + \sqrt{\bar{V}_t} \bar{S}_t d\tilde{W}_t, \quad \bar{S}_\tau = s_\tau \\ d\bar{V}_t &= \kappa^*(\xi^* - \bar{V}_t)dt, \quad V_\tau = v_\tau \\ d\bar{I}_t &= \bar{S}_t dt \end{aligned} \tag{42}$$

for $t \in [\tau, T]$ and $\tau \in [0, T]$, where the price process of the underlying asset follows the Geometric Brownian Motion. For this system of SDEs the solution $\bar{V}$ can be explicitly computed by dividing $\xi^* - \bar{V}_t$ on both side and is given by

$$\bar{V}_t = \xi^* + (v_\tau - \xi^*)\, e^{-\kappa(t-\tau)}. \tag{43}$$

Now consider the pricing equation for the Fixed Strike Arithmetic Asian option

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + S\frac{\partial V}{\partial I} + rS\frac{\partial V}{\partial S} - rV = 0, \tag{44}$$

by omitting the term $S\frac{\partial V}{\partial I}$, we have the Black-Scholes equation for a European call option, denoted by $BS(S_\tau, K, r, \sigma, T)$ with spot price S_τ, short rate r, volatility σ and maturity T. We utilize the explicit form of European call option to give the approximate function $\bar{u}$ in 39:

$$\begin{aligned}
\bar{u}(t, x) &= \tilde{E}((\frac{1}{T}\int_0^t \bar{S}_\tau d\tau - K)^+) \\
&\approx e^{r(T-t)} BS(S_t, K, r, \bar{\sigma}_t, T - t),
\end{aligned} \tag{45}$$

where

$$\begin{aligned}
\bar{\sigma}_t &= \sqrt{\frac{1}{T-t}\int_t^T \bar{V}_z dz} \\
&= \sqrt{\xi^* - (V_t - \xi^*)\frac{e^{-\kappa(T-t)} - 1}{\kappa(T-t)}}.
\end{aligned} \tag{46}$$

We derive the exact form of $\bar{\sigma}$ as in 46 for implementation in the Appendix 7. Evaluation of the expectation appearing in 37 requires initial value $\bar{u}(0, x)$ and the calculation of the values $(L^0 - \bar{L}^0)\bar{u}(t, X_t)$. For the initial value $\bar{u}(0, x)$, we apply crude Monte-Carlo method to compute the Asian option price under the diffusion process.

Based on 33 and the explicit form of $\bar{\sigma}_t$, we have

$$\begin{aligned}
(L^0 - \bar{L}^0)\bar{u}(t, x) &= \eta V_t e^{r(T-t)} \\
&\times \Bigg[\rho \frac{\partial^2 BS}{\partial S_t \partial \bar{\sigma}_t}(S_t, K, r, \bar{\sigma}_t, T - t)\frac{\partial \bar{\sigma}_t}{\partial V_t} \\
&+ \frac{1}{2}\eta \Bigg\{ \frac{\partial^2 BS}{\partial \bar{\sigma}_t^2}(S_t, K, r, \bar{\sigma}_t, T - t) \Bigg) \left(\frac{\partial \bar{\sigma}_t}{\partial V_t}\right)^2 \\
&+ \frac{\partial BS}{\partial \bar{\sigma}_t}(S_t, K, r, \bar{\sigma}_t, T - t)\frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2} \Bigg\} \Bigg].
\end{aligned} \tag{47}$$

19

Now we compute the partial derivatives $\frac{\partial BS}{\partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial S_t \partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial \bar{\sigma}_t^2}$, $\frac{\partial \bar{\sigma}_t}{\partial V_t}$ and $\frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2}$ using the Black-Scholes formula together with the expression for $\bar{\sigma}_t$ in 46. Then we have

$$\frac{\partial \bar{\sigma}_t}{\partial V_t} = \frac{1}{2}\bar{\sigma}_t^{-1}\frac{1 - e^{-\kappa(T-t)}}{\kappa(T-t)},$$

$$\frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2} = -\frac{1}{2}\frac{\partial \bar{\sigma}_t}{\partial V_t}\bar{\sigma}_t^2$$

For Rough Heston Model, consider the following diffusion process

$$d\bar{S}_t = r\bar{S}_t dt + \sqrt{\bar{V}_t}\bar{S}_t d\tilde{W}_t, \quad \bar{S}_\tau = s_\tau$$

$$d\bar{V}_t = \bar{V}_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa(\xi - V_s)ds \tag{48}$$

$$d\bar{I}_t = \bar{S}_t dt$$

for $t \in [\tau, T]$ and $\tau \in [0, T]$ Based on the same process, we need to find the explicit expression for $\bar{\sigma}_t$

4.1.2 Deriving DOI Greeks for SABR and Rough SABR Models

Recall the SABR model 17, we choose the mean-reverting process as the $\bar{V}_t$ in the diffusion process

$$d\bar{F}_t = r\bar{F}_t^\beta dt + V_t\bar{F}_t^\beta d\tilde{W}_t, \quad \bar{F}_0 = f_0$$
$$d\bar{V}_t = (V_\tau - V_t)dt, \quad V_\tau = v_\tau, \tag{49}$$

which makes sure that the forward price process follows a Geometric Brownian Motion, and $\mathbb{E}[V_t|\mathcal{F}_\tau] = V_\tau$. The Greeks can be derived based on previous discussion.

4.2 Antithetic Variates

The general integral

$$C_A(fz) = \int_A f(x)z(x)dx, \tag{50}$$

where $A \subseteq \mathbb{R}^k$ and z is a probability density function with $supp(z) \subseteq A$, can be estimated using Monte Carlo integration. The estimate is denoted by $\hat{C}_A(fz)$. Antithetic variates is a variance reduction technique used to improve Monte Carlo estimates that involves using each element from a sequence of random variates more than once. This allows a sample of n random variates to become a sample of size $2n$ using the fact that, if density z is symmetric about its mean, then X and

20

$2\mathbb{E}[X] - X$ are both distributed according to z (Glasserman, 2004).Here $2\mathbb{E}[X] - X_1, \ldots, 2\mathbb{E}[X] - X_n$ is referred to as the antithetic sample. Let $\hat{C}_{A,n}^{-}(fz)$ denote the Monte Carlo estimate using the new sample (of size n). It is clear that this is an unbiased estimate. Let $\hat{C}_{A,n}^{+-}(fz)$ denote the Monte Carlo estimate using both the antithetic and original samples. The improved Monte Carlo estimate is calculated as follows,

$$\hat{C}_{A,n}^{+-}(fz) = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i) + f(2\mathbb{E}[X] - X_i)}{2}. \tag{51}$$

It can be deduced that the variance of the new estimate is always smaller than or equal to the original crude Monte Carlo estimate. It is also clear that the original and antithetic sample are not independent and the variance of the new Monte Carlo estimate depends on the correlation between the two samples. The issue with comparison however is that $\hat{C}_{A,n}^{+-}(fz)$ requires $2n$ evaluations of $f(x)$, hence for the technique to produce an improved estimate we require $Var[\hat{C}_{A,n}^{+-}(fz)] \leq Var[\hat{C}_{A,2n}(fz)]$. This can be proved true with the aid of Chebyshev's Monotone Convergence Inequality (Glasserman, 2004). This method has been shown to be effective. For example, Broadie et al. (1997) find that antithetic branching alone is highly effective in reducing variance for American option price modelling, however the result is improved by using a European option price as a control variate.

4.3 Control Variate

Control variates is another variance reduction technique that leverages known information to improve the Monte Carlo estimate of an integral. The estimate is calculated as

$$\hat{C}_{A,n}^{*}(fz) = \hat{C}_{A,n}(fz) + \alpha\left(\hat{C}_{A,n}(gz) - C_A(gz)\right). \tag{52}$$

In this equation, it is assumed that $C_A(gz)$ can be evaluated exactly and $\hat{C}_{A,n}(gz)$ is the Monte Carlo estimate using the same sample used to estimate $\hat{C}_{A,n}(fz)$. The optimal alpha α^{opt} chosen in order to minimize the variance turns out to be

$$\alpha^{opt} = -\frac{\mathbb{C}ov[f(X), g(X)]}{\mathbb{V}ar[g(X)]}. \tag{53}$$

This method is often used because, provided $\mathbb{C}ov[f(X), g(X)] \neq 0$ its variance is strictly less than the crude estimator (Glasserman, 2004). The statistics required to obtain α^{opt} are usually estimated using a small sub-sample ($n_0 \ll n$) of the simulated data and thus α^{opt} can be estimated as

$$\hat{\alpha}_{n_0}^{opt} = -\frac{\hat{\mathbb{C}ov}_{n_0}[f(X), g(X)]}{\hat{\mathbb{V}ar}_{n_0}[g(X)]}. \tag{54}$$

The variance of this method can be calculated as

$$\mathbb{V}ar\left[\hat{C}^{*}_{A,n}(fz)\right] = \mathbb{V}ar\left[\hat{C}_{A,n}((f+\alpha g)z)\right]. \tag{55}$$

4.4 Endpoint Stratification

Stratified sampling is a further variance reduction method which limits the fraction of observations drawn from partitions of the sample space. These partitions satisfy supp$(z) = A = \cup_{i=1}^{d} A_i$ and $\forall \leq i \neq j \leq d, A_i \cap A_j = \emptyset$, where z is typically the uniform density. The Monte Carlo estimate for stratified sampling is as follows

$$\hat{C}^{\equiv}_{\overline{A},n}(fz) = \Sigma_{i=1}^{d} p_i \frac{1}{n_i} \Sigma_{j=1}^{n_i} f(x_{ij}), \tag{56}$$

where p_i represents the probability of being in the ith partition. $\hat{C}^{\equiv}_{\overline{A},n}(fz)$ is an unbiased estimator of the crude Monte Carlo estimate, and can only produce a lower variance than crude Monte Carlo given proportional stratification which can be proved via Jensen's inequality (Glasserman, 2004). The simplest form of this method is proportional sampling, however path dependent derivatives are better modelled using terminal, or endpoint, stratification. Endpoint stratification utilizes the fact that fixing the terminal value and generating random paths for Brownian motion using multivariate conditioning leads to lower variance of the Monte Carlo estimate of the option price. (Glasserman, 2004). Terminal stratification entails generation of a discretized Brownian path through Brownian bridges and the inverse transform method to generate intermediate values in each strata. This method can however be computationally expensive given the need to utilize Brownian bridges. The variance of this method can be calculated as

$$\mathbb{V}ar\left[\hat{C}^{\equiv}_{\overline{A},n}(fz)\right] = \sum_{i=1}^{d} p_i^2 \mathbb{V}ar\left[\hat{C}_{A_i,n_i}(fz)\right] \tag{57}$$

5 Numerical Simulation

Approximation of stochastic differential equations requires discretization of the time grid, with the simplest method being the Euler, or Euler-Maruyama, scheme (Glasserman, 2004). Let $(\pi_n)_{n \geq 1}$ denote a sequence of discrete-time grids on $[0, T]$, with $\pi_n = \{0 = t_0^n < t_1^n < t_2^n < ... < t_n^n = T\}$ for each $n \geq 1$. Denote $\delta_n := |\pi_n| := \max_{0 \leq k \leq n-1} \Delta t_{k+1}^n$, with $\Delta t_{k+1}^n := t_{k+1}^n - t_k^n$. In below we write t_k^n as t_k for simplicity.

5.1 Discretization of the Heston and SABR models

For SDEs of the form

$$dX_t = a(X_t)dt + b(X_t)dW(t), \tag{58}$$

22

where X_0 is a constant the conditions for uniqueness and existence of a strong SDE solution are satisfied by a and b. For $i = 0, \ldots, n$, $0 = t_0 < \cdots < t_n$ and $\hat{X}_0 = X_0$, the approximation for the one-dimensional case is as follows

$$\hat{X}_{t_i} = \hat{X}_{t_{i-1}} + a(\hat{X}_{t_{i-1}})\Delta t + b(\hat{X}_{t_{i-1}})\sqrt{\Delta t}Z_i, \tag{59}$$

where Z are independent standard normal random numbers and $\Delta t = t_i - t_{i-1}$. This approximation scheme has strong and weak orders of convergence of $\frac{1}{2}$ and 1 respectively (Glasserman, 2004). Now we give the Euler-type scheme for 7. Notice that the process V_t is $\mathbb{R}_+$-valued in the continuous-time setting, but it could become negative in a discrete-time simulation. For this reason, we use $(V_t)_+ := \max(V_t, 0)$ in the square root term $\sqrt{(V_t)_+}$ to define the discrete-time scheme.

Remark 5.1. Let us define $Y_t = lnS_t$, then one has

$$Y_t = Y_0 + \int_0^t (r - \frac{1}{2}V_s)ds + \int_0^t \sqrt{V_s} \, d\tilde{W}_t,$$

in the formulation 15

In view of remark 5.1, we would like to simulate the process Y in place of S in 15. As observed in the Black-Scholes model, the simulation of Y permits to avoid the time discretization of the process S in the dynamics of S, and one can expect a better performance for its simulation(Richard et al., 2022).
The approximation for the Heston model is

$$\begin{aligned}
V_{t_i}^n &= V_{t_{i-1}}^n + \kappa(\tilde{\theta} - V_{t_{i-1}}^n)\Delta t + \epsilon\sqrt{(V_{t_{i-1}}^n)_+}\Delta\tilde{W}_{t_i}, \\
Y_{t_i}^n &= Y_{t_{i-1}}^n + (r - \frac{1}{2}V_{t_{i-1}})\Delta t + \sqrt{(V_{t_{i-1}}^n)_+}\Delta\tilde{B}_{t_i}, \\
S_{t_i}^n &= \exp(Y_{t_i}^n),
\end{aligned} \tag{60}$$

where

$$\begin{aligned}
\Delta\tilde{W}_{t_i} &= \Delta W_{t_i} + \frac{\mu - r}{\sqrt{(V_{t_i}^n)_+}}, \\
\Delta\tilde{B}_{t_i} &= \rho\Delta\tilde{W}_{t_i} + \sqrt{1 - \rho^2}\Delta B_{t_i},
\end{aligned} \tag{61}$$

and ΔW_{t_i} and ΔB_{t_i} are normally distributed random numbers with mean zero and variance Δt.
The approximation for the SABR model is

$$\begin{aligned}
S_{t_i}^n &= S_{t_{i-1}}^n + (S_{t_{i-1}}^n)^\beta r\Delta t + (S_{t_{i-1}}^n)^\beta V_{t_i}^n\Delta W_{t_i}, \\
Y_{t_i}^n &= Y_{t_{i-1}}^n - \frac{1}{2}\alpha^1\Delta t + \alpha\tilde{B}_{t_i}, \\
V_{t_i}^n &= \exp(Y_{t_i}^n),
\end{aligned} \tag{62}$$

23

where W_{t_i} and B_{t_i} are correlated by ρ.

In both cases, maturities of 0.5 and 1 year were considered where the period was broken down into $N = 104$ and $N = 52$ time intervals respectively. Tables 1 and 2 contain the parameters used when simulating the Heston and SABR models respectively, each simulated using $M = 100000$ stock price paths and initial values.

Table 1: Parameter values for Heston Model

T	N	M	S_0	r	V_0
1.0(0.5)	104(52)	100000	100	0.04	0.2

κ	θ	η	μ	ρ	α
2.0	0.2	0.5	0.04	-0.15	1

Table 2: Parameter values for SABR Model

T	N	M	S_0	r	V_0
1.0(0.5)	104(52)	100000	100	0.04	0.2

β	α	ρ			
1.0	0.4	-0.05			
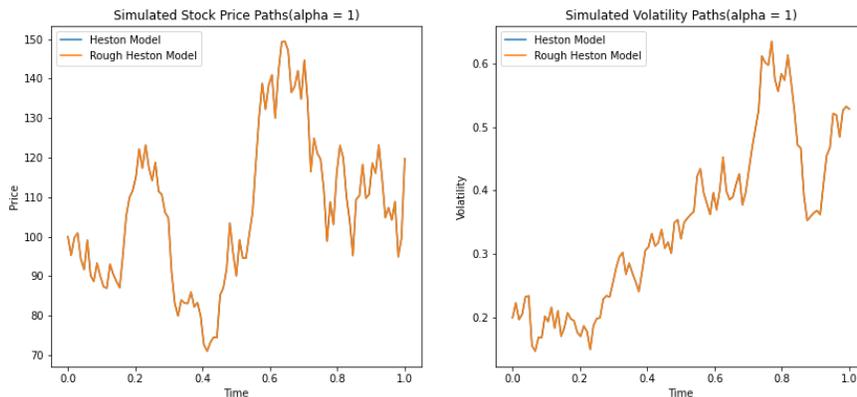
Figure 1: Heston and Rough Heston Model(alpha = 1)

5.2 Discretizing Roughness Kernel

Discretizing the integral approximation of the volatility process in 9 is a crucial and intricate step in numerically simulating SDEs in the rough Heston model. This
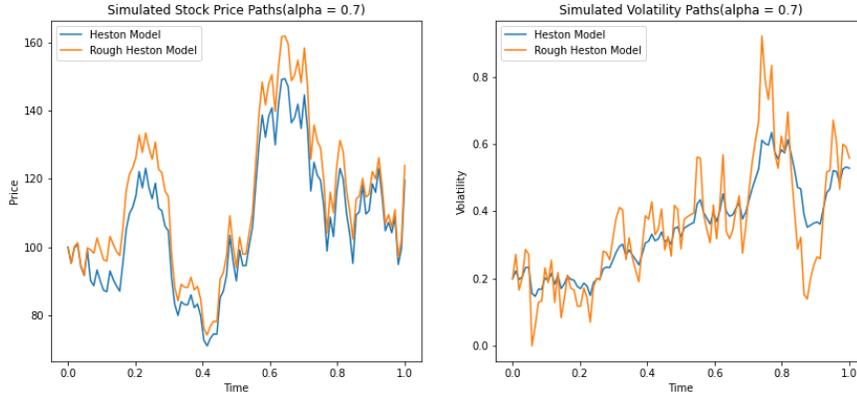
Figure 2: Heston and Rough Heston Model(alpha = 0.7)
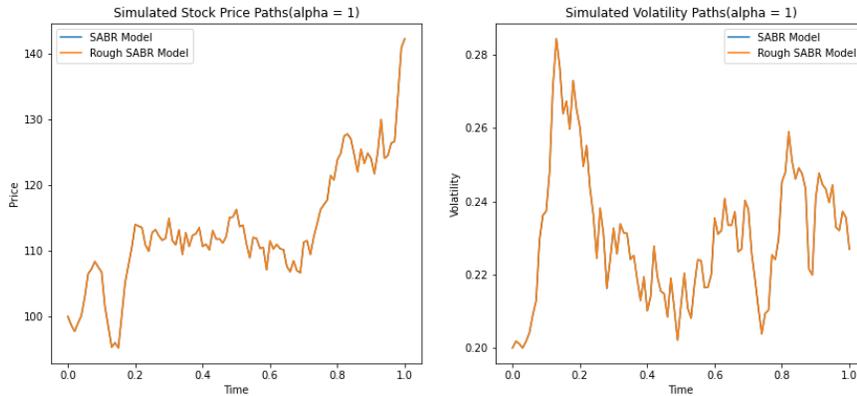
Figure 3: SABR and Rough SABR Model(alpha = 1)

process involves converting continuous-time integrals into discrete sums, making it possible to perform numerical simulations on a computer.

Now we give the Euler-type scheme for 15. Notice that the process V_t is $\mathbb{R}_+$-valued in the continuous-time setting, but it could become negative in a discrete-time simulation. For this reason, we use $(V_t)_+ := \max(V_t, 0)$ in the square root term $\sqrt{(V_t)_+}$ to define the discrete-time scheme.

Denote by $(S^n, V^n) = (S^n_{t_k}, V^n_{t_k})_{k=0,1,\cdots,n}$ the corresponding numerical solution,
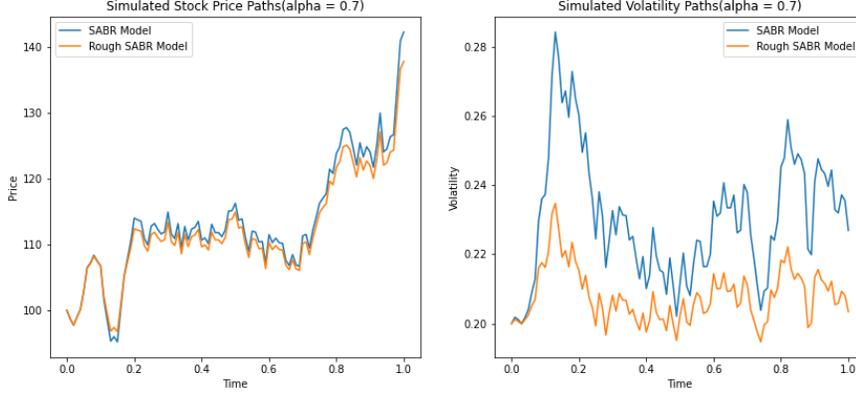
Figure 4: SABR and Rough SABR Model(alpha = 0.7)

which is given as $S_{t_k}^n := \exp(Y_{t_k}^n)$, $k = 0, 1, \cdots, n$, and

$$
\begin{aligned}
Y_{t_k}^n &= Y_0 + \sum_{i=0}^{k-1} \Big(r\Delta t_{i+1}^n - \frac{1}{2}(V_{t_i}^n)_+ \Delta t_{i+1}^n \\
&\quad + \sqrt{(V_{t_i}^n)_+} \Delta\tilde{W}_{t_{i+1}} \Big), \\
V_{t_k}^n &= V_0 + \sum_{i=0}^{k-1} \Big(K(t_k - t_i)\big(\xi - (V_{t_i}^n)_+\big)\Delta t_{i+1}^n \\
&\quad + K(t_k - t_i)\eta\sqrt{(V_{t_i}^n)_+}\Delta\tilde{B}_{t_{i+1}} \Big).
\end{aligned}
\tag{63}
$$

Similarly, the approximation for the rough SABR model in system 21 is

$$
\begin{aligned}
S_{t_i}^n &= S_{t_{i-1}}^n + \mu\beta(S_{t_{i-1}}^n)\Delta t + \sigma_{t_k}\beta(S_{t_{i-1}}^n)\Delta\tilde{W}_{t_{i+1}} \\
\xi_{t_k}^n(t_k) &= \xi_0^n(t_k) + \sum_{i=0}^{k-1} K(t_k - t_i)\xi_{t_i}^n(t_k)\Delta\tilde{B}_{t_{i+1}} \\
\sigma_{t_k} &= \sqrt{\xi_{t_k}^n(t_k)}
\end{aligned}
\tag{64}
$$

where

$$
\xi_0(t_k) = \alpha_0^2 \exp\{\frac{1}{4}\eta^2 t_k\},
$$

Figure 1, 2, 3, 4 show the trajectory of different models. By setting the seed for the randomness, when $\alpha = 1$, the rough heston/SABR model will recover the classical heston/SABR model.

26

5.3 Monte-Carlo Pricing Algorithms

We now introduce the different Monte-Carlo pricing algorithms using the four different variance reduction methods. The general crude MC pricing algorithms for arithmetic Asian options (without variance reduction) with a particular strike k and maturity T is given by Algorithm **??**.

Algorithm 1 Generic Crude Monte Carlo Arithmetic Asian Options Pricing Algorithm

---

1: **Input:** S_0, k, T, r, $\theta_{\mathbf{M}}$, N, M, **Model**
2: **Output:** Option price C
3: Initialize $C \leftarrow 0$
4: **for** $i = 1$ to M **do**
5: $\quad S \leftarrow S_0$, $A \leftarrow 0$
6: $\quad$ **for** $t = 1$ to N **do**
7: $\quad\quad \Delta t \leftarrow \frac{T}{N}$
8: $\quad\quad S \leftarrow \mathbf{Model}(S, \theta_{\mathbf{M}}, \Delta t)$
9: $\quad\quad A \leftarrow A + S$
10: $\quad$ **end for**
11: $\quad A \leftarrow \frac{A}{N+1}$
12: $\quad C \leftarrow C + \exp(-rT) \max(A - K, 0)$
13: **end for**
14: $C \leftarrow \frac{1}{M}C$
15: **return** C

Here, S_0 is the initial stock price, $\theta_{\mathbf{M}}$ is the parameter of underlying asset evolution model **Model**, N is the number of time steps, and M is the total number of Monte Carlo trials. Algorithms for the four asset dynamics models are given in the Appendix 7.

In addition, we present here the algorithm for the DOI variance reduction technique.

The other three alternative variance reduction techniques can be found in the Appendix 7.

5.4 Method Comparison Metrics

The standard error and variance reduction is used as metrics to test the accuracy of the models. Standard error is calculated as follows

$$SE = \frac{\sigma}{\sqrt{M}}, \tag{65}$$

Algorithm 2 Monte Carlo Arithmetic Asian Options Pricing Algorithm with the DOI Variance Reduction method

---

1: **Input:** S_0, V_0, T, r, N, M, θ, **Model**
2: **Output:** Option price C
3: Initialize $C \leftarrow 0$
4: **for** $i = 1$ to M **do**
5:     $S \leftarrow S_0$, $A \leftarrow 0$
6:     **for** $t = 1$ to N **do**
7:         $\Delta t \leftarrow \frac{T}{N}$
8:         $\bar{V} \leftarrow$ **Model**$(V, \theta, \Delta t, dW, dB)$
9:         $\bar{S} \leftarrow$ **Model**$(V, \theta, \Delta t, dW, dB)$
10:         $\bar{\sigma} \leftarrow$ **Model**$(V, \theta, \Delta t, T)$
11:         $d1 \leftarrow$ **Model**$(S, K, r, T, \bar{\sigma})$
12:         $d2 \leftarrow$ **Model**$(S, K, r, T, \bar{\sigma})$
13:         Calculate $\frac{\partial BS}{\partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial S_t \partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial \bar{\sigma}_t^2}$, $\frac{\partial \bar{\sigma}_t}{\partial V_t}$ and $\frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2}$
14:         $a = \rho \frac{\partial^2 BS}{\partial S_t \partial \bar{\sigma}_t} \frac{\partial \bar{\sigma}_t}{\partial V_t}$
15:         $b = \frac{1}{2}\xi \frac{\partial^2 BS}{\partial \bar{\sigma}_t^2} \frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2}$
16:         $c = \frac{\partial BS}{\partial \bar{\sigma}_t} \frac{\partial^2 \bar{\sigma}_t}{\partial (V_t)^2}$
17:         $u \leftarrow u_0 + \int_0^T \xi V_0 \exp(-rT) + a + b + c \, dt$
18:     **end for**
19:     $C \leftarrow C + \exp(-rT)u$
20: **end for**
21: $C \leftarrow \frac{1}{M}C$
22: **return** C

---

and the variance reduction is calculated as below

$$VR = \frac{\sigma_j^2 - \sigma_i^2}{\sigma_i^2}, \tag{66}$$

where σ_i^2 is the standard deviation in the crude Monte Carlo error. These methods only consider the accuracy, however time is another factor to consider when evaluating these models therefore a further metric described below is calculated.

5.5 Efficiency of the Monte Carlo methods

Monte Carlo simulations typically demand substantial sample sizes to ensure accurate estimates, which can result in lengthy computing times. To expedite these simulations, variance reduction techniques are employed. However, when evaluating their performance, it is essential to consider both the variance reduction achieved and the associated computing time. Simply reducing variance at the expense of significantly increased computing time due to a more complex algorithm

does not provide a practical benefit. What truly matters is the trade-off between accuracy and computation time for a given method. Therefore, we require a criterion that allows us to determine whether one variance reduction technique outperforms another.

We propose the introduction of a criterion to assess the efficiency of various simulation methods, whether they involve standard simulations or variance reduction techniques. This criterion takes into account the computing time required by each method during the simulation process.

- Efficiency of the method j with regard to the method i is defined by:

$$\varepsilon(i,j) = \frac{\sigma_{N_i}(i)}{\sigma_{N_j}(j)} \sqrt{\frac{t_{N_i}(i)}{t_{N_j}(j)}}$$

where N, t_N, σ_N respectively denote sample size, computing time and standard error of the estimate from N simulations. This value seems to be dependent on sample sizes N_i and N_j, but it is not if we assume that computing time is proportional to the sample size, that is there exists a factor k such that $t_{N_i}(i) = k_i N_i$. This hypothesis is very realistic. Then:

$$\varepsilon(i,j) = \frac{\sigma_i}{\sigma_j} \sqrt{\frac{k_i}{k_j}}$$

To obtain this formulation, we just use that $\sigma_N^2(i) = \frac{\sigma^2(i)}{N_i}$ where $\sigma^2(i)$ is the variance of the estimated function for the method i.

k exprimes complexity of the algorithm for the considered method.

- The value of $\varepsilon(i,j)$ can not be calculated, unless we know variances $\sigma^2(i)$ and $\sigma^2(j)$. This is not the case in general: when we estimate $I = E[\psi(X)]$ we don't know the parameter $\sigma^2 = \text{Var}[\psi(X)]$ too. But we can obtain an estimation of this efficiency within the Monte Carlo simulation.

$$\widetilde{\varepsilon}(i,j) = \frac{\widetilde{\sigma}_{N_i}(i)}{\widetilde{\sigma}_{N_j}(j)} \sqrt{\frac{t_{N_i}(i)}{t_{N_j}(j)}}$$

and

$$\lim_{N_i, N_j \to +\infty} \widetilde{\varepsilon}(i,j) = \varepsilon(i,j)$$

where $\widetilde{\sigma}_{N_i}$ is the estimated standard error for the estimator with the method i.

- The method j is considered to be more efficient than the method i if $\varepsilon(i,j) \geq 1$. We obviously have $\varepsilon(i,i) = 1$.

For instance, if $\varepsilon(i,j) = 3$, it means that method i requires $9 \left(= 3^2\right)$ times more time than method j to obtain the same accuracy. In other words, with the same computing time, standard error for method j is 3 times smaller than the one of method i. The higher $\varepsilon(i,j)$ is, the more efficient method j is with regard to the method i. The crude Monte Carlo method is chosen as the reference method, i, with $\varepsilon(i,j) < \varepsilon(i,k)$ indicating that method k is more efficient than the method j. Although the efficiencies are estimations, given a sample size of 100000 was used for each simulation the convergence of Monte Carlo is assumes to be sufficient to provide reliable estimates (Alfeus et al., 2024).

5.6   Results

Tables 3 to 6 show the results for the option price estimate, standard error, variance reduction and efficiency for each of the Monte Carlo methods used and maturities of 0.5 and 1 year while differing the strike from 80 to 120 in increments of 5.

5.6.1   Naïve and Rough Heston Model Results

The options price estimates across the varying Monte Carlo methods are relatively similar, with the exception of DOI estimate which is consistently lower than the crude Monte Carlo and other variance reduction techniques across strikes and maturities considered for the naïve Heston model.
Considering the naïve Heston model with a maturity of 0.5 years, the standard errors across models generally decrease as the strike is increased from $K = 80$ to $K = 120$, with the exception of the antithetic variate which increases until $K = 100$, followed by a steady decline. The antithetic technique produces a high variance reduction of 92% at $K = 80$ with a dramatic decline to 55% at $K = 120$, which is a relatively larger change in variance reduction as compared to the other variance reduction techniques with an increase in strike. End point stratification reaches a similarly low variance reduction at $K = 120$, however starting at 71% at $K = 80$, while control variates have a consistently higher variance reduction with a slight decline to 99% at $K = 120$. The DOI method on the other hand displays an increase in standard error as the strike is increased, however the standard error in this case is relatively small, being 230 times smaller than the crude Monte Carlo standard error at $K = 80$, and decreasing to over 25 times smaller at $K = 120$. The variance reduction achieved for this technique is 100% for the lowest strike and decreases to slightly below this as the strike is increased, which is greater than that achieved by the remaining three variance reduction techniques. The efficiency generally decreases as strike is increased, and comparison of technique efficiency demonstrates that the DOI method is to be superior across the varied strikes, with it being more than 170 times greater than the crude Monte Carlo at the lowest strike, and decreasing the over 20 times greater at the highest strike considered.

Table 3: Results for the naïve Heston model at maturities of 0.5 years (left) and 1 year (right)

Strike	Metric	T = 0.5					T = 1				
		crude MC	AV	CV	ES	DOI	crude MC	AV	CV	ES	DOI
80	Option Price	21.377	21.325	21.253	21.922	19.830	23.086	23.098	23.013	23.758	20.497
	Standard Error	0.054	0.015	0.002	0.029	0.000	0.073	0.073	0.005	0.043	0.001
	Variance Reduction	0%	92%	100%	71%	100%	0%	0%	100%	66%	100%
	Efficiency	1.000	3.262	22.084	2.284	195.597	1.000	0.901	14.625	1.770	60.934
85	Option Price	17.215	17.165	17.103	17.486	15.659	19.434	19.440	19.424	19.816	17.006
	Standard Error	0.051	0.017	0.002	0.028	0.000	0.070	0.030	0.005	0.041	0.001
	Variance Reduction	0%	89%	100%	70%	100%	0%	82%	100%	65%	100%
	Efficiency	1.000	2.753	22.986	2.381	131.240	1.000	2.108	15.033	1.788	48.194
90	Option Price	13.480	13.435	13.406	13.495	12.071	16.155	16.154	16.192	16.299	13.835
	Standard Error	0.048	0.019	0.002	0.026	0.000	0.066	0.031	0.004	0.039	0.001
	Variance Reduction	0%	84%	100%	69%	100%	0%	78%	100%	64%	100%
	Efficiency	1.000	2.284	22.586	2.343	105.485	1.000	1.885	14.984	1.779	45.503
95	Option Price	10.260	10.217	10.237	10.094	9.058	13.267	13.262	13.361	13.256	11.110
	Standard Error	0.043	0.021	0.002	0.024	0.000	0.062	0.032	0.004	0.037	0.001
	Variance Reduction	0%	77%	100%	68%	100%	0%	73%	100%	63%	100%
	Efficiency	1.000	1.918	22.273	2.074	81.788	1.000	1.729	14.653	1.700	35.791
100	Option Price	7.598	7.556	7.623	7.344	6.393	10.778	10.769	10.924	10.684	8.847
	Standard Error	0.038	0.021	0.002	0.022	0.000	0.057	0.032	0.004	0.035	0.001
	Variance Reduction	0%	70%	100%	67%	100%	0%	68%	99%	62%	100%
	Efficiency	1.000	1.666	20.665	2.260	60.594	1.000	1.568	13.713	1.725	26.771
105	Option Price	5.483	5.446	5.545	5.221	4.472	8.675	8.667	8.858	8.558	6.836
	Standard Error	0.033	0.020	0.002	0.020	0.001	0.052	0.031	0.004	0.033	0.002
	Variance Reduction	0%	64%	100%	65%	100%	0%	64%	99%	60%	100%
	Efficiency	1.000	1.501	18.579	2.218	48.679	1.000	1.487	12.702	1.655	22.748
110	Option Price	3.863	3.833	3.947	3.645	2.772	6.921	6.910	7.131	6.829	5.091
	Standard Error	0.028	0.018	0.002	0.017	0.001	0.047	0.030	0.004	0.031	0.002
	Variance Reduction	0%	60%	100%	63%	100%	0%	61%	99%	58%	100%
	Efficiency	1.000	1.513	17.370	2.256	36.674	1.000	1.442	11.976	1.640	19.832
115	Option Price	2.663	2.640	2.755	2.513	1.565	5.480	5.464	5.703	5.439	3.593
	Standard Error	0.024	0.016	0.002	0.015	0.001	0.043	0.028	0.004	0.029	0.002
	Variance Reduction	0%	57%	100%	59%	100%	0%	58%	99%	55%	100%
	Efficiency	1.000	1.395	14.258	2.050	27.699	1.000	1.395	10.990	1.553	16.892
120	Option Price	1.801	1.782	1.890	1.718	0.651	4.310	4.294	4.535	4.329	2.426
	Standard Error	0.020	0.013	0.002	0.013	0.001	0.038	0.025	0.004	0.027	0.002
	Variance Reduction	0%	55%	99%	55%	100%	0%	56%	99%	52%	100%
	Efficiency	1.000	1.346	12.316	1.930	21.060	1.000	1.348	9.814	1.529	13.613

Table 4: Results for the rough Heston model at maturities of 0.5 years (left) and 1 year (right)

Strike	Metric	T = 0.5					T = 1				
		crude MC	AV	CV	ES	DOI	crude MC	AV	CV	ES	DOI
80	Option Price	20.501	21.220	21.619	23.242	21.091	22.206	22.840	22.835	25.201	22.721
	Standard Error	0.053	0.018	0.007	0.036	0.000	0.076	0.033	0.010	0.059	0.000
	Variance Reduction	0%	89%	98%	53%	100%	0%	81%	98%	39%	100%
	Efficiency	1.000	2.139	5.926	1.429	2618.283	1.000	1.612	5.612	1.433	568.549
85	Option Price	16.351	17.010	16.576	18.798	16.956	18.584	19.171	18.010	21.421	19.371
	Standard Error	0.050	0.020	0.007	0.036	0.000	0.072	0.035	0.009	0.058	0.000
	Variance Reduction	0%	85%	98%	50%	100%	0%	77%	98%	36%	100%
	Efficiency	1.000	1.821	5.817	1.401	1946.117	1.000	1.456	5.669	1.413	446.539
90	Option Price	12.655	13.260	11.975	14.910	13.419	15.378	15.921	15.097	18.128	16.325
	Standard Error	0.047	0.022	0.006	0.035	0.000	0.068	0.037	0.007	0.057	0.000
	Variance Reduction	0%	78%	98%	46%	100%	0%	71%	99%	31%	100%
	Efficiency	1.000	1.517	5.783	1.347	1572.275	1.000	1.322	6.656	1.357	363.765
95	Option Price	9.506	10.066	9.165	11.688	10.334	12.601	13.103	11.992	15.319	13.529
	Standard Error	0.043	0.024	0.006	0.033	0.000	0.064	0.038	0.009	0.056	0.000
	Variance Reduction	0%	69%	98%	40%	100%	0%	66%	98%	25%	100%
	Efficiency	1.000	1.286	5.588	1.284	1146.582	1.000	1.204	5.114	1.283	293.181
100	Option Price	6.945	7.467	6.380	9.121	7.829	10.243	10.706	10.267	12.956	11.353
	Standard Error	0.038	0.024	0.006	0.032	0.000	0.060	0.038	0.008	0.054	0.000
	Variance Reduction	0%	59%	98%	31%	100%	0%	60%	98%	18%	100%
	Efficiency	1.000	1.109	5.134	1.186	828.266	1.000	1.121	4.937	1.224	248.479
105	Option Price	4.958	5.441	4.621	7.118	5.789	8.268	8.702	8.423	10.978	9.379
	Standard Error	0.033	0.023	0.004	0.030	0.000	0.055	0.037	0.006	0.053	0.000
	Variance Reduction	0%	52%	99%	18%	100%	0%	56%	99%	8%	100%
	Efficiency	1.000	1.020	6.395	1.090	657.152	1.000	1.068	6.002	1.158	198.999
110	Option Price	3.473	3.912	3.824	5.579	4.216	6.638	7.045	6.465	16.302	7.711
	Standard Error	0.028	0.021	0.005	0.028	0.000	0.051	0.035	0.006	0.030	0.000
	Variance Reduction	0%	45%	97%	0%	100%	0%	52%	98%	66%	100%
	Efficiency	1.000	0.958	4.525	0.993	526.855	1.000	1.027	5.541	1.894	171.097
115	Option Price	2.401	2.793	2.270	4.402	3.018	5.311	5.690	5.146	7.997	6.306
	Standard Error	0.024	0.019	0.005	0.027	0.000	0.047	0.033	0.006	0.050	0.000
	Variance Reduction	0%	39%	96%	-23%	100%	0%	49%	98%	-16%	100%
	Efficiency	1.000	0.914	4.141	0.896	370.990	1.000	0.999	5.040	1.030	141.407
120	Option Price	1.645	1.989	1.701	3.500	2.087	4.246	4.596	4.186	6.872	5.103
	Standard Error	0.021	0.017	0.004	0.026	0.000	0.043	0.031	0.007	0.049	0.000
	Variance Reduction	0%	34%	96%	-55%	100%	0%	47%	98%	-32%	100%
	Efficiency	1.000	0.875	3.842	0.797	310.689	1.000	0.973	4.547	0.984	122.628

Table 5: Results for the naïve SABR model at maturities of 0.5 years (left) and 1 year (right)

Strike	Metric	T = 0.5					T = 1				
		crude MC	AV	CV	ES	DOI	crude MC	AV	CV	ES	DOI
80	Option Price	21.637	21.610	21.330	21.515	20.053	23.241	23.263	23.333	1.196	20.122
	Standard Error	0.028	0.006	0.001	0.014	0.000	0.042	0.013	0.003	0.022	0.000
	Variance Reduction	0%	95%	100%	75%	100%	0%	90%	100%	74%	100%
	Efficiency	1.000	4.318	35.927	2.697	2843.801	1.000	2.948	16.893	2.094	757.770
85	Option Price	16.739	16.712	16.519	16.616	15.302	18.472	18.494	18.467	18.325	15.594
	Standard Error	0.028	0.006	0.001	0.014	0.000	0.042	0.014	0.002	0.022	0.000
	Variance Reduction	0%	95%	100%	75%	100%	0%	90%	100%	74%	100%
	Efficiency	1.000	4.100	36.519	2.897	1954.592	1.000	2.729	17.087	2.093	759.553
90	Option Price	11.928	11.902	11.953	11.790	10.717	13.938	13.958	13.961	13.757	11.506
	Standard Error	0.027	0.007	0.001	0.014	0.000	0.041	0.015	0.002	0.021	0.000
	Variance Reduction	0%	93%	100%	75%	100%	0%	86%	100%	73%	100%
	Efficiency	1.000	3.512	37.188	2.880	1575.085	1.000	2.374	16.468	1.997	650.258
95	Option Price	7.620	7.596	7.858	7.441	6.659	9.994	10.014	9.996	9.765	7.876
	Standard Error	0.025	0.009	0.001	0.013	0.000	0.038	0.018	0.002	0.020	0.000
	Variance Reduction	0%	86%	100%	74%	100%	0%	79%	100%	72%	100%
	Efficiency	1.000	2.837	43.804	3.375	1207.451	1.000	1.967	16.129	2.135	522.111
100	Option Price	4.355	4.338	4.742	4.156	3.464	6.892	6.908	6.973	6.629	4.983
	Standard Error	0.021	0.011	0.001	0.011	0.000	0.035	0.019	0.002	0.019	0.000
	Variance Reduction	0%	72%	100%	71%	100%	0%	70%	100%	71%	100%
	Efficiency	1.000	1.768	33.897	2.725	917.470	1.000	1.583	15.780	1.975	374.373
105	Option Price	2.282	2.273	2.585	2.101	1.696	4.640	4.651	4.670	4.375	2.901
	Standard Error	0.016	0.010	0.001	0.009	0.000	0.030	0.019	0.002	0.017	0.000
	Variance Reduction	0%	60%	100%	68%	100%	0%	62%	100%	69%	100%
	Efficiency	1.000	1.379	28.953	2.566	1038.932	1.000	1.391	13.951	1.946	590.981
110	Option Price	1.128	1.123	1.318	0.991	0.485	3.084	3.091	3.083	2.847	1.465
	Standard Error	0.012	0.008	0.001	0.007	0.000	0.026	0.017	0.002	0.015	0.000
	Variance Reduction	0%	55%	100%	64%	100%	0%	57%	99%	66%	100%
	Efficiency	1.000	1.377	22.595	2.387	387.104	1.000	1.330	13.275	1.954	465.322
115	Option Price	0.539	0.536	0.651	0.452	0.058	2.041	2.048	2.040	1.844	0.618
	Standard Error	0.008	0.006	0.000	0.005	0.000	0.022	0.015	0.002	0.013	0.000
	Variance Reduction	0%	53%	100%	58%	100%	0%	54%	99%	63%	100%
	Efficiency	1.000	1.344	17.083	2.021	237.134	1.000	1.343	11.158	1.844	174.168
120	Option Price	0.253	0.259	0.319	0.202	0.098	1.355	1.363	1.365	1.196	0.114
	Standard Error	0.006	0.004	0.000	0.004	0.000	0.018	0.013	0.002	0.012	0.000
	Variance Reduction	0%	52%	99%	50%	100%	0%	52%	99%	59%	81%
	Efficiency	1.000	1.293	12.727	1.917	159.391	1.000	1.233	10.311	1.778	95.199

Table 6: Results for the rough SABR model at maturities of 0.5 years (left) and 1 year (right)

Strike	Metric	T = 0.5					T = 1				
		crude MC	AV	CV	ES	DOI	crude MC	AV	CV	ES	DOI
80	Option Price	19.623	19.616	19.626	19.678	20.053	19.269	19.299	19.661	19.395	20.122
	Standard Error	0.025	0.002	0.001	0.013	0.000	0.035	0.005	0.005	0.018	0.000
	Variance Reduction	0%	99%	100%	75%	100%	0%	98%	98%	74%	100%
	Efficiency	1.000	8.297	13.153	1.799	2609.941	1.000	4.734	4.443	1.804	621.992
85	Option Price	14.775	14.767	14.709	14.811	15.302	14.720	14.747	14.347	14.796	15.594
	Standard Error	0.025	0.003	0.003	0.012	0.000	0.033	0.007	0.005	0.017	0.000
	Variance Reduction	0%	99%	99%	75%	100%	0%	96%	98%	74%	100%
	Efficiency	1.000	6.186	6.753	1.808	1770.433	1.000	3.351	4.648	1.772	603.497
90	Option Price	10.153	10.151	10.105	10.138	10.717	10.577	10.601	10.568	10.580	11.506
	Standard Error	0.024	0.005	0.003	0.012	0.000	0.031	0.009	0.005	0.016	0.000
	Variance Reduction	0%	96%	99%	74%	100%	0%	92%	97%	73%	100%
	Efficiency	1.000	3.251	6.571	1.784	1369.949	1.000	2.275	4.392	1.748	489.198
95	Option Price	6.148	6.158	6.126	6.074	6.659	7.083	7.104	7.235	7.021	7.876
	Standard Error	0.020	0.007	0.002	0.011	0.000	0.027	0.011	0.005	0.015	0.000
	Variance Reduction	0%	88%	99%	73%	100%	0%	83%	97%	71%	100%
	Efficiency	1.000	1.930	6.096	1.696	985.545	1.000	1.589	4.020	1.706	365.544
100	Option Price	3.188	3.203	3.248	3.089	3.464	4.399	4.413	4.773	4.306	4.983
	Standard Error	0.016	0.008	0.002	0.009	0.000	0.022	0.012	0.005	0.012	0.000
	Variance Reduction	0%	71%	98%	70%	100%	0%	70%	95%	69%	100%
	Efficiency	1.000	1.257	4.974	1.634	684.924	1.000	1.216	3.446	1.642	239.008
105	Option Price	1.397	1.406	1.461	1.312	1.696	2.528	2.532	2.087	2.445	2.901
	Standard Error	0.010	0.007	0.002	0.006	0.000	0.017	0.011	0.005	0.010	0.000
	Variance Reduction	0%	59%	96%	65%	100%	0%	61%	93%	65%	100%
	Efficiency	1.000	1.117	3.520	1.524	677.262	1.000	1.059	2.708	1.553	334.422
110	Option Price	0.513	0.518	0.519	0.467	0.485	1.349	1.345	1.286	1.291	1.465
	Standard Error	0.006	0.004	0.002	0.004	0.000	0.013	0.008	0.004	0.008	0.000
	Variance Reduction	0%	53%	89%	57%	100%	0%	56%	87%	61%	100%
	Efficiency	1.000	0.991	2.148	1.375	206.802	1.000	0.992	2.050	1.449	225.498
115	Option Price	0.158	0.160	0.157	0.142	0.058	0.669	0.664	0.835	0.635	0.618
	Standard Error	0.003	0.002	0.002	0.003	0.000	0.009	0.006	0.004	0.006	0.000
	Variance Reduction	0%	51%	64%	45%	100%	0%	53%	75%	54%	100%
	Efficiency	1.000	0.963	1.112	1.193	96.905	1.000	0.970	1.483	1.356	69.743
120	Option Price	0.042	0.043	0.047	0.037	0.098	0.312	0.307	0.335	0.297	0.114
	Standard Error	0.002	0.001	0.002	0.001	0.000	0.006	0.004	0.004	0.004	0.000
	Variance Reduction	0%	52%	-31%	29%	100%	0%	51%	45%	44%	100%
	Efficiency	1.000	0.971	0.607	1.063	47.017	1.000	0.953	0.974	1.234	30.477

The above trends are observed for the 1 year term option as well, however the standard error is generally higher for this higher maturity and the efficiency is typically lower with relatively similar variance reduction achieved. Additionally, the trend in standard error for the antithetic model sees a shorter increasing trend, from $K = 80$ to $K = 85$, followed by a longer decline.

Now considering the rough Heston model, the estimates for the variance reduction techniques for the were typically greater than the option price estimate for crude Monte Carlo, with the difference in estimates between endpoint stratification and crude Monte Carlo being greatest, and comparable estimates from the antithetic variate and DOI techniques.

The trends in method metrics observed for the naïve model are generally preserved, with some differences to be noted as follows. Considering a maturity of 0.5 years, the same pattern is observed for the variance reduction, however with a larger difference in variance reduction between the highest and lowest strikes for the antithetic method, the difference being 55% for the rough model and 37% for the naïve model. Additionally, endpoint stratification for the rough Heston model achieved poor variance reduction when compared to the naïve Heston model and in comparison to the other variance reduction techniques for the rough Heston model, with the variance reduction at strikes of 115 and 120 being lower than expected (Glasserman, 2004). Additionally, the DOI method achieves even greater efficiency for the rough Heston model as compared to the naive model, with the efficiency being nearly 2620 times greater than the crude Monte Carlo efficiency for the 0.5 year maturity option and as much as 565 times greater for the 1 year maturity option.

5.6.2   Naïve and Rough SABR Model Results

The values for the control variates for the SABR models are based on selected values of σ for the closed form of the geometric Asian option price which allow for a reasonable option price estimate. Although previous publications investigating the SABR model typically calculate σ as the implied volatility for a Black-Scholes European option, the implied volatility results in poor estimation which we conclude to be a scaling issue, given that use of the implied volatility for out-the-money options overestimates the option price, which underestimating the price of in-the-money options. The heuristic approach of selecting reasonable values for σ was therefore adopted, while a mechanical correction of the scaling issue or a different form of the implied volatility equation are further methods to potentially alleviate the issue.

With regard to the naïve SABR method, the estimates for the options prices are relatively similar for the differing Monte Carlo methods, with the DOI method being comparatively lower for both maturities and for most strikes considered.

For the half year term option, the pattern observed for the standard error in the

naïve SABR model is a general decrease with an increase in strike. The standard error in the antithetic variate method increases until a strike of 100 and thereafter decreases, as observed for the naïve Heston model, while the other models exhibit a steady decrease with increase in maturity, with the exception of the DOI method whose standard error is negligible, being to the power of negative 5 and 6 across all maturities. Regarding variance reduction, this metric for the antithetic method decreases from 95% at $K = 80$ to 52% at $K = 120$, while efficiency generally decreases. Endpoint stratification sees a decrease in variance reduction from 75% and $K = 80$ to 50% at $K = 120$, while the control variate method achieving better variance reduction, the highest being 100% at $K = 80$ and the lowest being 99% at $K = 120$. A variance reduction of 100% is consistently observed for the DOI method, with an efficiency over 2800 times greater than crude Monte Carlo for $K = 80$, and nearly 160 times greater at $K = 120$. The all the variance reduction techniques generally decreases with an increase in strike, however the efficiency of the DOI method starkly higher, being almost 80 times greater than the second most efficient method, the control variate method, at the lowest strike considered and over time times greater than this method at the highest strike considered.

Most patterns are preserved for higher term option modelled by the naïve SABR model, however the standard errors are greater for the higher term option, and lower variance reduction is achieved by the antithetic, control variate, and endpoint stratification methods, while the DOI method results in a variance reduction of 100% of the highest strike where the variance reduction is 81%, which is the only instance where the variance reduction is lower than that achieved by the control variate method.

Regarding the rough SABR model, the option price estimates are generally higher than the crude Monte Carlo estimate, and overall lower than the estimates from the naïve SABR model. The trends in standard error, variance reduction and efficiency are generally the same, with the standard errors and variance reduction being comparable when comparing the naïve and rough at the same strikes and maturities. One notable difference is the performance of the control variate method for the rough models. Where the naïve SABR model's control variate estimate had consistently high variance reduction, this technique applied to the rough model resulted in variance reduction comparable to the antithetic variate method, with a greater decrease in variance reduction, being 52% for the one-year option and an even greater reduction for the maturity of half a year. Additionally, while the DOI method yielded a 100% variance reduction for all cases of the naïve SABR model, values marginally below this are recorded for the rough SABR model at strikes of 115 and 120 for both maturities considered, however this still greatly outperforms the crude Monte Carlo method and the other variance reduction techniques. Lastly, unlike the Heston model, the efficiency achieved by the DOI method is lower for the rough SABR models than the naïve SABR models, being nearly 2610 times greater than the crude Monte Carlo efficiency for the 0.5 year maturity option

and as much as 620 times greater for the 1 year maturity option.

5.6.3 Results Discussion

The estimate for the DOI method is typically furthest from the crude Monte Carlo estimate as compared to the other variance reduction techniques, however this shift is likely a result of the model used for deriving the Greeks being Black-Scholes. Regarding the method comparison metrics, the DOI method outperforms the other variance reduction techniques in terms of standard error and variance reduction across all strikes and maturities considered and all four models investigated. Additionally, the efficiency of the DOI estimate is typically the highest across models and varied parameters, indicating that the DOI method is superior when considering both accuracy and model complexity.

The dramatic reduction in variance achieved by the DOI method can be explained by it being its variance being related to the variance of the approximation method used. For a well chosen approximation method, the variance of the operator term will be small thereby resulting in a low variance for the estimate (Heath and Platen, 2002). Given that we used Monte Carlo simulation to generate the initial option price for the DOI method, and calculated said price via the arithmetic option pricing formula, it is expected that the variance will be low. Additionally, the DOI method is similar to the martingale representation method which generally results in low variance (Heath and Platen, 2002), and can explain why the control variate estimate generally results in the next best variance reduction.

Lastly, the pattern observed with respect to maturity can be attributed to a longer term allowing for greater variation, and the patterns in metrics with respect to strike are attributed to being in-the-money, at-the-money, or out-the-money cases, where in-the-money options result in greater variance reduction and higher efficiency.

6 Conclusion

This paper investigates the application of Monte Carlo methods on non-stochastic volatility and stochastic volatility Heston and SABR models for arithmetic Asian option pricing. Variance reduction techniques were considered, given the poor variance associated with Monte Carlo estimation, namely the DOI method, control variates, antithetic variates, and endpoint stratification with the main focus being the application of the DOI method. Considering accuracy and efficiency metrics, our results indicate that the DOI method is superior in both regards, achieving 100% reduction in variance for most of the cases tested and achieving efficiencies as of over 2600 times greater than crude Monte Carlo for the rough models. The price estimate from the DOI method further from the crude Monte Carlo estimate

than desired, likely a result of the diffusion process chosen to approximate the dynamics.

Further research will focus on testing the DOI method with different diffusion process approximations, investigating the performance of importance sampling as a variance reduction technique, and applying the developed models to market data to check if the simulation results are reasonable.

7 Appendix

7.1 Deferred Proofs

7.1.1 Derivation of Risk-Neutral Naïve Heston Model

We consider the naïve Heston model under the historical measure $\mathbb{P}$ in 3. By Cameron-Martin-Girsanov Theorem (Girsanov, 1960), given Brownian motion $(W_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$, then their respective risk-neutral Brownian motion under the measure $\mathbb{Q}$ is

$$\tilde{W}_t = W_t - \lambda_1 t$$
$$\tilde{B}_t = B_t - \lambda_2 t,$$

where $\lambda_1 = \int_0^t \alpha_s^1 ds$, $\lambda_2 = \int_0^t \alpha_s^2 ds$ are scaling constants representing drift adjustments and $(\alpha_t^1)_{0\leq t\leq T}$ and $(\alpha_t^1)_{0\leq t\leq T}$ are predictable processes in L_2. The equivalent risk-neutral measure $\mathbb{Q}$ is given by

$$d\mathbb{Q} = Z_T d\mathbb{P},$$

where the process $(Z_t)_{0\leq t\leq T}$ is a martingale given by

$$Z_t = e^{\frac{-1}{2}\int_0^t \|\alpha_s^1\|^2 ds + \int_0^t \alpha_s^1 \cdot dW_s}.$$

Given riskless fixed interest rate as r, the risk-neutral stock price process is given by

$$
\begin{aligned}
dS_t &= rS_t dt + \sqrt{V_t} S_t d\tilde{W}_t \\
&= rS_t dt + \sqrt{V_t} S_t (W_t - \lambda_1 t) \\
&= (rS_t - \sqrt{V_t} S_t \lambda_1)t + \sqrt{V_t} S_t dW_t
\end{aligned}
$$

Therefore the following equation holds

$$(rS_t - \sqrt{V_t} S_t \lambda_1)dt = \mu S_t dt.$$

So

$$\lambda_1 = \frac{r - \mu}{\sqrt{V_t}}.$$

It follows that

$$\tilde{W}_t = W_t + \frac{\mu - r}{\sqrt{V_t}}.$$

39

For the risk-neutral process for the variance, we introduce a function $\lambda(S, V, t) = \lambda V_t$ into the drift of dV_t in Equation 3 as follows

$$
\begin{aligned}
dV_t &= [\kappa(\xi - V_t) - \lambda(S, V, t)]dt + \eta\sqrt{V_t}d\tilde{B}_t \\
&= [\kappa(\xi - V_t) - \lambda(S, V, t)]dt + \eta\sqrt{V_t}(dB_t - \lambda_2 dt) \\
&= .[\kappa(\xi - V_t) - (\lambda V_t + \eta\sqrt{V_t}\lambda_2)]dt + \eta\sqrt{V_t}dB_t
\end{aligned}
$$

It follows that the below equality must hold

$$
\lambda V_t + \eta\sqrt{V_t}\lambda_2 = 0.
$$

Thus we have

$$
\lambda_2 = \frac{-\lambda V_t}{\eta\sqrt{V_t}}
$$

Therefore we recover the risk-neutral Brownian motion

$$
\tilde{B}_t = B_t + \frac{\lambda V_t}{\eta\sqrt{V_t}}.
$$

For the sake of compactess of notation, we define transformed risk-neutral parameters $\kappa^* = \kappa + \lambda$ and $\xi^* = \kappa\xi/(\kappa + \lambda)$. Then we have

$$
dV_t = \kappa^*(\xi^* - V_t)dt + n\sqrt{V_t}d\tilde{B}_t.
$$

Verify that

$$
(\kappa + \lambda)(\kappa\xi/(\kappa + \lambda) - V_t)dt = \kappa(\epsilon - V_t) - \lambda V_t. \,\square
$$

7.1.2 Derivation of Naïve Heston Risk-Neutral Log-price Process

We consider the risk-neutral naïve Heston model stock process described in 4. Let $\xi_t = rS_t$ and $V_t = \sqrt{V_t}S_t$. By Ito's formula:

$$
df(S_t) = (\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S_t}\xi_t + \frac{1}{2}\frac{\partial^2 f}{\partial S_t^2}\sigma_t^2)dt + \frac{\partial f}{\partial t}V_t d\tilde{W}_t
$$

Using $f(S_t) = \ln(S_t)$, we have

$$
\begin{aligned}
d\ln S_t &= (\frac{1}{S_t}(rS_t) + \frac{1}{2}(-\frac{1}{S_t^2})(\sqrt{V_t}S_t)^2)dt + \frac{1}{S_t}(\sqrt{V_t}S_t)d\tilde{W}_t \\
&= (r - \frac{1}{2}V_t)dt + \sqrt{V_t}d\tilde{W}_t. \,\square
\end{aligned}
$$

7.1.3 Derivation of Risk-Neutral Rough Heston Model

Since the rough version of the naïve Heston Model only introduces a kernel function $K(s)$ as described in equation 11 to the volatility SDE but does not interfere with its drift adjustments, we expect to have the same risk-neutral parameters as the original Heston model. Let us show this rigorously. We assume the same riskless fixed interest rate r and volatility premium function $\lambda(S, V, t) = \lambda V_t$.

First, the stock price process is intact so we have the same risk-neutral formula as before:

$$dS_t = rS_t dt + \sqrt{V_t} S_t d\tilde{W}_t,$$

Next, the risk neutral stochastic integral of the variance process is given by

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\kappa(\xi - V_t) - \lambda(S, V, t)] ds$$
$$+ \frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} d\tilde{B}_s,$$

where we similarly have by Girsanov's theorem the below representation for $\tilde{B}_t$:

$$\tilde{B}_t = B_t - \lambda_2 t,$$

for some $\lambda_2 = \int_0^t \alpha_s^2 ds$ with predictable process $(\alpha_t^2)_{t \geq 0} \in L_2$. It follows that

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\kappa(\xi - V_s) - \lambda(S, V, s)] ds$$
$$+ \frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} (dB_s - \lambda_2 ds)$$
$$= V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\kappa(\xi - V_s) - \lambda V_s] ds$$
$$+ \frac{\eta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{V_s} (dB_s - \lambda_2 ds)$$
$$= V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} [\kappa(\xi - V_s) - (\lambda V_s + \eta \sqrt{V_s} \lambda_2)] ds$$
$$+ \sqrt{V_s} dB_s.$$

It follows that given any fixed $t \geq 0$, the below equality holds true for all $s \in [0, t]$:

$$\lambda V_s + \eta \sqrt{V_s} \lambda_2 = 0$$

or equivalently

$$\lambda_2 = -\frac{\lambda V_s}{\eta \sqrt{V_s}}$$

for all $s \in [0, t]$. As before, we define transformed risk-neutral parameters $\kappa^* = \kappa + \lambda$ and $\xi^* = \kappa\xi/(\kappa + \lambda)$. We can write the risk-neutral rough Heston model as

$$dS_t = rS_t dt + S_t\sqrt{V_t}d\tilde{W}_t, \quad S_0 = s_0$$

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\kappa^*(\xi^* - V_s)ds$$

$$+ \frac{\eta}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{V_s}d\tilde{B}_s, \quad V_0 = v_0$$

$$\langle d\tilde{W}_t, d\tilde{B}_t\rangle = \rho dt,$$

where

$$\tilde{W}_t = W_t + \frac{\mu - r}{\sqrt{V_t}}$$

$$\tilde{B}_t = B_t + \frac{\lambda V_t}{\eta\sqrt{V_t}},$$

for all $t \geq 0$. $\square$

7.1.4 Derivation of Risk-Neutral SABR Model

The risk-neutral SDE for the forward price process follows the same analysis as that for the stock process in Heston models described above.
For the stochastic volatility, since there are no drift adjustment for the volatility process, we have $\tilde{B}_t = B_t$. $\square$

7.1.5 Derivation of Risk-Neutral Rough SABR Model

As the SDE for process $\xi_t(s)$ is under the risk-neutral measure $\mathbb{Q}$ as before, the volatility σ, which is a deterministic function of $\xi_t(s)$ is also risk-neutral. To adjust to the risk-neutral version of System 24, we follow the same procedures as before:

$$dF_t = r\beta(F_t)dt + V_t\beta(F_t)d\tilde{W}_t, \quad F_0 = f_0$$
$$d\xi_t(s) = K(s-t)\xi_t(s)d\tilde{B}_t, \quad t < s, \; \xi_0(0) = v_0^2 \tag{67}$$
$$\langle d\tilde{W}_t, d\tilde{B}_t\rangle = \rho dt,$$

where $\tilde{W}_t, \tilde{B}_t$ follow the same dynamics as in 18. $\square$

7.1.6 Recovering SABR from Rough SABR

Let

$$\beta(F_t) = F_t^\beta, \quad 0 \leq \beta \leq 1$$

42

and

$$\xi_0(s) = \alpha_0^2 \exp\{\frac{1}{4}\eta^2 s\},$$

Then when $H = 1/2$, we recover SABR model in 16 from rough SABR model in 24. To see this, we set $H = 1/2$. We have

$$\xi_t(s) = \mathbb{E}^{\mathbb{Q}}[\xi_s(s)|\mathcal{F}_t]$$
$$= \xi_0(s) \exp\{\eta B_t - \frac{1}{2}\eta^2 t\}, \ \ 0 \le t \le s$$

and

$$K(s - t) = \eta.$$

It follows that

$$d\xi_t(s) = \eta \exp\{\eta B_t - \frac{1}{2}\eta^2 t\}dB_t$$

and

$$V_t = \sqrt{\xi_t(t)} = \sqrt{\xi_t(t)} \exp\{\frac{\eta}{2}B_t - \frac{1}{4}\eta^2 t\}.$$

Taking the natural logarithm:

$$\ln V_t = \frac{1}{2} \ln \xi_t(t)(\frac{\eta}{2}B_t - \frac{1}{4}\eta^2 t).$$

It follows that

$$d\ln(V_t) = \frac{d}{dt}(\frac{1}{2} \ln \xi_0(t))dt + \frac{\eta}{2}dB_t - \frac{1}{4}\eta^2 dt$$
$$= (\frac{1}{2}\frac{\xi_0'(t)}{\xi_0(t)} - \frac{1}{4}\eta^2)dt + \frac{\eta}{2}dB_t.$$

Denote stochastic process $Y_t = \ln(V_t)$. Let $f(Y_t) = \exp\{Y_t\}$. By Ito's formula:

$$d(\exp\{Y_t\}) = \exp\{Y_t\}dY_t + \frac{1}{2}\exp\{Y_t\}\langle dY_t, dY_t\rangle$$
$$= V_t((\frac{1}{2}\frac{\xi_0'(t)}{\xi_0(t)} - \frac{1}{4}\eta^2)dt + \frac{\eta}{2}dB_t) + \frac{1}{2}V_t(\frac{\eta^2}{4}dt)$$
$$= V_t((\frac{1}{2}\frac{\xi_0'(t)}{\xi_0(t)} - \frac{1}{4}\eta^2 + \frac{1}{8}\eta^2)dt + \frac{\eta}{2}dB_t)$$
$$= V_t((\frac{1}{2}\frac{\xi_0'(t)}{\xi_0(t)} - \frac{1}{8}\eta^2)dt + \frac{\eta}{2}dB_t),$$

(68)

43

since we note that the quadratic variation term simplifies to

$$\langle dY_t, dY_t \rangle = \langle \frac{\eta}{2} dB_t, \frac{\eta}{2} dB_t \rangle$$
$$= \frac{\eta^2}{4} \langle dB_t, dB_t \rangle$$
$$= \frac{\eta^2}{4} t.$$

Substituting $\xi_0(0)$ into equation 68:

$$d(\exp\{Y_t\}) = dV_t$$
$$= V_t((\frac{1}{2}(\frac{1}{4}\eta^2) - \frac{1}{8}\eta^2)dt + \frac{\eta}{2}dB_t)$$
$$= V_t \frac{\eta}{2} dB_t.$$

Hence, the system of SDE is given by

$$dF_t = \mu F_t^\beta dt + V_t F_t^\beta dW_t, \quad F_0 = f_0$$
$$V_t = \alpha V_t dB_t, \quad V_0 = v_0$$
$$\langle dW_t, dB_t \rangle = \rho dt,$$

where $\alpha = \frac{\eta}{2}$. Notice that this is exactly the vanilla SABR model described in 16. $\square$

7.1.7 Recovering Fractional Brownian Motion from the Rough SABR Model

We offer more details in recovering the Mandelbrot-Van Ness represented fractional BM from the rough SABR model 21 introduced by Fukasawa and Gatheral (2021).
First, we show that the rough SABR model 21 can be re-written into the form

$$V_t = V_0 \exp\{\gamma B_t^H - \frac{1}{2} \mathbb{V}\text{ar} B_t^H\},$$

where $\gamma = 1/2$ and

$$B_t^H = \eta \sqrt{2H} \int_0^t (s-u)^{H-1/2} dB_u.$$

Observe that

$$V_t = \sqrt{\xi_0(t)} \exp\{\frac{1}{2}(\eta\sqrt{2H} \int_0^t (s-u)^{H-\frac{1}{2}} dB_u - \frac{1}{2}\eta^2 t^{2H})\}$$

44

Since

$$\mathbb{V}\mathrm{ar}(B_t^H)$$
$$= (\eta\sqrt{2H})^2 \int_0^t \int_0^t (s-u)^{H-1}(s-v)^{H-1}\mathbb{E}[dB_u, dB_v]$$
$$= (\eta\sqrt{2H})^2 \int_0^t \int_0^t (s-u)^{H-1}(s-v)^{H-1}\delta(u-v)dudv$$
$$= (\eta\sqrt{2H})^2 \int_0^t (s-u)^{H-1}du$$
$$= (\eta\sqrt{2H})^2 \int_0^t w^{2H-1}du$$
$$= (\eta\sqrt{2H})^2 \frac{t^{2H}}{2H}$$
$$= \eta^2 2H \frac{t^{2H}}{2H}$$
$$= \eta^2 t^{2H}$$

where we have used the substitution $w = t - u$ and $\delta(\cdot)$ is the Dirac function, we can write

$$V_t = \sqrt{\xi_0(t)} \exp\{\frac{1}{2}(B_t^H - \mathbb{V}\mathrm{ar}B_t^H)\},$$

as desired.

Second, we show that B_t^H can indeed be represented as a fractional Brownian motion with Hurst parameter H via the Mandelbrot-Van-Ness formula 12 by setting the scaling constant

$$\frac{1}{c_1(H)} = \eta\sqrt{2H}.$$

Let us start by analyzing formula 12. Note that:

- For $u \leq 0$: $((-u)^+)^{H-1/2} = (-u)^{H-1/2}$.

- For $u > s$: $((s-u)^+)^{H-1/2} = 0$.

So for $0 \leq u \leq s$, we have that

$$f_s(u) = (s-u)^{H-1/2}$$

and $f_t(u) = 0$ otherwise. Since $t \leq s$, it follows that

$$\int_0^t (s-u)^{H-1/2}dB_u = \int_{-\infty}^t ((s-u)^+)^{H-1/2}dB_u.$$

45

If we further set the normalization constant

$$\frac{1}{c_1(H)} = \eta\sqrt{2H},$$

then we exactly recover the effective Mandelbrot-Van-Ness formula 12 on the domain $[0, t]$.

For a more rigorous discussion, we note that scaling constant $\eta\sqrt{2H}$ is exact in recovering its Mandelbrot-Van-Ness fractional BM representation when $\eta = 1$. To see this, recall that in this case

$$\mathbb{Var}(B_t^H) = t^{2H}.$$

If we vary η, then we will, as in the rough Heston model, obtain a scaled Mandelbrot-Van-Ness represented fractional BM. $\square$

7.1.8 Deriving $\bar{\sigma}$ for Rough Heston Model DOI Estimator

Recall $\bar{\sigma}$ defined in Equation 46:

$$\bar{\sigma}_t = \sqrt{\frac{1}{T-t}\int_t^T \bar{V}_z \mathrm{d}z}$$

and recall we use the following non-stochastic $\bar{V}_t$ to approximate V_t defined in the risk-neutral rough Heston model 15:

$$\bar{V}_t = V_0 + \frac{\kappa^*}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\xi^* - \bar{V}_s)ds, \quad \bar{V}_0 = \bar{v}_0$$

Thus we have

$$\bar{\sigma}_t = \sqrt{\frac{1}{T-t}\int_t^T \bar{V}_z \mathrm{d}z}$$

$$= \sqrt{\bar{v}_0 + \frac{1}{T-t}\int_t^T \left(\frac{\kappa^*}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\xi^* - \bar{V}_s)ds\right)dz}$$

We first examine the inner integral. Observe that integral

$$\frac{\kappa^*}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\xi^* - \bar{V}_s)ds \tag{69}$$

is a so-called Riemann-Liouville fractional integral of order α. We recall the definition.

Definition 7.1 (Riemann-Liouville Fractional Integral). We define the Riemann-Liouville fractional integral of order α as the following:

$$_aI_x^n f(x) := \frac{1}{\Gamma(n)} \int_a^x (x-t)^{n-1} f(t) dt. \tag{70}$$

where $n \in \mathbb{N} \cup \{0\}$ (Mathai, 2008).

Thus we have:

$$\frac{\kappa^*}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} (\xi^* - \bar{V}_s) ds = \kappa^* {}_0I_t^\alpha (\xi^* - \bar{V}_t) \tag{71}$$

To further simplify, we exploit the relation between fractional integrals and the so-called Mittag-Leffler function of parameter (α, β) is defined as the following Mathai (2008):

$$E_{\alpha,\beta}(x) = \sum_{k=0}^{\infty} \left(\frac{x^k}{\Gamma(\alpha k + \beta)} \right) \tag{72}$$

Moreover, for $(a, \lambda) \in (0,1) \times \mathbb{R}_+$, define

$$f^{\alpha,\lambda}(t) := \lambda t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^\alpha), \ \ t > 0,$$
$$F^{\alpha,\lambda}(t) := \int_0^t f^{\alpha,\lambda}(s) ds, \ \ t \geq 0 \tag{73}$$
$$= 1 - E_{\alpha,1}(-\lambda t^\alpha).$$

$f^{\alpha,\lambda}$ is known as the *Mittag-Leffler density function* and $f^{\alpha,\lambda} \in L^2$. Functions in 73 have the following asymptotic properties (Haubold et al., 2009):

$$\lim_{t \to 0^+} f^{\alpha,\lambda} = \frac{\lambda}{\Gamma(\alpha)} t^{\alpha-1},$$
$$\lim_{t \to \infty} f^{\alpha,\lambda} = \frac{\alpha}{\lambda \Gamma(1-\alpha)} t^{-(\alpha+1)} \tag{74}$$

and

$$\lim_{t \to 0^+} F^{\alpha,\lambda} = \frac{\lambda}{\Gamma(\alpha+1)} t^\alpha,$$
$$\lim_{t \to \infty} (1 - F^{\alpha,\lambda}) = \frac{1}{\lambda \Gamma(1-\alpha)} t^{-\alpha} \tag{75}$$

Using definition 7.1, 72, and 73, we observe that

$$_0I_t^{1-\alpha} f^{\alpha,\lambda} = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{n-1} \lambda t^{\alpha-1} E_{\alpha,\alpha}(s) ds$$
$$= \lambda(1 - F^{\alpha,\lambda}). \tag{76}$$

47

We are now ready to present a derivation for the DOI $\bar{\sigma}$ in the rough Heston model, which approximates $\bar{\sigma}$ by calculating the expected value of $\bar{V}_t$ [5].

First, instead of finding the exact integral representation, we take the expectation of $\bar{V}_t$ to approximate integral $\int_t^T \bar{V}_z dz$:

$$\mathbb{E}[\bar{V}_t] = v_0 + \mathbb{E}\left[\frac{\kappa^*}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(\xi^* - \bar{V}_s)ds\right],$$

which is well-defined, since by Jaisson and Rosenbaum (2015), we can show that for any $t \geq 0$,

$$\mathbb{E}\left[\int_0^t V_s ds\right] < \infty.$$

This implies that $t \to \mathbb{E}[V_t]$ is locally integrable. Further, building on results of Jaisson and Rosenbaum (2015), proposition 3.1 in Euch and Rosenbaum (2017) shows that [6]

$$\mathbb{E}[\bar{V}_t] = v_0 + (\xi^* - v_0) \int_0^t \kappa^* s^{\alpha-1} E_{\alpha,\alpha}(-\kappa^* s^\alpha) ds$$

$$= v_0 + (\xi^* - v_0) \int_0^t f^{\alpha,\kappa^*}(s)ds. \tag{77}$$

Then we can approximate the integral in Equation 46 as the following

$$\bar{\sigma}_t = \sqrt{\frac{1}{T-t} \int_t^T \bar{V}_z dz}$$

$$\approx \sqrt{\frac{1}{T-t} \int_t^T \mathbb{E}[\bar{V}_z] dz}$$

$$= \sqrt{\frac{1}{T-t} \int_t^T v_0 + (\xi^* - v_0) \int_0^z f^{\alpha,\kappa^*}(s)ds dz}$$

$$= \sqrt{v_0 + \frac{(\xi^* - v_0)}{T-t} \int_t^T \int_0^z f^{\alpha,\kappa^*}(s)ds dz}$$

$$= \sqrt{v_0 + \frac{(\xi^* - v_0)}{T-t} \int_t^T (1 - E_{\alpha,1}(-\kappa^* z^\alpha))dz}$$

$$= \sqrt{v_0 + (\xi^* - v_0) - \frac{(\xi^* - v_0)}{T-t} \int_t^T E_{\alpha,1}(-\kappa^* z^\alpha)dz}$$

$$= \sqrt{\xi^* - \frac{(\xi^* - v_0)}{T-t}(T E_{\alpha,2}(-\kappa^* T^\alpha) - t E_{\alpha,2}(-\kappa^* t^\alpha))},$$

---

[5]Stochastic process $\{\mathbb{E}[V_t]\}_{t \geq 0}$ is known as the *Forward Variance Curve*.

[6]Note that the η term in the original formulation scales to 1 as $\bar{V}_s$ no longer has the volatility term. Equation 77 is also referenced in Gatheral and Keller-Ressel (2019).

Here, the third to last equality is due to property 73 and the second to last equality is derived from the following:

$$
\begin{aligned}
\int_t^T E_{\alpha,1}(-\kappa^* z^\alpha) dz &= \int_t^T \sum_{k=0}^{\infty} \frac{(-\kappa^* z^\alpha)^k}{\Gamma(\alpha k + 1)} dz \\
&= \sum_{k=0}^{\infty} \int_t^T \frac{(-\kappa^* z^\alpha)^k}{\Gamma(\alpha k + 1)} dz \\
&= \sum_{k=0}^{\infty} \frac{(-\kappa^*)^k}{\Gamma(\alpha k + 1)} \int_t^T z^{\alpha k} dz \\
&= \sum_{k=0}^{\infty} \frac{(-\kappa^*)^k}{\Gamma(\alpha k + 1)} \frac{T^{\alpha k + 1} - t^{\alpha k + 1}}{\alpha k + 1} \\
&= T \sum_{k=0}^{\infty} \frac{(-\kappa^* T^\alpha)^k}{\Gamma(\alpha k + 2)} - t \sum_{k=0}^{\infty} \frac{(-\kappa^* t^\alpha)^k}{\Gamma(\alpha k + 2)} \\
&= T E_{\alpha,2}(-\kappa^* T^\alpha) - t E_{\alpha,2}(-\kappa^* t^\alpha),
\end{aligned}
$$

where we can interchange the order of summation and integration since the Mittag-Leffler function

$$
E_{\alpha,1}(-\kappa^* z^\alpha)
$$

converges uniformly for $\alpha > 0$ (Gorenflo et al., 2014). $\square$

7.1.9 Deriving $\bar{\sigma}$ for Rough SABR Model DOI Estimator

First, we note that since the rough SABR model 21 does not have a drift term, its DOI estimate should be the same as in the naïve SABR model 16.
Recall that the moments of the rough SABR model for $0 \le \beta < 1$ is given by the following (Musiela, 2016):

$$
\mathbb{E}[V_{t \wedge \tau_n}^m] \le a(t) + b \int_0^t \mathbb{E}[V_{s \wedge \tau_n}^m] ds,
$$

where X is a martingale in L^m ($1 < m < \infty$) and stopping time

$$
\tau_n = \inf\{s \ge 0 : V_s \ge n\}.
$$

In particular, a closed-form moments function is possible when $\beta = 1$, in which case we recover the log-normal rough SABR model:

$$
\mathbb{E}_{\text{lognormal}}[V_t^m] = v_0^m \exp\{\frac{1}{2}\alpha^2 t^{2\alpha - 1} m(m - 1)\} \tag{78}
$$

49

It follows that

$$\mathbb{E}_{\text{lognormal}}[V_t] = v_0.$$

Hence, we have

$$
\begin{aligned}
\bar{\sigma}_t &= \sqrt{\frac{1}{T-t} \int_t^T \bar{V}_z \mathrm{d}z} \\
&\approx \sqrt{\frac{1}{T-t} \int_t^T \mathbb{E}_{\text{lognormal}}[\bar{V}_z] \mathrm{d}z} \\
&= \sqrt{\frac{1}{T-t}(T-t)v_0} \\
&= \sqrt{v_0}.
\end{aligned}
$$

We note that for the best possible performance of the DOI technique on the SABR model, we present both the SABR model and the rough SABR model with $\beta = 1$. But in general, we note that using $\mathbb{E}_{\text{lognormal}}[V_t]$ to estimate $\bar{\sigma}_t$ for the SABR models (both naïve and rough) still perform well. □

7.2 Deferred Algorithms

We present in this section the deferred algorithms.

Algorithm 3 Heston Model

1: **Function Model**(S, θ, Δt)
2: Extract parameters: $\theta = \{\mu, \kappa, \xi, \eta, \rho, v\}$
3: Generate correlated random variables $W, B \sim \mathcal{N}(0, 1)$ with correlation ρ
4: $W_t \leftarrow W\sqrt{\Delta t}$
5: $B_t \leftarrow \rho W \sqrt{\Delta t} + \sqrt{1-\rho^2} B \sqrt{\Delta t}$
6: $S \leftarrow S + \mu S \Delta t + S\sqrt{v}W_t$
7: $v \leftarrow v + \kappa(\xi - v)\Delta t + \eta\sqrt{v}B_t$
8: $v \leftarrow \max(v, 0)$
9: **return** S

7.3 Deferred Plots

We present in this section the deferred algorithms.

7.3.1 Histogram of price frequency

In below we give the histogram of the frequency of price under crude Monte-Carlo method and DOI variance reduction method.

Algorithm 4 Rough Heston Model

1: **Input:** ρ, μ, r, κ, ξ, η, T, S_0, V_0, N, M
2: **Output:** M Simulated path $\{S_t\}_{t=0,\ldots,N-1}$, $\{V_t\}_{t=0,\ldots,N-1}$
3: Initialize $S_t, V_t, Y_t \leftarrow S_0, V_0, \log(S_0)$
4: Generate random variables ΔW and ΔB for each time step
5: $\Delta t \leftarrow \frac{T}{N}$
6: **for** $t = 1$ to n **do**
7:      **for** $i = 0$ to t **do**
8:          Compute adjusted $d\tilde{W}$ and $d\tilde{B}$
9:          $V_t \leftarrow V_t + K(t-i)(\xi - V_{max})\Delta t + K(t-i)\eta\sqrt{V_{max}}$
10:          $Y_t \leftarrow Y_t + (r - V_{max})\Delta t + \sqrt{V_{max}}\Delta W$
11:      **end for**
12:      $S_t \leftarrow \exp Y_t$
13: **end for**
14: **return** $\{S_t\}_{t=0,\ldots,N-1}$, $\{V_t\}_{t=0,\ldots,N-1}$

Algorithm 5 SABR Model

1: **Function Model**(S, θ, Δt)
2: Extract parameters: $\theta = \{\alpha, \beta, \mu, \rho, v\}$
3: Generate correlated random variables $W, B \sim \mathcal{N}(0, 1)$ with correlation ρ
4: $W_t \leftarrow W\sqrt{\Delta t}$
5: $B_t \leftarrow \rho W\sqrt{\Delta t} + \sqrt{1 - \rho^2}B\sqrt{\Delta t}$
6: $v \leftarrow v + \alpha v \Delta B_t$
7: $F \leftarrow F + vF^\beta \Delta W_t$
8: $S \leftarrow F\exp(-rt)$
9: $v \leftarrow \max(v, 0)$
10: **return** S

7.3.2 Trajectory

Here we give the trajectory of 10 simulations of heston/rough heston, SABR/rough SABR model.

Algorithm 6 Rough SABR Model

1: **Input:** ρ, μ, r, $\beta(\cdot)$,α_0, T, F_0, ξ_0, N, M
2: **Output:** M Simulated path F, V
3: Initialize $F_t, \xi_t \leftarrow F_0, \xi_0$
4: Generate random variables ΔW and ΔB for each time step
5: $\Delta t \leftarrow \frac{T}{N}$
6: **for** $t = 1$ to n **do**
7:     **for** $i = 0$ to t **do**
8:         Compute adjusted $d\tilde{W}$ and $d\tilde{B}$
9:         $\xi_t \leftarrow \xi_t + K(t-i)\xi_i \Delta B$
10:         $S_t \leftarrow S_t + \mu\beta(S_t)\Delta t + \sqrt{\xi_t}\beta(S_t)\Delta W$
11:     **end for**
12: **end for**
13: **return** $\{S_t\}_{t=0,\dots,N-1}$, $\{V_t\}_{t=0,\dots,N-1}$
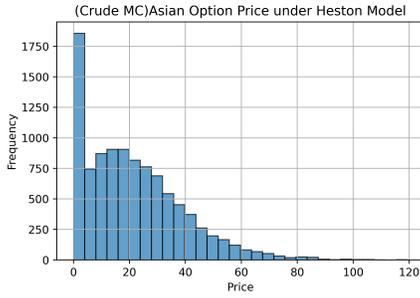
(a) (Crude MC)Asian Option Price under Heston Model
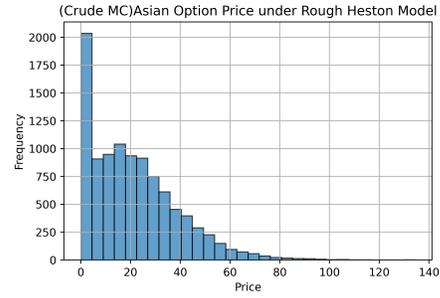
(b) (Crude MC)Asian Option Price under Rough Heston Model

Figure 5: Asian Option Prices under Heston and Rough Heston Models
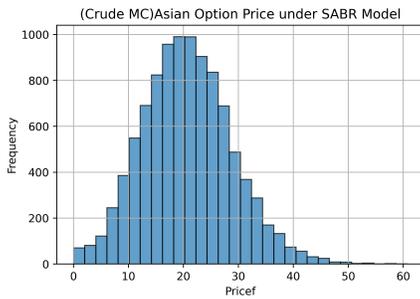
(a) (Crude MC)Asian Option Price under SABR Model
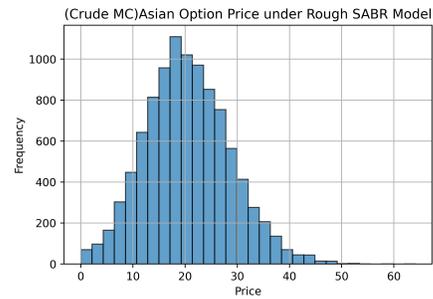
(b) (Crude MC)Asian Option Price under Rough SABR Model

Figure 6: Asian Option Prices under SABR and Rough SABR Models

Algorithm 7 Monte Carlo Arithmetic Asian Options Pricing Algorithm with an Antithetic Variate

---

1: **Input:** S_0, V_0, T, r, N, M, θ, **Model**
2: **Output:** Option price C
3: Initialize $\hat{C} \leftarrow 0$
4: **for** $i = 1$ to M **do**
5:      $S \leftarrow S_0$, $A \leftarrow 0$
6:      $S_a \leftarrow S_0$, $A_a \leftarrow 0$
7:      $dW_a \leftarrow -dW_a$, $dB_a \leftarrow -dB_a$
8:      **for** $t = 1$ to N **do**
9:          $\Delta t \leftarrow \frac{T}{N}$
10:          $V \leftarrow \textbf{Model}(V, \theta, \Delta t, dW, dB)$
11:          $V_a \leftarrow \textbf{Model}(V_a, \theta, \Delta t, dW, dB)$
12:          $S \leftarrow \textbf{Model}(S, \theta, \Delta t, dW, dB)$
13:          $S_a \leftarrow \textbf{Model}(S_a, \theta, \Delta t, dW_a, dB_a)$
14:          $A \leftarrow A + S$
15:          $A_a \leftarrow A_a + S_a$
16:      **end for**
17:      $A \leftarrow \frac{A}{N+1}$
18:      $A_a \leftarrow \frac{A_a}{N+1}$
19:      $C \leftarrow C + \exp(-rT)\max(A - K, 0)$
20:      $C_a \leftarrow C_a + \exp(-rT)\max(A_a - K, 0)$
21:      $m = \frac{M}{2}$
22:      $\hat{C} \leftarrow \frac{1}{m}\sum_{i=1}^{m}\left(\frac{C(1:m)+C_a(1:m)}{2}\right)$
23: **end for**
24: $C \leftarrow \frac{1}{M}\sum C$
25: **return** C

Algorithm 8 Monte Carlo Arithmetic Asian Options Pricing Algorithm with a Control Variate

1: **Input:** S_0, V_0, T, r, N, M, θ, σ, **Model**
2: **Output:** Option price C
3: Initialize $\hat{C} \leftarrow 0$
4: $\bar{\mu} = (r - 0.5\sigma^2)(T/2)$
5: $\bar{\sigma} = sqrt(((\sigma^2 T)/6)(T/(T + dt) + 1))$
6: $a = (\log(S_0/K) + \bar{\mu})/\bar{\sigma}$
7: $c = \exp(-rT)(S_0 \exp(\bar{\mu} + 0.5\bar{\sigma}^2)\phi(a + \bar{\sigma}) - K\phi(a))$
8: **for** $i = 1$ to M **do**
9:     $S \leftarrow S_0$, $A \leftarrow 0$
10:     $S_a \leftarrow S_0$, $A_a \leftarrow 0$
11:     **for** $t = 1$ to N **do**
12:         $\Delta t \leftarrow \frac{T}{N}$
13:         $V \leftarrow \textbf{Model}(V, \theta, \Delta t, dW, dB)$
14:         $S \leftarrow \textbf{Model}(S, \theta, \Delta t, dW, dB)$
15:         $A_f \leftarrow A_f + S$
16:         $A_g \leftarrow A_g \cdot S$
17:     **end for**
18:     $C_f \leftarrow C + \exp(-rT) \max(\frac{A_f}{N+1} - K, 0)$
19:     $C_g \leftarrow C + \exp(-rT) \max(A_g^{\frac{1}{N+1}} - K, 0)$
20:     $m = n/2$
21:     $\hat{C} \leftarrow C_f + \alpha^{opt}(C_g - c)$
22: **end for**
23: $C \leftarrow \frac{1}{M} \sum C$
24: **return** C

Algorithm 9 Monte Carlo Arithmetic Asian Options Pricing Algorithm with End-point Stratification

---

1: **Input:** $S_0, T, t, r, N, M, d\ \theta$, **Model**, **BM**
2: **Output:** Option price C
3: Initialize $C \leftarrow 0$
4: **for** $i = 1$ to M **do**
5:      $S \leftarrow S_0$, $A \leftarrow 0$
6:      $Z \leftarrow$ equally distributed between $d$strata
7:      **for** $t = 1$ to N **do**
8:          $\Delta t \leftarrow \frac{T}{N}$
9:          $W_{0,i} \leftarrow 0$
10:         $W_{end,i} \leftarrow \sqrt{(T)}Z$
11:         $W_{mid,i} \leftarrow \textbf{BM}(t, W_0, W_{end}, Z)$
12:         $W_i \leftarrow$ appended $W_{0,i}, W_{mid,i}, W_{end,i}$
13:         $V \leftarrow \textbf{Model}(V, \theta, \Delta t, W_1)$
14:         $S \leftarrow \textbf{Model}(S, \theta, \Delta t, W_2)$
15:         $A \leftarrow A + S$
16:      **end for**
17:      $A \leftarrow \frac{A}{N}$
18:      $C \leftarrow C + \exp(-rT)\max(A - K, 0)$
19:      $C \leftarrow$ reshaped to $\frac{M}{d}$ rows and d columns
20: **end for**
21: $C \leftarrow \frac{1}{d}\sum(\frac{1}{M/d}\sum C)$
22: **return** C
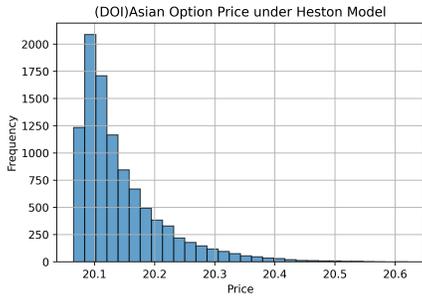
(a) (Crude MC)Asian Option Price under Heston Model
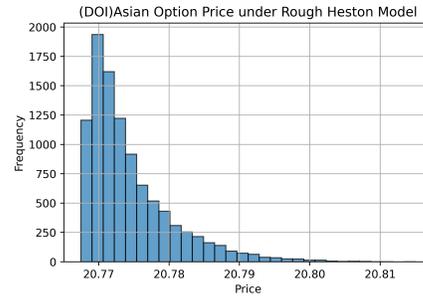
(b) (Crude MC)Asian Option Price under Rough Heston Model
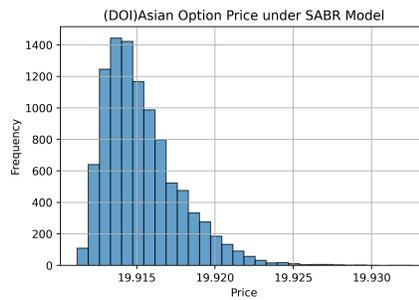
(c) (Crude MC)Asian Option Price under SABR Model
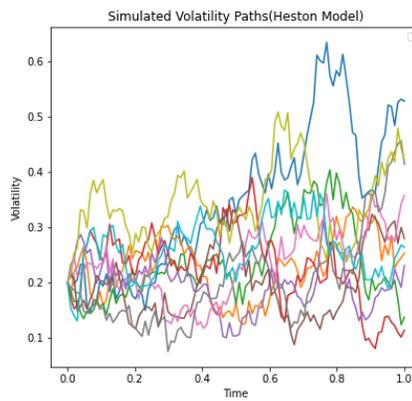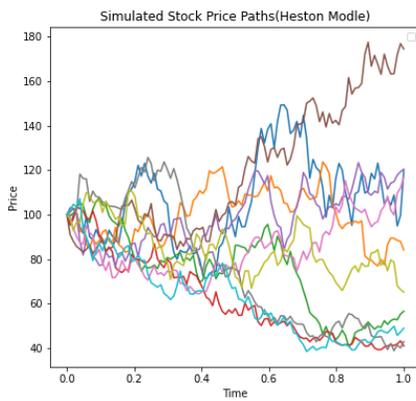
Figure 8: (Crude MC)Asian Option Price under SABR Model
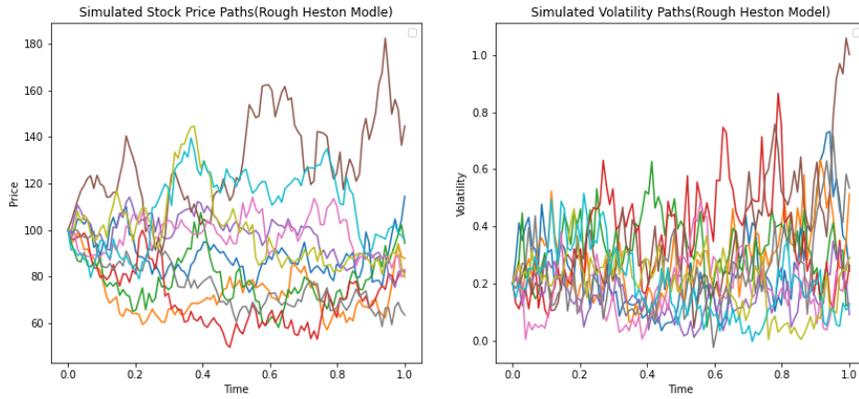
Figure 9: (Crude MC)Asian Option Price under SABR Model
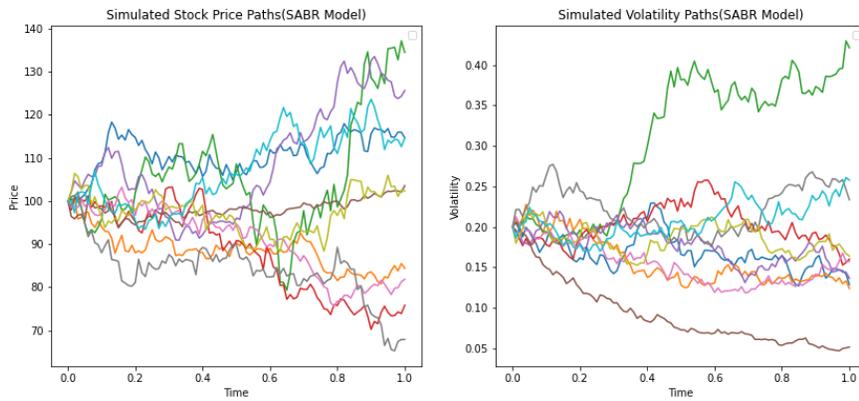
Figure 10: (Crude MC)Asian Option Price under SABR Model
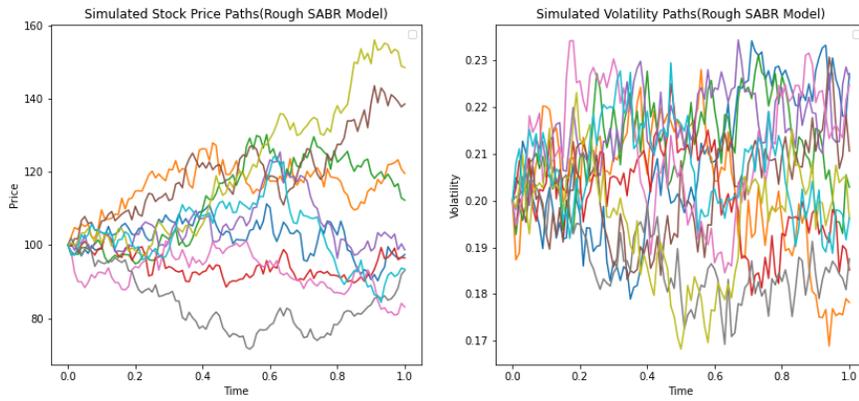
Figure 11: (Crude MC)Asian Option Price under SABR Model

Bibliography

Alfeus, M., Nikitopoulos, C.S., Overbeck, L., 2024. Implied roughness in the term structure of oil market volatility. Quantitative Finance 24, 347–363. `https://doi.org/10.1080/14697688.2023.2291081`.

Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. Quantitative Finance 16, 887–904. Available at SSRN: `https://ssrn.com/abstract=2554754` or `http://dx.doi.org/10.2139/ssrn.2554754`.

Breeden, D.T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7. URL: `https://ssrn.com/abstract=2642306`. available at SSRN: https://ssrn.com/abstract=2642306.

Broadie, M., Glasserman, P., Jain, G., 1997. Enhanced Monte Carlo estimates for American option prices. Journal of Derivatives 5, 25–44.

Cox, J.C., Jr., J.E.I., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407. URL: `https://doi.org/10.2307/1911242`. accessed 25 June 2024.

El Euch, O., Rosenbaum, M., 2019. The characteristic function of rough Heston models. Mathematical Finance 29, 3–38.

Euch, O.E., Rosenbaum, M., 2017. Perfect hedging in rough heston models. `arXiv:1703.05049`.

Fu, M., Laprise, S., Madan, D., Su, Y., Wu, R., 2001. Pricing American options: A comparison of Monte Carlo simulation approaches. Journal of Computational Finance 4, 39–88.

Fukasawa, M., Gatheral, J., 2021. A rough SABR formula. Frontiers of Mathematical Finance doi:`10.2139/ssrn.3844278`.

Gatheral, J., Jaisson, T., Rosenbaum, M., 2014. Volatility is rough. `arXiv:1410.3394`.

Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. Quantitative Finance 18, 933–949.

Gatheral, J., Keller-Ressel, M., 2019. Affine forward variance models. Finance and Stochastics 23, 501–533. doi:`10.2139/ssrn.3105387`. available at SSRN: `https://ssrn.com/abstract=3105387` or `http://dx.doi.org/10.2139/ssrn.3105387`.

Girsanov, I.V., 1960. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability and Its Applications 5, 285–301. URL: `https://doi.org/10.1137/1105027`.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer.

Gorenflo, R., Kilbas, A.A., Mainardi, F., Rogosin, S.V., 2014. Mittag-Leffler Functions, Related Topics and Applications. Springer. This book offers an in-depth treatment of the Mittag-Leffler function, its series representation, and convergence properties.

Hagan, P., Kumar, D., Lesniewski, A., Woodward, D., 2002. Managing smile risk. Wilmott Magazine , 84–108.

Haubold, H.J., Mathai, A.M., Saxena, R.K., 2009. Mittag-leffler functions and their applications. URL: `https://arxiv.org/abs/0909.0230`, `arXiv:0909.0230`.

Heath, D., Platen, E., 2002. A variance reduction technique based on integral representations. Quantitative Finance 2, 362–369. URL: `https://ideas.repec.org/a/taf/quantf/v2y2002i5p362-369.html`.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies 6, 327–343. URL: `http://www.jstor.org/stable/2962057`.

Ingersoll, J., 1987. Theory of Financial Decision Making. Rowman & Littlefield.

Jaisson, T., Rosenbaum, M., 2015. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed hawkes processes. URL: `https://arxiv.org/abs/1504.03100`, `arXiv:1504.03100`.

Kloeden, P., Platen, E., 2008. Numerical solution of stochastic differential equations. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council 19, 1991.

Mathai, A.M., 2008. Mittag-Leffler Functions and Fractional Calculus. Springer New York, New York, NY. pp. 79–134. URL: `https://doi.org/10.1007/978-0-387-75894-7_2`, doi:`10.1007/978-0-387-75894-7_2`.

Mudzimbabwe, W., Patidar, K., Witbooi, P., 2012. A reliable numerical method to price arithmetic Asian options. Applied Mathematics and Computation 218, 10934–10942.

Musiela, M., 2016. Analysis and development of the sabr framework. Workshop presentation. URL: `https://mail.qgroup.org.au/content/analysis-and-development-sabr-framework`. presented on Wed, 02/03/2016 at 5:30pm.

Richard, A., Tan, X., Yang, F., 2022. On the discrete-time simulation of the rough heston model. `arXiv:2107.07835`.

Yin, L., 2015. Monte Carlo Strategies in Option Pricing for SABR Model. Ph.D. thesis. e University of North Carolina.

Zhang, J., 2003. Pricing continuously sampled asian options with perturbation method. Journal of Futures Markets 23, 535–560.

Climate Risk Analysis of the South African Interest Rate Swap Market

TEAM 4 Riskworx

KEFENTSE FREDDY DIPUDI, University of Cape Town
ANDREW RANEY, University of Cape Town
KATHERINE VAN DER MERWE, Stellenbosch University
HANNES KRUGER, University of Cape Town

Supervisors:
ANDREA MACRINA, University College London
THOMAS MCWALTER, University of Cape Town

African Institute of Financial Markets and Risk Management

Contents

Abstract

This report introduces a robust framework for analyzing transition risk in the trading book, employing the Nelson-Siegel-Svensson model to parameterize the yield curve and a calibrated three-factor short rate model to estimate profit and loss distributions for interest rate derivatives. We find that transition risk shocks similar to those identified by ISDA align closely with impacts observed during the COVID-19 pandemic in the USA, as indicated by changes in swap and swaption distributions and Tail Value at Risk estimates. However, this correlation does not extend to the South African market. The framework is adaptable for potential extension to the banking book and inclusion of additional risk factors. This work contributes to the advancement of climate risk quantification, consistent with the Fundamental Review of the Trading Book (FRTB) through the use of Expected Tail Loss for comprehensive risk assessment.

Keywords: Transition Risk, Trading Book, Nelson-Siegel-Svensson Model, Three-Factor Short Rate Model, ISDA, COVID-19 Pandemic, Climate Risk, FRTB.

Acknowledgements

Chapter 1

Introduction

In an era where the impacts of climate change are becoming increasingly apparent, integrating climate risk analysis into the banking and trading books is no longer a choice but a necessity. To begin with, the focus of climate risk analysis has been on the long-term effects and climate shock implications in the banking book. However, this emphasis on the long term has paved the way for developing a robust framework for climate risk analysis in the trading book. By ensuring that both long-term and short-term climate risks are consistently addressed, this framework aligns the banking book and trading book, enabling traders and risk managers to better navigate the complexities of an evolving market landscape.

The International Swaps and Derivatives Association (ISDA) has commissioned practitioners to develop and implement climate risk scenarios specifically for the trading book. This framework addresses various critical use cases, including stress testing, internal capital adequacy assessment, risk management, and strategy development. It outlines coherent and plausible climate scenarios that produce climate shocks, which in turn generate shocks to macroeconomic variables. By incorporating these scenarios, financial institutions can better prepare for and manage the potential impacts of climate change on their trading activities.

The scenarios encompass physical risk, transition risk, and combined risk scenarios. The physical scenario is characterized by a sudden deterioration in the environment due to increased global emissions. This results in an instantaneous rise in average global surface temperature by 1.5 degrees Celsius above the current global average (Defra et al., 2013). Transition risks arise from how effectively and rapidly an organization responds to both internal and external pressures to cut greenhouse gas emissions and shift to renewable energy sources. This transition encompasses changes in policy and legal frameworks, technological advancements, and market dynamics necessary to meet climate change mitigation and adaptation goals. The financial impact of these risks on organizations depends on the speed, nature, and

focus of these changes. Combined risk scenarios incorporate both physical and transition risks, providing a comprehensive view of potential climate-related challenges. We shall focus only on transition risk.

This is important because there is some degree of agency in having a consistent framework for modeling the impact of transition risk in the trading book. Shocks due to policy change can occur at any time from the present. As an example, in Denmark, the country's coalition government agreed this week to introduce the world's first carbon emissions tax on agriculture [1]. Starting in 2030, an annual tax of 672 kroner ($96) per cow will be introduced to account for the planet-heating emissions generated by livestock (Ziady, 2024). Farmers will face higher costs due to the carbon tax on their cattle's emissions. These increased costs might be passed on to consumers through higher prices for beef and dairy products. Some farmers might reduce their cattle herds to lower their tax burden, potentially reducing the overall output of beef and dairy. Higher food prices could contribute to an increase in inflation, especially if beef and dairy products constitute a significant portion of the consumer price index. Central banks might respond to rising inflation by adjusting interest rates, which in turn affects the sovereign bond curve.

Another example of a policy shift is the European Union's Carbon Border Adjustment Mechanism (CBAM). This market friction is designed to ensure that imported carbon-intensive goods reflect their true environmental cost. By applying a fair price on the carbon emissions associated with their production, CBAM aims to promote cleaner industrial practices in countries outside the EU. In its initial phase, CBAM focuses on specific sectors. These include cement, iron and steel, aluminum, fertilizers, electricity, and hydrogen. Therefore, CBAM is identified as a transition risk that could have a significant influence on the financial markets of South Africa and the US. CBAM is planned to be implemented in January 2026, and it aims at eliminating carbon leakage.

In this report, we apply the framework to the interest rate markets of the United States of America (USA) and the Republic of South Africa (RSA), focusing particularly on South Africa's interest rate swap market known for its liquidity and susceptibility to macroeconomic shocks such as GDP and inflation fluctuations. We are aware of the spill-over risk between the US market and SA swap market, hence in our analysis we consider the two jurisdictions in isolation. This report adds to the ISDA working group's analysis, which previously covered European, Asian, and American markets but excluded Africa. We enhance the analysis of the US interest rate market by adopting a three-factor interest rate model instead of the traditional one-factor Hull-White model, applying the same approach to South Africa. Our model draws on the methodologies of Babbs and Nowman (1999); Duffie and Kan

---

[1]https://edition.cnn.com/2024/06/26/business/denmark-cows-carbon-tax/index.html

(1996); Pang (1997); Beaglehole and Tenney (1991) and Nunes et al. (1999).

We obtain the government yield curve using the Nelson-Siegel-Svensson (NSS) parametric representation (Svensson, 1995). This representation allows us to characterize the yield curve in terms of its level, slope, and curvature (Gilli et al., 2010) providing insights into how transition risk affects the sovereign bond curve. We then calibrate our three-factor short rate model to historical yield curve data using the Kalman filter (Nunes et al., 1999; Babbs and Nowman, 1999; Huang et al., 2009; Wells, 2013). Time series analysis of the NSS parameters is performed on historical data pre-, post, and at times of shocks, including the DotCom crash, Global Financial Crisis (GFC), and Covid-19. This analysis is used to develop the corresponding shocks on the sovereign bond curve. This is not a trivial exercise, as it enables the observation of how the different financial crises affect the sovereign bond curve and with what magnitude, therefore enabling us to analyse the coherence and plausibility of our framework.

Given scenarios, we then price swaps and swaptions using our computationally efficient pricing engine. We use our pricing engine to perform risk analysis on the fixed-income trading book by considering the distribution of profit and loss, given the policy shocks. Our framework is in line with the fundamental review of the trading book (FRTB), which forms part of the Basel Committee's broader agenda to reform regulatory standards for banks. We use the expected shortfall as a quantitative risk metric, as it accounts for the tail risk in a more comprehensive manner, considering both the size and likelihood of losses above a certain threshold (Porretta and Agnese, 2021). We calculate the expected shortfall for periods of significant financial stress.

Figure 1.1 serves as a road map for the framework that is implemented. The first part starts with the calibration of the three-factor Vasicek model on historical bond yield data, using the Kalman filter. The model parameters and the state variables are then used in the pricing engine to calculate probability distributions for profit and loss for interest rate derivatives. The second part pertains to the NSS parameterization of the yield curve. On the one hand we have the 'business as usual' case and on the other hand we have changed NSS parameters due to certain shocks to the yield curve. The estimated state variables, corresponding to these changes in the yield curve, are then estimated. Lastly, the pricing engine is used to price swaps and swaptions and obtain profit and loss distributions for each shock for these interest rate derivatives, facilitating risk analysis for the given scenarios.

The report proceeds as follows: in Section 1 we outline the parameterisation of the government bond curve using the Nelson-Siegel-Svensson function and we discuss the 3-factor model of Nunes et al. (1999); Huang et al. (2009) and its calibration

using the Kalman filter in Section 2. Sections 1 and 2 are the quantitative finance building blocks for climate risk analysis in the trading book framework discussed in Section 4.1. We then discuss the results in Section 3 and conclude this report in Section 4.

Model calibration

Historical yield curve data
Three-factor Vasicek model
Kalman filter

Model parameters
State variables

Pricing

Swaps and swaptions
PnL distributions

Risk metrics

NSS curve

Business as usual

Estimated state variables
corresponding to
the change in NSS yield

Change in NSS parameters

Shocked NSS curve

DotCom
Global financial crisis
COVID-19
Hypothetical transition risk

Figure 1.1: Our road map.

Chapter 2

Framework

This chapter discusses the necessary theory that is required for use in the rest of the document. It outlines the necessary yield-curve modeling framework, the Nelson-Siegel-Svensson paramaterisation, the Kalman filter algorithm for calibration of our interest-rate model, the pricing engine for generating profit and loss distribution of swaps and swaptions and, finally, explains the necessary machinery for climate risk analysis.

1 Yield Curve Modelling

The Nelson-Siegel-Svensson (NSS) model, an extension of the Nelson-Siegel (NS) model, is widely used in finance as a parametric representation of the term structure of interest rates. This term structure represents the relationship between bond yields and their maturities, which is crucial for valuing bonds, managing risk, and formulating monetary policy. Nelson and Siegel (1987) introduced a straightforward, parsimonious model capable of depicting the variety of forms typically associated with yield curves. These yield curves exhibit characteristics such as monotonicity, humps, or occasional S-shapes, which reflect diverse term structures of interest rates.

Let $y(\Gamma)$ be the zero coupon rate for maturity T. The yield curve as parameterised by Nelson and Siegel (1987) can be formulated as follows,

$$y(\Gamma) = \beta_0 + \beta_1 \left(\frac{1 - e^{(-\frac{T}{\tau_1})}}{\frac{T}{\tau_1}} \right) + \beta_2 \left(\frac{1 - e^{(-\frac{T}{\tau_1})}}{\frac{T}{\tau_1}} - e^{(-\frac{T}{\tau_1})} \right)$$

where the vector of constants $\Gamma = (\beta_0, \beta_1, \beta_2, \tau_1)$ characterises the model.

According to Gilli et al. (2010), the parameter β_0 represents the level of the yield curve and is independent of T, often interpreted as the long-run yield level. The

parameter β_1 influences the slope of the yield curve and is weighted by a function of time to maturity. This function is unity when $T = 0$ and decays exponentially to zero as T increases, meaning the impact of β_1 is primarily felt at the short end of the curve. The parameter β_2 accounts for the curvature of the yield curve. This parameter effect on the yield curve is zero at $T = 0$, increases to create a hump, and then decreases back to zero as T grows, thereby introducing a hump-shaped curvature to the yield curve. Finally, the parameter τ_1 determines the rate of decay and specifically controls the position of the hump on the yield curve.

The model constraints are,

$$\beta_0 > 0, \beta_0 + \beta_1 > 0, \tau_1 > 0.$$

Svensson extends the NS model by adding an extra term to capture the shapes of yield curves better, especially those with multiple humps or U-shapes. This allows more than one local extreme along the maturity profile and hence can be useful in improving the fit of the yield curves. The extended model is defined by

$$y(\Gamma) = \beta_0 + \beta_1 \left(\frac{1 - e^{\left(-\frac{T}{\tau_1}\right)}}{\frac{T}{\tau_1}} \right) + \beta_2 \left(\frac{1 - e^{\left(-\frac{T}{\tau_1}\right)}}{\frac{T}{\tau_1}} - e^{\left(-\frac{T}{\tau_1}\right)} \right) \tag{2.1}$$

$$+ \beta_3 \left(\frac{1 - e^{\left(-\frac{T}{\tau_2}\right)}}{\frac{T}{\tau_2}} - e^{\left(-\frac{T}{\tau_2}\right)} \right) \tag{2.2}$$

where $\Gamma = (\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$. The constraints are the same as the NS model, with $\tau_2 > 0$. The inclusion of β_3 and τ_2 enables the model to achieve more flexibility. β_3 adds an additional hump term to the NS model. According to Gürkaynak et al. (2007), the NSS model will accurately capture the shape of the yield curve whilst simultaneously smoothing out any idiosyncratic issues for specific financial securities, such as liquidity premia, hedging demand or market segmentation. The model allows for two humps in the yield curve at the short end and the long end. Monetary policy expectations are accounted for at the short end, while the long end takes convexity effects and market segmentation into account Gürkaynak et al. (2007).

Once NSS parameters have been estimated for (2.2), these parameters can be used to estimate yields for various maturities. Svensson (1995) describes methods for estimating the parameters; either minimising the sum of squared price errors or the sum of squared of yield errors. Minimising the price errors often results in fairly large yield errors for bonds and money market rates for short maturities due to the insensitivity of yields to prices for short maturities. Minimising yield errors directly minimizes the differences between observed yields and the yields predicted

by the model. We will minimize yield errors, as this approach provides a better fit for the term structure of interest rates across different maturities.

Gilli et al. (2010) describes the optimisation problem as

$$\min_{\Gamma} \sum (y - y^M)^2$$

where y are the model yields and y^M are the observed yields, subject to the constraints above.

2   Kalman Filter Algorithm and Three-Factor Vasicek model

Kalman filtering and Maximum Likelihood Estimation(MLE) are often used in combination to estimate optimal model parameters (Chatterjee, 2005). A brief overview of these two methods, as well as the calibration procedure, is given below.

Applying the Kalman filter requires us to formulate the transition and measurement equation and they are given as follows respectively,

$$x_k = A_{k-1} x_{k-1} + a_{k-1} + w_k,$$

$$y_k = H_k x_k + h_k + v_k.$$

The transition equation describes how the hidden state evolves while the measurement equation relates the observed values to the hidden states (Duan and Simonato, 1999). Here x_k is a hidden state vector (the short rate), y_k denotes a vector of observations (bond yields), w_k and v_k are zero-mean uncorrelated noise vectors, furthermore, they have covariance matrices Q_k and R_k, respectively. The Kalman filter uses the observation vector to determine the state vector by using a prediction step followed by an update step.

In the update step, the log-likelihood can be calculated. Here, the main idea is to find the set of parameters that are most likely to yield the given observations. This is achieved by maximising the log-likelihood function with respect to the model parameters, as suggested by James and Webber (2000).

The above is detailed by the following equations, providing a general summary of the algorithm that is implemented:

1. Prediction:

$$\hat{x}_k^- = A_{k-1}\hat{x}_{k-1}^+ + a_{k-1}$$

$$P_k^- = A_{k-1}P_{k-1}^+ A_{k-1}^T + Q_k$$

2. Update:

$$F_k = H_k P_k^- H_k^T + R_k$$

$$K_k = P_k^- H_k^T F_k^{-1}$$

$$\bar{y}_k = H_k \hat{x}_k^- + h_k$$

$$P_k^+ = (I - K_k H_k)P_k^-(I - K_k H_k)^T + K_k R_k K_k^T$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - \bar{y}_k)$$

$$LL_k = LL_{k-1} - \frac{d_y}{2}\log(2\pi) - \frac{1}{2}\log(\det(F_k)) - \frac{1}{2}(y_k - \bar{y}_k)^T F_k^{-1}(y_k - \bar{y}_k)$$

To apply this algorithm we initialise values for the prediction and update step, i.e., $\hat{x}_0^+$, P_0^+ and LL_k respectively. The Kalman filter is known to converge quickly, therefore this initialization does not have to be perfect, according to Chatterjee (2005).

We have outlined the Kalman filter procedure, now we consider the three-factor Vasicek model and represent it as a system of stochastic differential equations

$$dX_t = K(\bar{X} - X_t)dt + \Sigma dW_t^{\mathbb{Q}}, \tag{2.3}$$

where

$$X_t = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad K = \begin{bmatrix} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{bmatrix} \quad \bar{X} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}.$$

X_t is the hidden state process, κ_i for $i = 1, 2, 3$ are the rates of mean reversion of the hidden state variables to their long-term means, θ_i for $i = 1, 2, 3$ and W_t is a vector of independent Brownian motions, $W_t \in \mathbb{R}^3$. The martingale measure, $\mathbb{Q}$, denotes the probability measure obtained when the bank account is the numeraire asset underlying the model.

The Duffie and Kan (1996) model, presented in Nunes et al. (1999), gives the price of a risk-free zero coupon bond as

$$P(t, T) = \exp\left[A(\tau) + B^{\mathrm{T}}(\tau)X_t\right], \tag{2.4}$$

where

$$A(\tau) = \tau(G^{\mathrm{T}}a^{-1}b - f) + B^{\mathrm{T}}(\tau)a^{-1}b$$
$$+ \frac{\tau}{2}G^{\mathrm{T}}a^{-1}\Theta(a^{-1})^{\mathrm{T}}G$$
$$+ G^{\mathrm{T}}a^{-1}(I_n - e^{a\tau})\Theta(a^{-1})^{\mathrm{T}}G$$
$$+ \frac{\tau}{2}G^{\mathrm{T}}a^{-1}\Delta(\tau)(a^{-1})^{\mathrm{T}}G$$
$$B^{\mathrm{T}}(\tau) = G^{\mathrm{T}}a^{-1}(I_n - e^{a\tau})$$
$$\Delta(\tau) = e^{a\tau}Y e^{a^{\mathrm{T}}\tau} - Y$$

$$\Theta = \Sigma\Sigma^{\mathrm{T}}.$$

Y is computed the same way as in Nunes et al. (1999). The short rate r_t is given by

$$r_t = f + G^{\mathrm{T}}X_t. \tag{2.5}$$

Abu-Mostafa (2001) mentions that for the three-factor generalised Vasicek model, the short rate is simply the sum of the state variables. Which implies that $f = 0$ and $G = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

The SDE of X_t under the deterministic volatility formulation of the Duffie and Kan (1996) model has the following form

$$dX_t = [aX_t + b]dt + \Sigma dW_t. \tag{2.6}$$

From (2.3), it can be deduced that $a = -K$ and $b = K\bar{X}$. From the above, the bond yields can be calculated using

$$y(\tau) = -\frac{1}{\tau}[A(\tau) + B^{\mathrm{T}}(\tau)X_t]. \tag{2.7}$$

For the implementation of the Kalman filter, the transition equation based on Euler's discretization is

$$X_{t+\Delta t} = (I_3 - K\Delta t)X_t + K\bar{X}\Delta t + \sqrt{\Delta t}\Sigma Z_t \tag{2.8}$$

where $Z_t \sim N_3(0, I_3)$. Following from (2.7), the measurement equation is

$$y(\tau) = -\frac{1}{\tau}B^{\mathrm{T}}(\tau)X_t - \frac{1}{\tau}A(\tau) + v \tag{2.9}$$

where v is the noise vector.

9

3 The Pricing Engine

The pricing engine was developed to provide risk-neutral prices for fixed-interest financial instruments and derivatives making use of the estimated state process and optimised parameters provided by the calibrated interest rate model. These risk-neutral prices are then used in a one year profit and loss analysis under the 'Business as usual' scenario relative to the stressed scenarios considered. The engine provides the distributions of risk-neutral prices of interest rate swaps, bonds and swaptions for any inputed maturities for base and shocked scenarios one year after the shock.

The instantaneous shocks were implemented in the model by changing the initial state variable from which the discretised state process is evolved. This analysis considered maturities of half a year, one year, three years, five years, ten years and twenty years as these give insight into the impact of shocks on short, medium and long-term maturities and tenors. The derivative instrument considered was a two year swaption, valued one year after the shock, on swaps of the above-mentioned tenors. For this analysis, only payer swaps and swaptions on payer swaps with nominals of one million were considered, however, it could easily be extended to receiver swaps with any chosen nominal or portfolios of swaps. The probability distributions, using Kernel Density Estimation (KDE), of the Monte Carlo prices of the instruments were then plotted for our 'business as usual' scenario relative to each shocked scenario. This was done to see how the distribution of swap and swaption values changed under shocks to the yield curve for each maturity considered. The Monte Carlo procedure used $100\,000$ simulations for increased accuracy.

The value at risk (VaR) and tail value at risk (TVaR) were calculated for each distribution of swap prices and swaption prices. This facilitated comparison of these risk metrics and how they changed when shocked. VaR was calculated using equation (2.10) and TVaR was calculated using equation (2.11), where X is the price distribution of the instrument in question and α is the quantile at which losses are deemed extreme. The α parameter used was 0.25%, in order to achieve consistency with the confidence level of 97.5% recommended in BIS (2013).

$$\text{VaR}_\alpha(X) = -\inf\{x \in \mathbb{R} : F_X(x) > \alpha\} \tag{2.10}$$

$$\text{TVaR}_\alpha(X) = \mathbb{E}(-X|X \leq -\text{VaR}_\alpha(X)) \tag{2.11}$$

The use of TVaR is consistent with the recommendations in BIS (2013) as deficiencies have been pointed out concerning the use of VaR as a risk measure as it does not sufficiently quantify tail risk. TVaR quantifies both size and the likelihood of losses beyond a confidence level and is, thus, preferred BIS (2013).

4 Climate Risk Analysis in Trading Book

In this section we outline the general framework for analysis of transition risk in the trading book. Section 4.1 outlines the framework and in section 4.2 we discuss the application of this framework to historical events.

4.1 Climate Risk Analysis in Trading Book

We start by calibrating the 3-factor Vasicek model (2.3) to historical data using the Kalman filter, this calibration is under the risk neutral measure and we assume no transition risk. The price of the bond under the calibrated model is given by,

$$P_{tT} := P(t, T; X_t, \Theta), \tag{2.12}$$

where Θ is the set of parameters that characterise (2.3), given information up to time t.

To analyse the impact of transition risk on the trading book, we consider a finite set $\mathcal{P}$ of transition scenarios, defined by

$$\mathcal{P} := \{p_j | j = 0, 1, 2, ...\},$$

and the sovereign bond curve for our segmented sovereign jurisdiction corresponding to the transition scenario p_j, which characterized by NSS^j [1]. The yield of a zero coupon bond with maturity T_i is given by,

$$R^j_{t, T_i} := R(t, T_i; \Gamma^j), \tag{2.13}$$

where Γ^j is the set of NSS^j parameters corresponding to scenario j. The current scenario or 'business as usual' can be indexed by $j = 0$. To determine the corresponding Γ^j for j, which characterise the sovereign bond curve induced by scenario j, we first need to determine the magnitude of the scenario shock to Γ^0, that if applied to the bond curve will produce $\Gamma^j, j > 0$ of j scenario. Given the shocks corresponding to each transition scenario, we can calculate the corresponding NSS parameters Γ^j for the government bond curve, which can then be used to perform climate risk impact on the trading book.

To assess transition risk impact on the trading book for a given arbitrary transition scenario $j > 0$, we begin by estimating the initial state variables for our calibrated model with set model parameters Θ. To obtain the shocked state variables, shocks are first determined through scenario shocks of changes in yield curves through NSS parameters. The shocked yield curves can be obtained by minimising the error between yield curves obtained with state variables as described by (2.4) and

---

[1] We omit p and index scenario p_j by j

the shocked NSS yield curve. The state variables that minimise these are then regarded as the shocked state variables. Once we have the shocked initial state variable from the minimisation procedure in (2), we evolve the state process from this point to price interest-rate derivatives and determine their profit and loss distribution, given the stressed scenario. We calculate the expected tail loss for both 'business as usual' and the stressed scenario, j, for comparative analysis.

Types of shocks to the model include shocks to only the state variables and shocks to state variables and model parameters. Shocks to the state variables correspond to the case where the initial state variable, X_t, changes due to the policy shock induced by p_j, with model parameters in Θ remaining unchanged. Thus, the long-term mean of our model (2.3) remains the same, and our state variables revert to the same long-term mean. Shocks to state variables and model parameters result in a regime shift in our state variable process. Given Γ^j, we find the corresponding Θ, which means our 3-factor short rate model dynamics will change, prompting short rates to revert to a new long-term mean. For a given shock we then price zero-coupon bonds, and interest rates derivatives. This analysis we only consider shocks to the state variables as this corresponds with the one year time horizon of the trading book, however further analysis could consider changes to state variables and model parameters in response to a shock.

Risk analyses on the fixed-income book is done by considering the distribution of profit and loss, given the specific policy shocks. We apply this analysis to the SA and USA fixed-income market. To ensure that our analysis is consistent with the fundamental review of the trading book (FRTB), we use expected shortfall as a quantitative risk metric, as it accounts for the tail risk in a more comprehensive manner, considering both the size and likelihood of losses above a certain threshold (Porretta and Agnese, 2021).

4.2 Application of the Framework to Historical Events

We consider three historical crises that had an extensive impact on the financial markets, namely the Dotcom Bubble, the Global Financial Crisis (GFC) and the COVID-19 pandemic. We obtain the sovereign yield curve pre-event and post-event and perform an analysis of the NSS model parameters to determine the change in percentage in the model parameters. It is important to emphasize that we are not assuming climate change policy shocks are of an equivalent magnitude to the past financial crisis. This is not a trivial exercise, as it enables us to observe how the different financial crises affect the sovereign bond curve, gauge the magnitude of the effect and check the coherence and plausibility of our framework.

We perform risk analysis for each historical crisis, to the SA and USA fixed-income

market using the framework in Section 4.1.

Chapter 3

Numerical Results

1 Calibration Results

The calibration results of the three-factor Vasicek model on the historical USA and RSA sovereign bond data can be seen in the Appendix. For both of these jurisdictions, the model is capable of accurately predicting the yields of bonds with different maturities, using the estimated short rate and the estimated model parameters. Thus, validating its use in the pricing engine for swaps and swaptions. In the next section we discuss the results of the transition risk analysis in the trading book.

2 Risk Analysis on the Trading book

In this section, we examine the impact of financial shocks on the trading book. We begin by analyzing the shocks to the sovereign bond curve in Section **??**. We then observe how these shocks affect the interest rate swap and swaption markets in the USA (Section 2.2) and RSA (Section 2.3), with a primary focus on the latter.

2.1 Financial Shocks on the Sovereign Bond Curve

In this section, we analyse how different absolute spread changes (as proposed by ISDA and given in basis points) of government bond yields for different liquidity horizons affect the NSS parameterisation. We also discuss the results obtained from a time series analysis of the NSS model parameters performed on historical data pre-crisis, post-crisis, and at the time of the crisis. These crises include the Dot-Com Bubble, GFC, and Covid-19. We observe how these different financial crises affect the sovereign bond curve.

Table 3.1 and 3.2 summarise the basis point (bps) spread changes in yields of bonds with different maturities, for different liquidity horizons, for the USA and RSA. Table 3.1 was obtained from the ISDA (2023, 2024) documentation, while Table 3.2

was obtained by changing the parameters of the NSS paramaterisation (β and τ parameters) to replicate the spread changes reported in ISDA (2023, 2024).

Table 3.1: Absolute spread change in basis points of government bond yields for various liquidity horizons, as proposed by ISDA.

ISDA	USA			
	1D	10D	3M	1Y
GOV Bond 1D	30	70	80	160
GOV Bond 6M	25	60	70	150
GOV Bond 1Y	20	50	60	140

Table 3.2: Absolute spread change in basis points of government bond yields for various liquidity horizons possible with the NSS model.

This report	USA			
	1D	10D	3M	1Y
GOV Bond 1D	33	63	87	161
GOV Bond 6M	1.3	12	63	160
GOV Bond 1Y	1	10	53	137

From Table 3.1 and Table 3.2, it is clear that there is a discrepancy between the spread changes of the two longer maturities over the two shorter liquidity horizons. However, for the two longer liquidity horizons, the spreads agree within 10 basis points for all maturities. This also holds true for the one day bond and the shorter liquidity horizons. The reason for this can only be speculative as it is unknown how ISDA obtained their proposed shocks. Due to this restriction, and given that the NSS parameterisation does achieve the majority of the proposed spread changes, the NSS parameterisation is considered satisfactory for the purposes of this report.

Figure 3.3 illustrates the NSS parameters obtained for yield curves corresponding to 'business as usual' for both the USA and RSA. The observed changes in the shape of the yield curve due to shocks are contingent upon the initial values of the NSS parameters, as these values serve as the baseline from which percentage adjustments are made to derive new yield curves.

The percentage change in the NSS parameters for various shocks has driven significant shifts in yield curve dynamics. The percentage changes obtained for each shock can be found in Table 3.4. The magnitude and direction of these percentage changes, relative to the original parameters, play a crucial role in shaping the resultant yield curves. This dependency underscores the significance of understanding

both the initial parameterisation and the subsequent adjustments in elucidating the dynamics of the yield curve. This is illustrated by Figures 3.1 and 3.2, where the same shocks are applied to different 'business as usual' NSS parameters (USA and RSA), however very different changes in the yield curves are observed.

Table 3.3: Original NSS parameters - 'business as usual'.

NSS Parameters

	β_0	β_1	β_2	β_3	τ_0	τ_1
USA	2.6562	3.0513	-0.0001	4.8198	1.0437	17.3374
RSA	0.0752	-0.0303	-0.2915	0.4200	4.2727	6.3746

Table 3.4: Percentage change in the NSS parameters for different shocks.

		β_0	β_1	β_2	β_3	τ_0	τ_1
	DotCom	6.84	20.87	-8.6	29.06	22.29	-8.84
Shocks	GFC	2.51	30.17	27.68	18.84	-31.17	-29.78
	COVID-19	-15.42	89.99	61.09	57.91	-4.88	25.59
	ISDA	2.06	-1.93	228.19	68.07	0.58	123.16
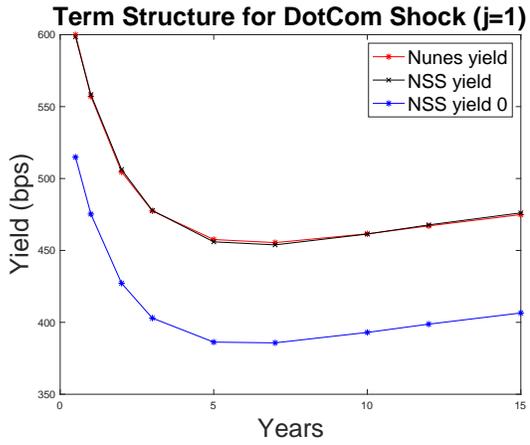
(a) Shock caused by DotCom.
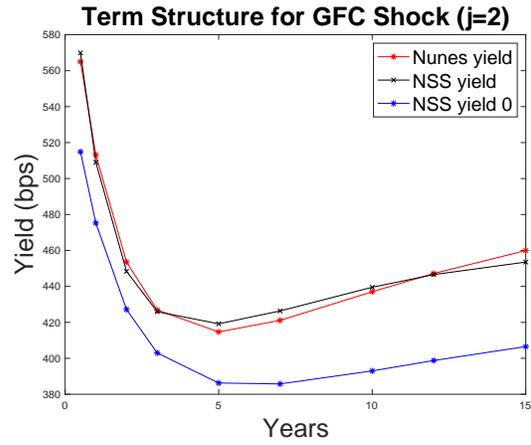
(b) Shock caused by GFC.
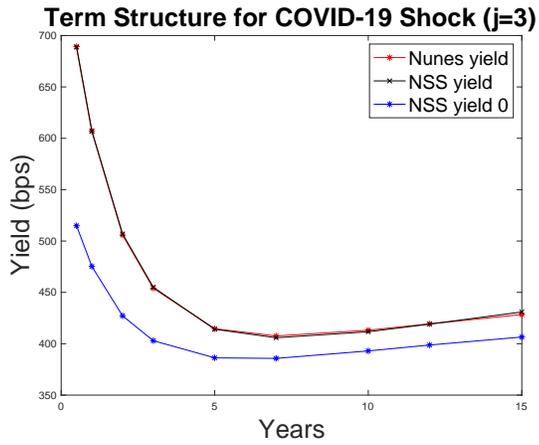
(c) Shock caused by Covid-19.

Figure 3.1: Shocks caused by Various Crises Applied to USA.
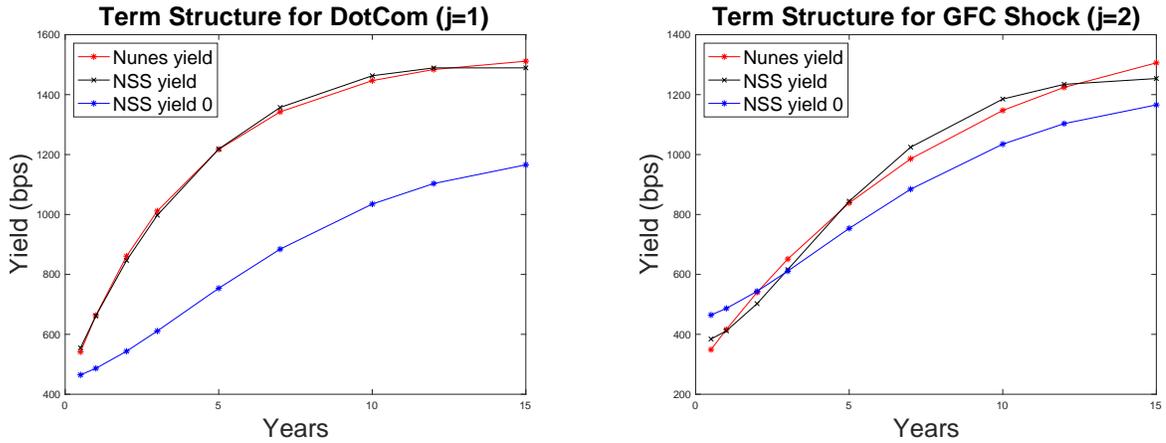
(a) Shock caused by DotCom.

(b) Shock caused by GFC.
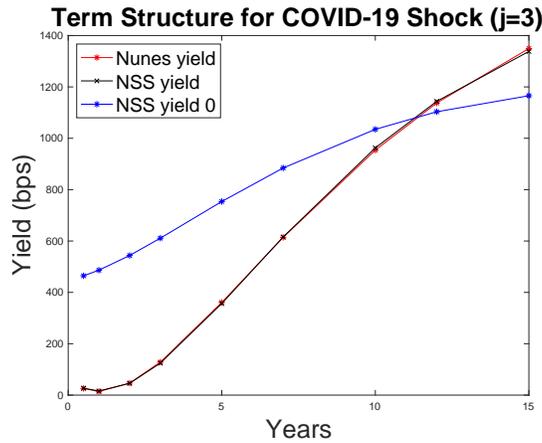
(c) Shock caused by Covid-19.

Figure 3.2: Shocks caused by various crises applied to RSA.

For the DotCom Bubble shock, the increase in β_0 by 6.84% suggests a rise in overall long-term interest rates, influencing yields across different maturities. A notable 20.87% increase in β_1 has steepened the yield curve, indicating higher short-term rates relative to longer-term rates. Conversely, a -8.6% change in β_2 adjusts the curvature, impacting medium-term yields. The substantial 29.06% increase in β_3 introduces complexity, affecting yield differentiation across maturities. Moreover, with τ_0 increasing by 22.29% and τ_1 decreasing by -8.84%, the adjustments imply prolonged effects on slope and curvature adjustments, influencing market expectations and economic conditions through dynamic interest rate dynamics. These changes can be observed in both the USA (Figure 3.1a) and RSA (Figure 3.2a), where the magnitude of the parameter percentage changes cause different shifts

18

in the original yield curve.

For the GFC shock, a slight increase of 2.51% in β_0 suggests a modest rise in overall long-term interest rates, affecting yields across different maturities. A substantial 30.17% increase in β_1 has steepened the yield curve notably, indicating higher short-term rates relative to longer-term rates. Moreover, a 27.68% increase in β_2 has intensified the curvature of the yield curve, impacting medium-term yields. The 18.84% increase in β_3 introduces additional complexity, influencing yield differentiation across various maturities. Additionally, with τ_0 decreasing by -31.17% and τ_1 decreasing by -29.78%, these adjustments suggest a faster decay in the effects of β_1 and β_3, affecting slope and curvature adjustments over time. These changes can be observed in both the USA (Figure 3.1b) and RSA (Figure 3.2b), where the magnitude of the parameter percentage changes cause different shifts in the original yield curve.

For the COVID-19 shock, a notable decrease of -15.42% in β_0 suggests a lower overall long-term interest rate environment, influencing yields across various maturities. A substantial 89.99% increase in β_1 has sharply steepened the yield curve, indicating significantly higher short-term rates relative to longer-term rates. Furthermore, a 61.09% increase in β_2 has enhanced the curvature of the yield curve, affecting medium-term yields. The 57.91% increase in β_3 introduces additional complexity, influencing yield differentiation across different maturities. Moreover, with τ_0 decreasing by -4.88% and τ_1 increasing by 25.59%, these adjustments suggest varying impacts on the decay rates of β_1 and β_3, influencing the persistence of slope and curvature adjustments over time. These changes can be observed in both the USA (Figure 3.1c) and RSA (Figure 3.2c), where the magnitude of the parameter percentage changes cause different shifts in the original yield curve.

A hypothetical transition risk shock is considered to be the NSS curve that best represents the bps spread changes provided in ISDA (2023, 2024). This assesses the magnitude of the ISDA-induced shock relative to historical shocks in terms of transition risk. The percentage changes resulting from these shocks are detailed in Table 3.4, while the corresponding alterations in the yield curve are depicted in Figure 3.3 for the USA and Figure 3.4 for RSA. β_0 experiences a modest 2.06% increase, indicating a slight rise in long-term interest rates. Conversely, β_1 decreases by -1.93%, suggesting a potential decline in short-term rates relative to longer-term rates. The most pronounced effect is seen in β_2, which surges by 228.19%, substantially altering the curvature of the yield curve and impacting medium-term yields prominently. Moreover, β_3 increases by 68.07%, introducing additional complexity and influencing yield differentiation among various maturities. The adjustments in τ_0 (-0.58%) and τ_1 (123.16%) further refine these impacts by altering the decay rates of β_2 and β_3 effects over time, respectively.
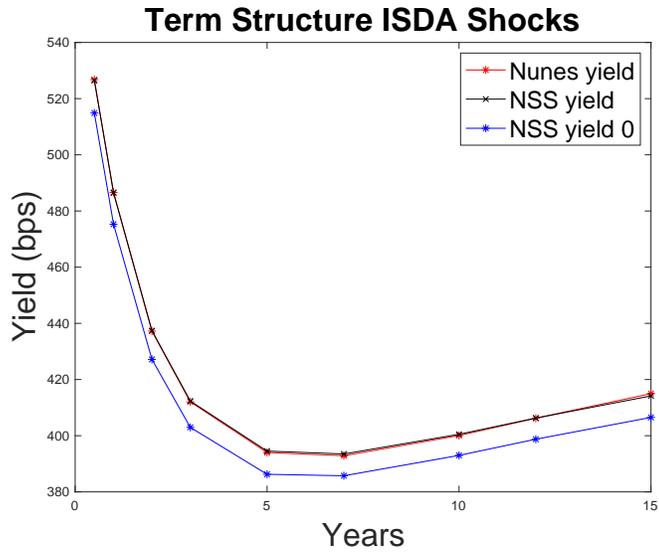
Figure 3.3: Shock produced by ISDA on USA.

Notably, the observed shift in the yield curve induced by the ISDA transition risk shock appears less pronounced compared to shifts caused by historical event shocks, but it can be argued is the most similar in shape and magnitude to the shock induced by the COVID-19 pandemic on the USA.

Term Structure for ISDA Shock (j=4)
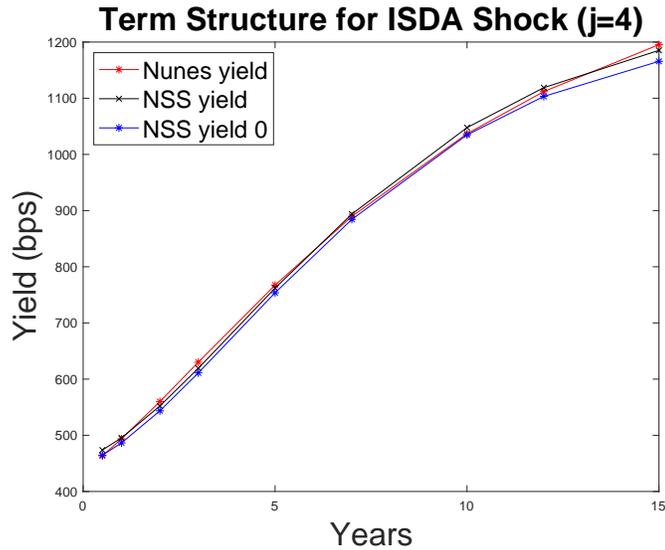
Figure 3.4: Shock produced by ISDA on RSA.

When applying the ISDA shocks to the RSA data, an even smaller shift in the yield curve is observed. Despite the smaller magnitude, the resulting curvature closely resembles that produced by the NSS yield curve under the COVID-19 shock. However, the slope generated by the ISDA shock is notably less steep than that observed in the COVID-19 shocked NSS yield curve.

After determining the shocked yield curves, the corresponding shocked state variables X_t^j are computed. This involves identifying the values of X_t^j required in the yield curve formulation of Nunes et al. (1999) (see equation (2.7)) that replicate the yield curves generated by the NSS yield curve shocks. Subsequently, these values of X_t^j are employed in the pricing engine for swaps and swaptions.

2.2   Risk Analysis on the Trading book: USA

This analysis assesses the change in swap value distributions under each stressed scenario, where the scenarios considered correspond to the Dot Com Bubble, the GFC and the Covid-19 pandemic. This is done by using the calibrated parameters from the interest-rate model, the shocked state process value provided by the NSS model and the pricing engine discussed in Section 3 of Chapter 2.

For example, this analysis will present the results of the GFC shock on the US swap and swaption markets, where the corresponding plots for the Dot Com Bubble can be found in Section 5 for brevity. We see in Figure 3.5 that there was a decrease in the value of swaps one year after the shock occurred, seen in the apparent mean

shift leftwards of the stressed scenario distribution. In this case, the shock was consistent with the magnitude and direction of the GFC. It can also be seen in Figure 3.11 that there is a marked increase in TVaR over all tenors with a shock consistent with the GFC. This indicates an increase in the tail risk of US swaps one year after the shock as well as a significant reduction in value.
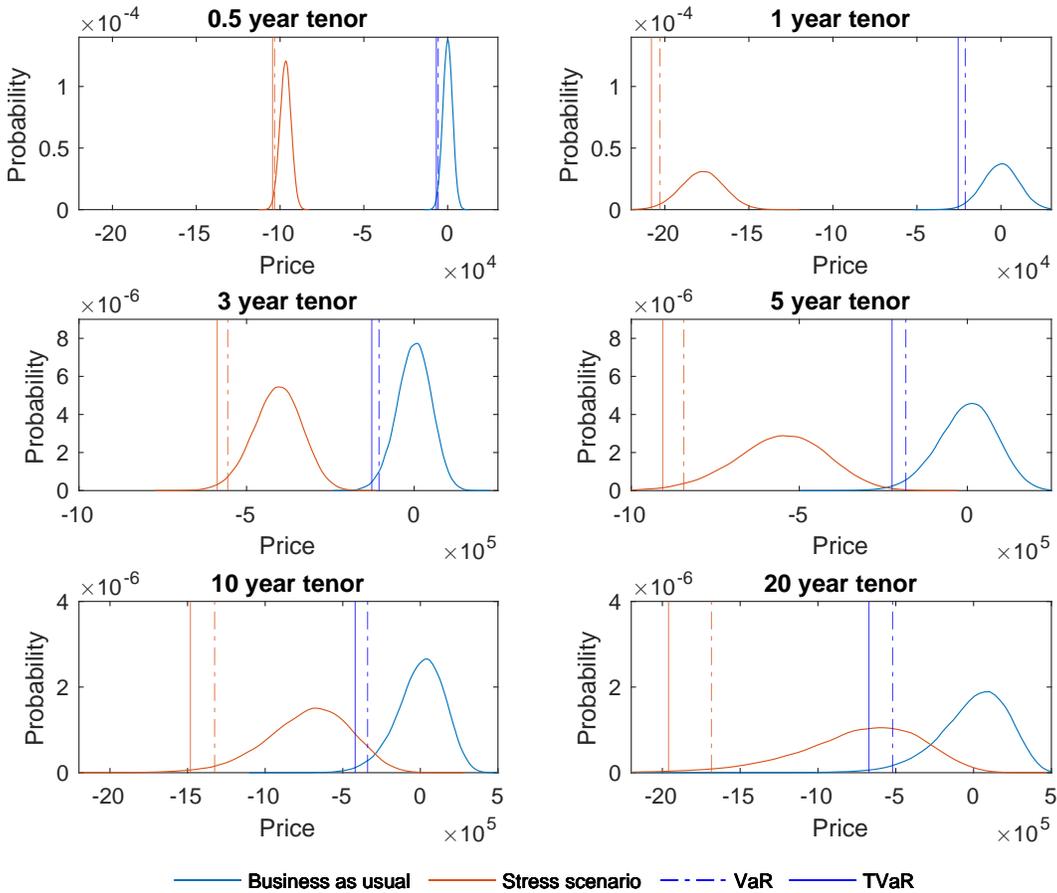
Figure 3.5: Shock to US swap prices caused by GFC.

Furthermore, there is a corresponding reduction in the value of two-year swaptions on swaps of the same tenor, seen in Figure 3.16. However, this reduction is capped, owing to their optionality, seen where there is a shift leftwards of the distribution but this becomes concentrated around zero rather than becoming significantly negative. There is a stark comparison between the significant increase in TVaR of the US swaps, relative to the capping of the TVaR of the swaptions to 0 in the stressed scenario, as seen in Figures 3.11 and 3.12. We also see significant distributional changes with increasing tenor of swaps, as the distributions become flatter with a reduction in mean and an increase in variance. However, the same behaviour does

not occur for swaptions, where distributions reduce in variance. The reduction in
the values of the swaps implies that many of the swaptions expire out of the money
and thus the distributions in the shocked scenario become concentrated around 0.
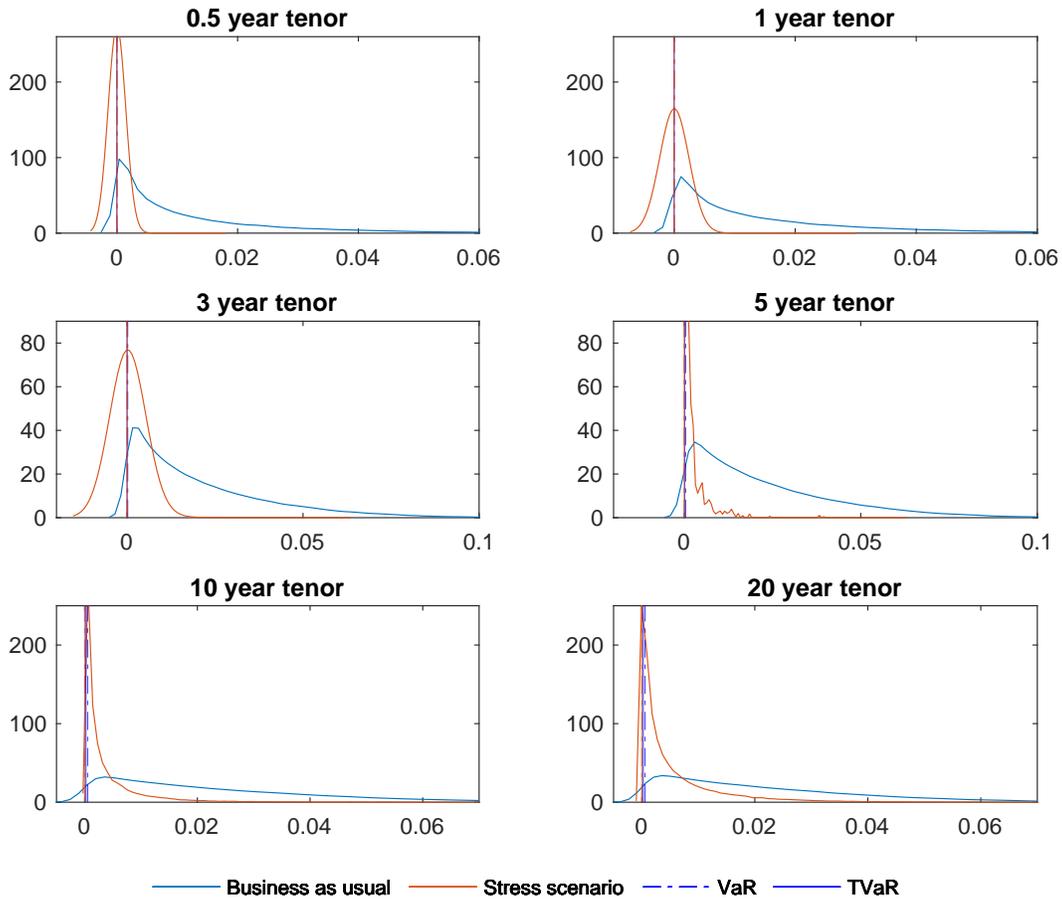
Figure 3.6: Shock to US swaption prices caused by GFC.
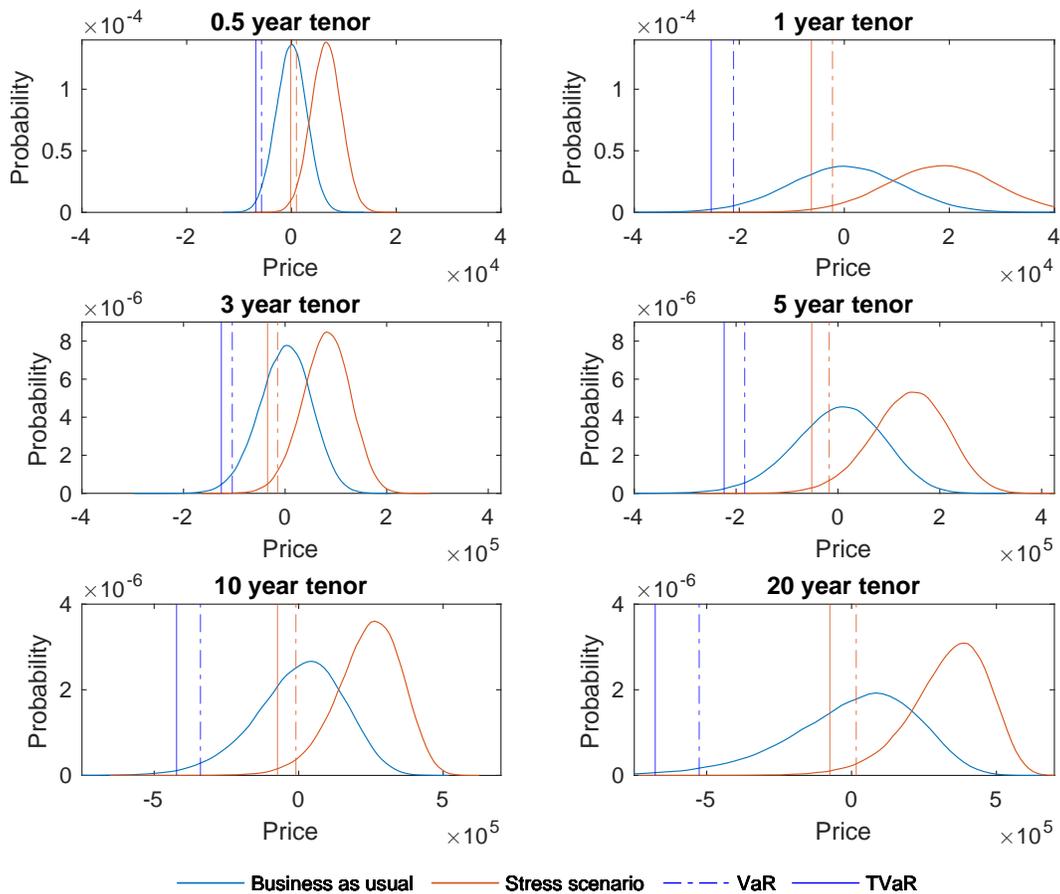
Figure 3.7: Shock to US swap prices caused by Covid-19.

We see in Figure 3.7 that there is an increase in swap prices for all maturities with a shift rightwards of all the distributions relative to the 'business as usual' scenario. For longer maturities there are greater distributional changes where there is a shift leftwards of the mean as well as an apparent reduction in variance as they are more peaked, as seen in Figure 3.7.
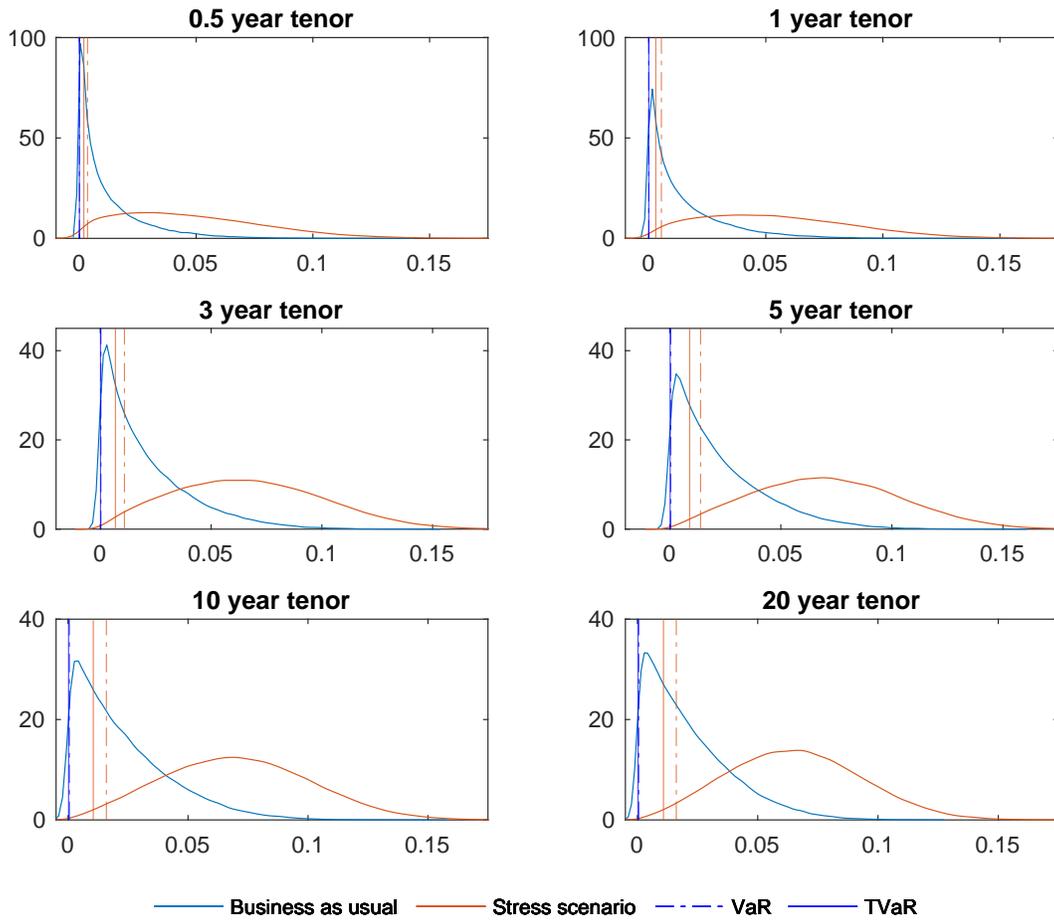
Figure 3.8: Shock to US swaption prices caused by Covid-19.

This result is the same for the swaption distributions under the Covid-19 shock scenario, as is seen in Figure 3.8, where all distributions shift to the left. There are, however, distributional changes at all maturities for the swaption prices with greater variance. The TVaR for all maturities of swaps and swaptions is also significantly reduced, as seen in Figures 3.11 and 3.18.
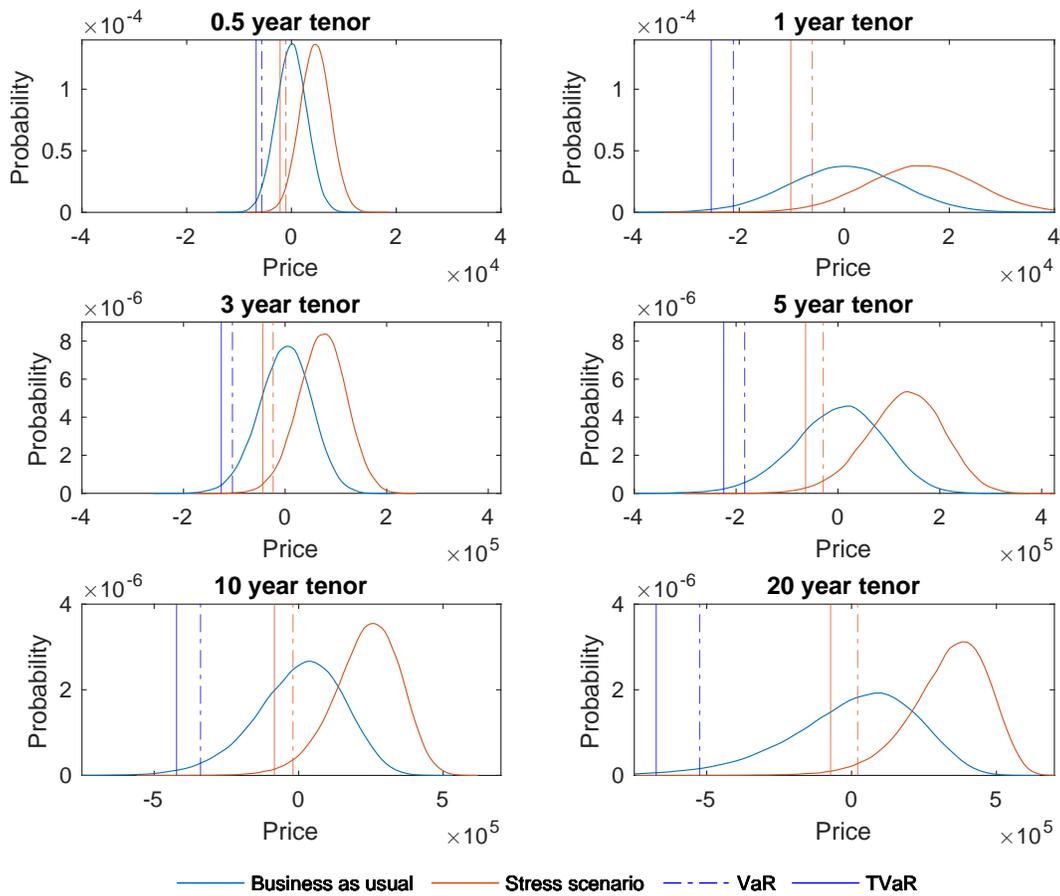
Figure 3.9: Shock to US swap prices caused by the ISDA shock.

The shock to the yield curve, consistent with the ISDA transition risk shock, seemingly had a small impact on the yield curve, as seen in Section 2.1. This can be seen in the short-term tenors being relatively unaffected, with small shifts in the distributions under the shocked scenario. However, for longer tenors, this small change to initial state has a compounded effect and precipitates a larger shift in distribution for tenors greater than three years, as seen in Figure 3.9.
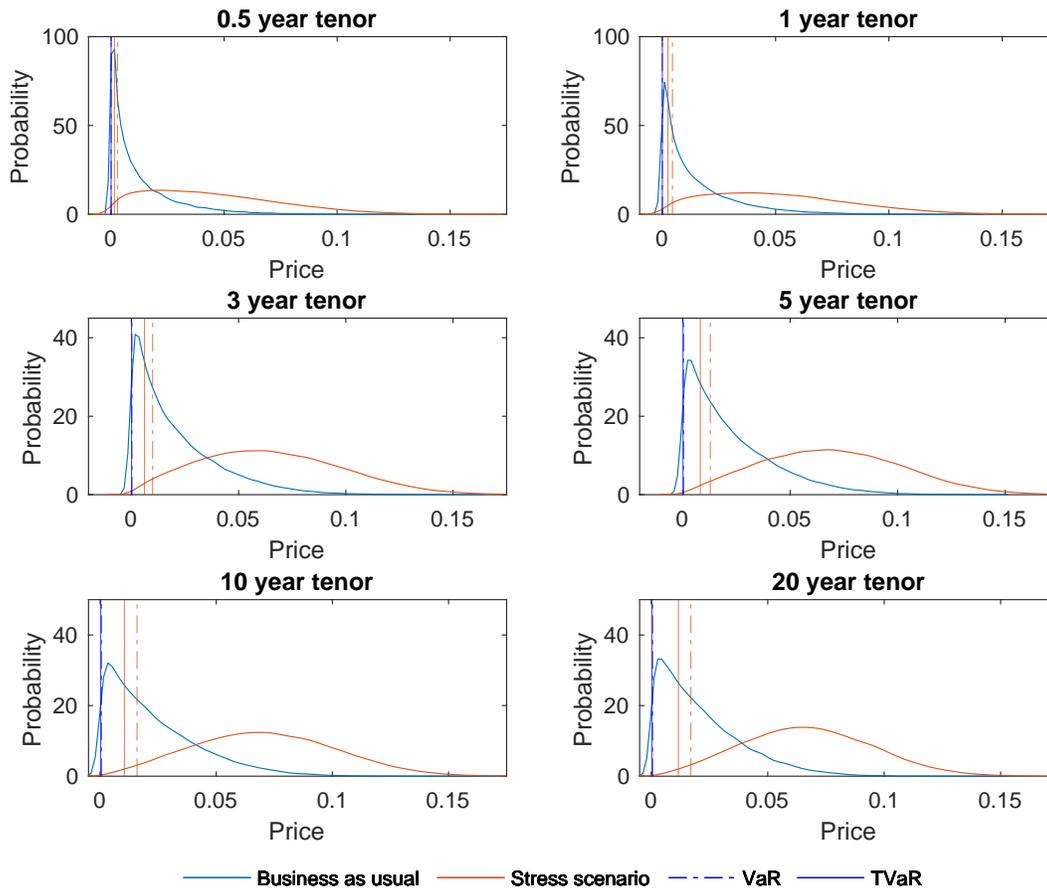
Figure 3.10: Shock to US swaption prices caused by the ISDA shock.

The first part of the previous result is also seen in swaption prices under the shock scenario proposed by ISDA. Figure 3.10 shows that, across all maturities, there is an increase in swaption prices corresponding to the increased swap prices seen in Figure 3.9. However, the distributional changes in the swaption prices are more significant than those seen in the swap prices. So the ISDA shock, whilst exhibiting a modest impact on the yield curve, does imply changes in the behaviour of swap and swaption prices.
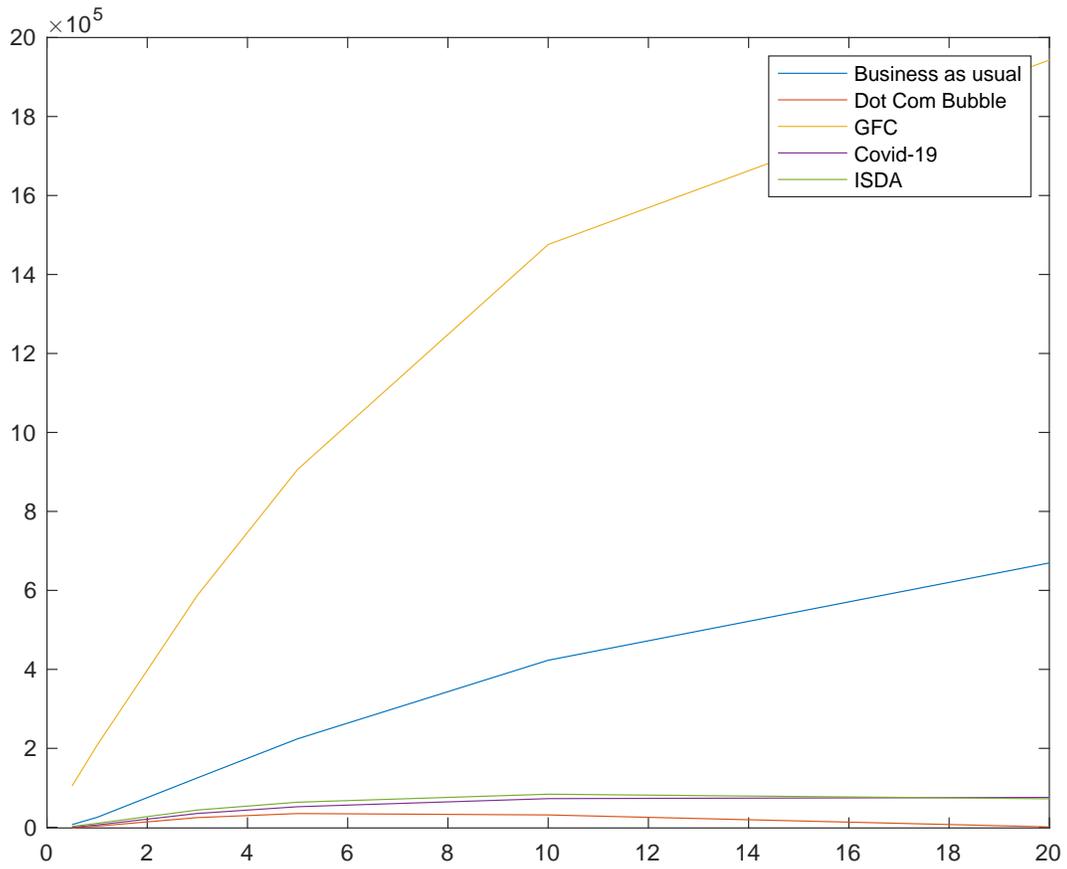
Figure 3.11: Changes in the TVaR of US swaps in each shock scenario.
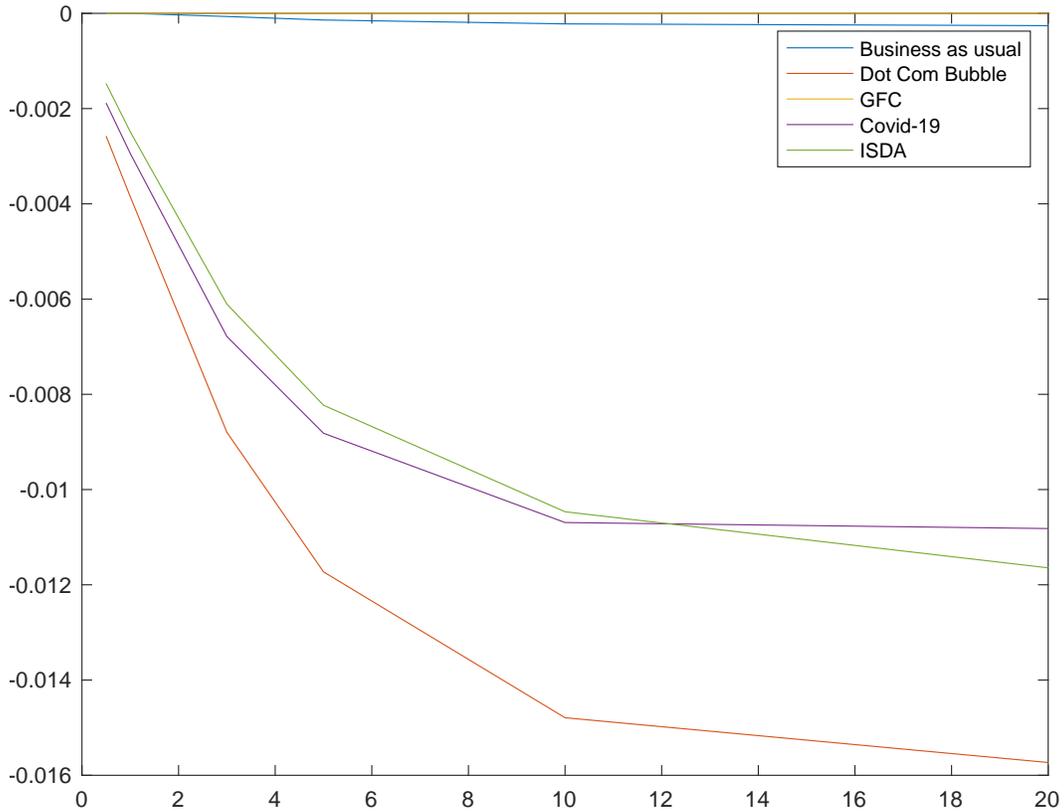
Figure 3.12: Changes in the TVaR of US swaps in each shock scenario.

We see in Figures 3.11 and 3.12 that the behaviour of TVaR under the ISDA shock behaves similarly to the behaviour of TVaR under a shock consistent with the Covid-19 pandemic. This is corroborated by the changes to swap distributions in Figures 3.7 and 3.9 behaving similarly for all maturities. Analogously, for swaptions, the distributional shifts seen in Figures 3.8 and 3.10 correspond. Covid-19 is an economic shock that was induced by a natural occurrence, this being the closest shock to that proposed by ISDA, therefore, intuitively makes sense. This allows a hypothesis that when ISDA releases transition shocks for South Africa, a starting point for an analysis of that shock on the South African swap and swaption market would be to compare the proposed shock with that of the Covid-19 pandemic.

2.3 Risk Analysis on the Trading Book: RSA

The analysis on the trading book for South Africa considers all three stress scenarios: the Dot Com bubble, the GFC and the Covid-19 pandemic. For each scenario, the swap distributions, the swaption distributions, the changes in TVaR for swaps and the changes in TVaR for swaptions are presented.
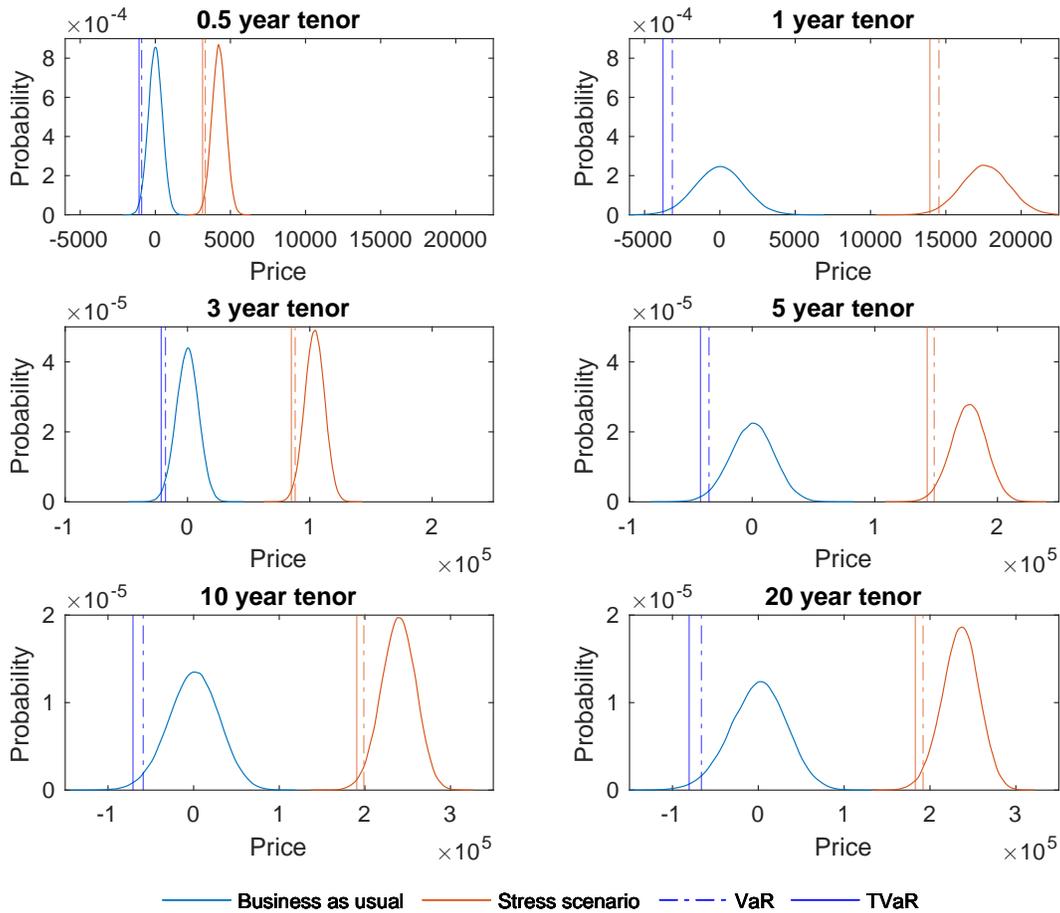
Figure 3.13: Shock to SA swap prices caused by the DotCom Bubble.

The DotCom shock presents a uniform increase in the mean of swap prices across all maturities, seen in Figure 3.13. The values of payer swaps actually increased using the shocked state variables with the interest rate models calibrated on the full data set. The TVaR decreased significantly, as seen in Figure 3.21. This increase in swap values is consistent with the increase in swaption values, seen in Figure 3.14. If receiver swaps had been considered, this shock would have decreased their values and so this downside risk would have reflected in the capping of TVaR of the swaption distribution, consistent with the behaviour of the US swaps and swaptions during the GFC discussed in Section 2.2.
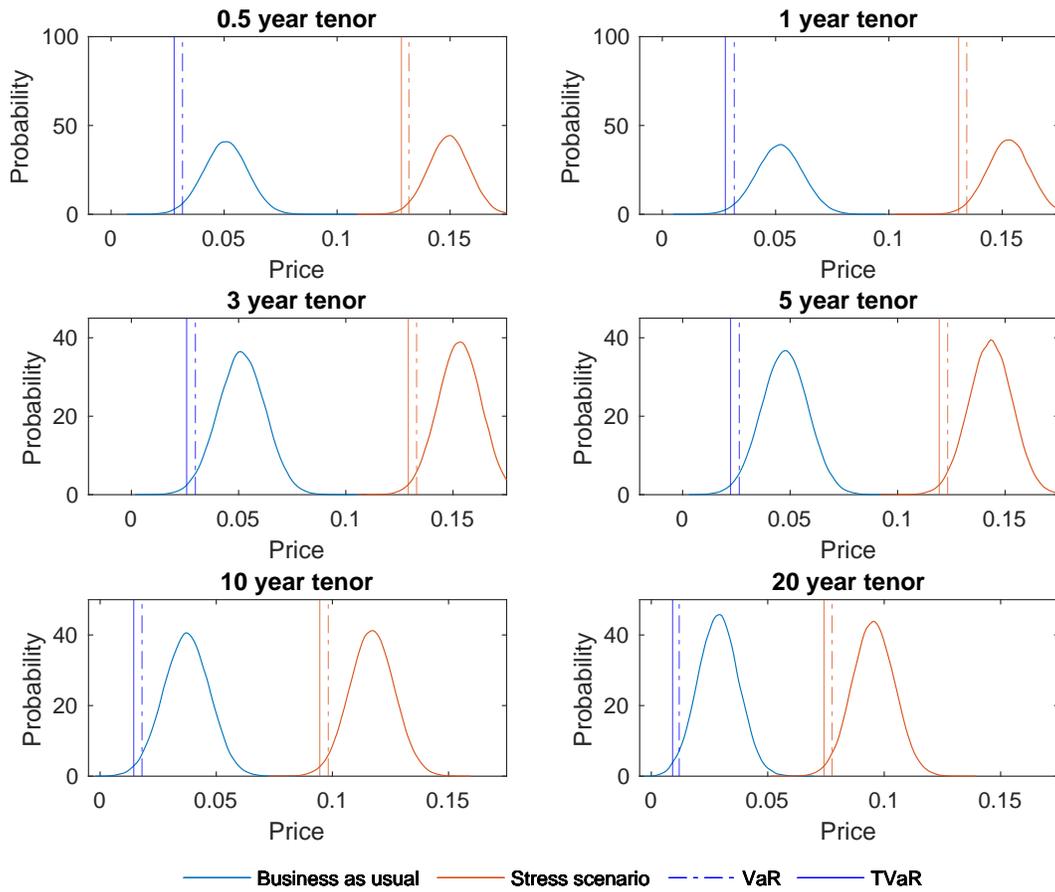
Figure 3.14: Shock to SA Swaption Prices caused by the Dot Com Bubble.
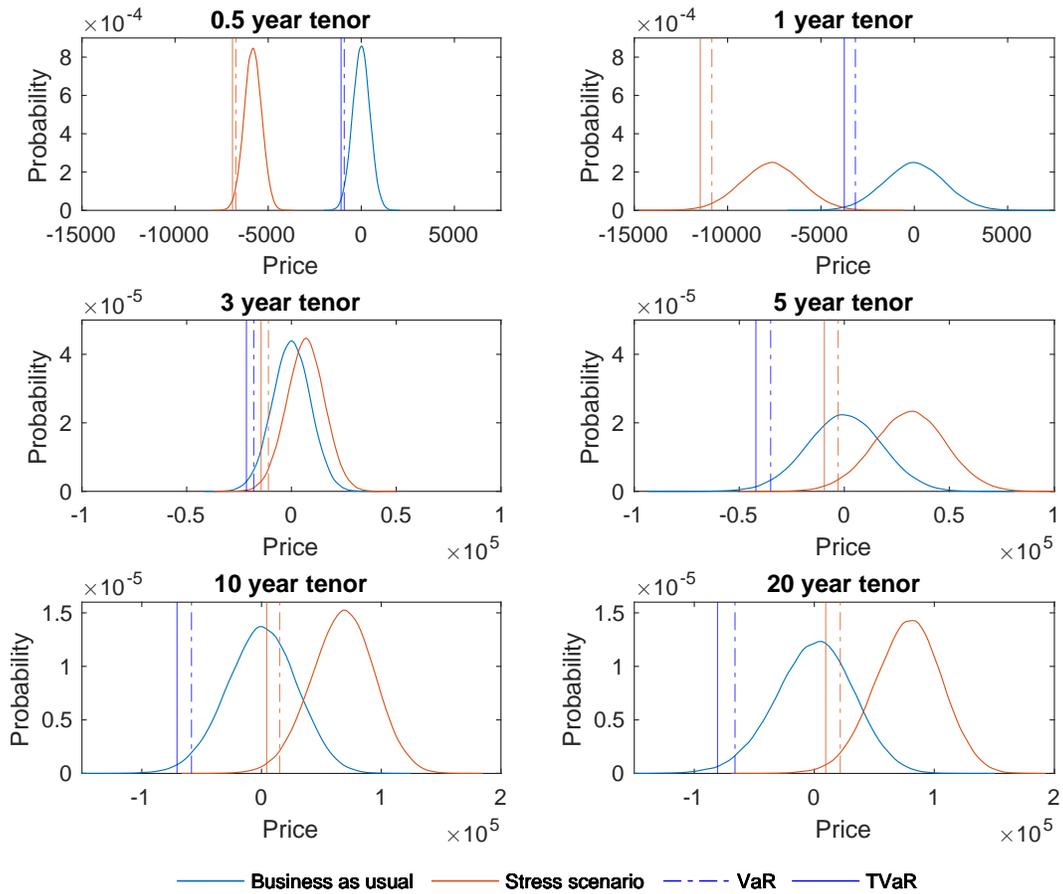
Figure 3.15: Shock to SA Swap Prices caused by GFC.

Conversely, the impact on swaps for the GFC stressed scenario differs significantly by tenor. The impact is negative for half a year and one year tenors, with the three year swap being relatively unaffected and the five year, ten year and twenty year swaps increasing in value, as seen in Figure 3.15. This corresponds to a decrease in TVaR for all maturities above one year, as seen in Figure 3.21.
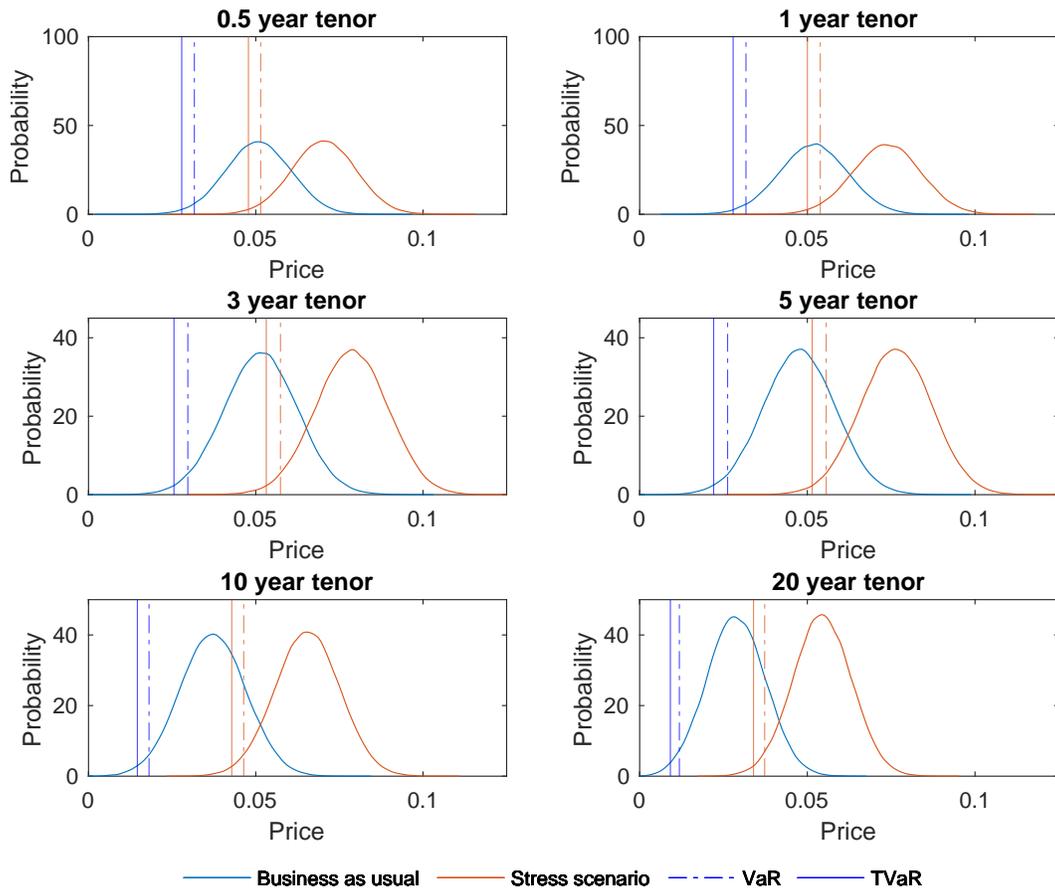
Figure 3.16: Shock to SA Swaption Prices caused by GFC.

Since the maturity of the swaption is beyond the tenors of the swaps that decreased in value during the GFC stress scenario, it can be seen in Figure 3.16 that the values of swaptions increased for all maturities. This corresponds to a decrease in TVaR for swaptions for each maturity, as seen in Figure 3.22.
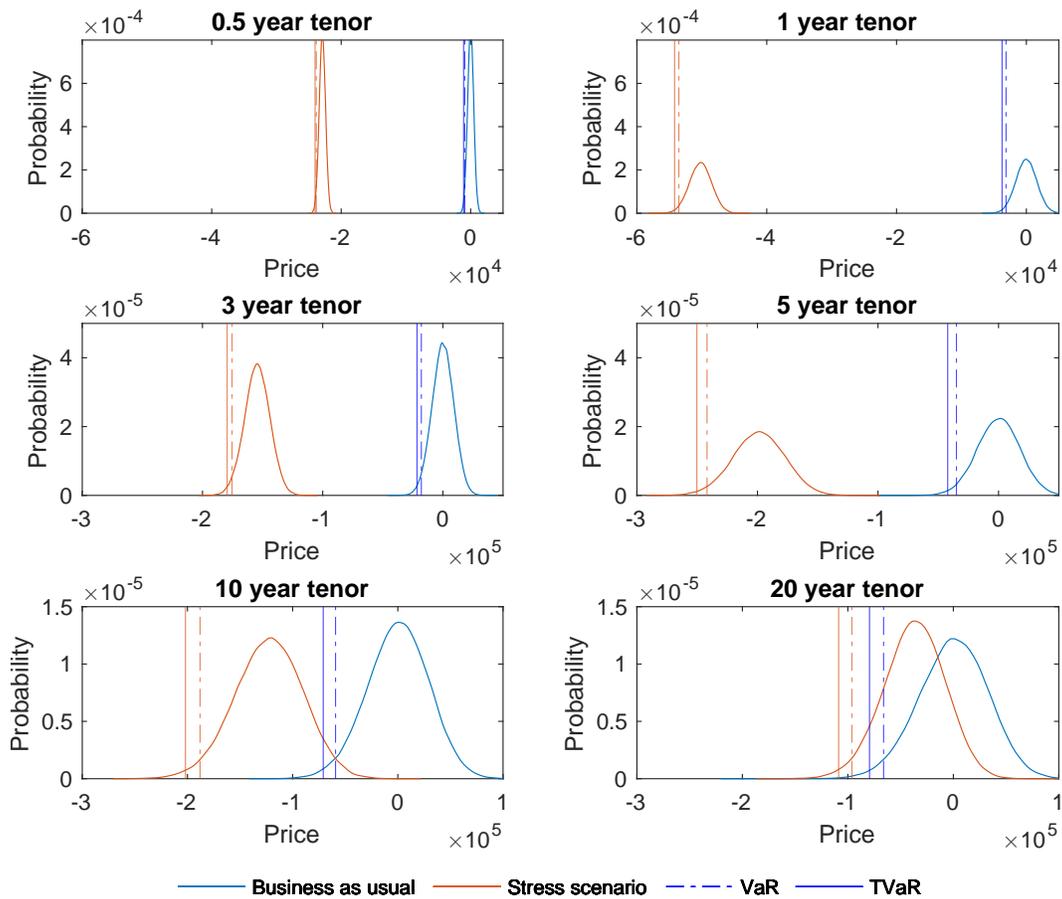
Figure 3.17: Shock to SA Swap Prices caused by Covid-19.

Figure 3.17 shows how Covid-19 negatively impacted all maturities, with all distributions showing a significant lefward shift of their means. However, this negative impact decreases with increasing tenor, where the distribution is reasonably unchanged for the twenty year tenor, perhaps because of a view of the market around how long the crisis was expected to persist.

This result is consistent with the swaption prices seen in Figure 3.18 as the swaption prices collapse to zero for short-term maturities, the distributions are heavily concentrated around zero and skewed to the left for the medium maturities and the longer-term maturities are relatively unchanged for the shocked and unshocked scenarios.

Moreover, this is seen in the changes in the TVaR of the swaps compared to the swaptions. Figure 3.21 shows how the tail risk increases considerably relative to the 'business as usual' scenario, whereas the TVaR for the swaptions is worse than

the 'business as usual' scenario, however it is capped to zero. This implies that the swaptions are an appropriate risk management instrument for a shock that is consistent with the Covid-19 pandemic.
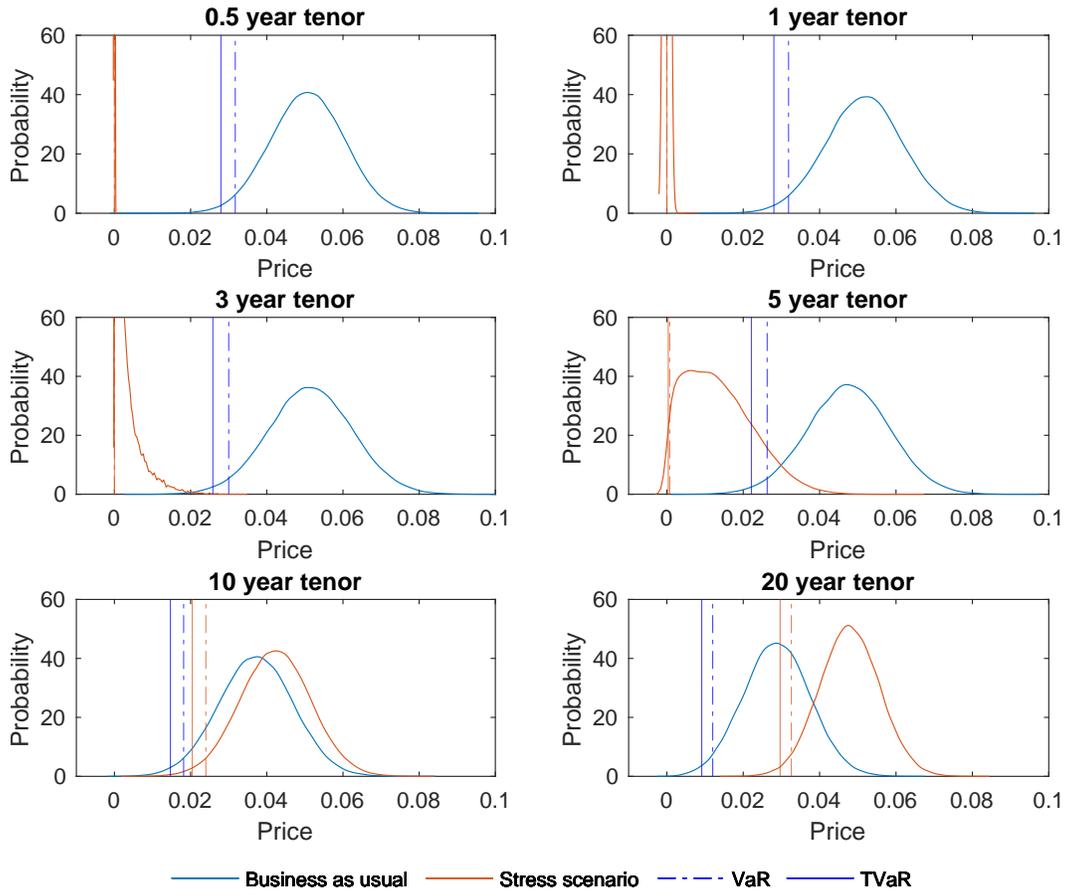
Figure 3.18: Shock to SA Swaption Prices caused by Covid-19.
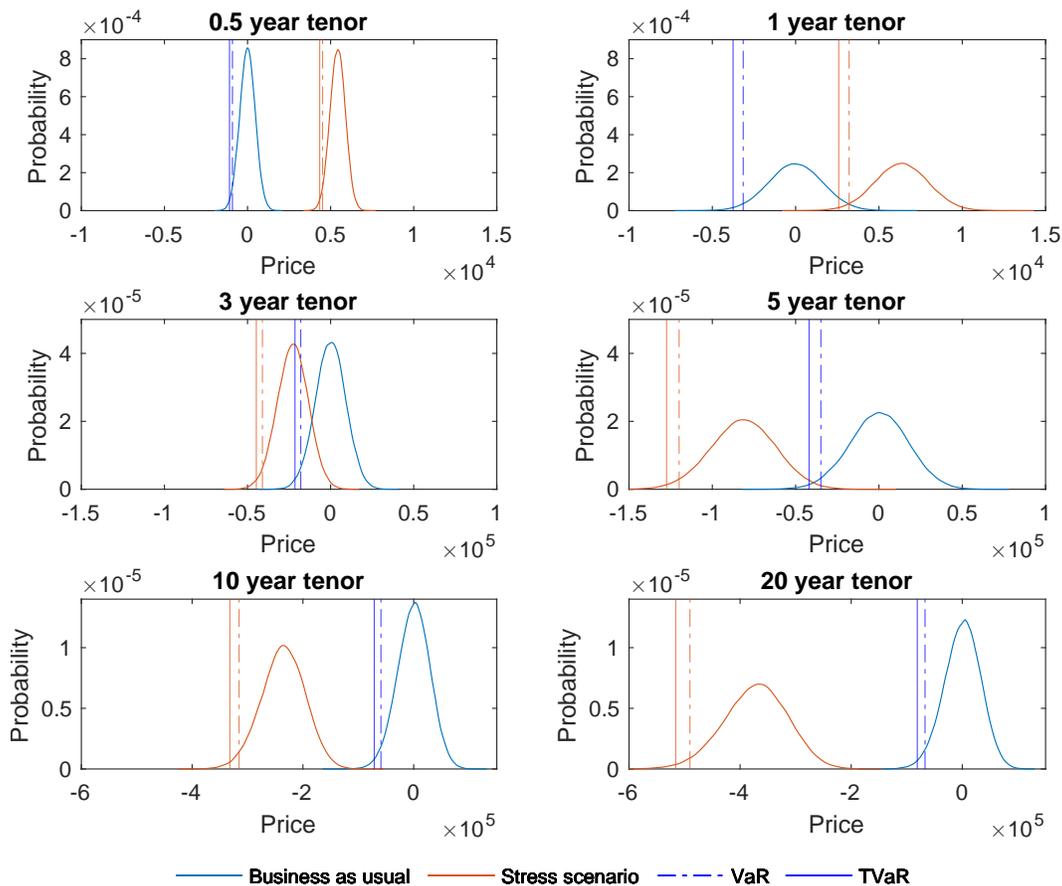
Figure 3.19: Shock to SA Swap Prices caused by the US ISDA shock.

We see in Figure 3.19 that when applying the ISDA transition risk shock, proposed for the US, on the South African swaps, we see a rightwards shift in distributions for short-term tenors and a leftwards shift in distributions for tenors greater than one year. The shape of the distributions remained reasonably similar, despite the changes in mean.

However, the swaption distributions remain virtually unchanged when shocked using the US ISDA transition risk shock, as seen in Figure 3.20. We do, however, see in 3.21 that the ISDA shock is behaves similarly to the Covid-19 pandemic shock in the behaviour of TVaR under the stressed scenario where, like Covid-19, there is an increase in TVaR across all maturities. The TVaR for the swaptions behaves almost exactly as the same as the 'business as usual' scenario, as seen in Figure 3.22. These results indicate that the ISDA transition risk shock proposed for the US does not behave as similarly to the Covid-19 pandemic shock scenario in the South African swap and swaption markets as it did in the US. The behaviour in TVaR for the swaps and swaptions had similarities but the distributional changes in swap
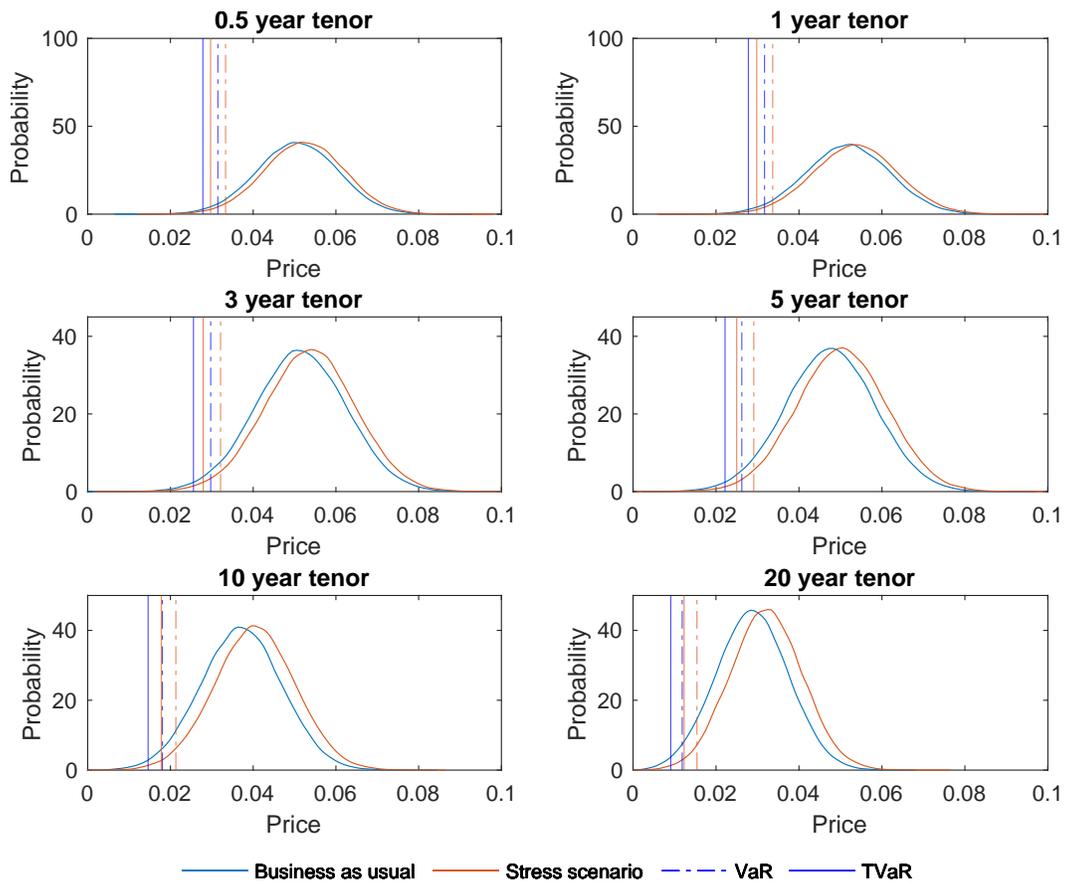
Figure 3.20: Shock to SA Swaption Prices caused by Covid-19.

and swaption prices were not the same as the Covid-19 pandemic shock in the South African market. However, upon the release of ISDA's transition risk shock appropriate to South Africa this analysis could be revisited to see whether the hypothesis that the transition risk shock behaves similarly to the pandemic shock is, in fact, true.
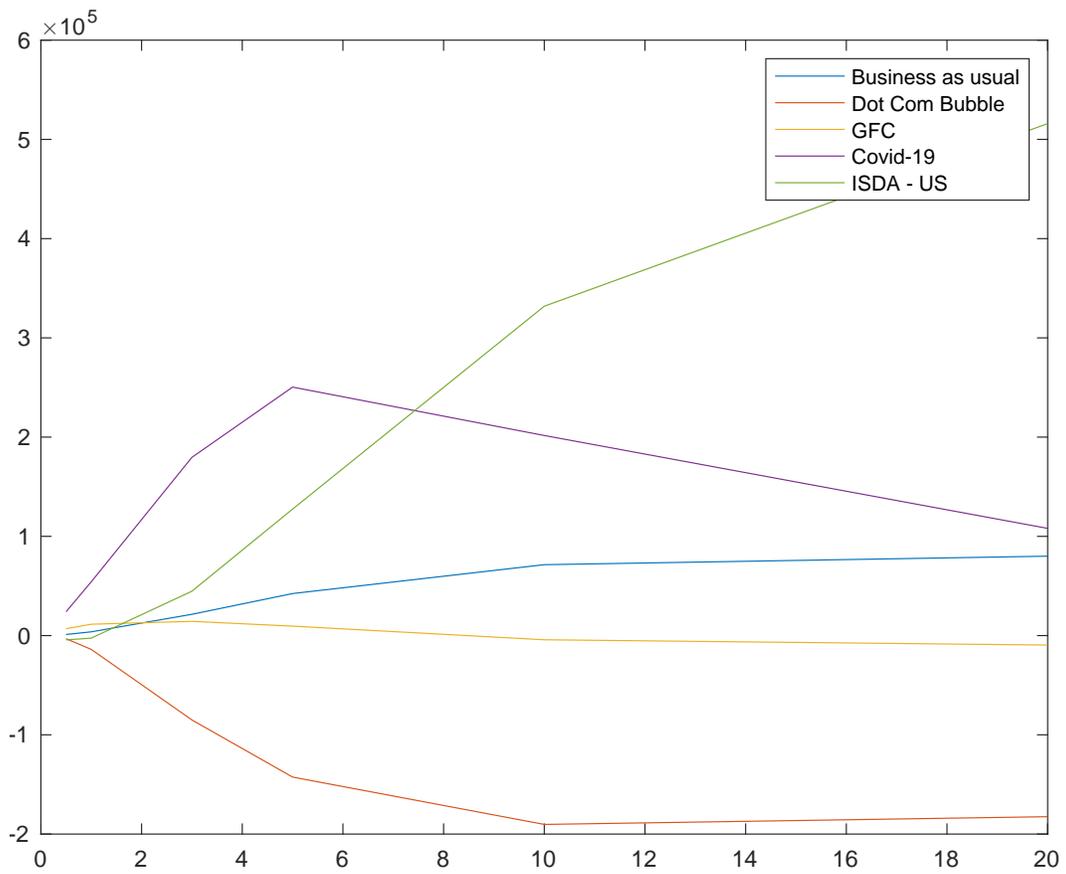
Figure 3.21: Changes in the TVaR of SA swaps in each shock scenario.
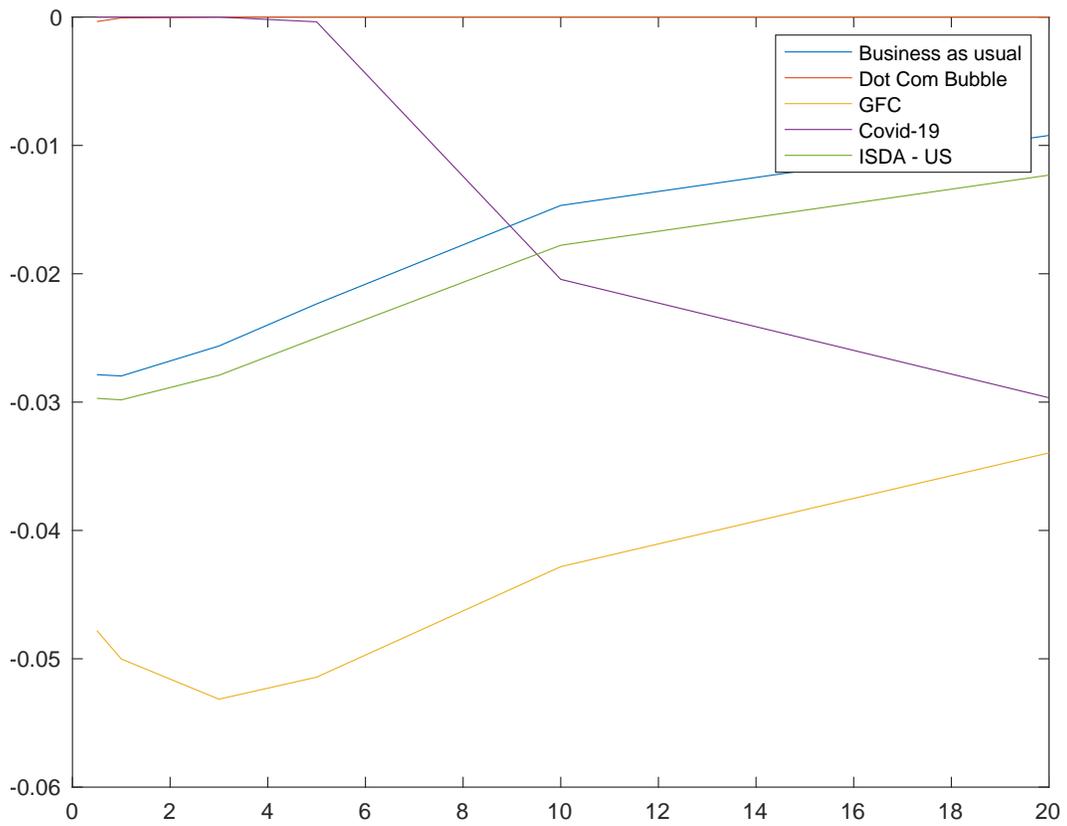
Figure 3.22: Changes in the TVaR of SA swaptions in each shock scenario.

Chapter 4

Conclusion

In conclusion, we have developed a framework that can be used to analyse the impact of transition risk in the trading book. We parameterized the yield curve using the Nelson-Siegel-Svensson model and calibrated the three-factor short rate model to historical data using the Kalman filter. This calibrated three-factor short rate model was used in a pricing engine to price and calculate probability distributions for profit and loss of interest rate derivatives. Our framework enables analysis of the impact of different transition risk shocks on the interest rate swap market. This is achieved by inducing a shock to the yield curve, parameterized by the NSS parameters, and using this shocked NSS yield curve to estimate the corresponding initial state variables. These estimated initial state variables can then be used in the pricing engine to produce profit and loss distributions for the considered interest-rate derivatives, for each transition risk shock.

A significant finding of this study revealed that the historical shock that most closely resembles the transition risk shock identified by ISDA is the COVID-19 pandemic in the USA. The shock of the NSS curve induced by the transition risk aligns notably with the shocks observed during the COVID-19 period. Moreover, the changes in distributions of swaps and swaptions under the transition shock exhibit similarities to those observed during the pandemic, further evidenced by the Tail Value at Risk estimates. Unfortunately, the same result was not seen for this shock applied to the South African market.

Looking ahead, once ISDA provides transition risk spreads for South Africa, it would be valuable to replicate the analysis conducted in this report to assess their implications for the South African market and an apt starting point would, therefore, be comparing its magnitude and behaviour to that of the COVID-19 pandemic in South Africa.

Additionally, future research could explore integrating the market price of risk into

a more sophisticated interest rate model and incorporating the potential counter-party risk associated with swaps and swaptions.

Our framework is flexible and could be extended to the banking book, however this was beyond the scope of this analysis. We have used previous historical crises that had an impact on financial markets to gauge the coherence and plausibility of our framework. Consistent with the Fundamental Review of the Trading Book (FRTB), our framework uses Expected Tail Loss as the quantitative risk metric for the trading book, ensuring robust and compliant risk assessment. This analysis has contributed to the existing discourse around how to quantify the impact of climate risk in the financial sector, specifically with respect to the South African trading book.

Bibliography

Abu-Mostafa, Y.S., 2001. Financial model calibration using consistency hints. IEEE transactions on neural networks 12, 791–808.

Babbs, S.H., Nowman, K.B., 1999. Kalman filtering of generalized Vasicek term structure models. Journal of financial and quantitative analysis 34, 115–130.

Beaglehole, D.R., Tenney, M.S., 1991. General solutions of some interest rate-contingent claim pricing equations. The Journal of Fixed Income 1, 69–83.

BIS, 2013. Fundamental review of the trading book: A revised market risk framework. URL: `https://www.bis.org/publ/bcbs265.pdf`. accessed on 2 July, 2024.

Chatterjee, S., 2005. Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany. In Computing in Economics and Finance 346.

Defra, et al., 2013. Understanding climate change .

Duan, J.C., Simonato, J.G., 1999. Estimating and testing exponential-affine term structure models by Kalman filter. Review of quantitative finance and accounting 13, 111–135.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical finance 6, 379–406.

Gilli, M., Große, S., Schumann, E., 2010. Calibrating the Nelson-Siegel-Svensson model. Available at SSRN 1676747 .

Gürkaynak, R.S., Sack, B., Wright, J.H., 2007. The US Treasury yield curve: 1961 to the present. Journal of monetary Economics 54, 2291–2304.

Huang, Z., Du, P., Kosterev, D., Yang, B., 2009. Application of extended Kalman filter techniques for dynamic model parameter calibration, in: 2009 IEEE Power & Energy Society General Meeting, IEEE. pp. 1–8.

ISDA, 2023. A conceptual framework for climate scenario analysis in the trading book. URL: `https://www.isda.org/a/boogE/A-Conceptual-Framework-for-Climate-Scenario-Analysis-in-the-Trading-Book.pdf`. accessed 2024/07/01.

ISDA, 2024. Climate scenario analysis in the trading book-phase ii. URL: `https://www.isda.org/a/YHlgE/Climate-Scenario-Analysis-in-the-Trading-Book-%E2%80%93-Phase-II.pdf`. accessed 2024/07/01.

James, J., Webber, N., 2000. Interest rate modelling. Wiley-Blackwell Publishing Ltd.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60, 473–489. URL: `http://www.jstor.org/stable/2352957`.

Nunes, J.P.V., Clewlow, L., Hodges, S., 1999. Interest rate derivatives in a Duffie and Kan model with stochastic volatility: An Arrow-Debreu pricing approach. Review of Derivatives Research 3, 5–66.

Pang, K., 1997. Calibration of interest rate term structure and derivative pricing models. Ph.D. thesis. University of Warwick.

Porretta, P., Agnese, P., 2021. The fundamental review of trading book: new standard approach and risk management impacts. Journal of Risk Management in Financial Institutions 14, 209–219.

Svensson, L.E., 1995. Estimating forward interest rates with the extended nelson & siegel method. Sveriges Riksbank Quarterly Review 3, 13–26.

Wells, C., 2013. The Kalman filter in finance. volume 32. Springer Science & Business Media.

Ziady, H., 2024. World's first carbon tax on livestock will cost farmers $100 per cow. URL: `https://edition.cnn.com/2024/06/26/business/denmark-cows-carbon-tax/index.html`. accessed on 28 June, 2024.

Chapter 5

Appendix
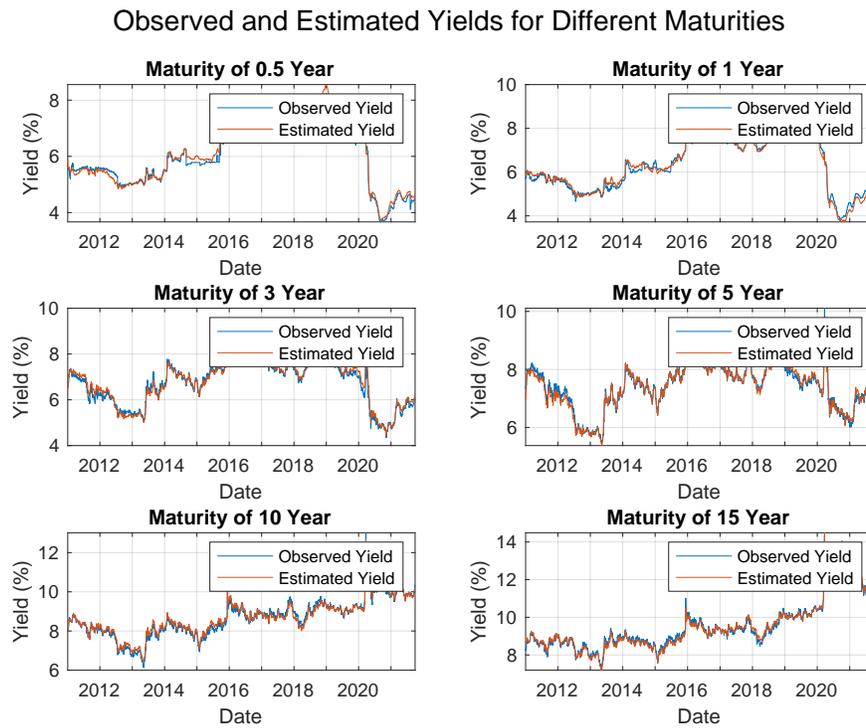
Figure 5.1: RSA Observed and Estimated Yield.

Errors for Different Maturities

Errors for 0.5 Year Maturity

Errors for 1 Year Maturity

Errors for 3 Year Maturity

Errors for 5 Year Maturity

Errors for 10 Year Maturity

Errors for 15 Year Maturity
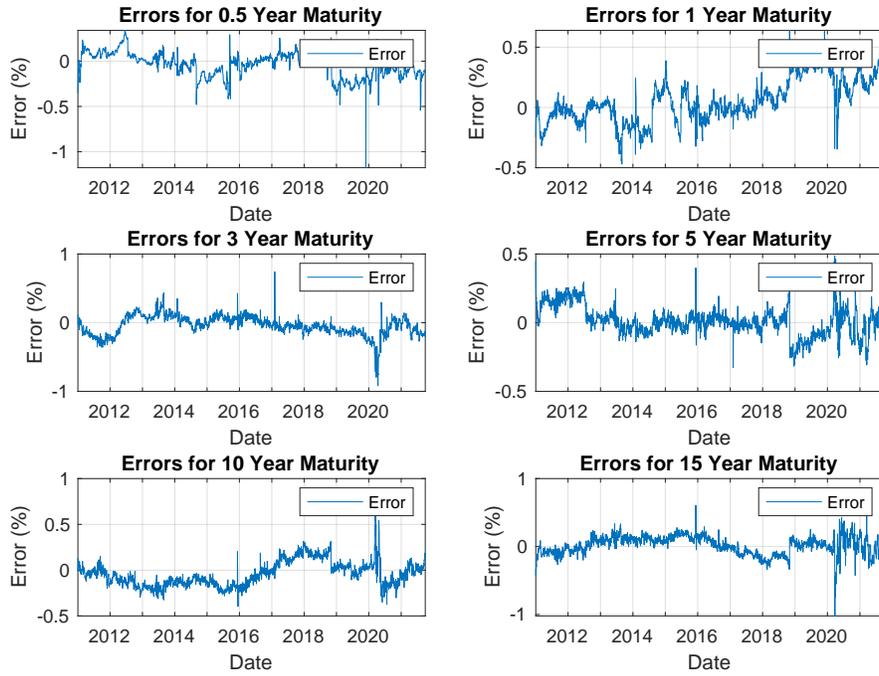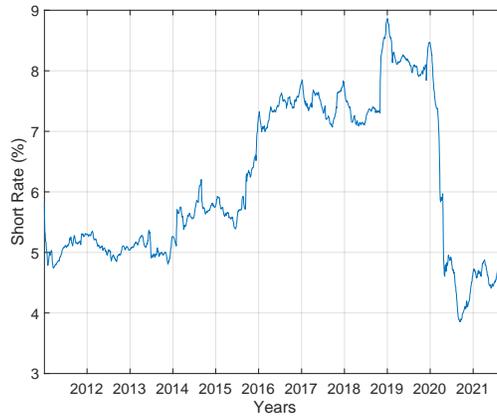
Figure 5.2: RSA Error.

Figure 5.3: RSA Estimated Short Rate.

45
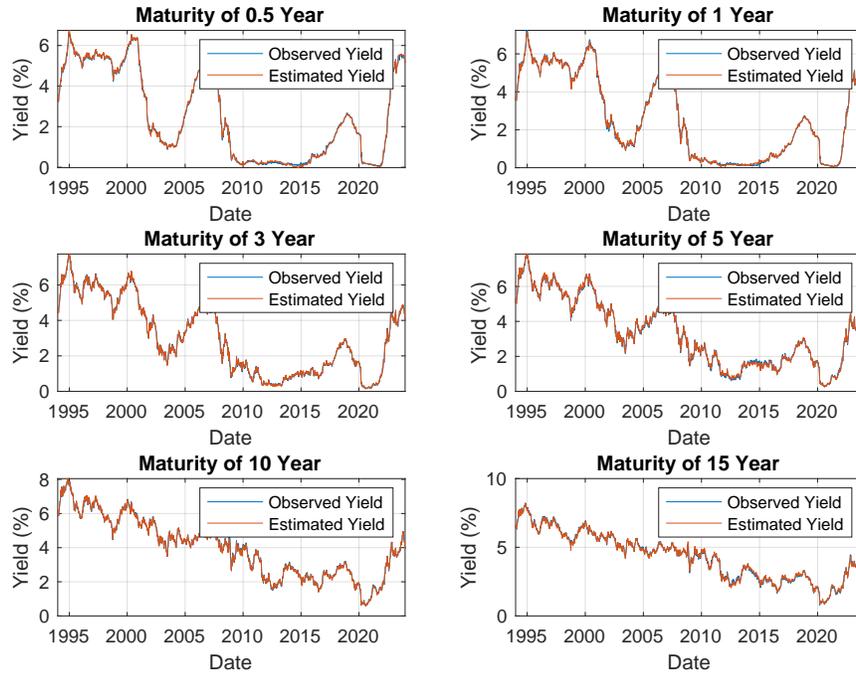
Figure 5.4: USA Observed and Estimated Yield.

Errors for Different Maturities

Errors for 0.5 Year Maturity

Errors for 1 Year Maturity

Errors for 3 Year Maturity

Errors for 5 Year Maturity

Errors for 10 Year Maturity

Errors for 15 Year Maturity
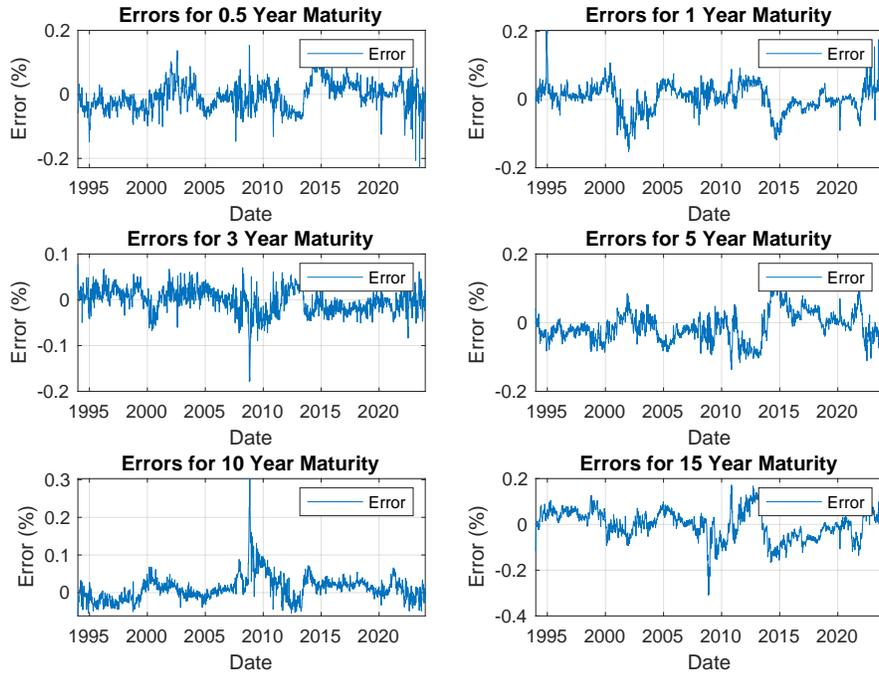
Figure 5.5: USA Errors.
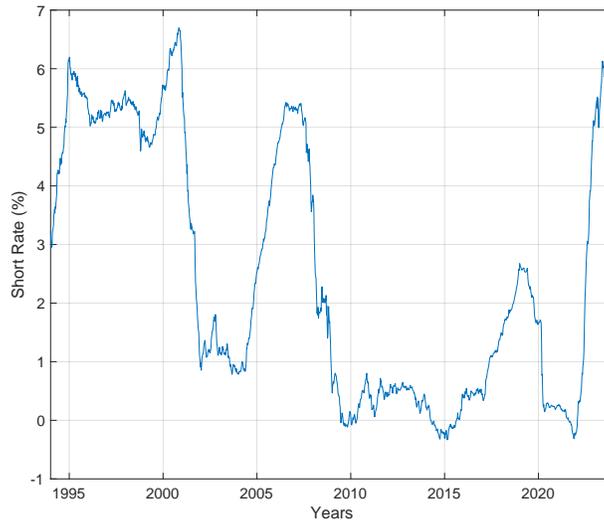
Figure 5.6: USA Estimated Short Rate.

47
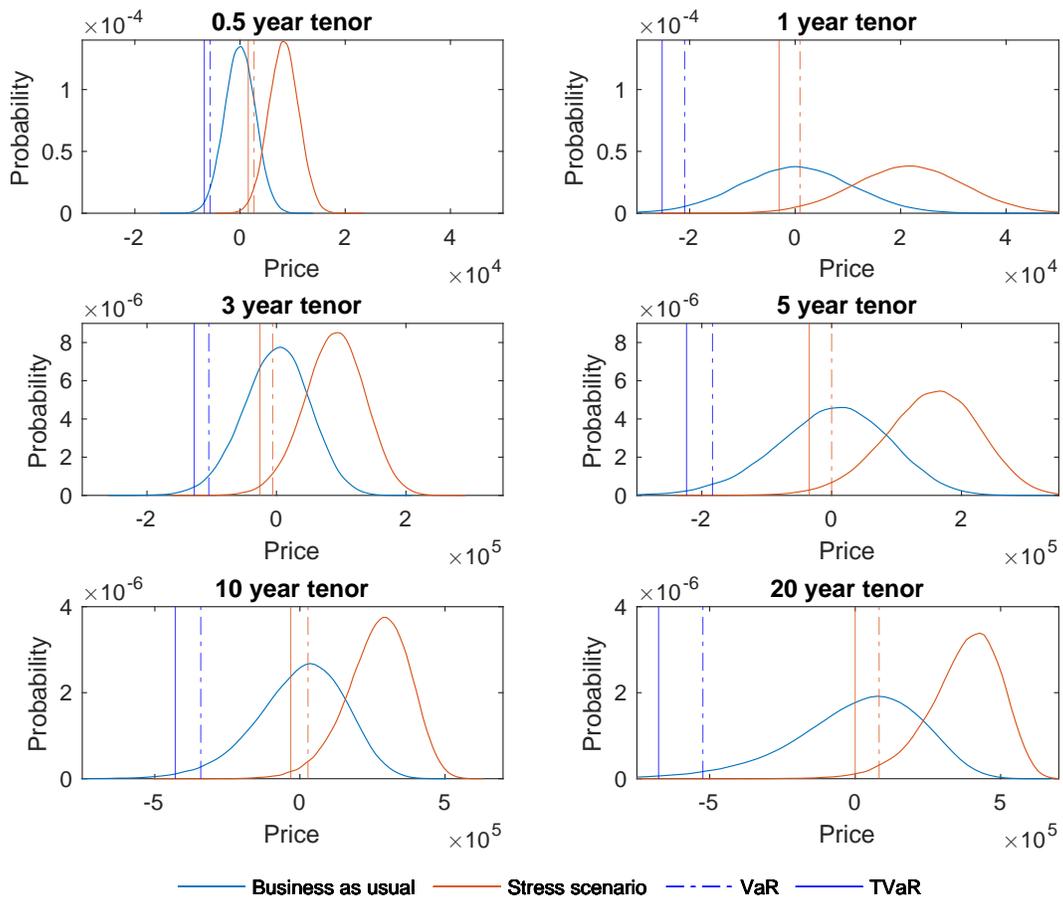
Figure 5.7: Shock to Swap Prices caused by the Dot Com Bubble.
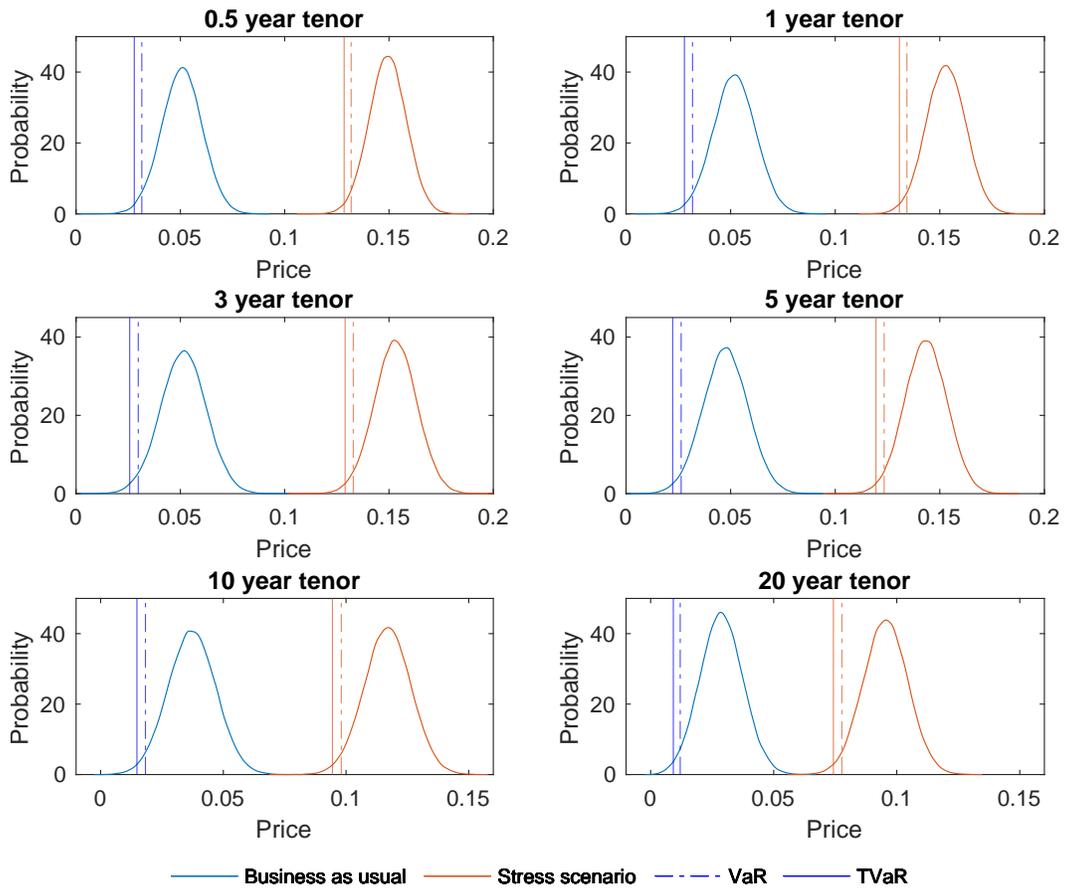
Figure 5.8: Shock to Swaption Prices caused by the Dot Com Bubble.