FINANCIAL MATHEMATICS TEAM CHALLENGE

A collection of the four reports from the 2019 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 purpose 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 Sixth Financial Mathematics Team Challenge was held from the 24th of June to the 5th of July 2019. The challenge brought together five teams of Masters and PhD students from France, Germany, China, Mexico, Zimbabwe, Zambia, South Africa and the UK to pursue intensive research in Financial Mathematics. Each team worked on a distinct research problem over the twelve days. Professional and academic experts from Canada, Australia, South Africa, and the UK individually mentored the teams; fostering teamwork and providing guidance. As they have in the past, the students applied themselves with remarkable commitment and energy.

This years research included topical projects on (a) robust pricing and hedging of basket options, (b) level-dependence of volatility and the CEV market model, (c) financing the green revolution, (d) option pricing and hedging with deep learning, and (e) inferring OIS discount factors in the South African market. These were either proposed directly by our 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 five 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 VI was a triumph, and we were particularly honoured when the UCT Vice-Chancellor Professor Mamokgethi Phakeng and the UCT Commerce Dean, A/Professor Linda Ronnie visited the FMTC. We greatly appreciate their encouragement to continue bringing to UCT many more FMTC editions. Meanwhile, FMTC VII, which will take place in July 2020, is already being organised!

Last, but by no means less important, we take the opportunity to communicate our great pleasure in acknowledging that the FMTC-BR took place, during 24 July 3 August 2019, for a second year running at the Fundaco Getulio Vargas (FGV) in Rio de Janeiro. It was another success!

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

Contents

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

Robust Pricing and Hedging of Basket Options

TEAM LSE

YICHENG CHEN, McMaster University
NIC COWEN, University of Cape Town
CHIPO GUNDANI, University of Zimbabwe
JACOB MWANZA, University of Cape Town

Supervisors:
BEATRICE ACCIAIO, London School of Economics

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

In this report we consider the problem of pricing and hedging basket options. The typical underlying of a basket option is a basket consisting of several stocks, that represent a certain economic sector, industry or region. In this study, SAFEX options have been used so the basket option is based on equity futures contracts (12).

The main advantage of a basket option is that it is cheaper to use such an option for portfolio insurance than to use the corresponding portfolio of plain vanilla options. Indeed, a basket option takes into account the correlation between the assets in the basket. Furthermore, the transaction costs are minimized because an investor has to buy just one option instead of several ones (6). We assume that call options are liquidly traded and can be used as hedging instruments. The classic way of addressing derivative pricing is to postulate a model for the traded asset and derive a fair price and associated hedging strategy from the model. Then by no arbitrage theory, the price of the option is the expectation of the discounted expected payoff of the option under the risk neutral measure. The popular assumption is that the assets are driven by correlated exponential Brownian motions.

For pricing simple options on one underlying, the celebrated Black Scholes model leads to a closed form solution, since the stock price at a fixed time follows a lognormal distribution. However, using the Black Scholes model for a collection of underlying stocks, does not provide us with a closed form solution for the price of a basket option. The difficulty stems primarily from the lack of availability of the distribution of a weighted sum of non-independent lognormals, a feature that has hampered closed-form basket option pricing characterization. Indeed, the value of a portfolio is the weighted average of the underlying stocks at the exercise date (6).

The Black Scholes model is often calibrated using the observed price of calls or other derivatives. This provides some level of consistency, however, the success of a replicating strategy is based on the fundamental truth of the model on which it is based. Although market risk can be eliminated, model risk remains in the form of misspecification. For example, prices of two traded calls can correspond to different implied volatilities. The question that the writer faces is which volatility to use for option pricing without creating arbitrage opportunities.

One can use Monte Carlo simulation techniques by assuming that the assets follow correlated geometric Brownian Motions, to obtain a numerical estimate of the price. Other techniques consist of approximating the real distribution of the payouts by another more tractable one. For instance, in industry it is common to use the lognormal distribution as an approximation for the sum of lognormals, although it is known that this methodology can lead sometimes to poor results. An extensive

discussion of different methods can be found in (1), (3) and (14).

An alternative approach in an environment in which there is model uncertainity is to infer information about the potential distribution of asset prices from the call prices available in the market rather than making explicit assumptions about the dynamics of the price process of the underlying asset. This idea goes back to Hobson (7), and the resulting methods are called model independent or robust. This second approach does not aim to derive a unique fair price but is robust in the sense that it is not dependent on the efficacy of the underlying model.

Two main settings have been considered in the mathematical finance literature:

(i) If the prices of call options on an underlying X for a fixed maturity T and for all strikes $K > 0$ are known, then by the Breeden-Litzenberger theorem the full marginals of X at time T are known. In this case the Optimal Transport and Skorokhod Embedding theories may be applied to study robust pricing as demonstrated in (9).

(ii) The second case of interest is when there are only a finite number of calls traded on each underlying. This is a more realistic situation. If we only know call prices for a few strikes then we only know partial information about the marginals. In this case the available market prices of call options can be used to find bounds on the prices of exotic derivatives (10).

The scope of this report is to investigate the procedure in (ii) for basket options and to provide an implementation of it based on available market data. The general aim is to provide bounds on the possible price of an exotic option which are consistent with no-arbitrage given the market prices of vanilla puts and calls available in the market. In essence, rather than using a single model, we consider the class of all models which are consistent with the observed call prices, and rather than quoting a single option price we give the range of prices which arise under models from this class. In this article we apply this philosophy to basket options in the setting of a one period static arbitrage model. We will focus on the case of upper bounds and leave the case of lower bounds to future research. In the early part of section 2 we assume a continuum of strikes and suppose that instead of hedging with just the underlying stocks and cash, investors are allowed to hedge a basket call option with calls on the constituent assets with the same maturity and all available strikes. This provides the investor with a greater range of hedging instruments. However, dynamic hedges based on stochastic calculus and delta hedging can no longer be expected to work since we do not postulate any model for the underlying asset.

We have in mind the scenario of an option writer, who is both skeptical and risk averse. Her risk aversion is such that she builds a portfolio that whatever happens

in the market at time T, she will be better off. We define a super-replicating strategy as a hedging strategy which generates a terminal fortune which dominates the option payout in every possible scenario. Moreover, her skepticism means she is not prepared to make any assumptions about how asset prices will behave. The upper bound on the price of a basket option represents the lowest price at which she is prepared to sell a basket option, and the lower bound the highest price which she is prepared to pay. The bounds represent the limits on the possible price of the basket option which are necessary for the absence of arbitrage (8).

The result in the case of finitely many strikes on each underlying has the advantage that the proof involves finding the cheapest super-replicating strategy for the basket option, and hence has an immediate interpretation as a hedging strategy (8). This strategy only involves the traded assets and call options. Since knowledge of a finite number of strikes only gives partial information about the marginals, it may be anticipated that there is no such straightforward characterisation of the least upper bound. The characterisation is based on an interpolation technique which basically fills in the missing values of the call price functions and hence completes the partial information about the marginal via Breeden and Litzenberger (4). This is achievable as a consequence of the observation that the largest convex function passing through n given points is the linearly interpolated function.

We investigate the problem of finding model independent bounds for basket options, but this can however also be posed for other exotic options. For example, Hobson (7) considers model independent bounds of the price of a lookback options. Brown et al (5) and Hodges and Neuberger (10) consider barrier options. In calculating the bounds for the price of a lookback option, Hobson (7) assumed that the call prices are derived within a complete market without transaction costs.

This report is motivated by the paper written by Hobson, Laurence and Wang (8). We applied their philosophies on the static-arbitrage upper bounds for the prices of basket options to a particular set of data derived from the JSE top 40 index which is discussed in section 3.1.

The remainder of this article is structured as follows: in section 2 we review the theoretical background with regard to moving from a model based approach to a model independent approach and the upper bound on the price of basket option. The main body of results and analyses is found in section 3 and the conclusion of the article follows in section 4.

2 Theoretical Background

2.1 Moving from a Model-based Approach to a Model-independent Approach

The $t = 0$ price, $C(K)$, of a European call option with strike K and maturity T on non-negative underlying asset X can be computed as the discounted value of the option's expected payoff under a risk neutral measure, which we may write as:

$$C(K) = e^{-rT}\mathbb{E}_{\mathbb{Q}}[(X_T - K)^+] = e^{-rT}\int_K^\infty (x-K)^+ f_{X_T}(x)dx$$

where r is the continuously compounded riskless rate (which is market observable) and $\mathbb{Q}$ is the risk-neutral measure, which is not directly observable. The method of Breeden and Litzenberger (4) recovers the distribution function $F(x) = \mathbb{Q}(X_T \leq x)$ and the density function of the asset price process at time T under the risk-neutral measure $\mathbb{Q}$ from market-observable liquid European call options.
No arbitrage constraints imply that $C(0) = X_0$ (i.e. we can think of the underling asset as a call option with strike $K = 0$), C is decreasing and convex. Thus C is bounded, $0 \leq C(K) \leq X_0$. In particular, left and right derivatives of C exist. For example, when we take the right derivative with respect to K, we get:

$$C'(K+) = -e^{-rT}\mathbb{Q}(X_T > K)$$

We let G denote the distribution of X_T, so that

$$G(K) = \mathbb{Q}(X_T \leq K) = 1 - \mathbb{Q}(X_T > K) = 1 + e^{rT}C'(K+)$$

Here G is the distribution of X_T driven by the option market data (no assumptions are made about the underlying asset price dynamics). Since we only know the call prices for a finite number of traded strikes, we cannot recover the full distribution of X_T. Hence G is piecewise linear (see figure 1) and the associated distribution is discrete, placing mass at only the finitely traded strikes (see figure 2).

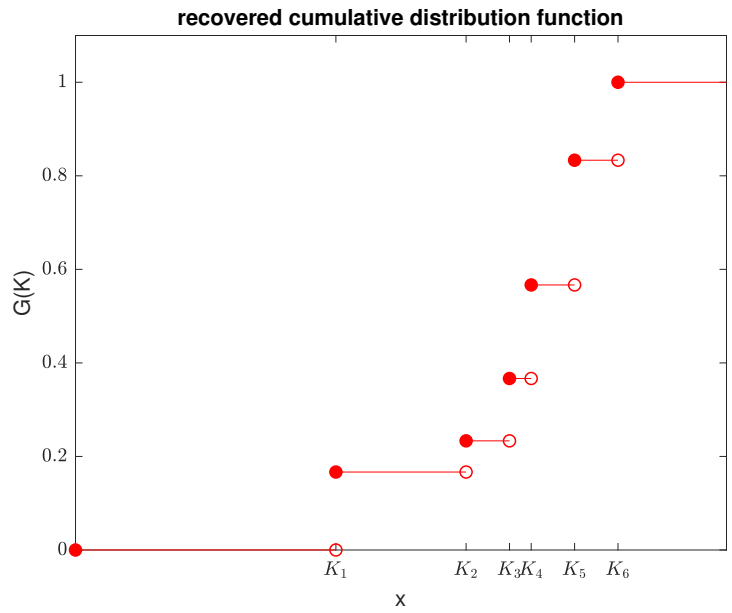


Figure 1: Cumulative distribution function of X_T

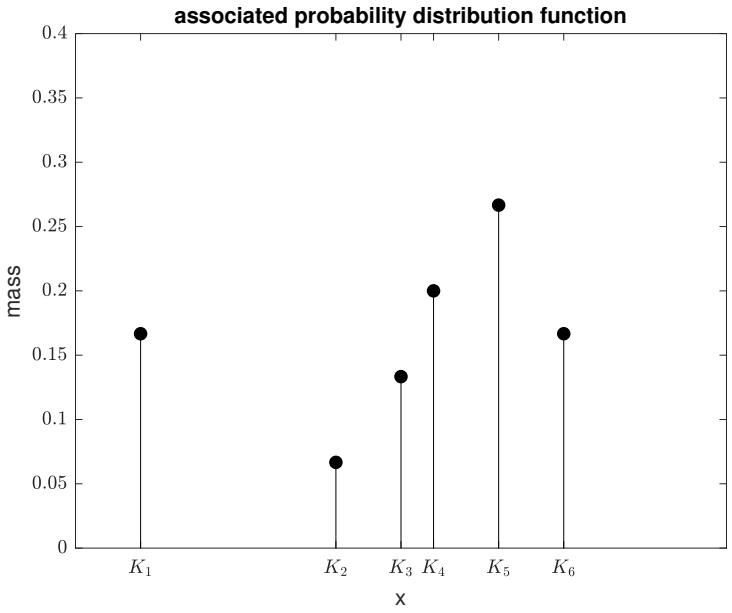


Figure 2: Probability distribution function of X_T

2.2 Upper Bound on Price of Basket Option

Here we outline the method used in Hobson (8), to find an upper bound for the price $B(K)$ of a basket option with strike K and maturity M with payoff defined on N traded underlyings, given N maturity-M calls on the individual stocks in the basket with a finite number of traded strikes. This is equivalent to the cheapest portfolio of calls and underlying assets which will super-replicate the basket option.
For any vector $\lambda = (\lambda_1, \ldots, \lambda_N)$ with $\lambda_i \geq 0$ and $\sum_{i=1}^N \lambda_i = 1$, we have

$$\begin{aligned}\Big(\sum_i w_i X_M^{(i)} - K\Big)^+ &= \Big(\sum_i w_i X_M^{(i)} - \sum_{i=1}^N \frac{\lambda_i K}{w_i} w_i\Big)^+ \\ &\leq \sum_i w_i \Big(X_M^{(i)} - \frac{\lambda_i K}{w_i}\Big)^+.\end{aligned} \tag{1}$$

i.e. the payoff of the basket option is bounded by a weighted sum of calls on the individual assets with strikes equal to $\lambda_i K / w_i$
No-arbitrage considerations imply that for arbitrary λ_i

$$B(K) \leq \sum_i w_i C^{(i)} \left(\frac{\lambda_i K}{w_i}\right). \tag{2}$$

This suggests a super-replicating strategy which consists of buying w_i calls with strike $\lambda_i K / w_i$ on underlying $X^{(i)}$. However, we have finite traded strikes in the market, thus options with desired strikes may not exist.
Now, since the λ_i are arbitrary, we have

$$B(K) \leq \inf_{\lambda_i \geq 0, \sum \lambda_i = 1} \sum_i w_i C^{(i)} \left(\frac{\lambda_i K}{w_i}\right). \tag{3}$$

Since we are minimising a bounded function over a compact interval, the infimum is attained. We define λ_i^* to be the minimising choice and write the upper bound as

$$\bar{\mathcal{B}}(K) = \sum_i w_i C^{(i)} \left(\frac{\lambda_i^* K}{w_i}\right). \tag{4}$$

To find the infimum of $\sum_i w_i C^{(i)} \left(\frac{\lambda_i K}{w_i}\right)$ over choices λ_i satisfying $\lambda_i \geq 0$ and $\sum_{i=1}^N \lambda_i = 1$, (8) defines the Langrangian function

$$L(\lambda, \phi) = \sum_i w_i C^{(i)} \left(\frac{\lambda_i K}{w_i}\right) + \phi \left(\sum_{i=1}^N \lambda_i - 1\right) \tag{5}$$

and shows that there exist ϕ^* and λ_i^*, so that

$$\inf_{\lambda:\lambda\geq 0} L(\lambda,\phi^*) = L(\lambda^*,\phi^*) = \sum_i w_i C^{(i)}\left(\frac{\lambda_i^* K}{w_i}\right). \tag{6}$$

For traded calls on asset $X^{(i)}$ with strikes $(k_j^{(i)})_{1\leq j\leq J^{(i)}}$ where $k_j^{(i)} < k_{j+1}^{(i)}$, (8) considers 3 cases below in the derivation of an upper bound for the basket option:
Case 1: $C(k_{J(i)}^{(i)}) = 0$ for all i (i.e. the call price of the largest traded strike is equal to 0), and $\sum_i w_i k_{J(i)}^{(i)} > K$.
Then for for $1 \leq i \leq N$, and $0 \leq j \leq J^{(i)}$, (8) defines $\bar{C}^{(i)}$ (see figure 3) the largest decreasing convex function which agrees with $C^{(i)}$ at the traded strikes is defined by:

- $\bar{C}^{(i)}(k_j^{(i)}) = C^{(i)}(k_j^{(i)})$ if $k_j^{(i)}$ is a traded strike.
- $\bar{C}^{(i)}(k_j^{(i)})$ = linearly interpolated value between neighbouring traded strikes of $k_j^{(i)}$.
- $\bar{C}^{(i)}(k) = 0$ for $k \geq k_{J(i)}^{(i)}$.

Then for the λ_i^* obtained from the optimisation of our Lagrange function, we let I be the set of stock indices such that $\lambda_i^* K/w_i$ is a traded strike and $\bar{j}(i)$ the index j such that $\lambda_i^* = w_i k_{\bar{j}(i)}^{(i)}/K$ is a traded strike. If I contains all the stock indices, then the upper bound, $\bar{\mathcal{B}}$ is given by:

$$\bar{\mathcal{B}}(K) = \sum_i w_i \bar{C}^{(i)}(k_{\bar{j}(i)}^{(i)}) = w_i C^{(i)}(k_j^{(i)}). \tag{7}$$

If $\lambda_i^* K/w_i$ is a not traded strike, we denote I^c the set of such i. For such i we define $\bar{j}(i)$ to be the index j such that $\lambda_i^+ = w_i k_{\bar{j}(i)}^{(i)}/K$ and $\lambda_i^- = w_i k_{\bar{j}(i)-1}^{(i)}/K$ where $w_i k_{\bar{j}(i)}^{(i)}/K$ and $w_i k_{\bar{j}(i)-1}^{(i)}/K$ are traded strikes, and $\lambda_i^* \in (\lambda_i^-, \lambda_i^+)$.
We define

$$\theta_i^* = \frac{(K\lambda_i^*/w_i) - k_{\bar{j}(i)-1}^{(i)}}{k_{\bar{j}(i)}^{(i)} - k_{\bar{j}(i)-1}^{(i)}} \tag{8}$$

and we have

$$\lambda_i^* K/w_i = (1-\theta_i^*)k_{\bar{j}(i)-1}^{(i)} + \theta_i^* k_{\bar{j}(i)}^{(i)}, \tag{9}$$

hence

$$w_i\Big(X_M^{(i)} - \lambda_i^* K/w_i\Big)^+ \leq w_i(1-\theta_i^*)(X_M^{(i)} - k_{\bar{j}(i)-1}^{(i)})^+ + w_i\theta_i^*(X_M^{(i)} - k_{\bar{j}(i)}^{(i)})^+. \tag{10}$$

Thus we have an upper bound given by

$$\bar{\mathcal{B}}(K) = \sum_{i \in I} w_i C^{(i)}\big(k^{(i)}_{\bar{j}(i)}\big) + \sum_{i \in I^c} w_i \Big\{ (1-\theta_i^*) C^{(i)}\big(k^{(i)}_{\bar{j}(i)-1}\big) + \theta_i^* C^{(i)}\big(k^{(i)}_{\bar{j}(i)}\big) \Big\}, \qquad (11)$$

which suggests a super-replicating strategy consisting of w_i calls on $X^{(i)}$ with strike $k^{(i)}_{\bar{j}(i)}$ for each $i \in I$ and, for $i \in I^c$, a combination of $(1-\theta_i^*)w_i$ calls with strike $k^{(i)}_{\bar{j}(i)-1}$ and $\theta_i^* w_i$ calls with strike $k^{(i)}_{\bar{j}(i)}$.
According to Hobson (8), the upper-bound in equation 11 is the price attained in the co-monotonic model (a model is one in which the underlying assets exhibit a perfect positive dependence) and it is a least model-independent upper bound.
Case 2: $C(k^{(i)}_{J(i)}) > 0$ for some i and $\sum_i w_i k^{(i)}_{J(i)} > K$.
In this case, for all i such that $C(k^{(i)}_{J(i)}) > 0$ Hobson (8) introduces a synthetic strike $k^{(i)+1}_{J(i)} > k^{(i)}_{J(i)}$ such that $C(k^{(i)}_{J(i)+1}) = 0$, which reduces the problem to case 1 above and shows that the associated super-replicating strategy does not involve investments in the synthetic calls.

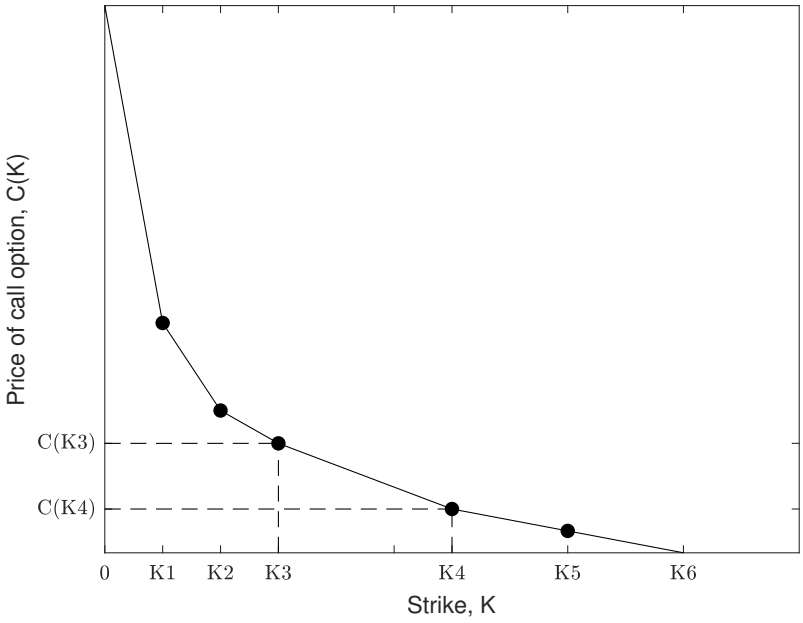


Figure 3: The largest convex call price function for finitely traded strikes

Case 3: $\sum_i w_i k^{(i)}_{J(i)} \leq K$.

In this case, we have

$$\Big(\sum_i w_i X_M^{(i)} - K\Big)^+ \leq \left(\sum_i w_i\Big(X_M^{(i)} - k_{J^{(i)}}^{(i)}\Big)\right)^+ \leq \sum_i w_i\Big(X_M^{(i)} - k_{J^{(i)}}^{(i)}\Big)^+. \quad (12)$$

Thus, the upper bound is then given by:

$$\bar{\mathcal{B}}(K) \leq \sum_i w_i C^{(i)}\big(k_{J(i)}^{(i)}\big). \quad (13)$$

This suggests a super-replicating strategy which consists of buying w_i calls with strike equal to the largest traded strike on underlying $X^{(i)}$.

The derivation of the upper bounds is summarised in the following theorem:

Theorem 2.1. Suppose maturity-M calls with strikes $(k^{(i)})_{\{1\leq j\leq J^{(i)}\}}$ are traded on assets $X^{(i)}$ and we wish to price a maturity-M basket option with payoff $(\sum_i w_i X_M^{(i)} - K)^+$. Then, for any model which is consistent with the observed call prices $C^{(i)}(k_j^{(i)})$, the fair price $B(K)$ for the option satisfies $B(K) \leq \bar{\mathcal{B}}(K)$ where, for $\sum_i w_i k_{J(i)}^{(i)} > K$,

$$\bar{\mathcal{B}}(K) = \sum_{i\in I} w_i C^{(i)}\big(k_{\bar{j}(i)}^{(i)}\big) + \sum_{i\in I^c} w_i\Big\{(1-\theta_i^*)C^{(i)}\big(k_{\bar{j}(i)-1}^{(i)}\big) + \theta_i^* C^{(i)}\big(k_{\bar{j}(i)}^{(i)}\big)\Big\}. \quad (14)$$

For $\sum_i w_i k_{J(i)}^{(i)} \leq K$,

$$\bar{\mathcal{B}}(K) = \sum_i w_i C^{(i)}\big(k_{J(i)}^{(i)}\big) \quad (15)$$

In both cases $\bar{\mathcal{B}}(K)$ is the smallest model-independent upper bound on the price of the basket option, in the sense that we can find models which are consistent with the observed call prices for which the fair price for the basket option is arbitrarily close to $\bar{\mathcal{B}}(K)$.

3 Numerical Simulation and Analysis

Volatility surfaces for various maturity dates were provided for the top 40 equities on the Johannesburg Stock Exchange (JSE). The moneyness for each of the maturities was also provided for each equity. The values of moneyness used ranged from 75% to 125% for all traded calls. In addition to this, the weighting of each of the top 40 constituents was provided along with the futures term structures for those equities. The data was gathered from the South African Futures Exchange (SAFEX) from the time period between 2009 and 2018.

The observed price of the basket option on the JSE Top40 was computed for a variety of moneyness values (from 75% to 125% of the spot price) and a variety of maturities. These were calculated using these values along with the implied volatilities from the volatility surface given. The basket option price observed in the market was plotted as a function of maturity and moneyness below:

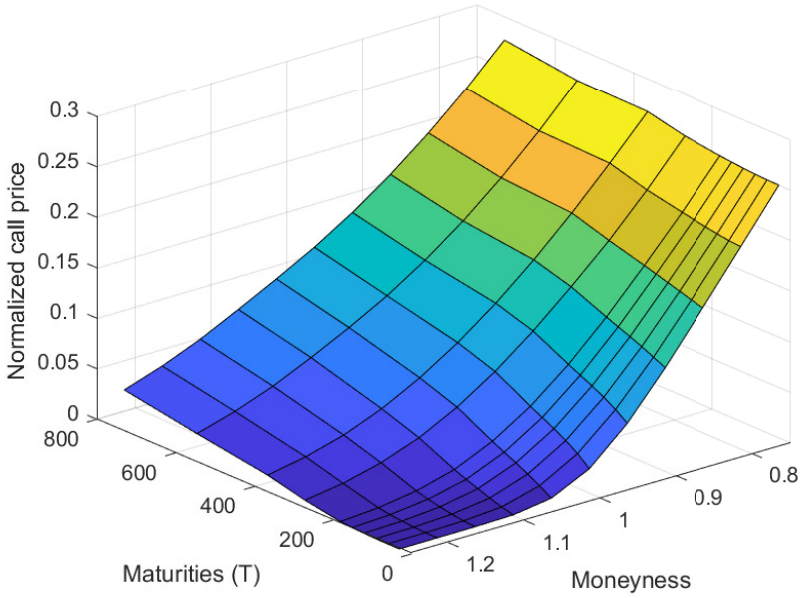


Figure 4: The surface curve of observed normalised call prices on the JSE Top40 as a function of moneyness and maturities (T)

As expected, figure 4 shows that the value of a call option on the JSE Top40 index (the basket option) is a decreasing function of strike and an increasing function of maturity. This simply shows that the given data behaves as we expected it to.

The smallest model-independent upper bound, $\bar{\mathcal{B}}(K)$, was then computed in line with the methodology outlined in (8). The observed strike prices $(k_j^{(i)})_{\{1\leq j\leq J(i)\}}$ of call options which are traded on equities $\left(X^{(i)}\right)_{1\leq i\leq N}$ have a maturity M. A basket option on all of the equities $X^{(i)}$ has a payoff of $\sum_i (w_i X_M^{(i)} - K)^+$. The model independent bounds developed in (8) imply that the fair price of the basket option, $B(K)$ is bounded as follows: $B(K) \leq \bar{\mathcal{B}}(K)$ where $\bar{\mathcal{B}}(K)$ is determined for two cases. The upper bound of the price of a basket option is determined using the equations outlined in theorem 2.1.

In theory, the weightings are determined by the respective market capitalisations of each of the equities in the JSE top 40 constituents. However, because of the way in which the JSE Top40 index is reported, the weightings are adjusted as outlined in section 3.1. The implied volatilities (from the volatility surfaces), equity prices $X^{(i)}$, maturities and strike prices $(k_j^{(i)})_{\{1\leq j\leq J(i)\}}$ are used as inputs into the Black-Scholes model to determine the observed market call prices $C^{(i)}$ for all 40 equities and maturity dates considered. The stock price on any particular day, denoted by time 0, is simply the futures prices on that day: $X^{(i)}$. The remainder of the inputs into the Black-Scholes model are gathered from SAFEX directly. Using the individual call prices $C^{(i)}\left(k_{J(i)}^{(i)}\right)$, the smallest model-independent upper bound, $\bar{\mathcal{B}}(K)$, is determined for the second case.

For the first case, the inputs used in theorem 2.1 are either directly observable or are calculated in line with the explanation in section 2. Specifically, the explanations of $C^{(i)}\left(k_{\tilde{j}(i)}^{(i)}\right)$ and $C^{(i)}\left(k_{\tilde{j}(i)-1}^{(i)}\right)$ as well as θ_i^* are discussed in this section.

3.1 JSE Top40 Index

The JSE Top40 is a weighted index in which the market capitalisation of each constituent of the index determines its weighting in the index. The index value therefore changes as the prices of the constituents change over time. The constituents having larger market capitalisation's affecting the index value changes more than price changes of constituents having lower market capitalisation's (13). According to (13), the index value is merely a number which shows the market value of the index constituents with reference to some starting point at a specific point in time.

In order to compare the share prices to the index value, the weightings, w_i, need to be changed from relative percentage values (that is, the fraction of the market capitalisation's) to values which, when summed up will give the total index value. These new weightings (now not percentage values any longer) will henceforth be denoted by $\tilde{w}_i$. Now, if a weighted index 'price' can be determined (for each T)

from the weighted prices of the constituents and denoted by Z_T as follows:

$$Z_T = \sum_i w_i F_T^i$$

where F_T is the forward price of the underlying for maturity T. If the index value for any particular T is denoted by Y_T then the 'conversion factor' between the index value and the index 'price', henceforth denoted by N_T, can be calculated as follows:

$$N_T = \frac{Y_T}{Z_T}.$$

$\tilde{w}_i$ is then simply the percentage weight converted to a weight with respect to index points:

$$\tilde{w}_i = w_i N_T$$

and is the weighting which will be used in all above calculations in place of w_i to account for the index being reported in index points as opposed to a weighted index 'price'.

3.2 The Call Price Function

As mentioned in section 2.2, there are only finitely traded strikes in the market. The optimisation exercise discussed in this section must make use of a call price function (with the call price given as an output based on a given strike) based on these strikes which are observable in the market. The call price function can be represented as a piece-wise linear function based on the discretely observable call prices for the strikes which are traded in the market as follows:

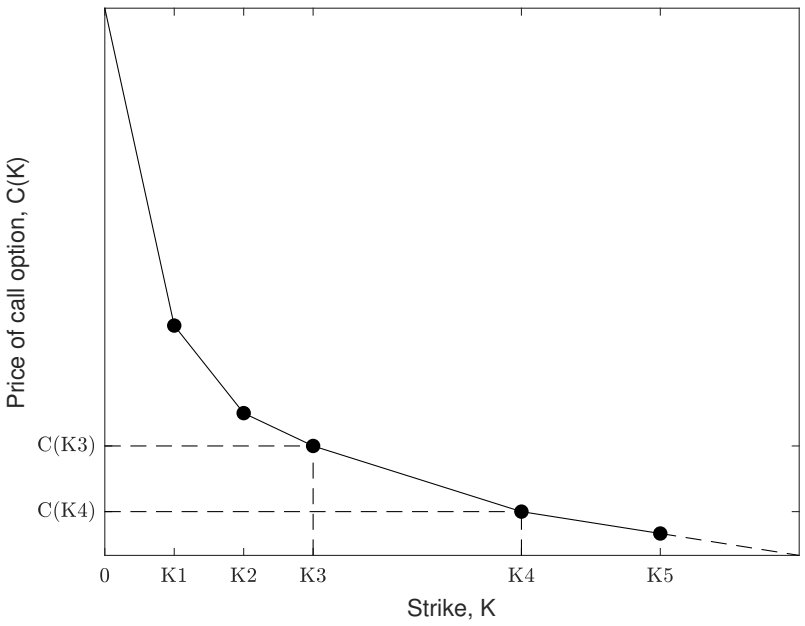


Figure 5: Piece-wise linear call price function based on market-observable call prices

From a practical point of view, one can think of the call price function as linearly interpolating between market-observable strikes and the corresponding call prices in order to determine the call price for a strike which lies between market observable strikes. It is important to note, however, that for strikes which are larger than the highest market-observable strike, an additional point is added to the call price function. This point is for a synthetically created strike of $k^{(i)}_{J(i)+1}$. On the figure above, this strike is given by extrapolating using the dotted line to get the strike for which the call price is zero. As discussed, the call price is expected to be a convex, decreasing function of the strike price. According to (8), the creation of this synthetic strike is simply a linear extrapolation of the existing call price function. It is calculated by first defining $\Delta_0 = e^{rT}$ and determining the gradient of each section of the function as follows:

$$\Delta^{(i)}_j = \frac{C^{(i)}\big(k^{(i)}_{j-1}\big) - C^{(i)}\big(k^{(i)}_j\big)}{k^{(i)}_j - k^{(i)}_{j-1}}$$

We then define Δ_∞ to be less than the minimum of these gradients (which is the final gradient if the function is convex). From this, the synthetic strike, $k_{J(i)+1}$, is then calculated as:

$$k_{J(i)+1} = k_{J(i)} + \frac{C(k_{J(i)}}{\Delta_\infty}$$

where $k_{J(i)}$ is the largest market-observable strike.

For the second case of theorem 2.1, the lowest call prices (corresponding to the highest observable strikes) are used. However, for the first part, the optimisation exercise described in section 2.2 is used. The objective function in equation 4, is minimised by using the piece-wise call price function shown above. If the strike isn't traded in the market (for example, the input into the call function in equation 4 is $\frac{\lambda_i^* K}{w_i}$ which is most likely not equal to the strike of a traded call option), then the value of the call option was linearly interpolated from the call price function as shown above. In this case, the value of the call option, $C\left(\frac{\lambda_i^* K}{w_i}\right)$, is given by the red 'x' on figure 5.

3.3 Numerical Implementation

The upper bound of a basket option was determined using theorem 2.1. In the second case, the computation of the upper bound, $\bar{\mathcal{B}}(K)$, is trivial. In the first case, however, an optimisation exercise is required. The objective function, equation 4, is minimised with the optimal choice of λ_i^* being the output of the exercise. The call price price function in figure 5 is then used to determine what $k^{(i)}_{j(\bar{i})-1}$ and $k^{(i)}_{j(\bar{i})}$ are. In other words, the market-observable strike prices which bound $\frac{\lambda_i^* K}{w_i}$ are referred to as $k^{(i)}_{j(\bar{i})-1}$ and $k^{(i)}_{j(\bar{i})}$. The corresponding values of $C_{(i)}(k^{(i)}_{j(\bar{i})-1})$ and $C_{(i)}(k^{(i)}_{j(\bar{i})})$ are determined from the observable call prices. The strike prices, $k^{(i)}_{j(\bar{i})-1}$ and $k^{(i)}_{j(\bar{i})}$, are used along with equation 8 to calculate the value of θ_i^* which is used in conjunction with equation 14 in theorem 2.1 to determine the upper bound of the basket option, $\bar{\mathcal{B}}(K)$.

It is important to note that call options are not liquidly traded on the top 40 equities on the JSE, especially not call options for a range of strike prices. For this reason, equations 14 and 15 in theorem 2.1 have to be adjusted such that the super-hedging strategy still involves holdings in the equities for which call options are liquidly traded. For the equities in which call options are not liquidly traded, the spot price was used as a call option price - this imitates a call option with a strike price of zero. For example, if call options for the top j equities were used (and positions in the underlying were used for the remainder of the equities) to create a super-hedging portfolio, equation 15 would be written as:

$$\bar{\mathcal{B}}(K) = \sum_{i=1}^{j} w_i C^{(i)}\left(k^{(i)}_{J(i)}\right) + \sum_{i=j+1}^{40} w_i F_0^{(i)}. \tag{16}$$

In our analyses, we investigate the effect of changing j on the resulting upper

bound of the option by varying the number of calls (as opposed to positions in the underlying) which are used in the super-hedging portfolio.

3.4 Results and Analysis

The sample data for the date 31-12-2018 was used to generate the following results. Data from a different date could have been used but the effect would be expected to be the same. The first point to be investigate was the effect of the number of equity call options used to estimate the upper bound of the price of the basket option as shown in equation 16 - that is, the effect of a changing j. It would be expected that increasing the number of equities for which the call option price is used (as opposed to the spot price) would more accurately estimate the minimum upper bound, $\bar{\mathcal{B}}(K)$. Because $\bar{\mathcal{B}}(K)$ is a minimum upper bound, increasing j should therefore decrease $\bar{\mathcal{B}}(K)$. If one were to think of equation 16 very simply, the call price of an option will be less than its equivalent spot price and therefore including the call price of more equities (as opposed to the equivalent spot price) should decrease the value of $\bar{\mathcal{B}}(K)$. When the number of call options used to estimate $\bar{\mathcal{B}}(K)$ was changed from call options on the top 5 to the top 15 equities, the following result was produced:

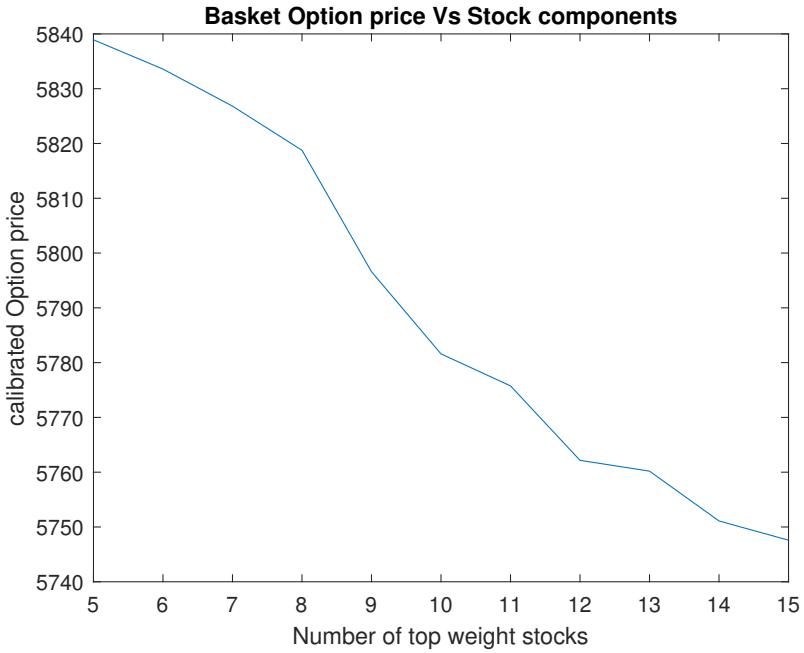


Figure 6: The minimum upper bound of a basket option based on a changing number of call options used to estimate the minimum upper bound from the top 5 to 15 equities. This plot is produced for an at-the-money option with a maturity of 1 year.

This plot was produced for an at-the-money basket option with a maturity of one year. As can be seen in figure 6, increasing the number of call options used does, as expected, decrease our minimum upper bound of the basket option price. Increasing the number of call options used to estimate $\bar{\mathcal{B}}(K)$ (as opposed to using the spot prices) would be expected to increase the accuracy of the estimate. This is because the use of the prices of more call options makes use of more market information than the use of the spot prices. As discussed in (7), robust hedging involves the use of observable call prices to "infer" a distribution of the equity price process. Therefore, the use of more call prices (and thus more information) should "infer" a more realistic distribution which should lead to a more accurate minimum upper bound.

It is worth noting, however, that the upper bound of $\bar{\mathcal{B}}(K)$ (or the cost of the super replicating portfolio) changes by less than $ZAR100$ from approximately $ZAR5840$ to just less than $ZAR5750$ when the number of equities changes from 5 to 15. This is a relatively small change of less than 2% of the original upper bound (using call prices on the top 5 equities). From this, it seems that the upper bound is relatively static as incorporating more market information doesn't change its value significantly. In fact, if the call options on the top 25 equities were used in the super-

hedging portfolio (implied volatility surfaces were only available for the top 25 equities which made up approximately 70% of the Top40 index in terms of market capitalisation), the effect can be seen below:

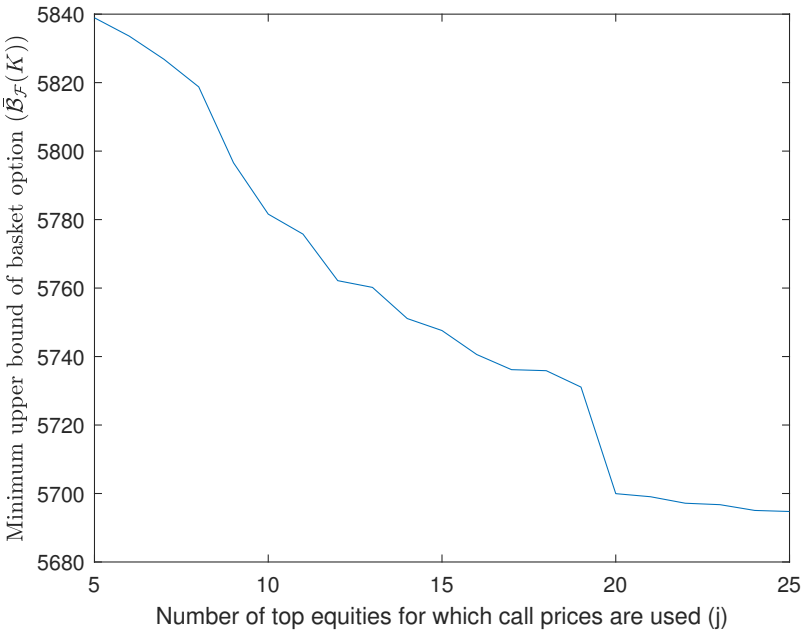


Figure 7: The minimum upper bound of a basket option based on a changing number of call options used to estimate the minimum upper bound from the top 5 to 25 equities. This plot is produced for an at-the-money option with a maturity of 1 year.

As can be seen in this plot, there is a decrease in $\bar{\mathcal{B}}(K)$ as the number of equities for which the call options are used in the super-hedging strategy (j in equation 16) increases as before. However, this effect only seems to be significant up to a point. After this point, the effect of j is relatively insignificant. We can postulate from this that the optimal number of equities for which the call options are used in the super-hedging strategy is 20. This is most likely because the weights of the equities outside the top 20 aren't high enough to significantly impact the value of $\bar{\mathcal{B}}(K)$. A table showing the total cumulative proportion $(\sum_i w_i)$ of the JSE Top40 index of the top j equities is given:

Table 1: The total proportion of the JSE Top40 index represented by the top j equities and the effect that this has on $\bar{\mathcal{B}}(K)$

Top equities j	Cumulative weight $\sum_{i=1}^{j} w_i$ (%)	Upper bound $\bar{\mathcal{B}}(K)$	Slope	Component weight w_i (%)
5	44.74	5838.927 703		
6	48.5	5833.583 232	−5.344 471 444	3.76
7	51.89	5826.816 953	−6.766 278 82	3.39
8	55	5818.759 841	−8.057 111 717	3.11
9	57.37	5796.594 567	−22.165 274 56	2.37
10	59.3	5781.594 376	−15.000 191 01	1.93
11	61.03	5775.756 541	−5.837 834 765	1.73
12	62.58	5762.162 483	−13.594 058 33	1.55
13	64.07	5760.192 422	−1.970 061 039	1.49
14	65.47	5751.096 922	−9.095 499 86	1.4
15	66.79	5747.593 434	−3.503 487 972	1.32
16	68.02	5740.589 977	−7.003 456 591	1.23
17	69.22	5736.171 092	−4.418 885 342	1.2
18	70.33	5735.869 455	−0.301 637 179	1.11
19	71.31	5731.077 664	−4.791 790 728	0.98
20	72.26	5699.956 881	−31.120 783 12	0.95
21	73.18	5699.055 09	−0.901 791 127	0.92
22	74.07	5697.160 559	−1.894 531 077	0.89
23	74.94	5696.735 952	−0.424 606 815	0.87
24	75.78	5695.054 456	−1.681 496 3	0.84
25	76.46	5694.745 007	−0.309 449 024	0.68

The estimate of $\bar{\mathcal{B}}(K)$ is given for a changing j. This is the data from which figure 7 is plotted. Importantly, the total weight that the top j equities held in the JSE Top40 are given in this table and can be seen to be increasing at a decreasing rate (as j increases). This is intuitive in that each added j has a lower market capitalisation. This table reinforces the idea that the effect of an increasing j on the value of $\bar{\mathcal{B}}(K)$ is insignificant for equities outside the top 20 (partly) because the weights which they have in the Top40 index are either 1% or less (see the values of w_i).

The value of $\bar{\mathcal{B}}(K)$ as a function of j is once again given in figure 8 below. The market price of the base is plotted along with this to show the relative value of the bound to the market price:

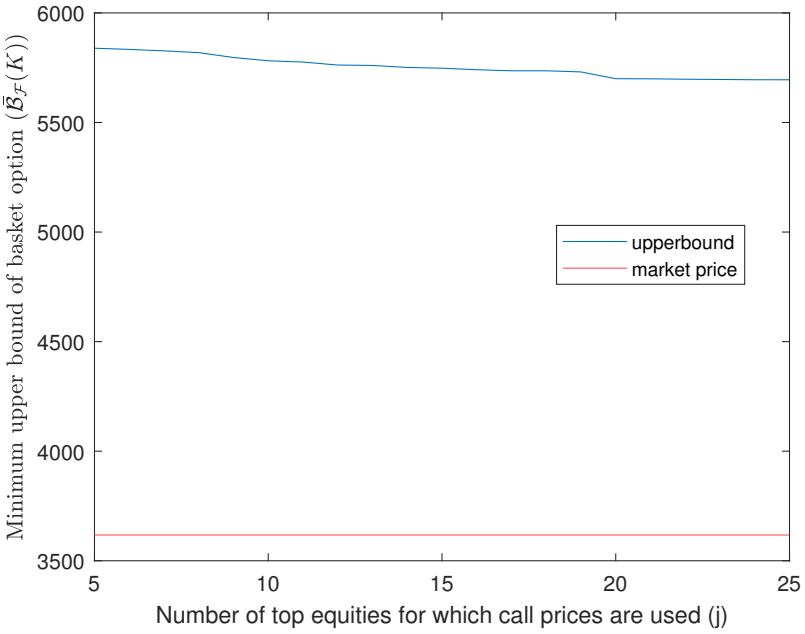


Figure 8: The minimum upper bound of a basket option based on a changing number of call options used to estimate the minimum upper bound from the top 5 to 25 equities along with the observed market prices of the JSE Top40 index option (the basket option). This plot is produced for an at-the-money option with a maturity of 1 year.

As can be seen in the figure, the observed market prices are significantly below the upper bound, $\bar{\mathcal{B}}(K)$. It is worth noting that, firstly, it is obvious that $\bar{\mathcal{B}}(K)$ is indeed an upper bound, albeit approximately 60% higher than the observed market value of the option. This high value of the upper bound implies that it may not be the $minimum$ upper bound. However, (5) states that the implemented super-hedging strategy is "less ambitious in scope" in that it doesn't aim to estimate a price of the option but rather a strict upper bound. The upper bound therefore may not necessarily approach the price of the actual option. The super-hedging strategy is expensive as it is required, by definition, to completely hedge the exotic derivative being priced (in this case a basket option). In practice, super-hedging is not used as a hedging strategy because it is often too expensive (11). For this reason, the upper bound of a basket option price determined from a super-hedging strategy will be high relative to the observed market price.

In addition to the above explanation as to why the value of $\bar{\mathcal{B}}(K)$ is so much higher than the observed market prices, it is important to consider that we haven't estimated the lower bound. Because the scope of our study excluded the computation

of the lower bound of a basket option, we are not able to hypothesize where this may be. A simplified sketch can be used to explain this:

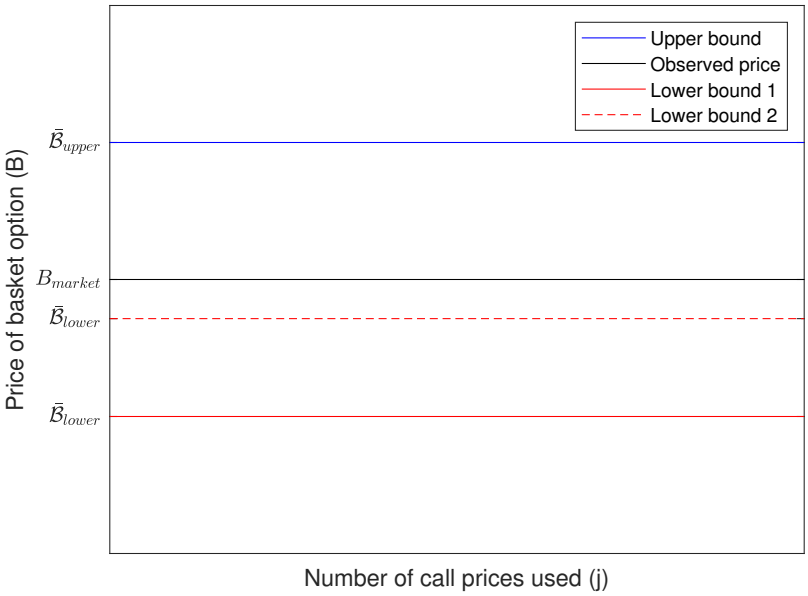


Figure 9: The minimum upper bound and possible maximum lower bounds of a basket option along with the observed market prices

In this sketch, the upper bound is presented as constant. This is obviously a simplification but as it changes very little, it is presented as constant for this explanation. If the lower bound is as far from the observed price (that is, "Lower bound 1" in the figure), then we can conclude that the nature of the super-hedging strategy is conservative for both the lower and upper bounds. However, if the observed price is closer to the lower bound than to the upper bound (that is, if the lower bound is "Lower bound 2" in the figure), then the JSE behaves in such a way that it is comparable to a model under which the lower bound holds.

In section 2, it is discussed that the minimum upper bound generally coincides with the price under a model in which the constituents are co-monotonic. This is substantiated by the "Numerical Results" of (5) which states that it is precisely when a model in which the correlation is close to 1 that the upper bound represents the price predicted by the model. This is intuitive in that, if the constituents behave in a way such that they have a correlation of 1, then if the price of a call option with a fixed maturity in a certain direction (when the strike changes), the prices of the call options on the other equities with the same maturities will all behave in an

analogous way. In this scenario, a weighted investment in the constituents of a basket option will perfectly hedge the basket option itself and the upper bound will represent the price of the basket option. From the results of this study, because the observed prices differ significantly from the upper bound, we can conclude that the constituents of the JSE Top40 index are not co-monotonic.

Investigating the error (difference between the derived upper bound and the observed price) as a percentage of the observed price (see table 2 below), we see that this increases as the time to maturity increases. This maybe because, liquidity in trading of futures contracts reduces as the time to maturity increases, as it is common for investors to use short-dated futures for hedging and then roll the hedge forward. This results in less liquidity in trading of call options written on these futures contracts. Hence dealers will be expected to demand an extra premium to be encouraged to trade in less liquid call options, resulting in higher prices and thus it becomes more costly to super-replicate the basket option using traded calls on the market. In addition to this, the model used for option pricing isn't co-monotonic. Thus as maturity increases, we would expect less correlation among the underlying assets in the basket, thus the model bound would move further away from the observed prices.

Table 2: The error (that is, the difference between the upper bound and the observed price) as a function of different maturities of a basket option on the JSE Top40

Time to maturity T	Error %
30	26.52
60	33.62
90	34.12
120	35.2
150	44.12
180	48.51
270	55.71
365	59.82
545	66.82
730	72

4 Conclusion and Proposed Future Work

Theorem 2.1, which is outlined in (8), was used to determine the robust upper bound, $\bar{\mathcal{B}}(K)$, of a basket option on the JSE top 40 equities. As discussed in section 3.4, the observed market prices lie far below the minimum upper bound. This indicates that the constituents of the JSE Top40 are not co-monotonic. It also implies that the super-hedging strategy is expensive which is why it is rarely used in practice according to (11).

However, as the number of equities for which the call prices were used in the computation of $\bar{\mathcal{B}}(K)$ was increased, the value of the upper bound seemed to decrease until the top 20 call prices were used. This indicated both that the use of the top 20 equity call prices in the computation is optimal and that the upper bound approaches some value (which is approximately $ZAR5700$) for an at-the-money basket option with a maturity of 1 year.

In addition to this, it appears that the observed price of the basket option is closed to the upper bound for smaller maturities. This indicates that the JSE top 40 equities are more co-monotonic for shorted maturities than for longer maturities. In addition to this, it may indicate that there is less liquidity in the market for shorter-dated options.

The predominant future work which is proposed based on this study is the computation of the lower bound of a basket option. In this way, the position of the observable market prices of the basket option within a band of model-independent can be determined. This will help in understanding whether the upper bound is an extreme estimate (if the lower bound is significantly lower than the observed market prices) or whether the observed prices are simply close to the lower bound. In addition to this, any method that can account for the correlation structure between the JSE Top40 constituents will probably improve the minimum upper bound in that it will be able to impose constraints on the co-monotonicity of constituents. This should lower the upper bound as it may no longer have to coincide with the situation in which the constituents have a correlation of 1.

Additionally, a time series of the data can be analysed as data from 2009 to 2018 was analysed. Although a major trend in this time series can be foreseen, it may provide insight into market changes over this period. For example, if the top 40 equities were more co-monotonic and therefore approached $\bar{\mathcal{B}}(K)$ at certain points in time.

Bibliography

[1] Arts, P.; Pricing the basket option and exploring its volatility structure. MSc thesis, Utretcht University, 1999

[2] Bakshi, G., Cao, C. and Chen, Z.; Pricing and hedging long term options. The Journal of Econometrics, vol. 94, 2000, pp. 277318.

[3] Beisser, J.; Topics in Finance- A conditional expectation approach to value Asian, Basket and Spread Options. PhD Thesis Johannes Gutenberg University Mainz, 2001.

[4] Breeden, Douglas T., and Robert H. Litzenberger.; Prices of State-Contingent Claims Implicit in Option Prices. The Journal of Business, vol. 51, no. 4, 1978, pp. 621651. JSTOR, www.jstor.org/stable/2352653.

[5] Brown, H.M., Hobson, D.G. and Rogers, L.C.G.; Robust hedging of barrier options. *Mathematical Finance*, **11**, 285-314, 2001.

[6] Deelstra, G., Liinev, J., Vanmaele, M.; Pricing of Arithmetic Basket Options by conditioning. Insurance: Mathematics and Economics, vol. 34, 2004, pp 1-23.

[7] Hobson, D.G.; Robust hedging of the lookback option. *Finance and Stochastics*, 1998, vol. 2, pp 329-347

[8] Hobson, D.G., Laurence, P. and Wang, T.H.; Static-Arbitrage Upper Bounds for the Prices of Basket Options. *Quantitative Finance*, 2005.

[9] Hobson, D.G.; The Skorokhod Embedding Problem and Model-Independent Bounds for Option Prices. *Finance and Stochastics*, 2009.

[10] Hodges, S.D., Neuberger, A.; Rational bounds on the prices of exotic options; FORC Preprint, 98/94, University of Warwick.

[11] Investopedia. (2019). Super-Hedging. [online] Available at: *https://www.investopedia.com/terms/s/super-hedging.asp*. [Accessed 1 Jul. 2019].

[12] Jse.co.za. (2015). JSE Equity Options Explained. [online] Available at: *https://www.jse.co.za/content/JSEEducationItems/JSEEquityOptionsExplained.pdf.* [Accessed 27 Jun. 2019]

[13] Jse.co.za. (2015). Guide to Calculation Methods for the FTSE/JSE Africa Index Series. [online] Available at: *https://www.jse.co.za/content/JSEIndexClassificationandCodesItems/Calculation%20Guide.pdf.* [Accessed 27 Jun. 2019]

[14] Van Diepen E. (2002). Baskets with a smile. An alternative model for pricing basket options. MSc Thesis, Free University Amsterdam.

On Level Dependence of Volatility and the CEV Market Model

TEAM UCT

ZEZHUN CHEN, London School of Economics
VALENTIN COURGEAU, Imperial College London
TINO MUCHABAIWA, University of Cape Town
ALAN YEUNG, University of Cape Town

Supervisors:
ALEX BACKWELL & RALPH RUDD, University of Cape Town

African Collaboration for Quantitative Finance and Risk Research

Contents

Chapter 1

Introduction

This report presents the study of volatility level-dependence in the LIBOR Market model using the CEV model during the Financial Mathematics Team Challenge, UCT, held from 25th June to July 2nd 2019.

The aim of the project is to uncover ways to calibrate the *constant elasticity of variance* (Andersen and Andreasen, 2000) or *CEV* model, to incorporate level-dependent volatility (volatility with dynamics depending on the changes but also the actual value of the underlying) as presented in empirical studies in Filipovic et al. (2017, Section IV, G) and Piazzesi (2010, Section 7.7). Informally, we model the forward rate process $(L_k(t),\ t \geq 0)$ defined by an initial value $L_k(0)$ and the stochastic differential equation

$$dL_k(t) = \sigma(L_k(t))^\gamma dW_k(t), \qquad t \geq 0,$$

for $k \in \{1, \ldots, M+1\}$ where $(W_k(t),\ t \geq 0)$ is a Brownian motion under the T_{k+1}-forward measure for $t \in [T_k, T_{k+1}]$ (see details in Section 2.1), $\gamma \geq 0$ is the CEV parameter and $\sigma > 0$ is the volatility. This parameter allows to go from a Gaussian model ($\gamma = 0$) to a log-normal model ($\gamma = 1$) via the so-called square-root process ($\gamma = 1/2$) (Piazzesi, 2010, Section 3).

An important motivation of this challenge comes from the conjecture of recent results (Filipovic et al., 2017) along with classical observations about financial data (Andersen and Andreasen, 2000). Those references explore further the impact of the underlying on the diffusion coefficient and precise the notion of *level dependence of volatility*. Simple vector autoregressive models find a positive correlation between their squared residuals and the interest rate levels (Piazzesi, 2010, Section 7.7). Alternatively, the approach that is studied herein is a simple regression of changes in implied volatility against change in interest-rate (Filipovic et al., 2017, Section IV.G).

We propose to investigate the role the CEV model might play in handling the volatility level-dependence as the CEV parameter γ can potentially be adjusted to do so. For that, we compare *four* calibration methodologies and their effect on

this issue. Those will span from fitting the market skew using different strikes to optimising the profit-and-loss of a CEV hedging strategy.

1.1 Report structure

The main task of the project is to consider the calibration of the Constant Elasticity of Variance (CEV) model (Andersen and Andreasen, 2000). Our report is subdivided into smaller tasks. In Section 1.2 and 1.3, we introduce the dataset provided on EURIBOR caps from 2013 and 2014. We also recall the definition of caplets and caps as they are the derivatives of interest in this project. Then, in Section 2, we present the LIBOR market model (LMM) and one of its extensions that is the CEV model as well as pricing formulas for caplets and caps. This is augmented by a practical and succinct understanding of the CEV model on volatility and its level dependence (Section 2.4). Section 3 comprises of details on the implementation of the implied volatility extraction and the caplet pricing formulas

1.2 Data overview

For this challenge, the data was two years worth of three daily EURIBOR 10-year caps log-normal implied volatilities and corresponding 20 semi-annual bonds prices (more precisely for the years 2013 and 2014). As required, the strikes for the three caps ($H \in \{0.01, 0.0175, 0.025\}$), the underlying forward rate, bond maturities and daily short rates were included in the dataset. This short rate is taken to be the simply-compounded instantaneous rate.

The daily bonds data show an increasing term structure of yields which is expected. Below is a plot of the term structure for bonds taken at days 1, 250 and 500 of the 504 days given in the data. The days were chosen at random.

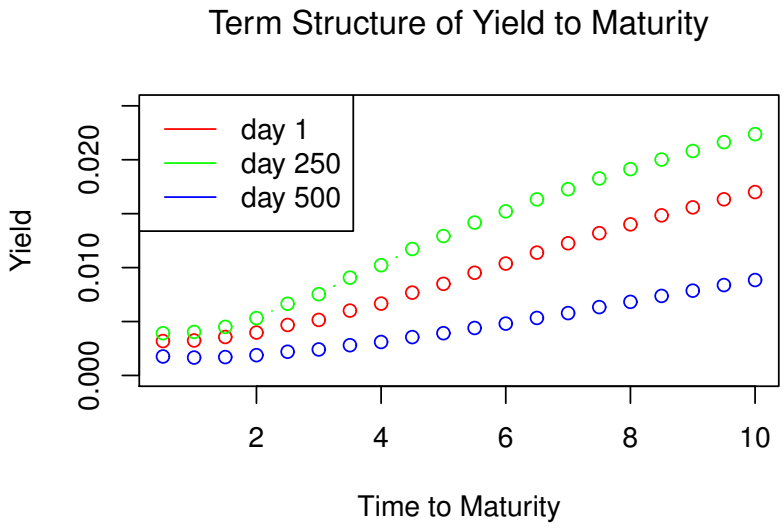


Figure 1.1: Yield plotted against term to maturity of bonds in the dataset.

The log-normal implied volatility of the three different strikes given in the dataset show that as the strike increases the volatility decreases. This is expected, caps with lower strike (in-the-money) have a greater price as to comparable caps with higher strike (out-the-money). Furthermore, price and volatility have positive relationship in the log-normal Model. Hence as strike increases, volatility decreases. Below are three histograms of the implied volatility of the log-normal model of the 504 days given under three different strikes $(0.01, 0.0175, 0.025)$.

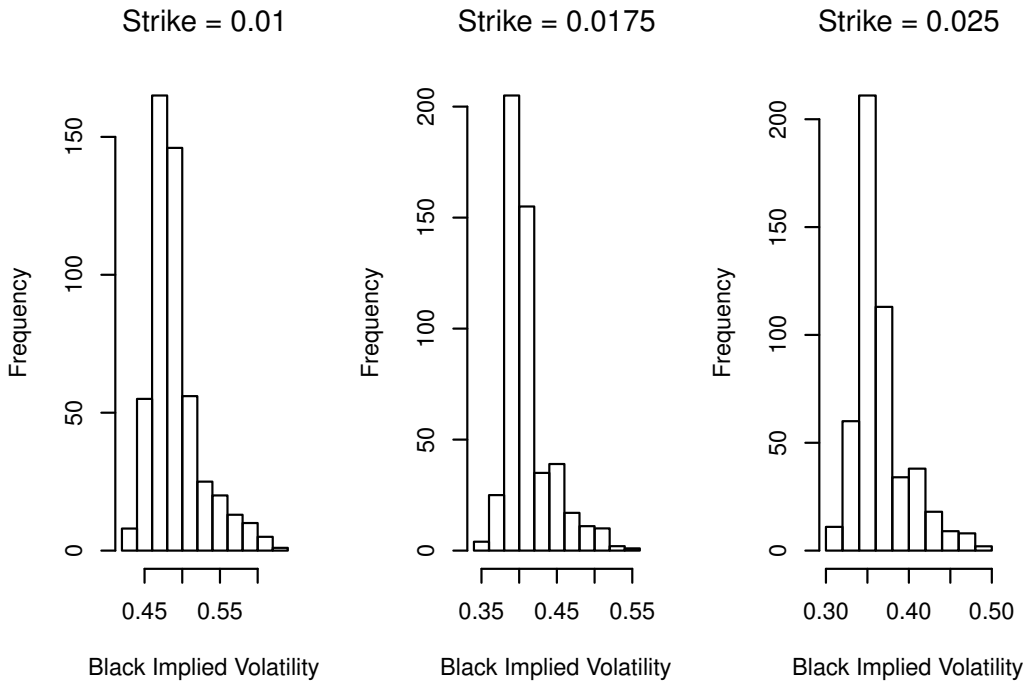


Figure 1.2: Histograms of daily implied volatility at different strikes.

Remark 1. *In the paper the strike $H_2 = 0.0175$ is referred to as the at-the-money strike, however it is not at-the-money but close to at-the-money.*

1.3 Cap/caplet definition

A cap is a contract which limits the the amount of interest which the buyer has to pay to a limit, called the strike. The buyer can be viewed as a payer of an interest rate swap where each exchange payment is executed only if it has positive value. A cap is made up of caplets just like an interest rate swap is made of forward rate agreements (Brigo and Mercurio, 2007). The caplet payoff is

$$\delta(L - H)^+,$$

where $x^+ := \max(0, x)$, δ is the difference between maturity time and when the caplet resets, L is the forward rate at its specific maturity time and H is the strike rate.

Chapter 2

CEV Market Model and caps

2.1 LIBOR Market Model

We consider the LIBOR Market Model (LMM) in this section. Similarly to Section 2, Andersen and Andreasen (2000), consider the sequence of increasing maturity structure $0 = T_0 < T_1 < \cdots < T_{M+1}$ and define the right-continuous mapping function $n : [0, \infty) \to \mathbb{N}$ such that $T_{n(t)-1} \leq t < T_{n(t)}$. For $t,\ T \geq 0$, denote the price at time t of a zero-coupon bond with maturity T as $P(t, T)$. The discrete forward rates on the aforementioned maturity structure is defined as

$$L_k(t) := \frac{1}{\delta_k}\left(\frac{P(t, T_k)}{P(t, T_{k+1})} - 1\right), \quad \text{with } \delta_k := T_{k+1} - T_k.$$

for $t \leq T_k$ and $k \leq M$. For each individual time set $[T_k, T_{k+1}]$, we fix the probability measured to be the T_{k+1}-forward measure $\mathbb{Q}^{k+1}$ as induced by the zero-coupon bond with T_{k+1} maturity as numeraire.
As per Andersen and Andreasen (2000), we assume $\mathbb{Q}^{k+1}$ to be unique which implies no arbitrage. This means that $P(t, T_k)/P(t, T_{k+1})$, and hence $L_k(t)$, are martingales.

Remark 2. *We will use both the $L_k(t)$ and $L(t, T_k)$ notations depending on which is easier to interpret given the context: for pricing theory, we shall mostly use $L_k(t)$ to denote the underlying rate and $L(t, T_k)$ when discussing properties of the cap option itself.*

2.2 CEV model

As defined in Andersen and Andreasen (2000), we define the forward rates dynamics as

$$dL_k(t) = \sigma(t) L_k^{\gamma}(t) dW_k(t), \qquad k \in \{1, \ldots, M\} \tag{2.1}$$

where $\gamma > 0$ and $t \mapsto \sigma(t)$ is a positive bounded deterministic function. Local Lipschitz continuity and linear growth of $x \mapsto x^{\gamma}$ are equivalent to the existence

and uniqueness of no-arbitrage solutions for $L_k(t)$ under the collection of measures $(\mathbb{Q}^{k+1},\ n(t) \leq k \leq M)$, respectively.

2.3 Cap and caplet under CEV

The price of a caplet is known exactly under the CEV model as follows

Theorem 1. *(Andersen and Andreasen, 2000, Theorem 3)*
Let $C_k(t)$ denote the price of a LIBOR caplet with strike H and payment time T_{k+1}. Let $\Phi(\cdot)$ be the standard normal cumulative distribution function, and $\chi^2(\cdot, \theta, \lambda)$ be the cumulative distribution of a non-central χ^2 distribution random variable with non-centrality parameter λ and θ degrees of freedom. Define

$$v_k(t,T_k) := \int_t^{T_k} \sigma(s)^2 ds; \qquad a := \frac{H^{2(1-\gamma)}}{(1-\gamma)^2 v_k(t,T_k)}; \qquad b := \frac{1}{1-\gamma};$$
$$c := \frac{L_k(t)^{2(1-\gamma)}}{(1-\gamma)^2 v_k(t,T_k)}; \qquad x_\pm := \frac{\ln\left(L_k(t)/H\right) \pm \frac{1}{2} v_k(t,T_k)}{\sqrt{v_k(t,T_k)}}.$$

Assuming the forward rate dynamics are as specified in Equation (2.1), the arbitrage-free value of $C_k(\cdot)$ is given by the following:

1. *For $0 < \gamma < 1$ and an absorbing boundary at the level $L_k = 0$;*

$$C_k(t) = \delta_k P(t,T_{k+1}) \left\{ L_k(t)(1 - \chi^2(a, b+2, c)) - H\chi^2(c, b, a) \right\}. \tag{2.2}$$

2. *For $\gamma = 1$:*

$$C_k(t) = \delta_k P(t,T_{k+1}) \left\{ L_k(t)\Phi(x_+) - H\Phi(x_-) \right\}. \tag{2.3}$$

3. *For $\gamma > 1$:*

$$C_k(t) = \delta_k P(t,T_{k+1}) \left\{ L_k(t)(1 - \chi^2(c, -b, a) - H\chi^2(a, 2-b, c) \right\}. \tag{2.4}$$

We also augment this result with the derivation of the Gaussian-distributed implied volatilities.

Proposition 1. *Consider the framework of Theorem 1 and assume the forward rate dynamics are as specified by an initial value $L_k(0)$ and*

$$dL_k(t) = \sigma dW_k(t), \qquad t \geq 0.$$

Then, the arbitrage-free value of $C_k(\cdot)$ is given by the following:

$$C_k(t) = \delta_k P(t, T_{k+1})\sigma(T_k - t)^{1/2} \left\{\eta(t, T_k)\Phi(\eta(t, T_k)) + \phi\left(\psi(t, T_k)\right)\right\},$$

where $\eta(t, T_k) := (L(t) - H)/(\sigma\sqrt{T_k - t})$ and ϕ, Φ are respectively the standard normal probability density function (PDF) and cumulative distribution function (CDF).

Proof. We have directly $L_k(t) = L_k(0) + \sigma W_k(t)$ for any $t \geq 0$. Now, we compute $\mathbb{E}_{\mathbb{Q}^k}\{(L_{T_k} - H)^+|\mathcal{F}_t\}$ as follows

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^k}\left\{(L_{T_k} - H)^+|\mathcal{F}_t\right\} &= \mathbb{E}_{\mathbb{Q}^k}\left\{(L_k(0) + \sigma W_t + \sigma(W_{T_k} - W_t) - H)^+|\mathcal{F}_t\right\} \\
&= \int_{-\eta(t,T_k)}^{\infty} \left\{(L_k(t) - H + \sigma\sqrt{T_k - t}x)(2\pi)^{-1/2}\exp\{-x^2/2\}\right\} dx \\
&= (L_k(t) - H)\Phi\left(\eta(t, T_k)\right) + \sigma(T_k - t)^{1/2}\phi\left(\eta(t, T_k)\right) \\
&= \sigma(T_k - t)^{1/2}\left\{\eta(t, T_k)\Phi(\eta) + \phi\left(\eta(t, T_k)\right)\right\}
\end{aligned}$$

We then evaluate this expectation at time T_k and discount back to time t from payment time T_{k+1} and conclude. □

2.4 Intuition behind the CEV model

Recall that LMM corresponds to the LIBOR Market Model and CEV to the Constant Elasticity of Variance model.
The main task of this challenge is to calibrate the CEV model and infer the value of the parameter γ which correctly prices the caps as per Equation (2.1). The LMM model is not the best model neither is the normal model. In the LMM model we assume that γ is one and normal model γ is zero. Below is a plot of the weekly and daily changes in the implied volatility and the change in the at-the-money strike. This is done for γ equal to 1, 0.005 and 0.

Remark 3. *We denote by diffusion coefficient the coefficient of the Brownian motion.*

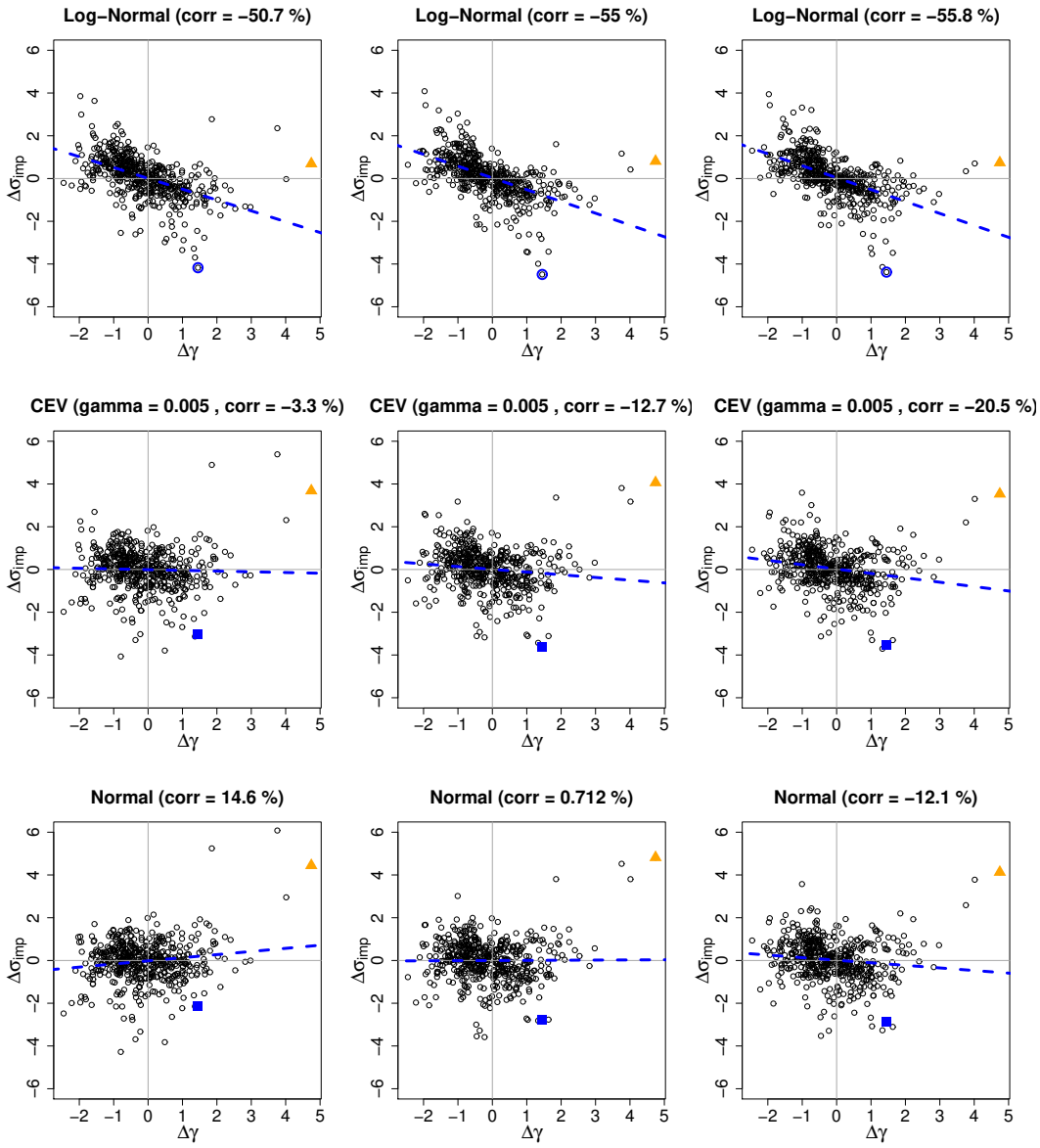


Figure 2.1: Regression of changes in implied volatility against change in the underlying interest-rate (given as at-the money strike) for the log-normal, CEV and normal cases, for respectively the strikes $\{0.01, 0.0175, 0.02\}$ column-wise.

The plots show that there is correlation between changes in implied volatility and changes in the forward rate.

2.4.1 Log-normal case ($\gamma = 1$)

Recall that the rate dynamics are now defined by

$$dL_k(t) = \sigma L_k(t) dW_k(t), \qquad t \geq 0.$$

The model features positive level-dependence between volatility and forward rate level. Indeed, the diffusion coefficient increases as L_k increase. The scatter plot of changes in implied volatility against changes in the forward rate might then feature negative correlation. Indeed, if the forward rate increases, the implied volatility decreases to accommodate the strong positive increase in the diffusion coefficient. Hence, observe the downward sloping dotted line for the log-normal model (first row, Figure 2.1). A positive change in the forward rate leads to a negative change in the volatility to compensate for strong positive level-dependency assumed by the model. We can conclude that the log-normal model would misinterpret an increase in the forward rate by having artificially low implied volatilities. This might affect a hedging portfolio for example.

2.4.2 Normal case ($\gamma = 0$)

In this case, one has that

$$dL_k(t) = \sigma dW_k(t), \qquad t \geq 0.$$

The normal model above assumes zero level-dependence between volatility and interest rate level as the forward is not apparent in the diffusion coefficient. Contrary to the log-normal case, an increase in the forward rate will potentially imply an increase in the implied volatility as it may try to replicate the presence of the underlying in the diffusion coefficient. Empirical evidence from the dataset (last row, Figure 2.1) shows that a positive change in the *at-the-money* strike leads to a positive change in the volatility. A blue dotted upward line makes this relationship more clear in the last row of Figure 2.1. This evidence shows that there is a need for some volatility level-dependence in our model as we observe positive correlation between changes in normal implied volatility and changes in the forward rate.

2.4.3 CEV case $(0 < \gamma < 1)$

The rate dynamics are now taken to be as follows

$$dL_k(t) = \sigma L_k^{\gamma}(t) dW_k(t), \qquad t \geq 0.$$

The correlation of implied volatility and forward rate level shown by the normal and the log-normal model show that the two models are not best fitted to price the caps as they promote ill-specified behaviour. On the other hand, the capacity to tune the impact of the underlying forward rate on the diffusion coefficient might

be helpful in taming the excessive of the underlying in the log-normal case or the absence of the underlying as per the normal case. Therefore, a CEV model above with an appropriate γ could address the level dependency in such a way that the correct amount of level dependency will give constant volatility. The second plot gives some insight, when γ is 0.005 the correlation is very low which means that there could be a γ between 0 and 1 which can give a CEV model with minimum correlation.

Chapter 3

Implementation

3.1 Testing results with the theory and intuition

During the project, we devised a series of tests for our implementation to test its behaviour in both an intuitive context or its asymptotic behaviour (e.g. as $\gamma \to 0$ or $\gamma \to 1$).

- Caplet pricing: caplets behave similarly to call options and have therefore an intuitive behaviour with respect to time-to-maturity, its strike rate and forward rate. We harness these properties to strengthen the foundations. As an example, we know that the value of the caplet should decrease as time to maturity increases. However, as the strike price increases a caplet with a small time-to-maturity would see its value decrease quicker than a caplet with larger time to maturity. See Section 3.2 for details.
- Implied volatility: we replicated the implied volatility tables presented in Andersen and Andreasen (2000) as featured in Section 3.3. An important test was to check the monotonicity of the pricing formula with respect to the volatility parameter σ. This allowed the computation times to speed up by a factor of approximately 100.
- Finally, we replicated the results from Figure 2, Andersen and Andreasen (2000) which sets the CEV implied volatility at $\sigma = 30\%$ for a strike $H = 6\%$ to obtain the corresponding price and computes implied volatilities for different strikes and γ values (Section 3.5).

3.2 Caplet Pricing

As defined above, we implement the closed-form solution of the caplet pricing formula to calculate the caplet (Theorem 1). The parameters from Section 4, Andersen and Andreasen (2000) for γ, the forward rate and volatility are used to price caplets.

The strike rate H ranges from 0.01 to 0.1 and the time-to-maturity T_k ranges from 1 to 30.

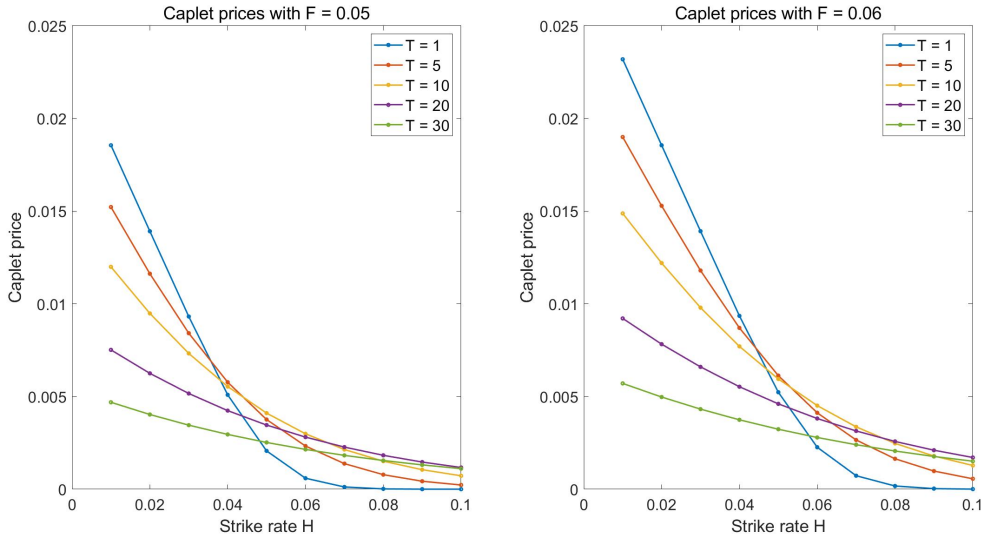


Figure 3.1: Caplet prices vs strike rate for $\gamma = 0.5$, forward rate of 0.05, implied volatility of 0.05 (RHS) and $\gamma = 1.5$, forward rate of 0.06, implied volatility of 0.83 (LHS).

Unfortunately, the seminal paper Andersen and Andreasen (2000) did not feature caplet prices as a benchmark hence we could not double-check those results. This said, we can infer the prices behaviour as follows: it is clear that as the strike rate increases, the caplet price decreases. This is as expected due to the fact that as strike rate increases, the caplet will be less in-the-money or more out-the=money which in turn drives the caplet price down. Moreover, when the strike rate is low (i.e., when the caplet is in-the-money), as time-to-maturity increases, the caplet price decreases. This is because the caplet is currently in-the-money, therefore if it is closer to maturity, the caplet will be realised in-the-money. On the other hand, when the strike rate is high (i.e., when the caplet is out-the-money), as time to maturity increases, the caplet price will decrease. This is because as time-to-maturity increases, the out-the-money caplet will have more time for the interest rate to increase so that it can be realised in-the-money.

3.3 Implied volatility of caplets

After calculating the caplet price from the CEV model, we could use this price to calculate the Black (log-normal) implied volatility. The relationship between the caplet price and volatility in the log-normal model is stated in section 2.3. We have used a bisection method to find the implied volatilities from the caplet prices. We used a tolerence level of 10^{-5} to implement the bisection method.
The tables below show the implied volatilities with the same parameters as the section above.

T_k/H	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1	0.3093	0.2655	0.2417	0.2256	0.2137	0.2042	0.1964	0.1899	0.1842	0.1792
5	0.3104	0.2662	0.2422	0.2261	0.2141	0.2045	0.1967	0.1901	0.1845	0.1795
10	0.3114	0.2669	0.2428	0.2266	0.2145	0.2049	0.1971	0.1905	0.1847	0.1797
20	0.3113	0.2677	0.2436	0.2273	0.2152	0.2056	0.1977	0.191	0.1853	0.1802
30	0.3084	0.267	0.2436	0.2276	0.2155	0.2060	0.1981	0.1914	0.1856	0.1806

Table 3.1: Implied volatilities with the bisection method with $\gamma = 0.5$, $\sigma_k(t) = 0.05$ and forward rate $F_k = 0.06$.

The table below shows the same table from Andersen and Andreasen (2000)'s paper.

T_k/H	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*	*0.06*	*0.07*	*0.08*	*0.09*	*0.1*
1	0.3109	0.2655	0.2417	0.2256	0.2137	0.2042	0.1964	0.1899	0.1842	0.1792
5	0.3104	0.2662	0.2422	0.2261	0.2141	0.2046	0.1967	0.1901	0.1844	0.1795
10	0.3114	0.2670	0.2428	0.2266	0.2145	0.2049	0.1971	0.1905	0.1847	0.1797
20	0.3113	0.2677	0.2436	0.2273	0.2152	0.2056	0.1977	0.1910	0.1853	0.1802
30	0.3084	0.2670	0.2436	0.2276	0.2155	0.2059	0.1981	0.1914	0.1856	0.1806

$\alpha = 0.5,\ \lambda_k = 0.05,\ F_k = 0.06$

Figure 3.2: Implied volatility as given in Figure 1A, Andersen and Andreasen (2000).

From the above two tables, we can see that all the entries are the same to 4 decimal places except the first entry where $H = 0.01$ and $T_k = 1$. The error might be caused by the non-central chi-squared distribution when the parameter is very small.
Table 3.2 below shows the implied volatilities with a different set of parameters and Figure 3.3 shows the same table from Andersen and Andreasen (2000). From the tables, it can be seen that when $\gamma = 0.5$, as the strike rate increases, implied volatility decreases. Whereas when $\gamma = 1.5$, as the strike rate increases, implied volatility also increases. This relationship is also shown graphically in Figure 3.4 which is presented in Andersen and Andreasen (2000).

T_k/H	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1	0.1526	0.1702	0.1835	0.1943	0.2034	0.2113	0.2184	0.2247	0.2305
5	0.1527	0.1704	0.1837	0.1945	0.2037	0.2117	0.2188	0.2252	0.2310
10	0.1529	0.1706	0.184	0.1949	0.2041	0.2121	0.2193	0.2257	0.2315
20	0.1532	0.1710	0.1845	0.1955	0.2048	0.2128	0.2201	0.2264	0.2323
30	0.1535	0.1714	0.1849	0.1958	0.2051	0.2132	0.2203	0.2267	0.2324

Table 3.2: Implied volatilities with the bisection method with $\gamma = 1.5$, $\sigma_k(t) = 0.83$ and forward rate $F_k = 0.06$.

T_k/H	*0.02*	*0.03*	*0.04*	*0.05*	*0.06*	*0.07*	*0.08*	*0.09*	*0.1*
1	0.1527	0.1702	0.1835	0.1943	0.2034	0.2113	0.2184	0.2247	0.2305
5	0.1527	0.1704	0.1837	0.1946	0.2037	0.2117	0.2188	0.2252	0.2310
10	0.1529	0.1706	0.1840	0.1949	0.2041	0.2121	0.2193	0.2257	0.2315
20	0.1532	0.1710	0.1845	0.1955	0.2047	0.2128	0.2200	0.2264	0.2323
30	0.1535	0.1714	0.1849	0.1958	0.2051	0.2132	0.2203	0.2267	0.2324

$\alpha = 1.5,\ \lambda_k = 0.83,\ F_k = 0.06$

Figure 3.3: Implied volatility as given in Figure 1B, Andersen and Andreasen (2000). In their setting, we have $\alpha = \gamma$ and $\lambda_k = \sigma_k$.

However the relationship between time-to-maturity and caplet prices depend on the parameters (γ, σ_k, F_k). With the first set of parameters, as time-to-maturity increases from 1 to 10, implied volatility increases. But, as time-to-maturity increases from 10 to 30, implied volatility decreases. However, with the second set of parameters, as time-to-maturity increases, implied volatility only increases.

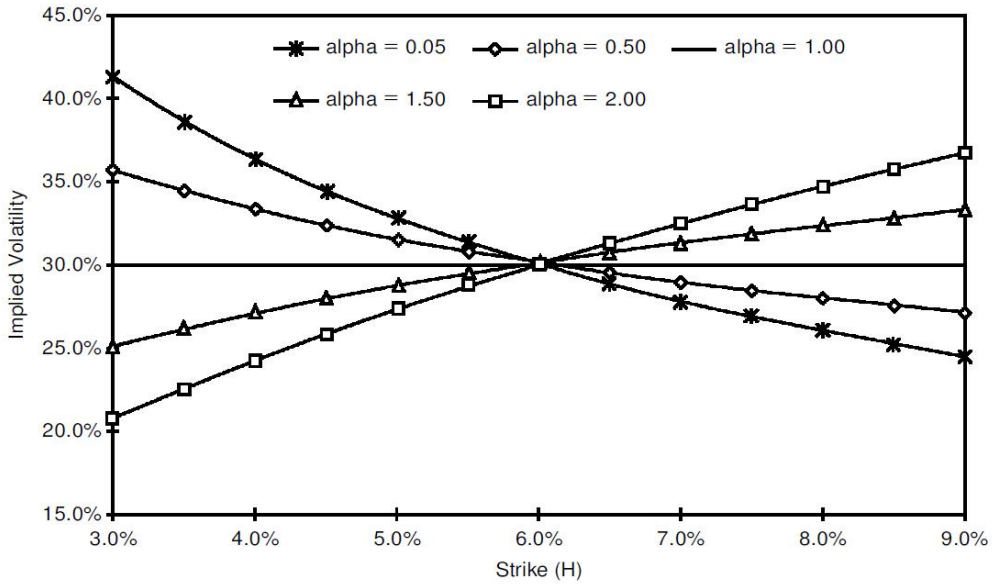


Figure 3.4: Implied volatility vs Strike rate for different CEV parameters (Andersen and Andreasen, 2000, Figure 2).

Remark 4. *It is noted that the strike rate of the second table starts from 0.02 as opposed to 0.01 in the first table. According to our implementation, we found that there will be an error when calculating the implied volatility when the strike rate is 0.01. Andersen and Andreasen (2000) also excluded the 0.01 strike rate. The option is well in-the-money which probably causes precision errors in the computation as the price is then much less sensitive to price moves and volatility.*

3.4 CEV implied volatilities as a function of γ

An additional check consisted of plotting the CEV implied volatilities as a function of γ to verify that the inversion procedure was working as expected. The argument being, using

$$dL(t) = \sigma L(t)^{\gamma} dW(t),$$

since $|L(t) < 1|$ in this project, as γ goes to 0, $L(t)^{\gamma}$ increases to 1 and therefore the implied volatility should decrease to keep the *diffusion coefficient* $\sigma L(t)^{\gamma}$ constant. This is what can be observed in Figure 3.5.

Remark 5. *Note that as γ gets close to 1, the parameters a, b and c from Theorem 1* explode *which makes the χ^2 CDF estimation cumbersome. For that reason, we use large thresholds of value 10^6.*

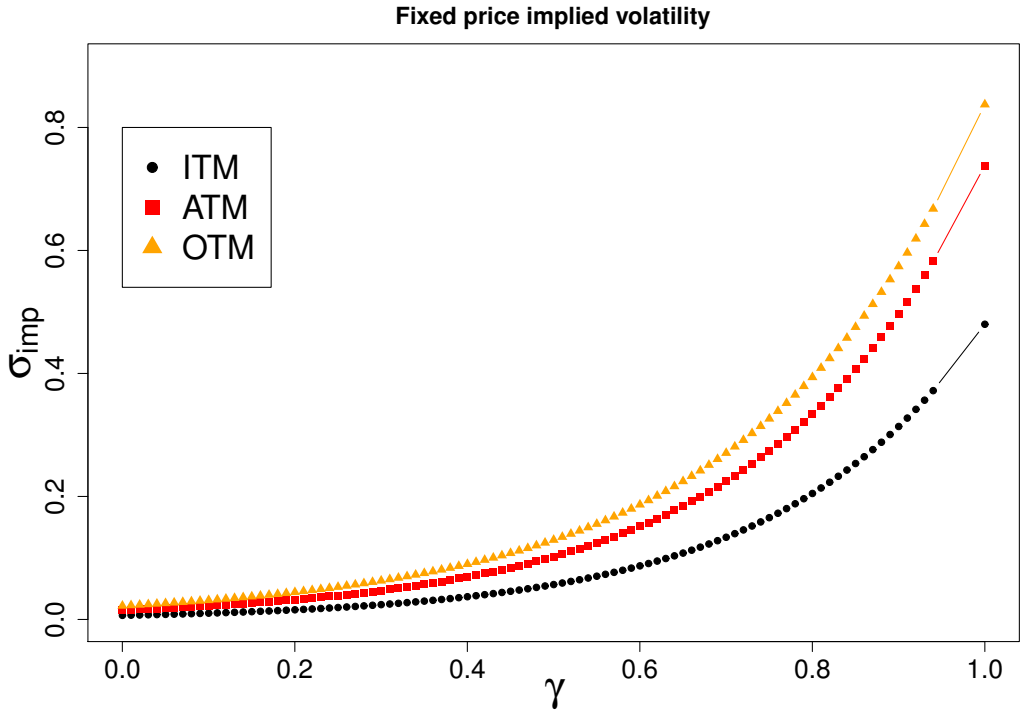


Figure 3.5: CEV implied volatility as a function of γ for a fixed cap price ≈ 0.01.

3.5 Cap pricing

A cap is simply a combination of caplets with different time to maturities. As a result, the cap price can be calculated as the sum of all the caplets. We calculated the cap prices based on the EURIBOR market data. The tenor of each caplet is 0.5 (half year) and the term of the cap is 10 years. When pricing the cap, special care needs to be taken when the first caplet is priced. At time 0, the interest rate is already known, therefore the payoff at time $t = 0$ (payment date of first caplet) is deterministic. The formula below shows thee price of the first *caplet*

$$C_0(0) = \delta_0 P(0, T_1) \left(L(0, T_1) - H\right)^+ .$$

The remaining caplets will be calculated as before using the formula mentioned in Section 2.3.

Chapter 4

Methodologies

In this section, we present the four methodologies explored in this project. We also present the theoretical foundations of the CEV hedging strategy.

4.1 Daily market skew calibration

To calibrate the model, firstly, a nave approach using daily cap prices was used to estimate the value of γ and σ across the three strikes available. This method is very crude because two unknown parameters are estimated using three data points. However, this method was carried out to have an insight of the range of the γ value and corresponding σ for the cross-sectional caps. This approach is indeed the industry standard when it comes to calibrating stochastic models, also known as fitting the market skew as we fit the model for different strikes on one specific day.

4.2 Parametric inference

Remark 6. *This approach suffers from the fundamental flaw that results are difficult to interpret. We include in the report for completeness with regards to the progress made during the week but do not rely on it to conclude anything. Indeed, it seems the CEV model does not have a global minimum in this context and leading to a non-identifiability issue.*

Consider the CEV model (Equation (2.1)) from a statistical perspective. Note that the likelihood function is difficult to obtain for a general $\gamma \in [0, 1]$ and quantifying its dependence structure is intractable. This said, the availability of market prices for different strikes allows for a parametric inference by simple sum of squares minimisation of the three different strikes as follows

$$\sum_{i=1}^{3} \sum_{k=1}^{M+1} \left(C_k\left(\gamma, \sigma; H_i, F_k\right) - \left(C_k\left(\gamma = 0, \sigma_{imp}^{B}; H_i, F_k\right)\right)^2,$$

where $\sigma_{imp}^{B,i}$ is Black's implied volatility as seen on the market for strike H_i.
This allows us to take the strikes' cross-sectional information into account whilst minimising the error on pricing.

4.3 Level-dependence minimisation

4.3.1 Stylised fact

We hope to capture level-dependence through this measure of absolute correlation as hinted by Filipovic et al. (2017).

Remark 7. *As given in (Filipovic et al., 2017, Section IV.G): "[..] there is a strong and positive relation between volatility and swap rate changes when swap rates are close to the zero lower bound."*

They first present a linear model in the form

$$\Delta\sigma = \beta_0 + \beta_1 \Delta L + \epsilon,$$

where σ are normal implied volatilities. They find a significant $\beta_1 \approx 0.2$. Subsequently, they found a more suitable formulation

$$\Delta\sigma = \beta_0 + \sum_{i=1}^{d} \beta_i I\{L \in [u_{i-1}, u_i)\}\Delta L + \epsilon,$$

where $I\{\cdot\}$ is the indicator function and $(u_i,\ i \in \{0, \dots, d\})$ is a collection of increasing thresholds for the interest rate. In their work, they take $d = 5$ and $u_i = i/100$ and compute, conditionally on each sub-categories of rate values, the coefficient of determination (R^2). They find sufficient proof that this conditional model fits the data well, especially in the low interest-rate domain. Indeed, they find

- Significant β_i for $i \in \{1, 2, 3\}$, insignificant for $i \in \{4, 5\}$.
- Positive and decreasing β_i for $i = 1, 2, 3$.

For this reason, we consider a framework which minimises the correlation between the two series of increments by navigating γ between 0 and 1.

Assumption 1. *We investigate the capacity of the CEV model to continuously interpolate (potentially) negative correlation of the log-normal model and (potentially) positive of the normal model.*

4.3.2 Level-dependence minimisation

We denote the level-dependence minimisation method as the method which finds $\gamma \in [0,1]$ such that changes in implied volatility and changes in the underlying forward rate are (empirically) uncorrelated.
This comes directly from the intuitive advantage of the CEV model over the normal and log-normal models as described in Section 2.4. More specifically, we look for

$$\widehat{\gamma}_{reg} := \arg \min_{\gamma \in [0,1]} \{|\mathrm{Cor}(\Delta\sigma_{imp,k}(\gamma), \Delta L_k)|\},$$

where $\Delta\sigma_{imp,k}(\gamma) := \sigma_{imp,k+1}(\gamma) - \sigma_{imp,k}(\gamma)$ with $\sigma_k(\gamma)$ the CEV-implied volatility with respect to γ at time T_k for $k \in \{1, \ldots, M\}$. Similarly, $\Delta L_k := L_{k+1} - L_k$. We take Pearson's empirical correlation estimator for the correlation. Note that, again as in Section 2.4 and Figure 2.1, a typical setting is to have positive correlation for the Normal model or negative correlation for the log-normal case (or both).

For that we proceed as follows: for a given γ, compute CEV implied volatilities. Then, compute the correlation between the weekly differences in the interest-rate and the implied volatility and then the optimiser finds new candidates which lower the absolute value of the correlation.

Remark 8. *Weekly differences are used instead of daily changes as external effects such as weekends, first day of week and liquidity issues need not be considered.*

As an example, the changes in implied volatilities for the log-normal case are negatively correlated with changes in the interest rate whereas we have negative correlation in the normal case.

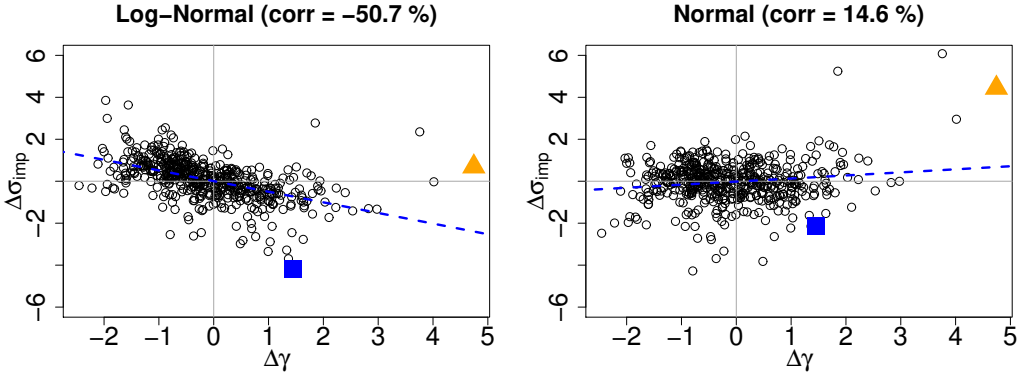


Figure 4.1: For the in-the-money strike $H_1 = 0.01$, different level-dependence showcased for log-normal and normal implied volatilities.

More precisely, with this approach we investigate the following assumption

Assumption 2. *The CEV model can continuously adjust the level-dependence between the (potentially positive) correlation of the normal model and the (potentially negative) correlation of the log-normal model.*

If this assumption is correct, we hope to achieve uncorrelation between the two by exploring the set $[0, 1]$ for the γ parameter.

4.4 Delta hedging

In this part, we use delta hedging to calibrate the parameter γ such that the profit and loss (P&L) of the hedging portfolio reaches its minimum.

4.4.1 Derivation for delta

Recall that for each caplet, starting at time T_k terminating at T_{k+1}, with strike H and forward rate L, there exists a unique risk-neutral measure $\mathbb{Q}^{k+1}$ such that, the price of caplet can be calculated as follows:

$$C_k(t) = \delta_k P(t, T_{k+1})\mathbb{E}[(L_k(T_k) - H)^+ | \mathcal{F}_t],$$

where $k = 1, 2, 3, \ldots$ and $\delta_k = T_{k+1} - T_k$. Similar to the stock market, the Fundamental Theorem for Asset Pricing (FTAP) and unique martingale measure $\mathbb{Q}^{k+1}$ implies that there exists a unique self-financing strategy (N_t, ψ_t) such that we can perfectly replicate European contingent claim $g(L) = (L - H)^+$, where (N_t^k, ψ_t) are the quantity invested in the underlying asset and bank account respectively. In the interest rate market, however, interest rates are not a tradeable asset. We can instead trade the forward contract on L and bonds instead.
Under $\mathbb{Q}^{k+1}$, $\widetilde{C}_k(t) = C_k(t)/P(t, T_{k+1})$ is a martingale such that the dynamic of L satisfies $dL_k(t) = g(L_k(t))dW_k(t)$ for some non-negative function g. Suppose that L follows log-normal model with $g(L_k(t)) = \sigma L_k(t)$, then we have a closed form solution for $C_k(t)$ which is stated in Equation 2.3.
From the martingale property, and Feyman-Kac Theorem, we have

$$d\widetilde{C}_k(t) = \frac{1}{P(t, T_{k+1})} \frac{\partial C}{\partial L_k(t)} dL_k(t) = \delta_k \frac{\partial F}{\partial L} dL_k(t), \tag{4.1}$$

where, $F(t, T_k) = C_k(t)/\delta_k P(t, T_{k+1})$. In the case where γ = 1 (log-normal model), the $\frac{\partial F}{\partial L}$ can be calculated from the for) which gives $\frac{\partial F}{\partial L} = \Phi(x_+)$. However, in the case where $0 < \gamma < 1$, we have to calculate this quantity by approximation: $\frac{\partial F}{\partial L} \approx (\Delta L)^{-1}(F(L + \Delta L) - F(L))$.

4.4.2 Delta hedging procedure

This hedging procedure was used by Fan et al. (2003, 2007); Driessen et al. (2003); Gupta and Subrahmanyam (2005) and we adapt it to our notation. Since caplets and floorlets are essentially options on the forward interest rate, they can be hedged with appropriate positions in the LIBOR forward market
(Gupta and Subrahmanyam, 2005). In practice, they are most commonly hedged using Eurodollar future contracts, due to the liquidity of the futures markets. A vanilla call option has a payoff function of $(S_T - K)^+$. It can be hedged by shorting $(S_t - K)$ since K is just a constant. In a similar way, a caplet which has a payoff function $(L_T - H)^+$ can be hedged by using a FRA which has a payoff function $(L_T - H)$.
Furthermore, a FRA can be decomposed into a *short* bond $P(t, T_k)$ and a *long* bond $P(t, T_{k+1})$. In addition, it is also necessary to hedge the *long* bond which acts as the discounting factor for the payoff. This discounting factor can be found in the pricing formula for a caplet. Therefore, in order to hedge a caplet, a combination of the *short* and *long* bond is used.
Referring to the (reference), delta hedging the caplet can be done by building a self-financing portfolio $(N_t^k, \psi_t^k - N_t^k)$ consisting of two bonds: $P(t, T_k), P(t, T_{k+1})$. $N_t^k = \frac{\partial F}{\partial L}$ should be held in $(P(t, T_k) - P(t, T_{k+1}))$ and

$$\psi_t = \frac{C_k(t) - N_t^k(P(t, T_k) - P(t, T_{k+1}))}{P(t, T_{k+1})},$$

should be held in $P(t, T_{k+1})$.
In our model, the delta is calculated numerically through finite difference method by

$$\frac{\partial F}{\partial L} \approx \frac{F(L + \Delta L) - F(L)}{\Delta L} \tag{4.2}$$

In theory, this delta-neutral hedge can perfectly replicate the porfolio but requires continuous rebalancing. In practice, however, only discrete rebalancing is possible. The accuracy of a delta hedge depends on how well the model's assumptions match the actual movements in interest rates.
In our data, we have caps with a 10 years maturity which consists of 20 caplets with 0.5 year maturities. Except for the first caplet which is deterministic at the beginning, we need to invest in 20 bonds $P(t, T_k),\ k = 1, ..., 20$ with weights $(\omega_1, \omega_2, ..., \omega_{20})$ calculated as:

$$\omega_1 = N_t^1,\ \omega_{20} = \psi_t^{19} - N_t^{19}$$

$$\omega_{k+1} = \psi_t^k + N_t^{k+1} - N_t^k \ with\ k = 1, ...18$$

For each cap, we build the hedging portfolio on the first day and hold it for 5-days. To assess the hedging portfolio performance, we compare the hedging portfolio

value after one week with the actual cap price after one week. This new cap, which is used to compare against the new hedging portfolio value, has a maturity time of 10 years still instead of 10 years less one week. This is because the cap market does not quote the evolution of the cap price over its term. The reason for using a cap with the same maturity is that if the volatility and underlying interest rate remains unchanged, the actual cap price will not change much after one week. Hence, to calculate the new cap price, instead of using interpolation on the cap prices, we just use the current term structure of volatilities (from the current prices of caps/floors) and put them into the pricing formula to calculate the current cap prices. This hedging procedure was used by Fan et al. (2003, 2007); Driessen et al. (2003); Gupta and Subrahmanyam (2005).

The hedging portfolio value will change due to changes in the bond prices. In our research, we only have data of bond prices for a fixed term. This means that we do not have prices of bonds with 1 year to maturity on day 1 and bonds with 1 year less 1 week to maturity after one week. Therefore, to overcome this problem, we used interpolation to calculate the price of bonds with maturity time being the original maturity time less 1 week. This is calculated as:

$$log\left\{P\left(t+\frac{1}{50},T_k-\frac{1}{50}\right)\right\} = log\left\{P\left(t+\frac{1}{50},T_{k-1}\right)\right\}\times\frac{1}{50} + log\left\{P\left(t+\frac{1}{50},T_k\right)\right\}\times\frac{49}{50}. \tag{4.3}$$

For a hedging portfolio on the first cap, we record the P&L after a week. So on day 2, we compute the new hedging portfolio. Then 1 week after day 2, a new portfolio value and a new cap price is calculated. The difference is then again recorded as P&L.

For comparison purposes, we also computed the unhedged position to assess the relative performance of hedging. This is done similarly to the hedging strategy. The only difference is that instead of investing in a delta hedged portfolio, the value of the cap is simply put into the bank account to earn simple interest. Therefore, after 1 week, the difference between the value in the bank account and the new cap price is recorded as P&L for unhedged portfolio.

In our research, we repeat this procedure over 2 years and record the P&L for the hedging portfolio $PL_{hedged}(i)$ and unhedged portfolio $PL_{unhedged}(i)$ on the i-th day. To assess the overall hedging performance, we use a hedging-equivalent R^2 as a performance measure. This is calculated as

$$R^2 = 1 - \frac{SS_{hedged}}{SS_{unhedged}} \tag{4.4}$$

where SS_{hedged} is calculated as

$$SS_{hedged} = \sum_i (PL_{hedged}(i))^2 \tag{4.5}$$

where i is the day where we calculate profit and loss.

$SS_{unhedged}$ is calculated as

$$SS_{unhedged} = \sum_i (PL_{unhedged}(i))^2. \tag{4.6}$$

If R^2 is closer to 1, this means that the hedging strategy performs well under a certain model.

4.4.3 Hedging algorithm

Our hedging procedure can be summarised as follows (the subscript B denotes the log-normal pricing formula and corresponding implied volatility): V_0^H V_0^U are, respectively, the current values of the hedging portfolio and the unhedged portfolio. V_1^H V_1^U are, respectively, the values of the hedging portfolio and unhedged portfolio after a week.

Data: 2013 - 2014 EURIBOR Caps market data
Fixing γ and strike H
for *day i* $= 1...$ **do**
  Calculate the corresponding implied volatility v_i;
  Initial amount for hedging $V_0^H = C_0$ = market price of cap on day i ;
  Calculate (N_t^k,ψ_t^k) for each caplet and deriving $W = (W_1, W_2, ..., W_{20})$;
  if *First caplet* > 0 **then**
    putting corresponding money Cp^1 into bank account;
  else
    No need to hedge
  end
  Build unhedged portfolio $V_0^U = C_0$, i.e., put into bank account to earn interest;
  After a week;
  Interpolate between the term structure P_{inp} and the term structure data on day i+5;
  Where $P_{inp} = (P(\frac{1}{50}, T_1 - \frac{1}{50}), ...P(\frac{1}{50}, T_{20} - \frac{1}{50}))^T$;
  Profit and Loss;
  $PL_{Hedge}(i) = V_1^H(P_{inp}) - C_1^B(P_{inp}, v_{i+5}^B) = WP_{inp} + Cp^1 e^{r\frac{1}{50}} - C_1^B(P_{inp}, v_{i+5}^B)$;
  $PL_{Unhedge}(i) = V_1^U - C_1^B(P_{inp}, v_{i+5}^B) = C_0 e^{r\frac{1}{50}} - C_1^B(P_{inp}, v_{i+5}^B)$;
  Calculate the R^2.
end

Chapter 5

Results on EURIBOR data

In this section, we focus on presenting the numerical results and performances of the different approaches on the EURIBOR dataset.

5.1 Daily calibration with cap prices

The estimation was done through minimising the squared errors between the actual cap prices and prices obtained using the optimum parameters across all the strikes. *L-BFGS-B* optimising function was used to minimise the squared residuals but there were computational difficulties due to a lack of data.

Remark 9. *Minimising the error sum of squares function is a difficult problem numerically as we only have three prices to work with, making the error very small in any case. To resolve this, we re-standardise this function with the Euclidean norm of the true prices and multiply by* 10^6.

The plots below show the different gamma values that were obtained through daily calibration.

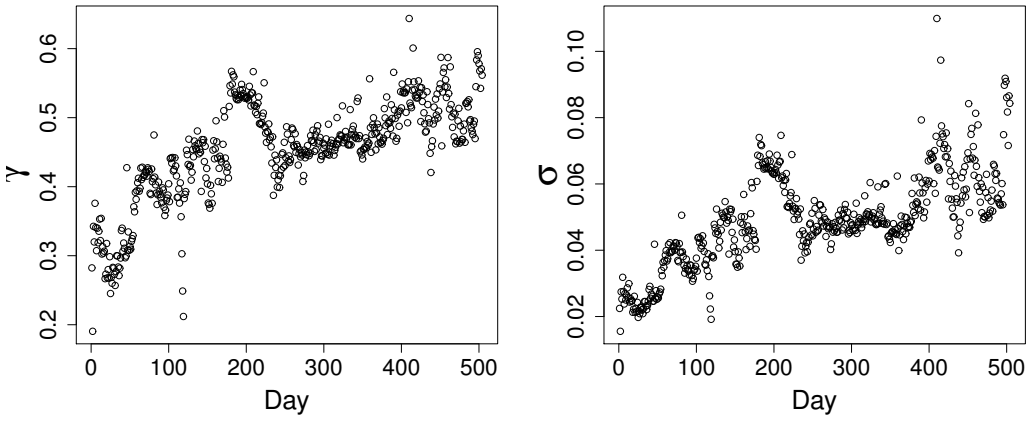


Figure 5.1: CEV parameters fitted each day on three cap prices with strikes $H_1 = 1\%$, $H_2 = 1.7\%$ and $H_3 = 2\%$.

	Minimum	**Lower quartile**	**Median**	**Upper quartile**	**Max**
γ	0.1905	0.4194	0.4632	0.5003	0.6437
σ	0.0155	0.0412	0.0486	0.0573	0.1099

Table 5.1: Distributional summaries of daily-fitted parameters (γ, σ).

The table above shows the quantiles for the γ and σ. The daily calibration suggest a γ between 0.2 and 0.64, a σ between 0.016 and 0.11. The sample mean for the γ is 0.45 and 0.05 for σ. This suggests that if there exist single γ and σ parameters which price the CEV model across different strikes, the parameters should be close to the values obtained above. As expected, we conclude that daily calibration to the market skew is useful as the market dynamics seem non-stationary as γ grows with time.

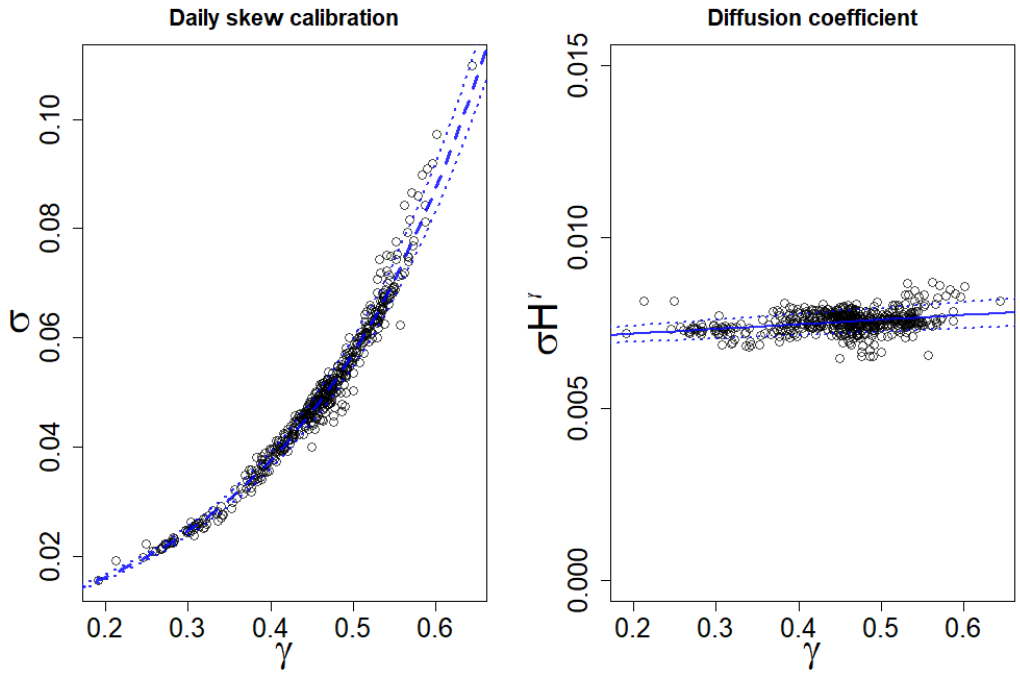


Figure 5.2: LHS: Scatter plot of daily calibrated (γ, σ) on market skew with fitted parametric exponential model $\sigma(\gamma) = \exp\{\theta_1\gamma + \theta_2\}$ with $(\theta_1, \theta_2) = (4.224\ (1.1e-2), -4.971\ (2.4e-2))$ (standard errors in parenthesis and $R^2 = 0.98$ on log-scale). RHS: CEV diffusion coefficient $\gamma \mapsto \gamma L^\gamma$ where we fix L to have the value of the ATM strike $H = 0.175$; the linear regression yields an intercept of $6.91e{-}3\ (8.4e{-}5)$ and a coefficient of $1.36e-3\ (1.83e-5)$, which is about 7 times as small.

In Figure 5.2, the scatter plot shows the exponential shape of σ as a function of γ which seems to be a reasonable fit. This is complemented with a computation of the fitted diffusion coefficient σL^γ as a function of γ. This is shown to be constant across time and values of γ. This is an essential result as it shows that the CEV model can be adapted to fit market data daily whilst keeping its stationarity with respect to volatility.

5.2 Calibration via parametric inference

The $BFGS$ optimisation routine was used to perform the minimisation which yielded **results** for the strike $H_2 = 0.0175$.

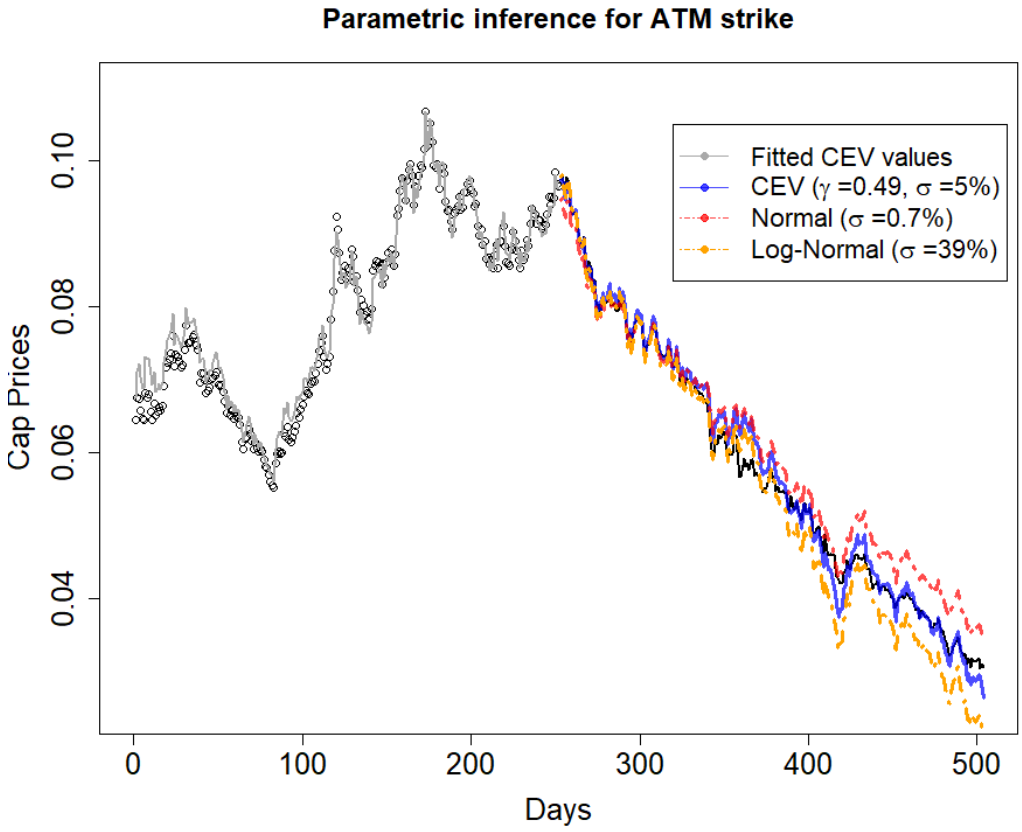


Figure 5.3: Parametric inference via error sum of squares for CEV, normal and log-normal models on one year of data and inference on the second year for the cap with strike $K_2 = 0.0175$.

We complement Figure 5.3 by calculating three performance measures: the mean average percentage error (MAPE), the coefficient of determination R^2 (percentage of explained variance) and the proportion of correctly predicted price movements.

	MAPE	R^2	PM
CEV ($\gamma = 0.49$)	2.76%	98.9%	86%
Normal	6.71%	97.8%	86%
log-normal	6.46%	97.4%	86%

Table 5.2: Performance metrics for out-of-sample data with strike $K_2 = 0.0175$ (one year worth of data).

We observe that both from Figure 5.3 and Table 5.2, the best model is the CEV model on *out-of-sample* data.
Note that although the three models have the same performance with respect to price move prediction, they do not always agree on that for a particular day.

5.3 Calibration via level-dependence minimisation

5.3.1 Results

Using the EURIBOR data and as featured in Figure 2.1 and Figure 5.4, we observe negative correlation between changes in implied volatilities and changes in the interest rate for the log-normal model. On the other hand, the normal is characterised by a small, yet significant, positive correlation between those aforementioned quantities.

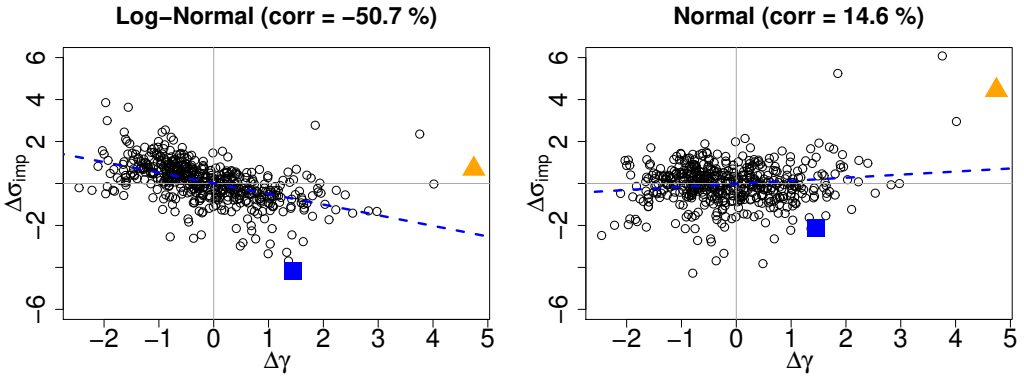


Figure 5.4: For in-the-money strike $H_1 = 0.01$, positive correlation is showcased between changes in implied volatilities and changes in the underlying for log-normal. Normal implied volatilities are positively correlation to change in the underlying.

The fitting procedure found that it would reach its optimal solution by finding the smallest value of γ one would allow. Indeed, Figure 5.6 with $\gamma = 0.002$ found a minimal correlation of -3.21%. Although this seems to be indicating that the correct model with respect the volatility level-dependence should very close to the

normal model. However, one notices the *discontinuity* this implies for the correlation as it should jump from slightly negative to positive as it reaches zero.

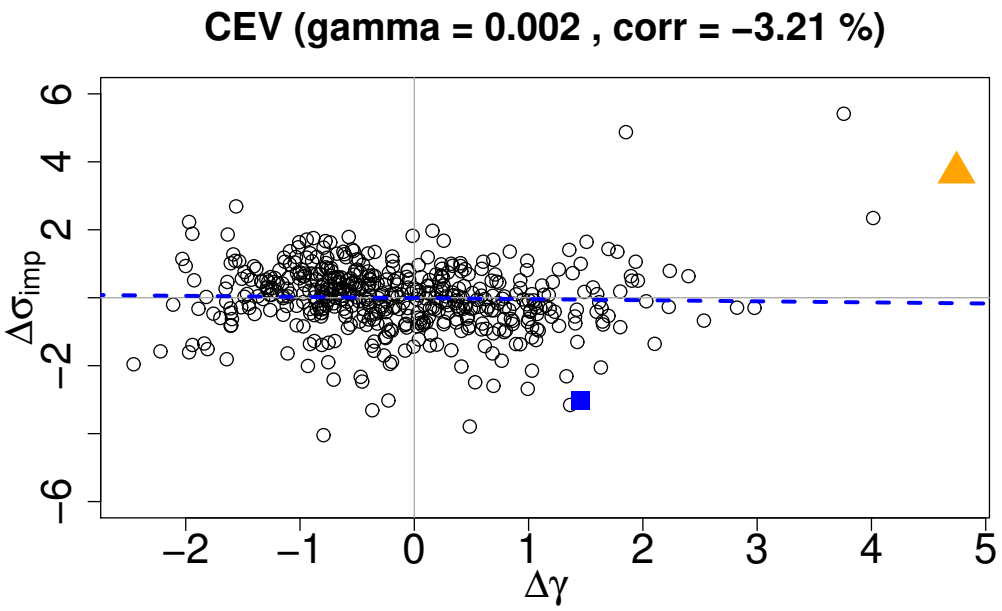


Figure 5.5: For in-the-money strike $H_1 = 0.01$, γ close to 0 seem to be adjusting for level-dependence.

5.3.2 Interpretation

It remains to explain the discontinuous behaviour of the CEV model as $\gamma \to 0$. Consider the CEV equation

$$dL(t) = \underbrace{\sigma L(t)^{\gamma}}_{\text{diffusion coefficient}} \, dW(t), \qquad t \geq 0.$$

The diffusion coefficient is $\sigma L(t)^{\gamma}$ for $\gamma \geq 0$. In particular, for $\gamma = 0$, we have

$$dL(t) = \sigma \times \mathbf{1} \times dW(t), \qquad t \geq 0.$$

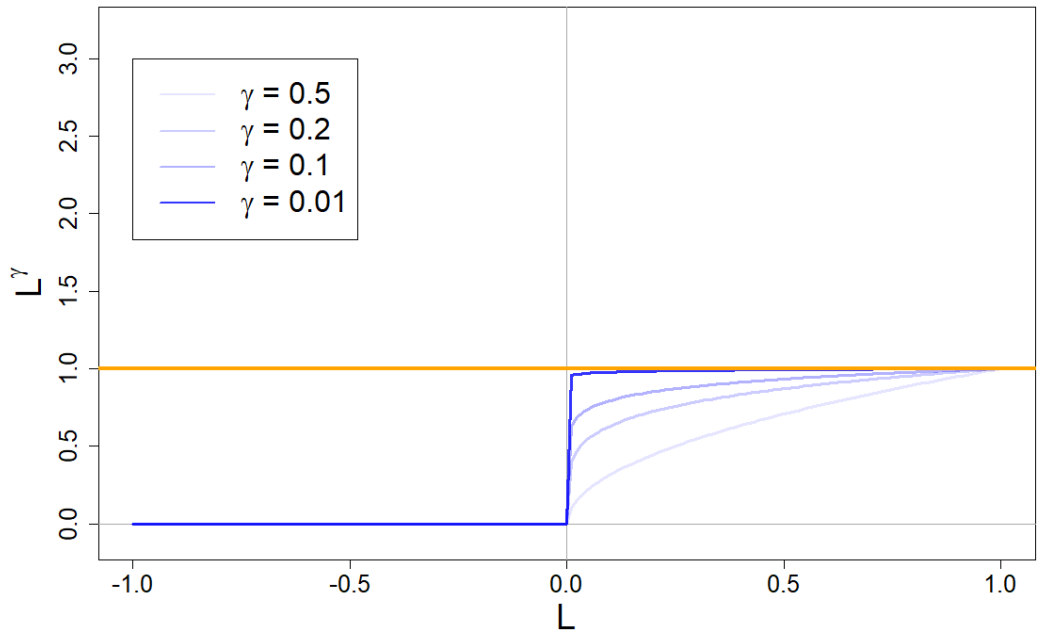


Figure 5.6: Discontinuity in the diffusion coefficient.

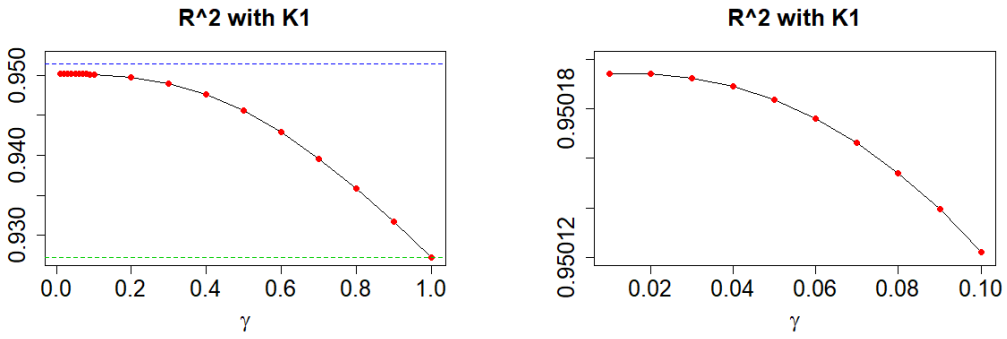
The discontinuity at $L = 0$ of $L \mapsto L^\gamma$ as γ approaches 0 involves that the underlying always play the role features in the log-normal model albeit in a less impacting way. This is the fundamental difference between any CEV model with $\gamma > 0$ and the normal model. This is particularly difficult to diagnose given implied volatility data since implied volatility does behave smoothly as $\gamma \to 0$ (Figure 3.5); which is explained by the scaling role σ has in the diffusion coefficient which is continuous with L.

5.4 Calibration via hedging

In this part, we are going to present the results from implementing the delta hedging procedure to calibrate the parameter γ such that the profit and loss (P&L) of the hedging portfolio reaches its minimum.
In order to check whether our hedging algorithm works, we use the hedging algorithm described in Section 4.4.2 but with a slight difference. The difference is: after a week, instead of using $C_1^B(P_{inp}, v_{i+5}^B)$, we are now using $C_1^{CEV}(P_{inp}, v_i)$.
The reason being, if the volatility remains unchanged for a week, then we would expect the delta hedging to almost perfectly catch the change in the underlying term structure in short term (a week in our case). Then the hedging R^2 should be around 99%. The reason for the R^2 not being 100% is that there exists discretisation error. If we delta hedge daily rather than weekly, then R^2 will increase closer to 100%.

LNM	H_1	H_2	H_3
Hedging test	99.94%	99.87%	99.78%
Real Hedging	92.72%	81.72%	59.60%
CEV	H_1	H_2	H_3
Hedging test	99.93%	99.85%	99.77%
Real Hedging	95.01%	89.69%	77.87%
NM	H_1	H_2	H_3
Hedging test	99.86%	99.77%	99.68%
Real Hedging	95.14%	90.33%	79.22%

Table 5.3: Hedging results for both the testing procedure and its optimal performance.



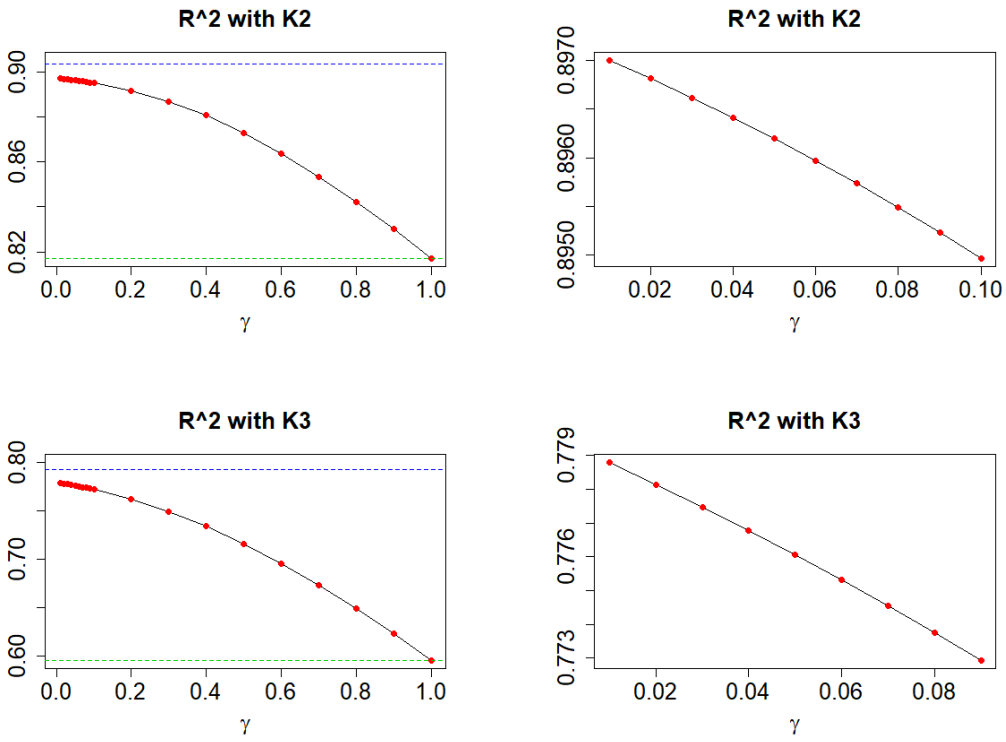


Figure 5.7: Hedging R^2 for CEV model with respect to γ

In table 5.3, we could compare those results in the column, it seems that the normal model always performs the best, and the log-normal model always perform the worst. The CEV performance is in between the two models.
As we can see from the above figures, the R^2 in the three cases tends to increase when $\gamma \in (0.1, 1)$ decreases. However, the R^2 keeps increasing when γ decreases in $(0.01, 0.1)$. We haven't investigated the limited distribution of the CEV model when $\gamma \in (0, 0.01)$. There seems to be a jump when setting a gamma close to 0. The results shown in the table for CEV are achieved by setting $\gamma = 0.01$.
Of the three strikes, only the normal model flattens the market skew the most which indicates there is almost no level-dependence between the underlying and normal-implied volatility. In other words, for the normal model, the change in the volatility mainly comes from randomness rather than coming from the change in the underlying. The hedging performance is therefore the best for the normal model

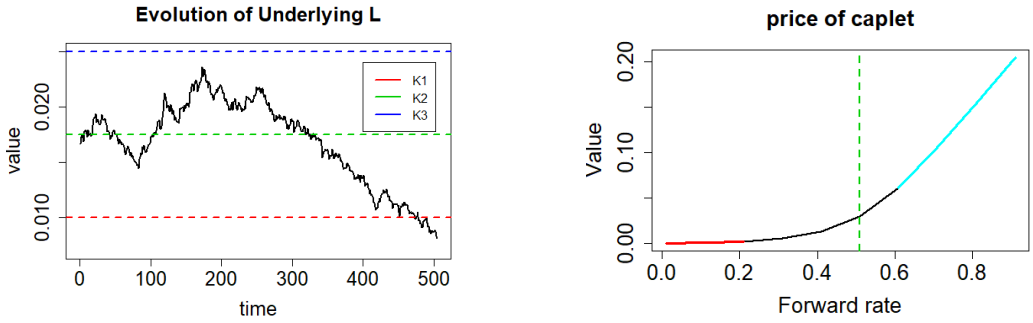


Figure 5.8: Price of caplet and Underlying movement

Sum of square ($\times 10^4$)	H_1	H_2	H_3
Delta Hedging (log-normal)	6.878100	9.516891	11.32189
Delta Hedging (CEV)	4.707577	5.362554	6.200317
Delta Hedging (normal)	4.584600	5.031673	5.824122
Cash Hedging	94.51882	52.06307	28.02997

Table 5.4: Statistics from hedging R^2

It is also worth noting that no matter which model we are using, the hedging performance for caps of strike H_3 is always the worst.
From figure 5.8, we could see that the underlying is actually decreasing from around the 200^{th} day to the end. Hence, it is clear that for most of the time, the caplet with Strike H_1 stays in the blue part of the right hand side plot. Although, price changes can be caused by both changes in the underlying and the volatility, in the short term the price change caused by underlying movement is much more significant than volatility movement.
On one hand, for the caplet with strike H_3, it will stay in the red part of the plot for most of the time. The delta in this case, is almost 0, hence price changes in the caplet mainly come from the change in the volatility. Hence in this case, the delta hedge is not going to hedge the change of price caused by volatility movements.
As a result, from table 5.4, the cash-hedge performs much better (sum of squares decrease) as the strike increases from H_1 to H_3. and the delta-hedge performs the worst.

Chapter 6

Conclusion

In this report, we explored the CEV model for caps as featured in Andersen and Andreasen (2000) through the lens of volatility level-dependence. The latter is the effect of an increase in the underlying on implied volatility motivated in (Piazzesi, 2010, Section 7.7) or (Filipovic et al., 2017, Section IV. G) and which is a mispecified feature of the normal, CEV and log-normal models. Given EURIBOR cap price data, we observed a significant negative correlation between changes in volatility and changes in the forward rate for the log-normal case - which has strong positive volatility level-dependence. On the other hand, the normal model is characterised by the absence of level-dependence and we observed a positive correlation between the two aforementioned quantities. The CEV does provide a flexible link between those two models and provides close to no correlation.

To this regard, we calibrate the CEV model according to three different approaches. The classical market skew fit using daily implied volatilities across strikes does not take volatility level-dependence into account and the resulting model does feature significant ill-specified behaviour as it overestimates the impact the underlying has.

Purpose-built calibration approach such that the *level-dependence minimisation* and *effective hedging* methodologies do provide alternatives and agree on their recommended values of the CEV parameter.

However, the fitted CEV parameters with respect to level-dependence are very close to zero with the EURIBOR dataset and we explain this by the discontinuity at $L = 0$ of $L \mapsto L^\gamma$ as $\gamma \to 0$. Although the marginal effect of the underlying on the diffusion coefficient becomes small, this discontinuity is very much present between the CEV and the normal model.

We also do the calibration by minimising the delta hedging error. For three different strikes, the log-normal model is as we expected from the previous part results that it overemphasises the relationship between the underlying and volatility, and hence is the worst performing model with respect to the hedging error. The normal model, on the other hand, shows little and even no-level dependence in the market

data and hence performs the best in the hedging results. Then, we confirm that the CEV model can adjust the level dependence between underlying and volatility by tuning the parameters γ, which is shown by the changing R^2 with respect to γ. Finally, discontinuity also appears in the hedging results such that it seems to have a jump between CEV model with $\gamma \in (0, 0.01)$ and normal model.

Chapter 7

Further Research

From the results, we know that our dataset shows little volatility level-dependence for the normal model. As next step, we could investigate even lower interest rate markets to check whether CEV model does feature to same asymptotic behaviour as $\gamma \to 0$ with respect to volatility level-dependence. Finally, we may try other models with different modelling methods on the level-dependence. More specifically, stochastic volatility models might be better suited to handle level-dependence as one would define extra degrees of freedom which could offset the re-balancing behaviour between volatility and the underlying as studied in this report.

Bibliography

Andersen, L., Andreasen, J., 2000. Volatility skews and extensions of the libor market model. Applied Mathematical Finance 7 (1), 1–32.

Brigo, D., Mercurio, F., 2007. Interest rate models-theory and practice: with smile, inflation and credit. Springer Science & Business Media.

Driessen, J., Klaassen, P., Melenberg, B., 2003. The performance of multi-factor term structure models for pricing and hedging caps and swaptions. Journal of Financial and Quantitative Analysis 38 (3), 635–672.

Fan, R., Gupta, A., Ritchken, P., 2003. Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. The Journal of Finance 58 (5), 2219–2248.

Fan, R., Gupta, A., Ritchken, P. H., 2007. On pricing and hedging in the swaption market: How many factors, really?

Filipovic, D., Larsson, M., Trolle, A. B., 2017. Linear-rational term structure models. The Journal of Finance 72 (2), 655–704.

Gupta, A., Subrahmanyam, M. G., 2005. Pricing and hedging interest rate options: Evidence from cap–floor markets. Journal of Banking & Finance 29 (3), 701–733.

Piazzesi, M., 2010. Affine term structure models. In: Handbook of financial econometrics: Tools and Techniques. Elsevier, pp. 691–766.

Financing the Green Revolution

TEAM McMaster

VEGAN PATHER, University of Cape Town
JOSÉ M. PEDRAZA-RAMÍREZ, London School of Economics
NOMVELO SIBISI, University of Cape Town
KILIAN WOHLLEBEN, University College London

Supervisor:
MATHEUS GRASSELLI, McMaster University

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

Climate risk poses a significant long-term risk for the financial industry. Many sectors are either directly or indirectly exposed to the effects of rising global temperatures, ranging from high losses for the insurers of natural disasters (growing by 30% annually) to the potential plummet of carbon-intensive asset values. Along with these risks, there is a substantial need for funds in order to migrate to a low-carbon environment. The most recent value for global climate finance is $455 billion, whereas required annual investment ranges from $450 billion to $2,000 billion by 2035 in order to reach the desired 2°C temperature increase above pre-industrial emissions (the goal of the 2015 Paris Agreement).

There are thus 2 key aspects to this problem: The actions needed to manage climate risks, and the funding needed to finance such actions. In order to assess these fundamental aspects, we make use of a stock-flow consistent integrated assessment model (IAM), which is a model that incorporates various elements from different spheres into a single model. This paper will focus on analysing the effect of different financing strategies on (i) debt accumulation and (ii) global warming.

Previous IAMs have assumed private debt as an aggregate of bank loans, but in this report we examine the effect of another financing tool in transitioning to a low-carbon economy, namely green bonds, which are bonds used for green purposes. Green bonds, which have a rapidly growing issuance and increasing demand, add another form of debt to the model at a different return to bank loans. Given that green bond funds are to be used in reducing greenhouse gas (GHG) emissions, they also play a role in abatement costs (costs of reducing emissions), and further affect total nominal profit in the productive sector, which impacts real investment. There is, thus, a significant interplay between components of the model due to the addition of these green bonds. It is therefore worthwhile to assess the effect of certain parameters on debt and average temperatures. These parameters are: the convexity of the abatement cost function θ, the rate of decline of backstop technology prices $\delta_{p_{BS}}$, and the spread between the rate of returns on green bonds and bank loans. Further assessments are done on three key policy mechanisms: Carbon Tax, which is advocated by the Stern-Stigltiz Commission; Public Subsidies, which are recommended to be implemented with Carbon Taxes by the High Level Commission on Carbon Pricing; and a Cap-and-Trade System. The methodology is consistent with that done by Bovari et al. (2018b).

We conduct sensitivity analyses on these three parameters, and a scenario analysis with these policy tools to assess the effect on remaining below a threshold of 2°C above pre-industrial emissions and a debt-to-output ratio of 2.7 (the threshold when aggregate equity in the private sector approaches zero, with correspondingly

high probability of systemic default) (Feenstra et al., 2015).

The report is organised as follows: Section 2 provides an overview of the important risks of climate change along with information on current decarbonisation strategies. Section 3 will review the current landscape of green finance with particular attention to green bonds. Section 4 provides a review on hedging strategies found in recent literature and in practice. Section 5 is an outline of the original Integrated Assessment Model used in Bovari et al. (2018b) which forms the base of our methodology. Section 6 then presents the modifications to this model along with results from scenario and sensitivity analyses. Main conclusions and recommendations for future research are provided in the last section.

2 Risks and Mitigation

2.1 Risks of Climate Change

Financial institutions are facing significant risks due to climate change. These risks can be summarised into three key channels: risks that arise from the physical impact of climate change; risks that arise from devaluation of assets; and risks related to the transition to a low-carbon economy.

2.1.1 Physical

The most apparent physical risks that are associated with climate change are caused by extreme weather events such as droughts, floods and earthquakes. Almost every sector is susceptible to short- and long-term physical implications, likely resulting in huge financial losses. These implications include possible damage to inventory, equipment, agriculture or infrastructure; disrupting the production process.

Climate-related disasters occurring in 2017 amounted to losses valued at $320 billion (GCEC, 2018). It is, therefore, in the best interest of the economy as a whole to impede the rising global temperatures, which in turn lowers the probability of such extreme weather events.

A further consequence relates to insurance and coverage. A rise in the frequency of naturally-
occurring disasters related to climate change leads to an increase in premiums and deductibles. Insurers and reinsurers may also decide to adjust the terms of coverage based on which areas are more subject to severe weather (Gardiner and Associates, 2011).

Not all physical risks are directly related to the market. Climate change can pose a risk not only to human health, but to the ecosystems. It is difficult to quantify the potential loss associated with non-market-related risks, however each one indirectly affects the market, and thus the economy. The Intergovernmental Panel on Climate Change (IPCC) reports that a 4°C increase in global temperatures will lead to damages accounting for 1-5% of GDP.

In our model, we define the total damages function, $\mathbf{D}$, as the sum of the distribution of damages over output, $\mathbf{D}^Y$, and the distribution of damages over capital, $\mathbf{D}^K$:

$$\mathbf{D} = \mathbf{D}^Y + \mathbf{D}^K \tag{1}$$

$$= 1 - \frac{1}{1 + \pi_1 T + \pi_2 T^2 + \pi_3 T^{\zeta_3}}, \qquad \pi_1, \pi_2, \pi_3 \in \mathbb{R} \tag{2}$$

$$\mathbf{D}^Y = \mathbf{D} - \mathbf{D}^K \tag{3}$$

$$\mathbf{D}^K = f_K \mathbf{D} \tag{4}$$

where f_K represents the share of total damages corresponding to the capital.

2.1.2 Stranded Assets

In order to keep climate change in accordance with the Paris Agreement of a 2°C target limit above pre-industrial temperature levels, a large portion of current fossil fuel reserves must remain unused. New governmental regulations around GHG emissions, or a decrease in the demand for fossil fuel reserves will cause these assets to lose value. It is for this reason that companies are being encouraged to disclose any risks to which they may be exposed due to climate change.

Stranded assets pose possible systemic implications: an abrupt adoption of greener policies may have a destabilising effect on the economy. The Global Commission on the Economy and Climate estimated a potential loss of $12 trillion in stranded fossil fuel assets by 2035, compared to the $250 billion loss in stranded mortgage assets in 2008. Minimising this risk can be achieved by a slow and smooth transition to a low-carbon economy.

2.1.3 Transitional

There are several repercussions that industries and investors face as a result of the subsequent transition to a low-carbon economy. These include regulatory risks, and mitigation and
adaptation costs.

Climate change is a major problem that should be addressed immediately and

earnestly.
Implementing several abatement strategies concurrently is likely to have the most promising outcome. The challenge then becomes how to transition to a low-carbon economy without jeopardising the global economy. As the risks of climate change become more noticeable, industries will have to start allocating a portion of their capital to backstop technologies and renewable energy. In the DICE model, Nordhaus showed that abatement costs, A, vary with the price of backstop technology, as well as the emission reduction rate, as follows

$$A = \frac{\sigma p_{BS} n^{\theta}}{\theta}, \quad \theta > 1 \tag{5}$$

where p_{BS} is the price of backstop technology; n is the emission reduction rate; and θ controls the convexity of the costs.

Currently, there is a lack of stringent global policies regarding carbon emissions. In fact, according to (Andersson et al., 2016), investors with low-carbon portfolios are projected to benefit financially more than those with standard portfolios as soon as the market begins to regulate CO_2 emissions. Thus, there are also potential losses associated with not anticipating the transition, and acting accordingly.

2.2 Decarbonisation Strategies

As climate change continues to present an increasing problem for the global economy, various researchers and economists are probing into potential methods of decarbonisation. These include, but are not limited to, energy conservation, carbon pricing, carbon mitigation, and transitioning to other, environmentally-friendly sources of energy.

Pollution that occurs as a result of transportation could be reduced by switching to electric vehicles or other modes of transportation with minimal negative environmental effects, as well as increasing energy efficiency to reduce usage. The transition to renewable energy sources such as hydro-, solar and wind power; while safer for the environment; can be costly and prove difficult to execute on a global scale. Unfortunately, current levels of research and development in renewable energy are relatively low, despite the International Renewable Energy Agency (IRENA)'s claim that renewable energy coupled with energy-efficiency improvement strategies could account for up to 90% of emission reductions by 2050.

Conserving energy is easily implementable; it simply involves the preservation of current resources, as well as investing in energy-efficient technology. The downside of this strategy is the cost of conversion to the aforementioned technology, which requires capital (at least $50 billion of new capital by 2020 to allow changes beyond the energy sector (GCEC, 2018)). Furthermore, it is not easily enforceable and does not fully place the burden on the sectors that cause the most pollution, but rather on the economy as a whole. On the other hand, introducing a penalty on carbon dioxide emission (carbon pricing) aids in decreasing global emissions by decentivising the use of CO_2-emitting technologies and making cleaner technologies a more appealing investment, particularly to said sectors. Carbon pricing typically falls into two categories:

1. Cap-and-trade system.
2. Carbon tax.

The former involves enforcing a cap on the volume of CO_2 emissions. This is done by creating CO_2 permits and distributing them to the various industries. A permit is required for every ton of CO_2 emitted; enabling some industries to sell their excess permits, given that they are able to reduce their level of emissions. Moreover, industries that exceed their permissible CO_2 emissions have to buy the excess permits from other industries, hence the term "cap-and-trade". This system presents a problem in the form of how to allocate the permits. The EU Emissions Trading Scheme has faced scrutiny regarding the effectiveness of its cap-and-trade scheme, with critics claiming that the permit allocation scheme was too lenient (Sterner and

Köhlin, 2015). In the first two years of the initial implementation (2005-2007), emissions reduced by 2.4-4.7% (Muûls et al., 2016). However, the European price of carbon has dropped significantly from €30 in 2008 to under €5 in 2016. There is a trade-off between the price of the permits and the severity of the cap: if the cap is too low, then the price of permits will skyrocket; whereas if the price of permits is kept low, then the cap will have to be less strict. As a consequence of the latter case, carbon emissions would not be significantly reduced.
Proposing a tax on carbon is arguably one of the most efficient and easy ways of regulating CO_2 emissions. We define the payment of carbon tax, pT_f, as

$$pT_f = pp_cE_{ind} \tag{6}$$

where p and p_c are the price of commodities and carbon, respectively; and E_{ind} is the industrial emissions level (defined in Equation (23)).

Carbon pricing is currently, or will be, in effect in over 70 countries and jurisdictions globally. Carbon pricing initiatives account for approximately 20% of global GHG emissions. While this figure is an improvement on the 15% coverage in 2017, half of these jurisdictions have policies that charge less than \$10/tCO_2-e (GCEC, 2018). According to a report released by the High-Level Commission of Carbon Prices, meeting the 2°C target demands that major economies (i.e. G20 countries) impose a carbon tax of at least \$40-80 by 2020; and \$50-100 by 2030. These figures are based on the Effective Carbon Rate, which is the culmination of taxes on fossil fuels, carbon taxes, and estimated prices of pollution permits.

The **carbon pricing gap** measures the difference between the benchmark[a] prices and the actual carbon prices per percentile of emissions. The following table compares the 2015 carbon pricing gap for a select number of countries.
The carbon pricing gap measured against the EUR30t/CO_2 benchmark was 76.5% in 2018, compared to 79.5% and 83% in 2015 and 2012 respectively. Hence, the carbon gap has decreased by 3% since 2015 and 6.5% since 2012. Although the price of carbon is rising over time, assuming the carbon pricing gap continues to fall at its current rate of $\pm 1\%$, carbon costs would only equate to the benchmark prices in 2095. (OECD, 2018).

Unlike the cap-and-trade system, a carbon tax will encourage industries to continue to look for ways to reduce their CO_2 emissions, rather than just reduce them enough to meet the cap. This also ensures that the sectors responsible for the most pollution get charged the most. The revenue from the tax could then be invested in research into more effective decarbonisation strategies, or green technology. The

[a] Two benchmark prices are applied; the EUR30t/CO_2, a low-end estimate of current carbon prices; and the EUR60t/CO_2, a low-end estimate of carbon prices in 2030 and a midpoint estimate of carbon prices in 2020.

Country	Carbon Pricing Gap (%)
Switzerland	27
Norway	34
United Kingdom	42
Germany	53
Sweden	63
Canada	65
Mexico	68
United States of America	75
South Africa	89
China	90
Russia	100

Table 1: Table depicting the carbon pricing gap in 2015 (%)

price of backstop technology decreases according to

$$\frac{\dot{p_{BS}}}{p_{BS}} = \delta_{p_{BS}} < 0 \tag{7}$$

The 2018 costs of various backstop technologies, as well as their respective changes in price, are tabulated below (IRENA, 2019).

	Global Weighted-Average Cost of Electricity (USD/kWh)	Decrease in the Cost of Electricity 2017-2018 (%)
Bioenergy	0.062	14
Geothermal	0.072	1
Hydro	0.047	11
Solar PV	0.085	13
Offshore wind	0.127	1
Onshore wind	0.056	13

Table 2: Table depicting the 2018 prices of renewable energy and the percentage change in price

It should be noted that if the carbon tax is set too low, then the respective industries will simply pay the amount owed and not waste any resources on transitioning to eco-friendly technology. Many decarbonisation strategies are most effective when used in conjunction with one another; for example, a carbon tax may lead to an increased investment in green technology. However, this does not affect the current emission levels. In other words, although the economy is moving towards a greener means of production, we still need to address the current pollution problem, by way of carbon mitigation.

Cumulative carbon emissions totalled 545GtC* between 1870 and 2014, with the current CO_2 concentration in the atmosphere averaging approximately 400ppm** (McGee, 2019). The graphs below illustrate global CO_2 emissions from 2000-2018; and the respective percentage of total CO_2 emissions induced by several countries (Wang, 2019), respectively.

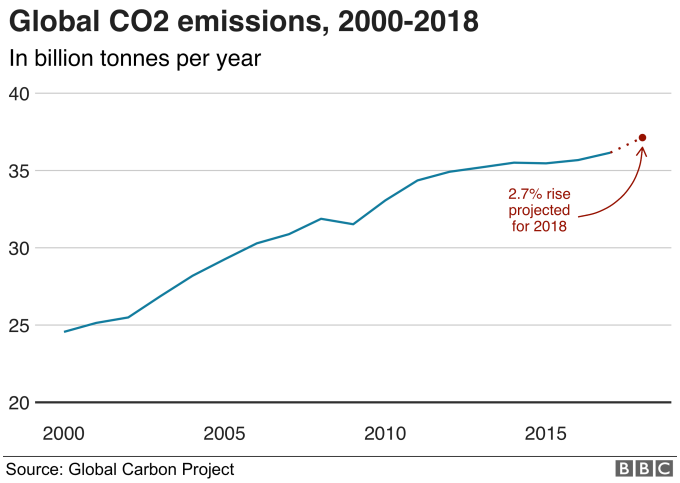


Figure 1: Global CO_2 emissions from 2000-2018 in billion tons per year

The graph in Figure 2 pinpoints China and the USA as the primary contributors to worldwide CO_2 emissions. Given the ongoing increase in emissions, it is clear to see that carbon mitigation technology is essential.

*Gigatons of carbon
**Parts per million

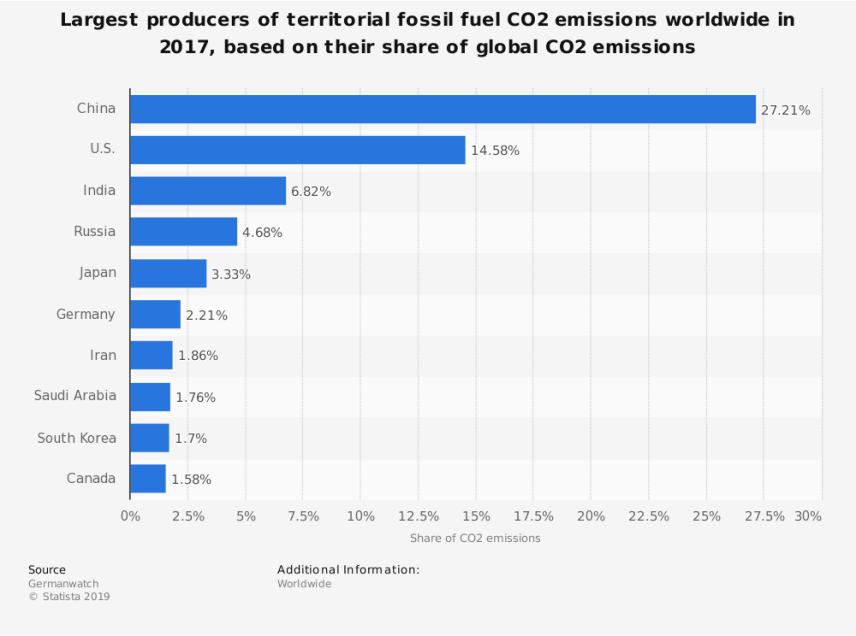


Figure 2: Worldwide CO_2 emissions in 2017

Carbon capture and storage (CCS) is a geoengineering technology that allows for negative emissions. CCS technology can capture up to 90% of CO_2 emissions in a plant (Leung et al., 2014). The process involves three steps:

1. Capture CO_2.
2. Transport CO_2.
3. Store CO_2 emissions/reuse carbon industrially.

Naturally, CCS technology has its limitations. Once captured and transported, the CO_2 is typically stored several hundred metres below the ocean. While this method of storage could keep the CO_2 contained for extended periods of time, there is a risk of ocean acidification, which may have adverse effects on the marine ecosystem (Leung et al., 2014). There are several global projects that make use of a CCS system presently. However, it is estimated that over 3000 CCS projects need to be initiated by 2050 in order to meet the 2°C limit. This, of course, places a further financial burden on the economy. It is estimated that the current cost of a CCS plant is \$60/tCO_2; and that by 2025, the cost will reduce to \$40/tCO_2(Institute, 2019). These figures form part of the mitigation costs.

Research into other breakthrough technologies is underway, with nuclear fusion energy a strong contender, provided it can be made cheap and scalable. Fusion,

the energy that powers the sun, is produced when two light atoms (hydrogen) fuse into one another under intense pressure and temperature. Recreating the conditions necessary for the atomic nuclei to coincide and fuse is not an easy task; the main challenges arise from the cost and complexity required. Presently, nuclear plants can only produce fission energy. Radioactive waste is a byproduct of fission energy, whereas generating fusion energy results in a non-toxic inert gas (helium). Consequently, radioactive waste is not a risk associated with nuclear fusion. Once technology has evolved enough to allow us to harness this type of clean energy on a large scale, nuclear fusion could prove to be the long-awaited, long-term solution for mitigating climate change.

The debate on which means of decarbonisation is the most cost- and energy-efficient is still up for discussion. However, there is a serious lack of funding required to finance the changeover to a low-carbon economy. It is estimated that between $9 and $40 trillion will need to be invested in green infrastructure by 2040; close to half of the entire current stock of capital, which was estimated to be $84.4 trillion in 2018. (GCEC, 2018). Considering that global green investment in 2018 totalled $332.1 billion, it is clear to see that there is a significant funding gap. Furthermore, the United Nations Environment report states that adaptation costs could leap to $280-500 per year by 2050. Adaptation costs are set to escalate rapidly unless fervent efforts are made to regulate climate change in the near future. Thus, in terms of our model, the cost of abatement will increase as a result.

3 Green Finance

Green Finance is a broad term which encapsulates a large system of cash-flows related to environmental sustainability goals. It refers to the flow of funds into: public and private green investments, the financing of policies, and the financing of financial instruments with all of these aimed at the mitigation of environmental damages and to the promotion of environment friendly initiatives (Lindenberg, 2014).

Table 3 below provides a breakdown of funding towards climate change mitigation in recent years as per the Climate Policy Initiatives Organisation 2018 Finance Update (Oliver et al., 2018).

3.1 Public Funding

Public funding can have a huge impact on climate change and the world economy. It is affecting climate change in two ways: on the one hand by supporting renewable energy, and on the other hand by the still-existing subsidies on fossil fuels. In the year 2016, subsidies for renewable power generation amounted to $140 billion,

Actors	2015	2016
Private	267	230
Commercial FI	54	42
Corporates	46	28
Households	39	44
Institutional Investors	3	2
Private Equity	2	1
Project Developers	124	113
Public	205	224
Governments and their agencies	17	19
Climate Funds	2	3
Public FI (Bilateral)	17	14
Public FI (Multilateral)	44	48
Public FI (National)	124	140
Total	472	455

Table 3: Breakdown of global and private investment (USD Billions)

with most of them issued by developed countries, while most fossil fuel subsidies were in developing countries (International Energy Agency, 2017). According to the IEA, the net fossil fuel subsidies sum up to $493 billion. In fact, the subsidies and damages of fossil fuels accumulated to 6.3% of the world GDP in 2015 according to the IMF. The same study predicts that by pricing fossil fuels optimally; by cutting all subsidies; and imposing a tax on CO_2 that covers the losses from the damages of fossil fuels, the global CO_2 emissions could decrease by 28% (Coady et al., 2019).

3.2 Green Bonds

In 2006, the Swedish pension fund was looking for an easy way to invest in sustainable projects. Together with the World Bank, they developed the green bond, a financial instrument that works just like a normal bond except that it has dedicated funding (Bank, 2019). So what is a green bond? Or in other words: what classifies a bond to wear the green label? Since there is no regulation in place, there is no clear definition in most major markets. Several approaches have been brought forward where the (probably) most important ones are the Green Bond Principles; the CICERO shades of green; and the Climate Bonds Initiative criteria. Even though no certification is required, 89% of the bonds labelled as green do have some sort of certification (Filkova et al., 2019). Today, the total value of outstanding green bonds is $644 billion with an expected issuance of $200 billion in 2019. Whilst the market issuance has been growing rapidly in the past from only $2 billion in 2012

to $168.5 billion in 2018 (Filkova et al., 2019), the demand is still high as indicated by the bid-ask-spread of green bonds being about 10bps tighter than for comparable non-green bonds (Harrison, 2019), and the yield being 7bps lower on average Pronina (2019).

Year	2012	2013	2014	2015	2016	2017	2018	2019 (proj)
Green Bonds Issued($bn)	2.0	11.0	36.6	41.8	81.0	155.5	168.5	(250)

Table 4: Number of green bonds issued each year (https://www.climatebonds.net)

Recent studies looked for a mathematical explanation for this phenomenon and found a simple justification similar to the CAPM: Let us assume that there are two types of investors in the market. The first one is the classical mean variance maximiser, i.e. following the utility function

$$w_1^t r - \frac{\gamma}{2} w_1^t \Sigma w_1$$

with w_1 the portfolio weights, r the returns of the assets, γ the risk aversion parameter, and Σ the covariance matrix. The second one is also looking at the environmental impact, e, of the assets where a positive value of e corresponds to a positive environmental impact. Therefore the utility function is

$$w_2^t r + w_2^t e - \frac{\gamma}{2} w_2^t \Sigma w_2.$$

Simple optimisation gives us:

$$w_1^* = \frac{1}{\gamma} \Sigma^{-1} r$$

and

$$w_2^* = \frac{1}{\gamma} \Sigma^{-1} (r + e).$$

Rewriting these formulas gives us

$$r = \gamma \Sigma w_1^*$$

and

$$r = \gamma \Sigma w_2^* - e$$

respectively. Now, the investors have a capital of a_1 and a_2 which they invest according to their preferences. The total market return is then

$$\begin{aligned} r &= \frac{a_1}{a_1 + a_2} \gamma \Sigma w_1^* + \frac{a_2}{a_1 + a_2} (\gamma \Sigma w_2^* - e) \\ &= \frac{\gamma \Sigma (a_1 w_1^* + a_2 w_2^*)}{a_1 + a_2} - \frac{a_2 e}{a_1 + a_2}. \end{aligned}$$

This enables us to conclude that assets with a positive environmental impact have a lower return (Baker et al., 2018). This gives an incentive for investors to finance their green projects with green bonds and also shows why the demand is not met (Zerbib, 2018). We will use this in the implementation to justify a spread between the rate of green and normal bonds and that we need not consider lack of demand. It is worth mentioning that, besides green bonds, green loans for smaller companies and individuals also exist. Similar to green bonds, they follow the green loan principles and are usually monetarily beneficial for the recipient[†].

[†]https://www.bbva.com/en/the-green-loan-principles-updated-to-offer-a-wider-range-of-sustainable-products

4 Hedging Climate Change

Due to high uncertainty around climate change, there are risk factors which can affect financial parties. Financial players face exposure to physical impacts, for example, insurers correctly need to model, price and reserve funds for natural disasters and extreme weather events. If not, insurers suffer severe losses such as the full economic loss of $340 billion that arose from natural disasters in 2017 (the highest recorded figure) (IAIS, 2018). Another example is businesses directly exposed to carbon intensive assets. Another risk comes from exposure to carbon intensive firms and industries. Carbon Delta estimates that the coal industry faces $221 billion in losses by 2050. Investors, such as pension funds and asset managers, also face possible losses from stranded assets, which is projected to be $12 trillion in 2035 for fossil fuel assets. It is therefore important for such parties to protect their financial positions. In this section, we introduce several strategies suggested in recent literature and examine current schemes of well-performing investors.

4.1 Mitigation and Adaptation Investment

Two types of investments have been proposed in recent literature: mitigation investment actions and adaptation investment actions. Mitigation actions involve investments which are linked to the reduction of GHG emissions and an increase in exposure to the low-carbon economy (IIGCC, 2015). Therefore mitigation can include:

- Direct investment into environmentally-sustainable infrastructure and projects, which can further reduce the price of backstop technologies, that are found in the model.
- Divestment from stranded assets.
- Divestment or minimisation of exposure to sectors and companies with high carbon footprints.
- Portfolio optimisation and regression techniques for carbon efficiency.

Such actions which spark more funding into renewable energy are also likely to reduce the price of backstop technologies in the future. Adaptation actions involve the reduction of an entity's vulnerability to the physical effects of climate change (IIGCC, 2015). In implementing mitigation strategies, investors implicitly make use of adaptation finance. Therefore adaption investment can include:

- Investments into financing tools such as green bonds, green funds, equities in green companies and green ETFs. Green Exchange Traded Funds (ETF) are investment funds which can be traded on an exchange and are based

on companies involved in the support and development of environmentally-friendly technology. Many green ETFs focus on sources of renewable energy such as wind energy, solar energy and nuclear energy. Therefore these funds, much like standard ETFs, provide a method for investors to diversify their portfolios across green sectors. This also allows investors to gain exposure to the green economy, which is a common approach found in practice.

- Partnerships with development banks, which are banks that can provide financing to developing groups who often have higher exposures to climate risks and struggle to adapt.

Mitigation actions have been more popular with only 5% directed to adaptation finance in 2016 (Oliver et al., 2018). This is because it is often easier to measure the mitigation of emissions than to quantify the long-term impact of climate change. Indeed in Engle et al. (2019), it is discussed that choosing the correct index for a news source to hedge climate news is non-trivial and subjective. However, adaptation investment is vital as there are many financial sectors with more exposure to geophysical risk events, such as agriculture and the insurance industry. These risk events are due to extreme weather events such as floods, hurricanes, and heat waves. For insurers, current losses are estimated to be approximately 0.5% of global GDP which amounts to $400 billion; losses are expected to increase by 6% annually. The United Nations Framework Convention on Climate Change also estimates that these losses can grow to 1-1.5% of GDP by 2030 (Liska and Holley, 2014). So far, most financial institutions hedge climate risk by reducing portfolio exposure to carbon-intensive firms, but recent literature recommends to go a step further by applying optimisation and regression techniques.

4.2 Optimisation Techniques

Andersson, Bolton and Samama (2015) have proposed a dynamic hedging strategy, for passive investors, which seeks to hedge climate risk without sacrificing returns. Passive long-term investment strategies often involve the tracking of benchmark green indices as opposed to tracking standard benchmark indices. These are either pure-play indices, which are indices related to environmentally-friendly businesses, or 'decarbonised' indices which are
constructed by removing or under-weighting companies with high carbon footprints from benchmark market indices (e.g. NASDAQ 100) (Andersson et al., 2016). Changes in climate mitigation policies across different nations (e.g. Spain and the US), and changing expectations regarding delays to carbon limits have worked in favour of carbon-intensive firms, and this can explain how pure-play indices have under performed compared with their market benchmarks
(e.g. S&P500). Table 5 below shows the under performance of such indices by considering the indices that track: the Market Vectors Environmental Services Fund (PP1) and the Market Vectors Global Alternative Energy ETF (PP2).

	S&P 500	NASDAQ 100	PP1	PP2
Annualised Return	4.79%	11.40%	5.02%	-8.72%

Table 5: Pure-play clean energy indices vs global indices 09/01/2015

It has thus been more appropriate to use decarbonised indices in such hedging portfolios. Decarbonised indices also provide protection against the timing risk of mitigation policies. This is an important risk, as an investor can decide on tracking a green index if they believe in a short-term implementation of such policies, only to have further delays which can result in underperformance (Andersson et al., 2016). An example of such an index is the S&P 500 Carbon Efficient Index, whose performance relative to the S&P500 index is shown in Figure 3.

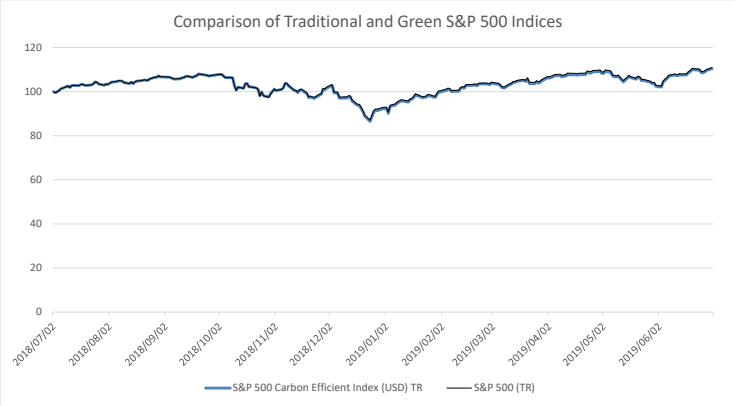


Figure 3: Comparison of green and standard benchmark index
(source: spindices.com)

Andersson, Bolton and Samama (2015) show how such a decarbonised index is constructed. Two optimisation schemes which differ by the treatment of companies with high carbon footprints have been proposed. Let N be the number of stocks in the benchmark index. Define weights $w_i^b = \frac{MarketCap(i)}{TotalMarketCap}$. Let R^g, R^b be the return from the green index and the return from the standard benchmark index, respectively. Each company is then ranked in decreasing order of carbon intensity q_l^i, where $l = 1$ has the highest intensity (Andersson et al., 2016). The first approach is then to minimise the tracking error (the standard deviation of the difference in benchmark and green returns) when the companies with the highest intensities are excluded. Say if k are excluded, then the following function is minimised: $minTE = stdev(R^g - R^b)$ subject to

$$w_j^g = 0, \quad j = 1, ..k0 \leq w_i^g \quad i = k+1, ..N$$

The second approach instead under-weights the companies with the highest inten-

sities, in which case the first constraint is replaced by the constraint that: $\sum_{l=1}^{N} q_l w_l^g \leq Q$. Both approaches allow us to obtain a decarbonised index, so as then to construct a decarbonised portfolio through the following:

Minimise

$$\sqrt{(W^p - W^b)'(\beta \Omega_f \beta' + \Delta^{AR})(W^p - W^b)}$$

subject to

$$w_j^g = 0, \quad j = 1, ..., k \quad \text{and} \quad 0 \leq w_i^g, \quad i = k+1, ..., N$$

$(W^p - W^b)$ is a vector of differences in the weights of the decarbonised and the benchmark portfolio respectively. Ω_f is the variance-covariance matrix of the risk factors, Δ^{AR} is a diagonal matrix of risk variances and β is a vector of risk factor exposures/sensitivities. However, a key concern around this approach is how the carbon footprints of companies are measured. Inaccurate measurements can place implicit bias when filtering high-carbon companies.

4.3 Regression Techniques

Instead of applying optimisation algorithms to minimise tracking errors, one can also apply regression techniques to construct hedging portfolios. In this way, investors can incorporate climate change news as a new risk factor. Engle et al. (2019) propose a dynamic hedging strategy using the mimicking portfolio approach. Consider n assets and $p+1$ risk factors. Let r_t be an $n \times 1$ vector of returns above the risk-free rate at time t. Asset returns are driven by p risk factors given by ν_t and by climate change news CC_t. Returns are given by the linear factor model:

$$r_t = (\beta_{CC}\gamma_{CC} + \beta_{CC}(CC_t - E[CC_t])) + \beta\gamma + \beta\nu_t + \mu_t$$

where γ_{CC}, γ are the returns in excess of the risk-free rate for the climate news factor and other risk factors. The error term is given by μ_t. In this mimicking portfolio approach, a regression is conducted on the climate risk factor CC_t using excess returns $\tilde{r_t}$ given by $CC_t = \xi + w'\tilde{r_t} + e_t$ (i.e. the climate risk factor is projected onto these excess returns). The estimated weights, $\hat{w}$, are then used to construct the hedge portfolio.

A set of projection portfolios are then chosen which are well-diversified and have constant risk exposures to the underlying risk factors. This can be done by sorting assets based on characteristics such that the exposures are dependent on these characteristics. Hence, a matrix of such firm-level characteristics Z_t is chosen such that $\tilde{r_t} = Z'_{t-1} r_t$.
This results in the following

$$CC_t = \xi + w' Z_{t-1} r_t + e_t$$

A fundamental aspect of this approach is determining how to construct an index that measures changes in climate change news. New sources reflect information that may not be material to investors across sectors. For example, investors exposed to fossil fuel industries may be more concerned with regulatory news, while investors with exposure to agricultural sectors may be concerned with news on physical impacts.

Another approach comes in the form of cross-sectional regressions used in the Fama-MacBeth approach (Engle et al., 2019). Firstly, the risk exposures, β_{CC} and β, are measured by a time-series regression on their corresponding factors (CC_t and ν_t respectively). Therefore, the estimates $\hat{\beta}^i_{CC}$, $\hat{\beta}^i$ for asset i are estimated from the following time-series regression:

$$r^i_t = \alpha^i + \beta^i_{CC} CC_t + \beta^i \nu_t + \mu_t.$$

Hedge portfolios are then determined via the following cross-sectional regression of returns

$$r_t = h^{CC}_t \hat{\beta}_{CC} + h_t \hat{\beta} + e_t.$$

4.4 Practical Implementations

The PDC update report 2017 provides key insights into how companies are decarbonising their portfolios. The most common approach relies on the exclusion or reduction of exposure to companies with high carbon footprints such as the Frontera Resources Corporation (FRR), who have excluded companies whose thermal coal mining business exceeds 20% of those companies' revenues (PCD, 2017). This is also coupled with preferential investing into particular sectors and companies that are involved in decarbonisation technologies such as renewable energy, recycling and electric cars. There is also greater emphasis towards developing the green investment system. Companies like AP4 have been involved in the development of low-carbon indices and other products, in the method described above. The critical finding is that companies do not use a single-track strategy to decarbonise their portfolio, but instead have a variety of such strategies which are implemented holistically. There has also been significant improvements to the divestment from fossil fuels to hedge against stranded asset risks. Figure 4 shows that more than 1000 institutions have divested assets amounting to approximately $8 trillion in 2018.
However, the key problem is the lack of global finance. The UNEP adaption report also states that $150 billion is required per year in adaption investment by 2035 in order to meet the 2°C target, while $23 billion was the total amount of adaptation investment in 2016. Furthermore, the IEA estimated that $2.3 trillion of mitigation investment is required by 2035 to achieve the 2°C target and total global mitigation investment finance is $427 billion (Bender et al., 2019).

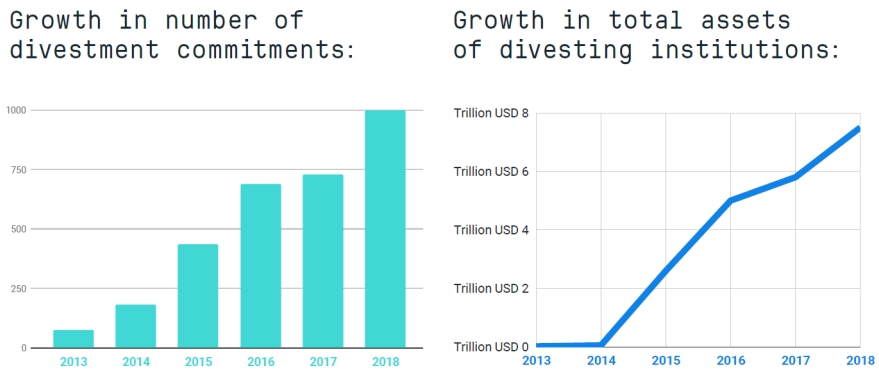


Figure 4: Fossil fuel divestment flow and number of committed companies *(source: gofossilfree.org)*

5 Stock-Flow Consistent IAM Model

In this section, we provide a brief overview of the ecological macroeconomic model proposed in Bovari et al. (2018b). This continuous-time model is based on three main pillars that interact with each other: the monetary macrodynamics, the climate module, and the damages and mitigation costs which links the former two.

5.1 Macroeconomic Dynamics

This module aims to model the global monetary economy. As a consequence of the mitigation efforts of climate change, the need for funding coming from the private sector and some other financial strategies results in over-indebtedness. The main purpose of this framework is to measure the long-term economic deflation and degrowth on this scenario.

5.1.1 Production, Damages and Abatement

We assume the productive sector produces a real amount of a homogeneous consumption and investment good which will be denoted by Y^0. This good is generated by a combination of labour and capital, L and K, respectively, in the following way

$$Y^0 = \frac{K}{\nu} = aL \tag{8}$$

where ν is a constant that represents the capital-output ratio, and a is the labour productivity which is assumed to grow exogenously at a constant rate

$$\frac{\dot{a}}{a} = \alpha \tag{9}$$

where $\alpha \geq 0$ (see Table 8 for value implemented in the model). We assume that full capital is used and Say's law is postulated, namely that supply equals demand at all times.

We further assume that a decarbonisation strategy policy is adopted and a carbon tax is paid when CO_2-e is emitted as a consequence of the economic activities. Then the productive sector might engage in abatement strategies to lower its CO_2-e emissions and hence the tax burden is reduced. That is, the productive sector deflects directly a proportion, $A \in [0, 1]$ (defined in Equation 5), of Y^0 to lower the CO_2-e emissions. Moreover, similar to Nordhaus (2018), a proportion, $\mathbf{D}^Y \in [0, 1]$, (defined in Equation 3) is damaged by global warming and is considered lost. The production available in the commodity market is then given by

$$Y := (1 - \mathbf{D}^Y)(1 - A)Y^0. \tag{10}$$

In addition, we assume that the public sector subsidises a fraction, $s_a \in [0, 1]$, of the abatement cost[3]. Therefore a real transfer

$$S_f^c := s_a A Y^0 \tag{11}$$

is performed to the productive sector.

5.1.2 Profits, Investment and Inflation

We define the variables: D, Π, p, w, r, p_c, E_{ind} and δ according to Table 6.

Variable	Definition
D	Private debt burden.
Π	Nominal net profit.
p	Price level.
w	Unitary money wage.
r	Short-term interest rate.
p_c	Real price of a ton of CO_2-e expressed in USD.
E_{ind}	Volume of industrial emissions in GtC.
δ	Depreciation rate of capital.

Table 6: Profit, investment and inflation variables

Recall, the payment of the carbon tax, pT_f, and the global depreciation rate of capital, δ_D are given by

$$pT_f = pp_c E_{ind}, \tag{12}$$

$$\delta_D = \delta + \mathbf{D}^K, \tag{13}$$

[3]We analyse in Section 6 different scenarios in which $s_a \in \{0, 1/2\}$

where $\mathbf{D}^K$ is the rate of depreciation induced by climate change (defined in Equation 4). In this report we assume that the carbon price p_c is piece-wise linear (see Figure 5).

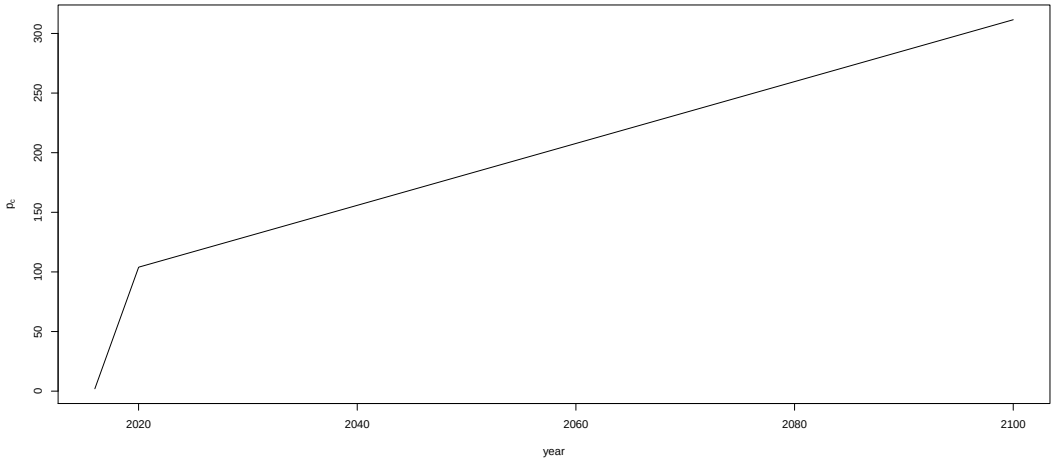


Figure 5: Carbon price path with respect to time (note $p = 1$ in 2016)

The nominal net profit is then

$$\Pi = pY - wL - rD + pS_f^c - pT_f - p\delta_D K. \tag{14}$$

In words, the nominal net profit is the nominal output minus production costs (wage bills, private debt burden, payment of the carbon tax and depreciation of capital) plus subsidies. Thus, the private debt ratio, d, and the nominal profit share, π, can be expressed by

$$d = \frac{D}{pY} \qquad \text{and} \qquad \pi = \frac{\Pi}{pY}.$$

In addition, we define the public debt ratio d_{pub}[4] as

$$d_{pub} = \frac{pS_f^c - pT_f}{pY} \tag{15}$$

that is, the amount paid in subsidy minus the amount received in tax divided by the output. Equation (15) represents the increment in public debt ratio induced by green policies. Let κ be an increasing real-valued function depending on the profit share, π.

[4]We assume that, apart from green subsidies and carbon taxes, the public sector runs a balanced budget.

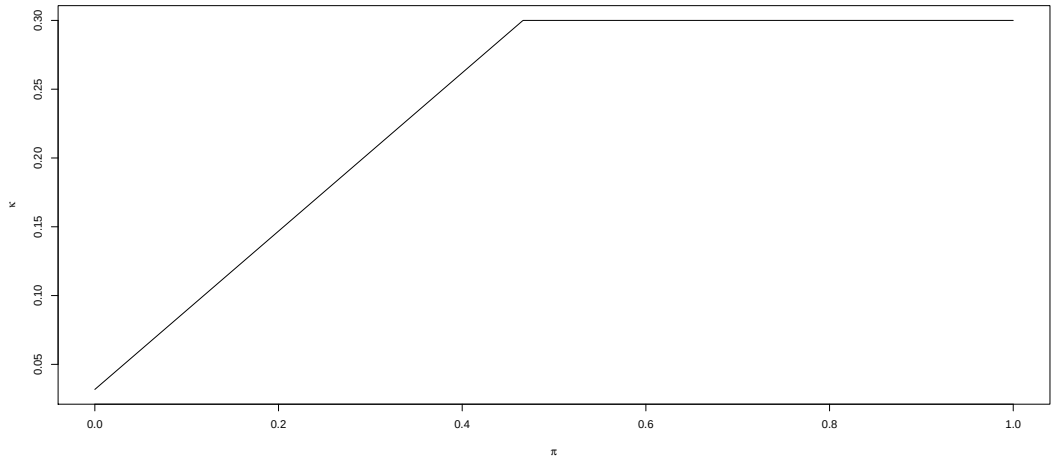


Figure 6: Graph of $\kappa(\pi)$

Such a function will determine the percentage of the production, Y, destined for re-investment, I (see Figure 6), i.e.

$$I = \kappa(\pi)Y, \tag{16}$$

$$\dot{K} = I - \delta_D K. \tag{17}$$

The latter equation tells us that the increment of capital depends on the amount invested minus the depreciation of the current capital. The increment of corporate debt is calculated in terms of the amount of investment, the dividends paid to the shareholders, the profits, and the depreciation of capital according to the following formula

$$\dot{D} = pI + \Pi_d(\pi) - \Pi - p\delta_D K,$$
$$\Pi_d(\pi) = \Delta(\pi)pY,$$

where Δ is a non-increasing function (see Figure 7) that specifies the proportion of Y that
will be paid as a dividends to the shareholders.

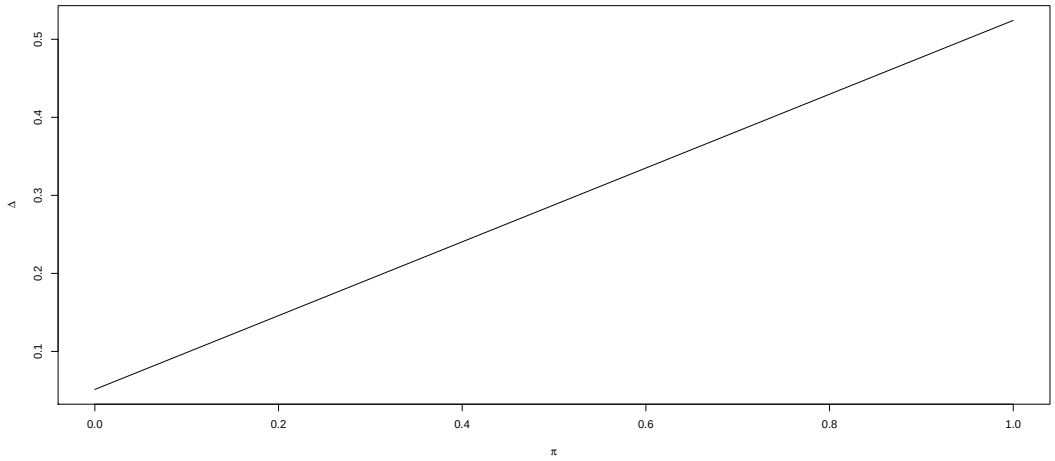


Figure 7: Dividends function

Increments of the price level, i.e. inflation i, are given by the equation

$$i := \frac{\dot{p}}{p} = \eta_p(mc - 1) \tag{18}$$

where $\eta_p \geq 0$ is a fixed parameter, $m \geq 1$ is a markup and c is the labour cost of production given by

$$c = \frac{wL}{pY}.$$

5.1.3 The Labour Market

The workforce, N, is assumed to be bounded by $P^N \geq 1$, and increases in line with the following formula

$$\frac{\dot{N}}{N} = q\left(1 - \frac{N}{P^N}\right) \tag{19}$$

where q is the rate of convergence of N to P^N. The employment rate is defined as the ratio between the number of employed workers, L, and the global labour force, N, i.e.

$$\lambda := \frac{L}{N}. \tag{20}$$

Hence, we assume that the increment of the wage is directly linked to the employment rate via

$$\frac{\dot{w}}{w} = \phi(\lambda) \tag{21}$$

where ϕ is an increasing, real-valued function. From Figure 8 we easily deduce that the wage decreases exponentially when the employment rate is below $1/2$ and increases exponentially when $\lambda \geq 1/2$.

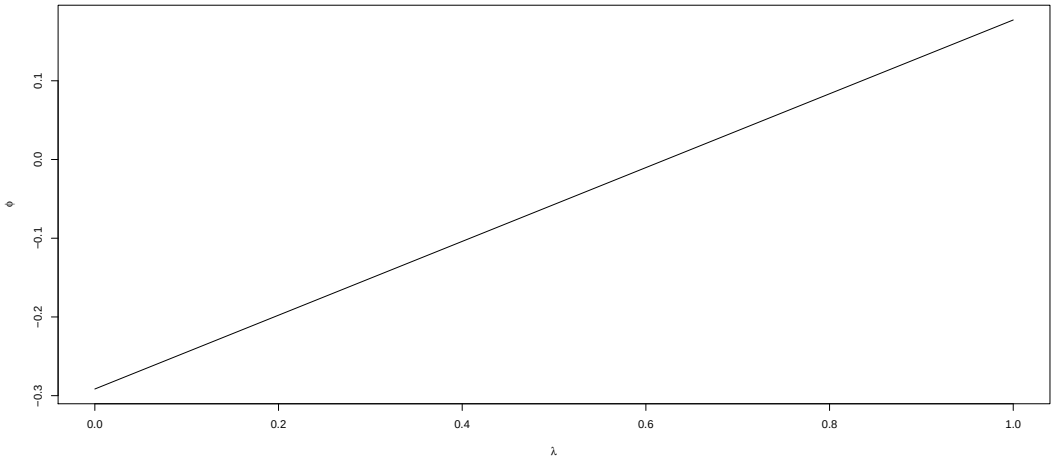


Figure 8: Wage function

5.2 Climate Dynamics

It is important to mention that the climate module is mainly based on the DICE model proposed by Nordhaus (2018) which is defined on a discrete-time framework. Figure 9 represents a simplified version of the DICE model. The red arrows indicate the effect of climate on the economy; the blue arrows indicate the effect of the economy on the climate; the green arrows correspond to the effect of climate policy; and the orange arrows show the purely economic component of the model.

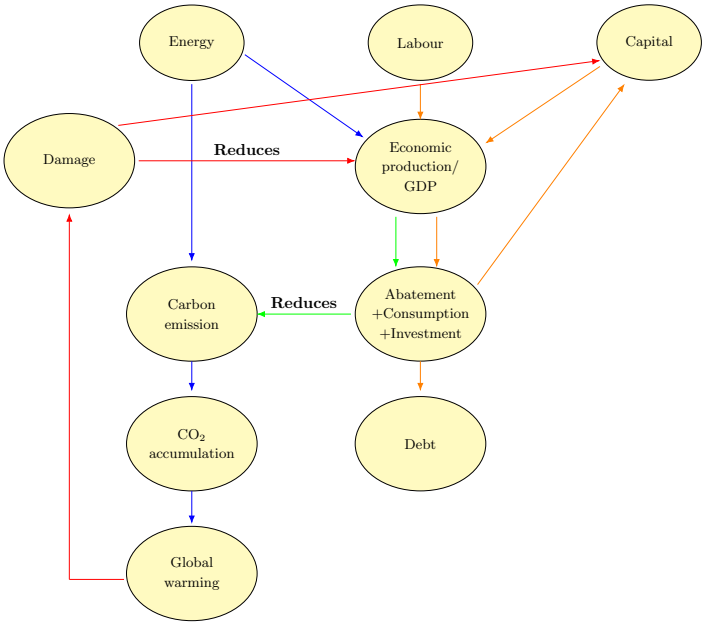


Figure 9: Simple schematic of the DICE model

Bovari et al. (2018a) adapted the DICE model to a continuous framework as follows: We assume that the total CO_2-e global emissions, E, come from two main sources: industrial and land (e.g. implied release of CO_2-e induced by deforestation), E_{ind} and E_{land}, respectively, i.e.

$$E = E_{ind} + E_{land}. \tag{22}$$

Industrial emissions, E_{ind}, are assumed to be directly proportional to the real produced output, Y^0, and is given by the formula

$$E_{ind} = Y^0 \sigma (1 - n) \tag{23}$$

where σ is the current emission intensity of the economy and n is the emission reduction rate chosen by the productivity sector as a function of the carbon price and the abatement technology (defined in Sections 2.2 and 5.3). The value of σ is driven by the following
differential equation where $\delta_{g_\sigma} \leq 0$ is a fixed parameter (see Figure 10):

$$\frac{\dot{\sigma}}{\sigma} = g_\sigma, \tag{24}$$

$$\frac{\dot{g_\sigma}}{g_\sigma} = \delta_{g_\sigma}. \tag{25}$$

On the other hand, we assume that the land-use emissions are exogenous and decrease at a rate of $\delta_{E_{land}} \leq 0$, i.e.

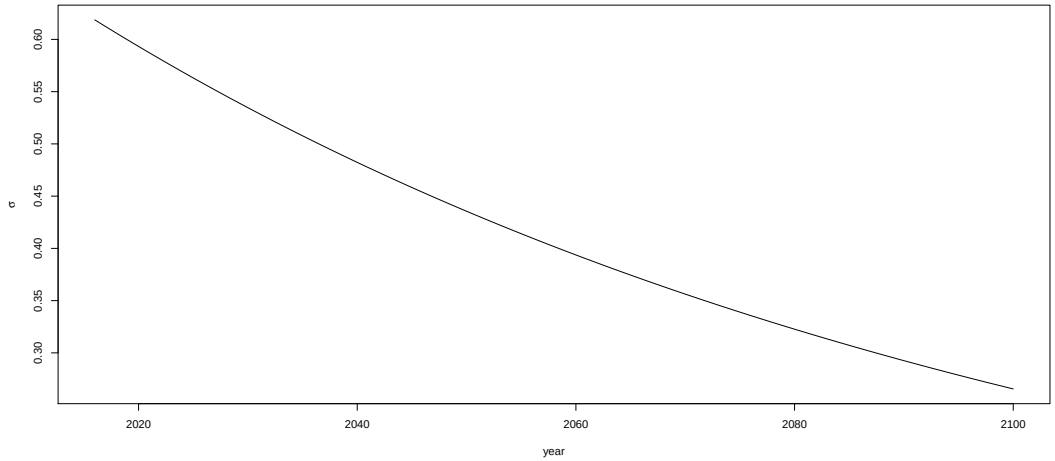


Figure 10: Function σ

$$\frac{\dot{E_{land}}}{E_{land}} = \delta_{E_{land}}. \tag{26}$$

Denote CO_2^{AT}, CO_2^{UP} and CO_2^{LO} as the CO_2-e concentrations in the: atmosphere, upper ocean and biosphere and deep ocean, respectively. In addition, define $C_{i_{pind}}$ as the relative pre-industrial concentrations for $i \in \{AT, UP, LO\}$ in GtC.

As a consequence of the accumulation of GHGs in the atmospheric layer, we get a rise in the radiative forcing, F. This is considered to be a sum of two terms:

$$F := F_{ind} + F_{exo} \tag{27}$$

where F_{exo} is a residual forcing which is assumed to be exogenous and linear with respect to time up to 2100, followed by a plateau (see Figure 11). On the other hand, F_{ind} represents the industrial forcing (from CO_2-e) and is modelled by

$$F_{ind} = \frac{F_{2\times CO_2}}{\log(2)} \log\left(\frac{CO_2^{AT}}{C_{AT_{pind}}}\right) \tag{28}$$

where $F_{2\times CO_2}$ represents the increase in the radiative forcing resulting from a doubling of the pre-industrial CO_2-e concentration.

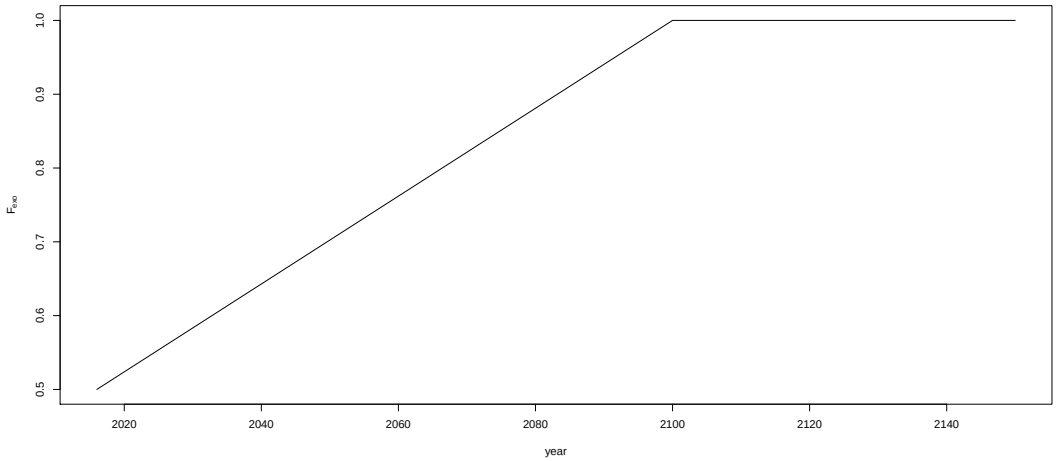


Figure 11: Exogenous radiative forcing function

The change of CO_2^i, $i \in \{AT, UP, LO\}$ concentrations are given by an interacting three-layer model

$$\begin{pmatrix} \dot{CO_2^{AT}} \\ \dot{CO_2^{UP}} \\ \dot{CO_2^{LO}} \end{pmatrix} = \begin{pmatrix} E \\ 0 \\ 0 \end{pmatrix} + \Phi \begin{pmatrix} CO_2^{AT} \\ CO_2^{UP} \\ CO_2^{LO} \end{pmatrix}, \tag{29}$$

where

$$\Phi := \begin{pmatrix} -\phi_{12} & \phi_{12} C_{UP}^{AT} & 0 \\ \phi_{12} & -\phi_{12} C_{UP}^{AT} - \phi_{23} & \phi_{23} C_{LO}^{UP} \\ 0 & \phi_{23} & -\phi_{23} C_{LO}^{UP} \end{pmatrix}, \tag{30}$$

and

$$C_i^j := \frac{C_{j_{pind}}}{C_{i_{pind}}}, \qquad i, j \in \{AT, UP, LO\} \tag{31}$$

where ϕ_{12} and ϕ_{23} are fixed parameters. We denote T (resp. T_0) as the mean of global temperature anomaly in the atmosphere and upper ocean (resp. the lower ocean). Then the dynamics of T and T_0 are assumed to be given by

$$\dot{T} = \frac{1}{C} \left(F - \rho T - \gamma^* (T - T_0) \right), \tag{32}$$

$$\dot{T_0} = \frac{\gamma^*}{C_0} (T - T_0), \tag{33}$$

where $\rho = \frac{F_{2\times CO_2}}{S}$ is the radiative feedback parameter, γ^* the heat exchange coefficient between the two layers, C the heat capacity of the atmosphere, and C_0 the heat capacity of the deep ocean layer.

5.3 Damages and Mitigation

Recall from Section 2, we define $\mathbf{D}$ as the total damages function. Such a function aims to summarise the economic impacts caused by climate change. It is assumed the changes of the damages are a function of the temperature anomaly according to the next equation

$$\mathbf{D} := 1 - \frac{1}{1 + \pi_1 T + \pi_2 T^2 + \pi_3 T^{\zeta_3}} \tag{34}$$

where π_1, π_2, π_3 and ζ_3 are fixed parameters (see Table 8 and Figure 12). A fraction of the total damages correspond to damages to the capital, $\mathbf{D}^K$. Then,

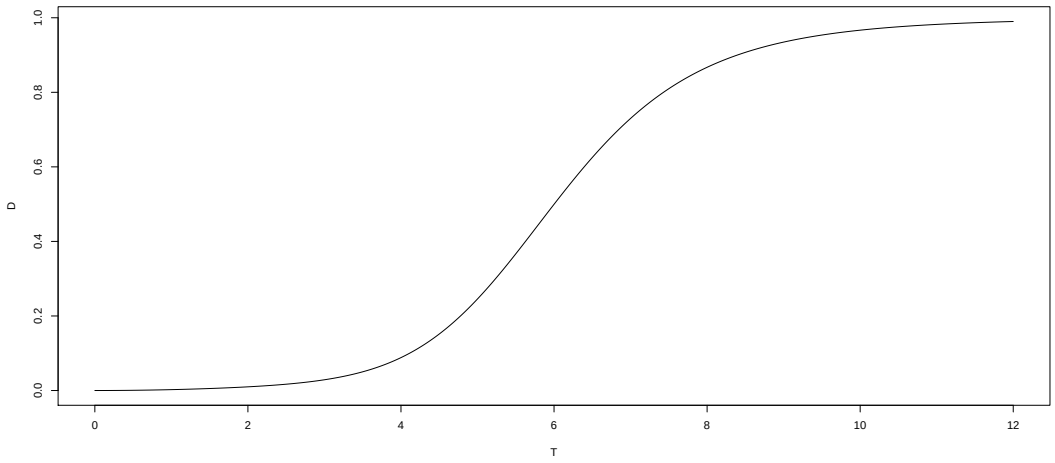


Figure 12: Damages function.

$$\mathbf{D}^K := f_K \mathbf{D} \tag{35}$$

where $f_K \in [0, 1]$ represents the share of total damages corresponding to the capital. Analysis on the sensitivity is carried out in Section 6. On the other hand, damages to the total output denoted by $\mathbf{D}^Y$ are defined as

$$\mathbf{D}^Y := \mathbf{D} - \mathbf{D}^K. \tag{36}$$

As an attempt of reduction of the carbon tax (see Equations (6) and (12)), some efforts on reduction of the emissions are made. This cost is a proportion of the total output given by AY^0. We assume the existence of a "backstop technology", which is a technology that produces energy services with zero greenhouse emissions. Hence, the proportion of abatement, A, is calculated according to the abatement formula

$$A := \frac{\sigma p_{BS}}{\theta} n^{\theta} \tag{37}$$

where p_{BS} is the price of the backstop technology, θ is a parameter that controls the convexity of the cost and n is the resulting emission reduction rate as a consequence of the minimisation efforts of the carbon tax and abatement. In other words, n is such that it minimises $T_f + S_f^c$. From equations (5), (6), (12), (23) and (37) we get that

$$\begin{aligned}\min_{n\in[0,1]} \left(T_f + S_f^c\right) &= \min_{n\in[0,1]} \left(p_c E_{ind} + s_a A Y^0\right) \\ &= \min_{n\in[0,1]} \left(p_c Y^0 \sigma (1-n) + s_a \frac{\sigma p_{BS}}{\theta} n^{\theta} Y^0\right).\end{aligned}$$

Taking derivatives with respect to n and equating to zero we obtain that n must solve the equation

$$-p_c Y^0 \sigma + s_a \sigma p_{BS} n^{\theta-1} Y^0 = 0.$$

Hence, solving the equation and assuming that $\theta > 1$[5], we obtain that n then takes the form

$$n = \min\left\{\left(\frac{p_C}{(1-s_a)p_{BS}}\right)^{\frac{1}{\theta-1}}, 1\right\} \tag{38}$$

The price of the backstop technology is assumed to be available at a declining rate $\delta_{p_{BS}} \leq 0$:

$$\frac{\dot{p_{BS}}}{p_{BS}} = \delta_{P_{BS}}. \tag{39}$$

Section 6 provides a sensitivity analysis taking into account variations on θ and $\delta_{p_{BS}}$.

[5]Checking the second derivative we conclude that n is a minimum only when $\theta > 1$

5.4 Benchmark and No-Policy scenarios

In this subsection we provide a brief overview of the results obtained on the model described above. This with the objective to provide a benchmark for the modifications to the model and results obtained in the next section. In addition, we present when no public policy is assumed, i.e. no carbon tax. The parameters and initial values can be found in Tables 7 and 8.

Variable	Description	Value
CO_2^{AT}	CO_2-e concentration in the atmosphere layer	851 Gt C
CO_2^{UP}	CO_2-e concentration in the biosphere and upper ocean layer	460 Gt C
CO_2^{LO}	CO_2-e concentration in the deep ocean layer	1,740 Gt C
d	Private debt ratio of the economy	1.53
E_ind	Industrial CO_2-e emissions	35.85 Gt CO_2-e
E_{land}	Exogenous land use change CO_2-e emissions	2.6 Gt CO_2-e
F_{exo}	Exogenous radiative forcing	0.5 $\frac{W}{m^2}$
g_σ	Growth rate of the emission intensity of the economy	-0.0152
p	Composite good price level	1
p_{BS}	Backstop price level	547.22
n	Emissions reduction rate	0.03
N	Workforce of the economy in billions	4.83
NG	Total population in billions	7.35
T	Temperature in the atmosphere, biosphere and upper ocean layer	0.85 ° C
T_0	Temperature in the deeper ocean layer	0.0068 ° C
Y	Gross domestic product (at factor prices) in trillions USD	59.74
λ	Employment rate of the economy	0.675
ω	Wage share of the economy	0.578

Table 7: Initial values of the model

Parameter	value	Parameter	value
α	0.02	δ	0.04
η_p	0.3	m	1.2
q	0.0274	P^N	12
δ_{g_σ}	-0.001	$\delta_{E_{land}}$	-0.022
$C_{AT_{pind}}$	588	$C_{UP_{pind}}$	360
$C_{LO_{pind}}$	1720	$F_{2\times CO_2}$	3.6813
ϕ_{12}	0.024	ϕ_{23}	0.0013
S	3.1	γ^*	0.0176
C	49.758	C_0	3.52
π_1	0	π_2	0.00236
π_3	0.00000507	ξ_3	6.754
θ	2.6	$\delta_{p_{BS}}$	-.005

Table 8: Parameters of the model

5.4.1 Benchmark Scenario

In this subsection we present the main results extracted from the model originally presented in Bovari et al. (2018b). This model assumes a carbon policy scenario with no subsidy ($s_A = 0$) as described above. Figure 13 and Table 10 present the trajectories and final values of some variables from 2016 to 2100. The variables presented are GDP, Y^0, debt-output ratio, d, temperature anomaly, T, and CO_2-e total emissions. We note that in this simulation, GDP maintains a growth over the years to get a final value of 458.35 in 2100. The debt-output ratio presents a constant growth on the first 40 years and stabilises over the year 2070 on a level around 1.85; reaching a value of 1.811 at 2100. The temperature has a constant growth rate over the years reaching a level of 3.1368 in 2100. Finally, we notice that the total carbon emissions present an important decay in the first 6 years. From 2020 to 2050, we observe a slight increase, and from 2060 we see that the emissions drop rapidly to get a value of 10 $GtCO_2$-e emissions.

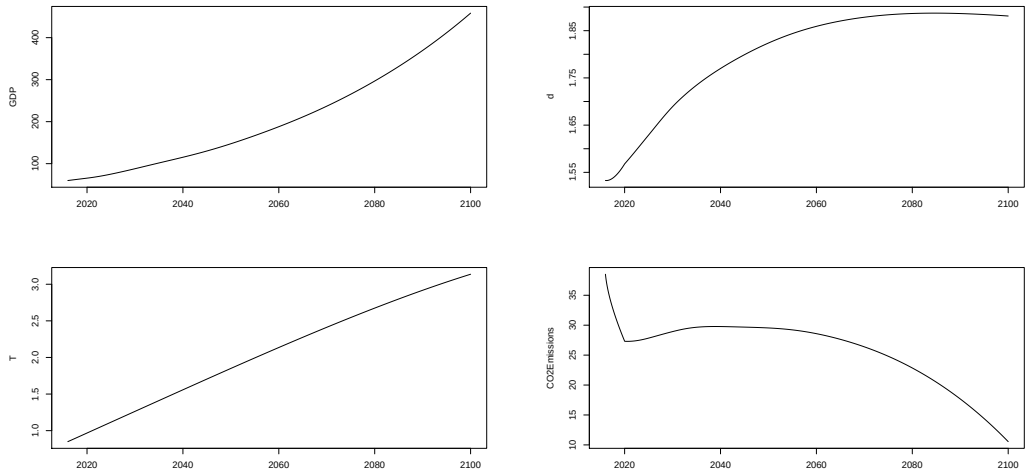


Figure 13: Simulation of the model presented in Bovari et al. (2018b)

Year	Y^0	d	T	E
2100	458.35	1.8811	3.1368	10.547

Table 9: Benchmark scenario: variable values in 2100

5.4.2 No-Policy Scenario

In this subsection, we present a brief description of the main results obtained by simulating the model presented in Bovari et al. (2018b) with a no carbon tax pol-

icy. That is, we assume that $p_C = 0$ and the carbon tax, pT_F, (defined in (12)) is zero. As a consequence, $n = 0$ (see Equation (38)) and no abatement is generated. In addition, there is no effort in reducing the industrial emissions and E_{ind} has a decreasing rate according to σ (see Equation (23) and Figure 10) and an increasing rate according to Y^0. The latter leads to a rapid growth of the temperature which implies an increase of the total damages.

Figure 14 and Table 10 present the trajectories of GDP Y^0, debt-output ratio, d, temperature anomaly T and CO_2-e total emissions. In comparison to the benchmark scenario the growth on GDP presents a slower trend reaching a level of 338 in 2100 with a downward tendency. The debt remains relatively low until the damages caused by the increasing temperature start to become too high. The temperature increases with no restrain since the CO_2 emissions increase unbounded as well until the GDP stops growing.

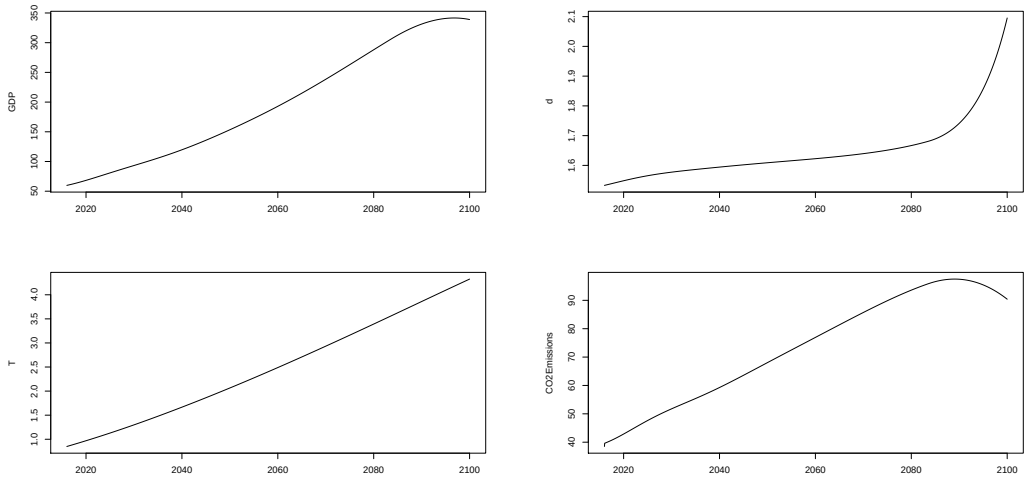


Figure 14: Simulation of the model presented in Bovari et al. (2018b) assuming no carbon tax

Year	Y^0	d	T	E
2100	338.94	2.0954	4.3261	90.396

Table 10: No policy scenario: variable values at 2100

6 Results

6.1 Proposed Modifications to the AFD Model

The AFD Model presented by Bovari, Giraud and McIsaac (2018) make key assumptions on the financing of investments in the productive sector. The variable D represents total debt and is assumed to be raised via bank loans at an assumed constant rate of $r = 0.03$. Green bonds are an alternative source of debt that have been shown to have a lower yield and can be used in constructing hedging portfolios for climate risk. We therefore investigate the effect of incorporating green bonds into the model. This is done by splitting total debt, D, into debt from bank loans, D_B, and debt from green bonds, D_G. Modifying the model in this way changes the nominal profit by subtracting a further cost term of $r_G D_G$, where r_G represents the yield on green bonds

$$\Pi = pY - wL - rD_B - r_G D_G - pT_f - p\delta_D K.$$

These green bonds, by definition, are used for 'green' purposes; hence they play a role in reducing emissions and thus fund some portion of the abatement cost, A. This portion is given by $f \leq 1$. The dynamics of green bonds is then given by

$$\dot{D_G} = f(1 - s_a)pAY^0 \tag{40}$$

and the full dynamics of debt is given by

$$\dot{D_B} = pI - \Pi_r - p\delta_D K - \dot{D_G}.$$

We define δ_G as the interest rate spread between r and r_G, i.e.

$$\delta_G = r - r_G. \tag{41}$$

and use, for now, today's real spread of 7bps or 0.0007.

6.2 Sensitivity Analysis

6.2.1 Deterministic analysis

In this section, we analyse the sensitivity of the model in terms of changes of some of the parameters. The parameters to consider are listed in Table 11.

θ	Parameter controlling the convexity of the cost (see Equations (5), (37) and (38))
δ_G	Interest rate spread (see Equation (41))
$\delta_{p_{BS}}$	Rate of decrease of the backstop technology (see Equations (7) and (39))
s_a	Subsidy of the abatement cost (see Equation (37))
f_K	Share of the total damages to the capital K (see Equations (4) and (35))

Table 11: Parameters for sensitivity analysis

In the subsequent subsections, we calculate the temperature and the debt-output ratio in 2100 for different values of θ, δ_G, δ_{pBS}, s_a and f_K. The simulated scenarios incorporate the modifications presented in Subsection 6.1 with $f = 1$ (see Equation (40)), i.e. we assume that debt generated by abatement costs are fully-covered with green bonds.

6.2.2 Sensitivity to θ

We present in Tables 12 and 13 below the results of varying the value of θ. Recall from Equations (5), (12), (23), (37) and (38) that θ intervenes in the total amount (a fraction of the total output) that is spent as a consequence of CO_2-e emissions.

Note that both the temperature anomaly and the debt-output ratio decreases when the value of θ increases in all the simulated scenarios. In the case in which the value of θ is closer to 1 and there is no subsidy, an increase of damages of capital imply an increase in the debt-output ratio. In all cases, incorporating a 50% of subsidy leads to a better temperature anomaly and debt levels. Another important fact that is worth mentioning is that under the presence of a 50% subsidy, the debt ratio does not show a significant increase when the damages of capital increases. These results are consistent with the fact that for large values of θ, we obtain low abatement costs for fixed level of emission reduction $n < 1$.

θ	°C	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$
1.1		4.2751	4.2050	4.1665		3.4066	6.8585	10.7882
2.8		3.0789	3.0680	3.0625		2.0897	2.0638	2.0552
5		2.6239	2.6186	2.6159		1.8651	1.8338	1.8162
20		2.1751	2.1740	2.1734		1.6991	1.6706	1.6549

Table 12: Variation of temperature anomaly and debt-output ratio in 2100 for different values of θ and f_K with no subsidy

θ	°C	$f_K = 0\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 50\%$	d_{pub}	$f_K = 0\%$	$f_K = 50\%$
1.1		3.3991	3.3797		2.0184	2.0856		0.049	0.048
2.8		2.4307	2.4270		1.7045	1.6506		0.018	0.018
5		2.2368	2.2349		1.6761	1.6279		0.01	0.01
20		2.0730	2.0726		1.6569	1.6139		0.0024	0.0024

Table 13: Variation of temperature anomaly, debt-output ratio and public debt in 2100 for different values of θ and f_K with a 50% subsidy

6.2.3 Sensitivity to δ_G

Recall that δ_G is defined as the spread rate between r and r_B (see Equation (41)). In Tables 14 and 15, we analyse the simulation for the temperature anomaly, T, and the debt-output ratio, d, for different values of δ_G. The studied values for δ_G are $\{1$, 7, 20 and $200\}$ bps. Although the preceding value might be an unrealistic comparison to reality, it is included since it still provides a good idea of how the model will react under notable changes in the spread.

We note that in the no-subsidy case, changes in δ_G do not induce significant changes in T and d. The extreme case of 200bps presents similar results. For the 50% subsidy case, we draw similar conclusions. This is likely due to the fact that abatement efforts remain a relatively small proportion of the total output Y^0 (see Figure 15). Moreover, similar to before, we realise that the presence of subsidy produces slightly better results in terms of the temperature anomaly and debt.

δ_G	° C	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$
1 BPS		3.0788	3.0680	3.0624		2.0926	2.0674	2.0597
7 BPS		3.0789	3.0680	3.0625		2.0897	2.0638	2.0552
20 BPS		3.0791	3.0682	3.0626		2.0833	2.0561	2.0456
200 BPS		3.0816	3.0707	3.0652		1.9986	1.9546	1.9287

Table 14: Variation of temperature anomaly and debt-output ratio in 2100 for different values of δ_G and f_K with no subsidy

δ_G	° C	$f_K = 0\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 50\%$	d_{pub}	$f_K = 0\%$	$f_K = 50$
1 BPS		2.4307	2.4270		1.7071	1.6538		0.018	0.018
7 BPS		2.4307	2.4270		1.7045	1.6506		0.018	0.018
20 BPS		2.4308	2.4270		1.6989	1.6438		0.018	0.018
200 BPS		2.4316	2.4278		1.6264	1.5557		0.018	0.018

Table 15: Variation of temperature anomaly, debt-output ration and public debt in 2100 for different values of δ_G and f_K with a 50% subsidy

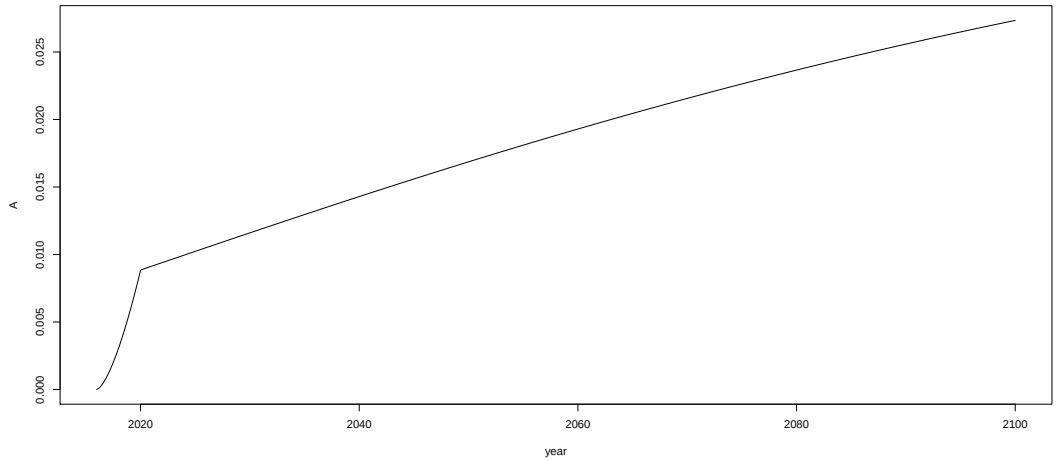


Figure 15: Proportion of abatement in the original scenario

6.2.4 Sensitivity to $\delta_{p_{BS}}$

Lastly, in this subsection we analyse the sensitivities when we vary the parameter $\delta_{p_{BS}}$. Recall that the parameter $\delta_{p_{BS}}$ measures the rate in which the price of the backstop technology decreases (see Equation (39)). That is, for bigger values on $\delta_{p_{BS}}$, the price of the backstop technology will decrease faster to zero over a fixed period of time. We test different values of $\delta_{p_{BS}}$ and summarise our findings in Tables 16 and 17. It is important to remark that assuming a decreasing rate of -10% and -50% is unrealistic. However, this simulation will be helpful to measure how efforts to obtain cheaper technologies to achieve zero GHG emissions will help mitigate the temperature anomaly and debt.

As intuition indicates, we conclude that with a higher decrease rate of the price of backstop technology, we obtain a significant decay in the temperature emissions and the debt ratio. Notice that in the no-subsidy scenario: when $\delta_{p_{BS}}$ from $-.5\%$ to -1% (i.e. we duplicated the rate), we saw a reduction in the temperature anomaly of about 6.50%; whereas an increase from -10% to -50% (a 5 times increment) led to a reduction in temperature of about 5.5%. That drives us to the conclusion that there is a threshold in which big efforts on the the reduction of the price will result in a small reduction in temperature. The same conclusions are drawn for the subsidy case.

It is also worth noting that when $\delta_{p_{BS}}$ goes from -10% to -50% the debt ratio decreases on a rate less than 1%.

δ_{pBS}	°C	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$
-0.1 %		3.2212	3.2074	3.2001		2.1974	2.1749	2.2385
-0.5%		3.0809	3.0700	3.0643		2.0911	2.0652	2.0571
-1%		2.8780	2.8708	2.8671		1.9478	1.9189	1.9023
-10%		2.2042	2.2035	2.2032		1.6720	1.6412	1.6243
-50%		2.0815	2.0814	2.0814		1.6557	1.6273	1.6118

Table 16: Variation of temperature anomaly, debt-output ratio and public debt in 2100 for different values of δ_{pBS} and f_K with no subsidy

δ_{pBS}	°C	$f_K = 0\%$	$f_K = 50\%$	d	$f_K = 0\%$	$f_K = 50\%$	d_{pub}	$f_K = 0\%$	$f_K = 50\%$
-0.1 %		2.5503	2.5445		1.7480	1.6936		0.025	0.025
-0.5%		2.4319	2.4281		1.7048	1.6510		0.018	0.018
-1%		2.3474	2.3448		1.6814	1.6286		0.012	0.012
-10%		2.1130	2.1127		1.6556	1.6101		<0.001	<0.001
-50%		2.0642	2.0641		1.6538	1.6106		<0.001	<0.001

Table 17: Variation of temperature anomaly, debt-output ratio and public debt in 2100 for different values of δ_{pBS} and f_K with a 50% subsidy

6.3 Scenario Analysis

As in Bovari et al. (2018b), we conduct a scenario analysis on different public policy scenarios. The first is a scenario with a carbon tax which is used in Bovari et al. (2018b) and this was calibrated in the same way based on data from the High Level Commission on Carbon Prices (2017). This also assumes optimistic conditions by focusing on high carbon prices such as achieving $80. The second adds a subsidy for backstop technologies in order to deal with the potential obstacles in pricing carbon, recommended by the Commission (Stiglitz et al., 2017). This implies that the abatement cost, AY^0, is partially subsidised, given by

$$A = \frac{\sigma(1 - s_a)p_{BS}n^\theta}{\theta}$$

Furthermore, we analyse the effect of changes to the parameter f_K, which is the share of total damages allocated to capital. The reason behind this analysis is due to the potential contagion effect to which Bovari et al. (2018b) refers. Damages to capital have the potential to affect, radically, many channels within the model such as investment and debt. As per Bovari et al. (2018b), we explore values of this parameter for the set $\{0, 33.3, \text{and } 50\%\}$, for consistent comparisons.

It should be noted that Bovari et al. (2018b) tested a baseline scenario of no policy action. In our analysis, this implies a zero reduction rate, n, which leads to zero abatement costs and hence, a scenario analysis on its convexity, θ, would not be

material.

Another scenario is also considered, namely the introduction of a cap-and-trade policy instead of a carbon tax. Thus, instead of fixing a price and leaving the quantity of emissions uncertain (as is the case in a carbon tax regime), we fix the level of emissions and allow the market price of emissions to vary. The cap and trade scenario is presented in a later section.

One concern with the model is the high dimensionality of interactions in the system (Nordhaus, 2018). This also proves computationally difficult for simulations, however, by focusing on these three parameters, there is a reduction in dimensionality which allows for Monte Carlo simulations without approximations. We therefore present approximate densities for the three parameters: $\theta, \delta_{p_{BS}}, \delta_G$.

We conducted 10,000 Monte Carlo simulations with joint draws from each of the three
parameters, and simulated the temperature and debt-to-output trajectories from 2016 to 2100. The model parameters assume the optimistic scenario of using the polynomial damage function with $\pi_3 = 0.00000507$.

d	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$
Carbon Tax	2.242299	2.28198	2.339741
Carbon Tax & 50% subsidy	1.7672	1.732044	1.712238

Table 18: Median debt-to-output ratio in 2100

From Table 18, we see that with a carbon tax, the debt-to-output ratio increases, as the
allocation of damages to capital increases. This can be explained by the increase in debt that occurs as more finance is required to implement the repair and replacement of capital. We also see that the debt can be mitigated through the implementation of subsidies for backstop technologies. This is due to the reduction in abatement costs by this subsidy, and since we assume all abatement costs are to be financed by green bonds, we have that the debt burden should also reduce. Furthermore, adding a subsidy reduces the debt ratio for higher damages to capital and this is owed to the subsidy, shifting the debt burden more towards public debt.

Table 19 shows that the temperature anomaly declines over increasing allocation of damages to capital, which is consistent with the trends found in Bovari et al.

Temperature	$f_K = 0\%$	$f_K = 33\%$	$f_K = 50\%$
Carbon Tax	3.284138	3.270338	3.262077
Carbon Tax & 50% subsidy	2.601612	2.597607	2.593778

Table 19: Median temperature value in 2100

(2018b). Comparing this with Table 18 on the Carbon Tax regime, we see a trade-off between increasing debt and lowering temperature. However, with a subsidy, this trade off does not occur.

It is also worth noting that in this optimistic scenario, the debt-to-output ratio remains below the 2.7 threshold. However, the temperature anomaly remains consistently above the desired 2°C target. This shows that even adding green bonds to the model results in no significant improvement to attaining the goal, even in the most optimistic scenario.

6.4 Cap-and-Trade

For the Cap-and-Trade mechanism, we have to fix the emission function first. We assume an exponential decay. The industrial emissions then follow

$$\dot{E}_{ind} = E_{ind}\delta_{E_{ind}}$$

with $\delta_{E_{ind}} < 0$. As in the original model we have that

$$E_{ind} = Y^0\sigma(1-n)$$

and therefore we get

$$n = 1 - \frac{E_{ind}}{\sigma Y^0}.$$

Note that n is always less than 1, so we can transform

$$n = \min\left\{\left(\frac{p_C}{(1-s_a)p_{BS}}\right)^{\frac{1}{\theta-1}}, 1\right\}$$

to

$$p_C = n^{\theta-1}(1-s_a)p_{BS}.$$

This gives us the implied price of carbon necessary to reach the desired emissions. In Figures 16 and 17, we see plots of both the Cap-and-Trade, and the Carbon Tax scenarios. In this example, we look at an exponential decay of two percent annually starting with today's industrial emissions, and compare it to the AFD model with a linear pricing function.

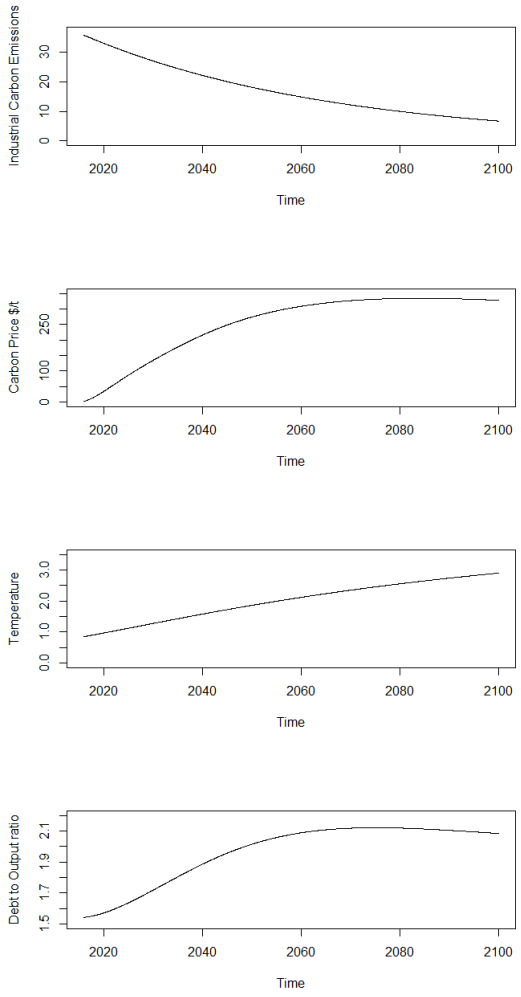


Figure 16: Industrial emissions in $GtCO_2$-e, Carbon Price in $\$/t\,CO_2$-e, temperature increase in °C, and debt-to-output ratio for Cap-and-Trade with 2% decay

Looking at these graphs, we see that we can stay in the range of 3 °C-warming without many problems as we do not come close to the assumed critical debt-to-output ratio of 2.7 (Global Default). While it is slightly higher in the beginning of the Cap-and-Trade model, we see that it is already decreasing, while the Carbon Tax model seems stationary. On the other hand, the temperature increase is slightly lower in the Cap-and-Trade than in the Carbon Tax model and the slope appears less steep. Further calibration could bring these results closer together, but at this point we are satisfied with the realisation that both approaches lead to similar results.
Now we look at the sensitivity of our model to the reduction rate of emissions.

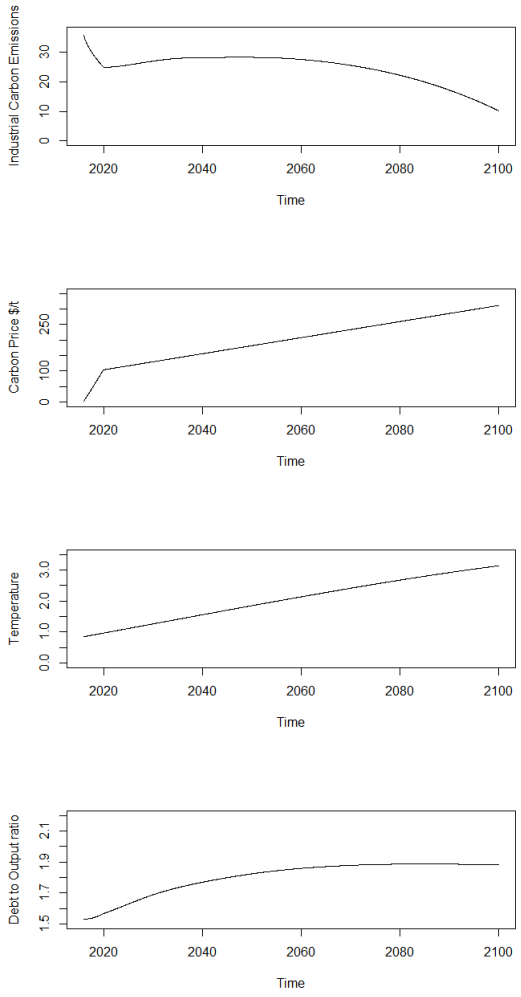


Figure 17: Industrial emissions in $GtCO_2$-e, carbon price in $\$/tCO_2$-e, temperature increase in °C, and debt-to-output ratio for Carbon Tax with linear price function as in the AFD model

Therefore, we look at $\delta_{E_{ind}}$ between 0-5% with steps of 0.5, and see how it affects the temperature; the debt-to-output ratio; the total inflation; and the price of carbon.

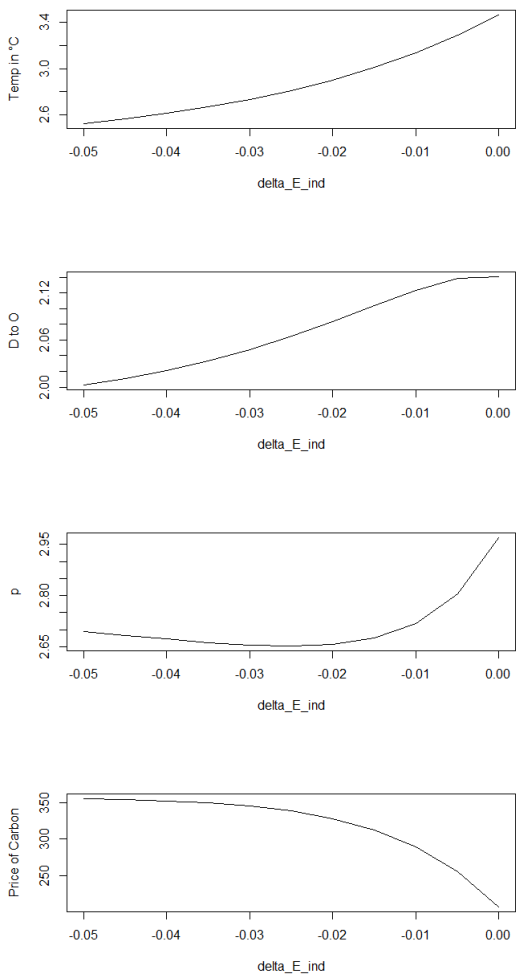


Figure 18: Temperature, debt-to-output ratio, prices and price of carbon for varying rates of emission decrease

Unsurprisingly a faster decrease in emissions leads to a lower increase in temperature and a higher price of carbon. We also observe that the debt-to-output ratio is lower for higher emission reductions, which enables to us conclude that the avoided damages outweigh the mitigation costs. The slope of the price curve is included to illustrate that we stay within reasonable economical bounds with these parameters.

7 Conclusions

After giving an overview of the threats of climate change and the methods to limit it and to mitigate the damages, we looked at means to fund the transition to a carbon-free economy. There, we looked at the current state of public funding and green bonds, both means of financing sustainable projects. After that we analysed ways to hedge against climate risks both in theory and in practice. Then we summarised the Stock-Flow Consistent IAM Model in Bovari et al. (2018b) and gave two scenarios in which we see the drastically different outcomes between a scenario with carbon tax and one without. Finally we started our own analysis where we implemented the green bonds as an extension of the model. Furthermore we analysed how the model's outcome would change for differences in some parameters. We also conducted a Monte-Carlo simulation with various combinations of some of these parameters to see the effect of subsidies. Finally, we changed the model from a carbon tax to a cap and trade model to see whether it would result in different outcomes.

We hence reached the conclusion that, even though green bonds are an attractive instrument for sustainable investors, they alone will not make a big difference on the climate. The results obtained from our proposed modifications of the model are similar to those presented in Bovari et al. (2018b). That is, even with an optimistic carbon tax curve price, the 2 °C target anomaly set in the Paris Agreement seems unreachable with high probability. However, is important to mention that adding a 50% subsidy to investment in the backstop technology increases the probability of achieving such a goal.

It is important to mention that the sensitivity analysis carried out focused on variables that are directly linked with abatement of the carbon emissions. We hence reached two main conclusions on that matter. With the current calibration of the model, the parameter θ, which controls the convexity of the abatement efforts, and the decay rate $\delta_{p_{BS}}$ of the price of the backstop technology were the most sensible parameters; whereas the spread rate δ_G does not play an important role in the final results. So a careful calibration of θ and $\delta_{p_{BS}}$ is important to obtain results close to reality. A surprising result is that there is a threshold for which significant efforts on getting cheaper backstop technology will stop leading to a significant improvement on the temperature anomaly. We estimate that when the decay rate of the backstop technology is below -20%, the temperature anomaly and the debt ratio will remain essentially the same.

Lastly, we have seen that, under an exponential decay assumption of the industrial emissions and the current calibration of the model, a cap-and-trade system for emissions leads to similar results.

Bibliography

Andersson, M., Bolton, P., Samama, F., 2016. Hedging climate risk. Financial Analysts Journal 72 (3), 13–32.

Baker, M., Bergstresser, D., Serafeim, G., Wurgler, J., 2018. Financing the response to climate change: The pricing and ownership of us green bonds. Tech. rep., National Bureau of Economic Research.

Bank, W., 2019. 10 years of green bonds: Creating the blueprint for sustainability across capital markets.
URL https://www.worldbank.org/en/news/immersive-story/

Bender, J., Bridges, T. A., Shah, K., 2019. Reinventing climate investing: building equity portfolios for climate risk mitigation and adaptation. Journal of Sustainable Finance & Investment, 1–22.

Bovari, E., Giraud, G., Mc Isaac, F., 2018a. Coping with collapse: a stock-flow consistent monetary macrodynamics of global warming. Ecological economics 147, 383–398.

Bovari, E., Lecuyer, O., Mc Isaac, F., 2018b. Debt and damages: What are the chances of staying under the 2°c warming threshold? International Economics 155, 92–108.

Coady, D., Parry, I., Le, N.-P., Shang, B., May 2019. Global fossil fuel subsidies remain large: An update based on country-level estimates. IMF Working Papers 19 (89), 1.

Engle, R. F., Giglio, S., Lee, H., Kelly, B. T., Stroebel, J., 2019. Hedging climate change news. Available at SSRN 3317570.

Feenstra, R. C., Inklaar, R., Timmer, M. P., 2015. The next generation of the Penn World Table. American Economic Review 105 (10), 3150–82.

Filkova, M., Frandon-Martinez, C., Giorgi, A., 2019. Green bonds the state of the market 2018.

URL https://www.climatebonds.net/files/reports/cbi_gbm_final_032019_web.pdf

Gardiner, D., Associates, 2011. Physical risks from climate change: A guide for companies and investors on disclosure and management of climate impacts. Oxfam America.

GCEC, 2018. Unlocking the inclusive growth story of the 21st century: Accelerating climate action in urgent times.
URL https://newclimateeconomy.report/2018/wp-content/uploads/sites/6/2019/04/NCE$_$2018Report$_$Full$_$FINAL.pdf

Harrison, C., 2019. Green bond pricing in the primary market July-December 2018.

IAIS, 2018. Issues paper on climate change risks on the insurance sector.
URL https://www.iaisweb.org/page/supervisory-material/issues-papers/

IIGCC, 2015. Portfolio investment in a carbon constrained world.
URL https://www.iigcc.org/resource/

Institute, G. C., 2019. Carbon capture cost.
URL https://hub.globalccsinstitute.com/publications/global-status-ccs-2014/74-carbon-capture-cost

International Energy Agency, 2017. World Energy Outlook 2017.
URL https://www.oecd-ilibrary.org/content/publication/weo-2017-en

IRENA, 2019. Renewable power generation costs in 2018. International Renewable Energy Agency.

Leung, D. Y., Caramanna, G., Maroto-Valer, M. M., 2014. An overview of current status of carbon dioxide capture and storage technologies. Renewable and Sustainable Energy Reviews.

Lindenberg, N., 2014. Definition of green finance.

Liska, A., Holley, E., 2014. Climate change and its implications for the insurance industry.

McGee, M., 2019. Global carbon emissions.
URL https://www.co2.earth/daily-co2

Muûls, M., Colmer, J., Martin, R., Wagner, U. J., 2016. Evaluating the eu emissions trading system: Take it or leave it? an assessment of the data after ten years. Grantham Institute Briefing Paper 21.

Nordhaus, W., 2018. Projections and uncertainties about climate change in an era of minimal climate policies. American Economic Journal: Economic Policy 10 (3), 333–60.

OECD, 2018. Effective carbon rates 2018: Pricing carbon emissions through taxes and emissions trading.

Oliver, P., Clark, A., Meattle, C., 2018. Global climate finance: An updated view 2018. Climate Policy Initiative, November.

PCD, 2017. Portfolio investment in a carbon constrained world.
URL https://unepfi.org/pdc/wp-content/uploads/PDC-Progress-Update-2017.pdf

Pronina, L., 2019. What are green bonds and how 'green' is green? Bloomberg Businessweek.
URL https://www.bloomberg.com/news/articles/2019-03-24/what-are-green-bonds-and-how-green-is-green-quicktake

Sterner, T., Köhlin, G., 2015. Pricing carbon: The challenges, 251–266.

Stiglitz, J. E., Stern, N., Duan, M., Edenhofer, O., Giraud, G., Heal, G. M., la Rovere, E. L., Morris, A., Moyer, E., Pangestu, M., et al., 2017. Report of the high-level commission on carbon prices.

Wang, T., 2019. Largest producers of territorial fossil fuel co2 emissions worldwide in 2017, based on their share of global co2 emissions.
URL https://www.statista.com/statistics/271748/the-largest-emitters-of-co2-in-the-world/

Zerbib, O. D., 2018. The green bond premium. SSRN Electronic Journal.

Option Pricing and Hedging with Deep Learning

TEAM 4

HOLLY BRANNELLY, University College London
ROHIN JAIN, University of Cape Town
PEACE SIBANDA, University of Zimbabwe
MATTHEW SYLVESTER, University of Cape Town

Supervisor:
MIKKO PAKKANEN, Imperial College London

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

Interest in the field of artificial intelligence (AI), in particular machine learning (ML), has been growing significantly over the past decades. Developments in computer hardware have allowed researchers in this area to successfully apply these models to a wide array of problems in both industry and academia, often where the problems prove too complex for human beings. Essentially, these models provide solutions by training a computer to gather knowledge from experience, such as past data, and use that knowledge to construct and train high dimensional functions that produce optimal outputs, such as the future value of some variable or a decision. Applications range from speech, image and video recognition, language generation, learning human habits such as those of a viewer on YouTube or Netflix to provide recommendations, or forming nonlinear relationships between variables that may conventionally be modelled by traditional regression techniques. Google have been key in championing the development and use of machine learning technologies, such as the free and open source software library TensorFlow, discussed by Abadi et al. (2016), with these technologies first gaining traction with the Google Brain project (a deep learning AI research team) in 2011, and the announcement that they built a neural network that could successfully identify images of cats in 2012. Google engineer, Francois Chollet developed an open-source neural network library written in Python called Keras, the documentation of which is given by Chollet et al. (2015). The focus of Keras was to allow for user-friendly experimentation with deep neural networks.

This report focuses on an application of machine learning - deep learning or neural networks - to quantitative finance through the use of Keras. Financial applications of deep learning are gaining more traction as people explore ways that these techniques can bridge the gaps where more traditional financial models fail due to the complexity, nonlinearity and scale of data-focused problems in the financial domain. Neural networks have proven to be valuable in pricing or risk management problems, predicting market variables such as future asset prices or implied volatilities, portfolio construction, risk management and high frequency trading. Referring to and drawing upon the ideas explored by Buehler et al. (2019) throughout this study, our focus will be 'deep hedging', that is implementing a deep learning algorithm to provide hedging strategies for liquid financial instruments such as vanilla European call options. Due to the algorithms being model-free, this approach can be used to overcome the issues associated with real market factors - such as transaction costs, liquidity constraints, market impact and the risk preferences and capital constraints of a trading desk - not being taken into consideration under classical models. Additionally, the feature sets of the neural networks used to model the hedging strategies can also contain external information such as trading signals, news events, or past hedging decisions. In other words, these models

aim to produce hedging strategies that more realistically mirror what traders do when they trade, that is, as stated by Buehler (Global Head of Analytics, Automation and Optimisation at JP Morgan Chase, London) in a podcast by Quantcast (2019), ask the question "What do I need to do in order to minimise my hedging error in the sense of P&L uncertainty?".

We train a deep neural network to obtain optimal hedging strategies of European call options, first in a Black-Scholes setting, to allow for comparison of the results to the classical delta hedge that is obtained by computing the *Greeks*, and then extend this to the Heston stochastic volatility model. All the data we use to train and test the model will be obtained from Monte Carlo simulation. Initially, we use a custom loss function that accounts for producing the optimal hedge in terms of P&L uncertainty, and then we choose a loss function to account for risk preferences, captured through the risk aversion parameter in the exponential utility function. In the second of these cases, we also obtain prices for the derivatives we are hedging using the indifference pricing methodology.

Alternative studies that explore techniques for approximating the value of contracts as well as hedging strategies using deep learning include that by Shin and Ryu (2012) where artificial neural networks (ANN) are used to enhance the performance of a dynamic option hedging strategy; Aggarwal and Aggarwal (2017), where deep learning hierarchical decision models are introduced for prediction analysis in pricing securities and portfolio selection; and Doyle and Groendyke (2019), where neural networks are used to reduce the computational cost of pricing and hedging variable annuity guarantees, where the problem with using traditional pricing approaches here lies in the large number of Monte Carlo simulations, and hence computational time, required for convergence.

We begin by providing an overview to theoretical underpinnings of deep neural networks, including the backpropogation training method and the stochastic gradient descent method that we will use to update the model parameters in order to optimize the objective function. In this section, we also give an overview of the Keras library. Section 3 introduces the idea of deep hedging, making comparisons to traditional hedging techniques, and discusses the outline of the problem including the architectural structure of the network we implemented and the main challenges faced. The second half of this report is tasked with obtaining indifference prices of the financial instruments using deep hedging. Our numerical results under both the Black-Scholes and Heston models are presented and discussed in Section 5. Finally, in Section 6 we provide the reader with areas in which this project may be explored further, and in Section 7 we conclude.

2 Deep Neural Networks

In this section we introduce artificial neural networks and the theoretical underpinnings of deep learning. Throughout, we refer to Higham and Higham (2018) and Goodfellow et al. (2016). Deep learning is a branch of ML based on ANN, that is the construction of networks with the intent of modelling an animal brain. The term 'deep' stems from the fact that many layers are stacked upon each other, and the network feeds learning data from the first layer – the input layer – to the last layer - the output layer. Each layer is a set of nodes, or neurons, where the nodes are connected by some nonlinear functions. These connections are called 'edges', where each edge has a set of weights and bias parameters that the algorithm is trying to optimise over, i.e. these are the parameters that are adjusted as the learning proceeds. The overall function that maps from the inputs to the outputs, i.e. the function that the model is trying to optimally estimate, can be viewed as the composition of these intermediate functions where each application of an intermediate function provides a new representation of the input which will be passed to the next layer, and each intermediate function need not be the same. The network maps elements component wise, and at each layer, each neuron receives one real value from every neuron at the previous layer and produces one real value that is passed to every neuron in the next layer. In other words, assume the network has L layers where layer l for $l \in 1, 2, \ldots, L$ has $n_l \in \mathbb{R}^+$ nodes. The layer with $l = 1$ is the input layer, the layer with $n = L$ is the output layer, and all intermediate layers are referred to as hidden layers because the network does not show the desired output for any of these layers. If we denote the function that the network is trying to model by F^*, the feedforward network defines a mapping $F : \mathbb{R}^{n_1} \to \mathbb{R}^{n_L}$ that tries to learn the parameters of F that give the best approximation for F^*. The training data is what gives approximate examples of $\boldsymbol{y} = F^*(\boldsymbol{x})$, evaluated at each of the different training points. We refer to a feedforward network throughout, meaning that the information flows forward though the network from the inputs $\boldsymbol{x}$ to the outputs, however recurrent neural networks where the networks also contain feedback connections can also be constructed. The model can be summarised by the following:

- $x \in \mathbb{R}^{n_1}$: a vector of inputs;
- $\boldsymbol{W}^{[l]} \in \mathbb{R}^{n_l \times n_{l-1}}$: the matrix of weights at layer l, i.e. those applied to the outputs of layer $l-1$ (inputs to layer l) to produce the outputs of layer l (and hence the inputs to layer $l+1$). Element-wise, $w_{jk}^{[l]}$ is the weight applied to the value passed from neuron k from layer $l-1$ to give the output from neuron j at layer l;
- $\boldsymbol{b} \in \mathbb{R}^{[l]}$: the vector of biases at layer l. Element-wise, $b_j^{[l]}$ is the bias used by neuron j at layer l;

- $\sigma_i(x) : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{i+1}}$: the nonlinear map that takes the outputs from layer i as inputs, and produces the outputs of layer $i+1$ for $i = 1, 2, \ldots, L-1$;
- $\boldsymbol{a}^{[l]} = \sigma_{l-1}(\boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]} + \boldsymbol{b}^{[l]}) \in \mathbb{R}^{n_l}$: the output from layer l for $l = 2, 3, \ldots, L$. Element-wise, $a_j^{[l]} \in \mathbb{R}$ is the output, or *activation*, from neuron j at layer l. The output of the network will be

$$\begin{aligned}\boldsymbol{a}^{[L]} &= F(x) \\ &= \sigma_{L-1}\left(\boldsymbol{W}^{[L]}\sigma_{L-2}\left(\boldsymbol{W}^{[L-1]}\sigma_{L-3}\left(\boldsymbol{W}^{[l-2]}\ldots\sigma_1\left(\boldsymbol{W}^{[2]}x + \boldsymbol{b}^{[2]}\right)\right.\right.\right. \\ &\qquad \left.\left.\left. + \boldsymbol{b}^{[3]}\right) + \boldsymbol{b}^{[4]}\right)\ldots + \boldsymbol{b}^{[L]}\Big) \in \mathbb{R}^{n_L}.\end{aligned} \tag{1}$$

In order to train the network using the training data $\boldsymbol{x}$, we must specify a cost, or *objective*, function that is minimised to drive F to F^*. The cost function will be a function of all the parameters, that is the weights and biases, and the data points are fixed. If, for example, we were to use the mean squared error (MSE) cost function, we would look to minimise

$$C(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathbb{R}^{n_l}} \left(F^*(x) - F(x, \boldsymbol{\theta})\right)^2, \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{W}^{[2]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[2]}, \ldots, \boldsymbol{b}^{[L]})$.

Note that if $\boldsymbol{y} = F^*(\boldsymbol{x})$ is a vector of 'labels' associated to the inputs that the network is trying to train the model to target, in order to go on to predict outcomes for unforeseen data, this will be referred to as *supervised learning*. In the case where the model is left to work on its own to learn information about or patterns in the data, without being fed target labels, this is referred to as *unsupervised learning*. Unsupervised learning allows for more complex processing tasks to be done, however can also lead to more unpredictable results being produced.

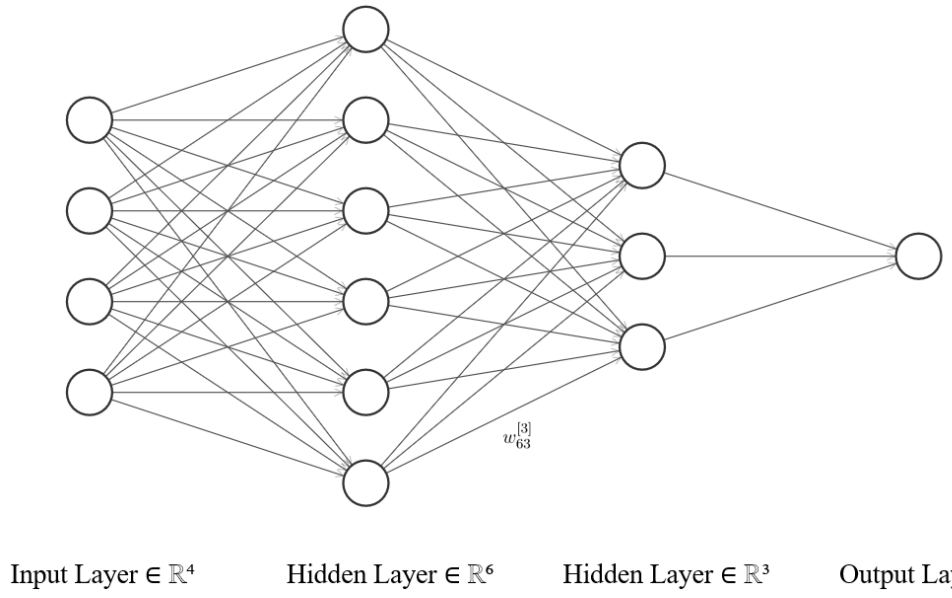


Figure 1: A network with 4 layers, with $n_1 = 4, n_2 = 6, n_3 = 3, n_4 = 1$. The edge corresponding to the weight $w_{63}^{[3]}$ is highlighted to show the weight that the output from neuron 6 in layer 2 is weighted by when is is fed into neuron 3 in layer 3.

2.1 Stochastic gradient descent

Whilst a number of optimisation algorithms may be used to train the network, we will focus on a method that is referred to as the *stochastic gradient descent* method. The aim when training the network is to choose values of the parameters that minimise the cost function. If we begin by storing the parameters as a single vector $p \in \mathbb{R}^s$ where $s = \sum_{i=2}^{L}(\dim(\boldsymbol{W}^{[i]}) + \dim(\boldsymbol{b}^{[i]}))$, instead of a set of matrices and vectors, then the algorithm computes a sequence of vectors iteratively until it converges to some vector that minimises the cost function $C(\boldsymbol{p})$. Beginning with some initial parameter vector $\boldsymbol{p}$, we want to choose some quantity $\Delta\boldsymbol{p}$ such that $C(\boldsymbol{p} + \Delta\boldsymbol{p}) < C(\boldsymbol{p})$. Taking a Taylor series expansion, we have

$$\begin{aligned} C\left(\boldsymbol{p} + \Delta\boldsymbol{p}\right) &\approx C(\boldsymbol{p}) + \sum_{r=1}^{s} \frac{\partial C(\boldsymbol{p})}{\partial p_r} \Delta p_r + \mathscr{O}\left(||\Delta\boldsymbol{p}||^2\right) \\ &= C(\boldsymbol{p}) + (\nabla C(\boldsymbol{p}))^\top \Delta\boldsymbol{p}, \end{aligned} \tag{3}$$

where $(\nabla C(\boldsymbol{p}))_r = \frac{\partial C(\boldsymbol{p})}{\partial p_r}$ denotes the first partial derivative of the cost function with respect to the r^{th} parameter. To reduce the cost function as much as possible with each iteration, it follows from Eq. (3) to make $(\nabla C(\boldsymbol{p}))_r^\top \Delta\boldsymbol{p}$ as negative as possible which, from the Cauchy-Schwarz inequality, will occur when $\Delta\boldsymbol{p}$ lies

in the direction of $-\nabla C(\boldsymbol{p})$ (whilst also ensuring that $\Delta\boldsymbol{p}$ remains small in order for the Taylor series truncation to remain relevant). This leads to updating $\boldsymbol{p}$ to $\boldsymbol{p} - \eta\nabla C(\boldsymbol{p})$ at each iteration, where η is the *learning rate*. Iterating in such a way until some stopping criterion (or computational capacity limit) is met is called the steepest descent method. The stochastic gradient descent method draws upon this, however since the cost function is a sum over all training points $\{x_i\}_{i=1}^N$, where this data set may be large, instead of computing

$$\nabla C(\boldsymbol{p}) = \frac{1}{N}\sum_{i=1}^{N} \nabla C_{x_i}(\boldsymbol{p}), \tag{4}$$

for $C_{x_i}(\boldsymbol{p})$ being the cost function evaluated at one data point $\{x_i\}$, a single training point can be chosen at random, leading to the parameter vector being updated from $\boldsymbol{p}$ to

$$\boldsymbol{p} - \eta\nabla C_{x_i}(\boldsymbol{p}).$$

The name *stochastic gradient descent* comes from the fact that the training point used is chosen at random from the full training set at each iterative step. Note that the training point $\{x_i\}$ can be sampled with replacement, so that it is possible for the same training point to be selected again with the same probability of any other training point being selected, or without replacement. If selected without replacement, after N iterations the algorithm will have cycled through the entire training set (note, this should be done in a random order), which is referred to as completing an *epoch*.

Alternatively, the data may also be organised into k *mini-batches* of size m, where $k \times m = N$, and the parameter vector update from $\boldsymbol{p}$ will become

$$\boldsymbol{p} - \eta\frac{1}{m}\sum_{i=1}^{m} \nabla C_{x_{k_i}}(\boldsymbol{p}).$$

It is important in this method to ensure that the learning rate η is chosen appropriately, as too high a learning rate may cause the learning to jump over minima in the cost function, and too low a learning rate may lead to it taking too long to converge or getting stuck in a local minima. It is always the case that $\eta < 1$, and the 'sgd' optimizer in the Keras library takes $\eta = 0.01$. Alternatively, and often something that is necessary in practice, the learning rate can be decreased over time in order to reduce the noise induced from the random sampling from the training data as the algorithm arrives at a minimum. If we denote the learning rate at iteration k by η_k, Algorithm 1 shows how we can apply the mini-batch stochastic gradient descent

method. The conditions

$$\sum_{k=1}^{\infty} \eta_k = \infty, \tag{5}$$

$$\sum_{k=1}^{\infty} \eta_k^2 < \infty, \tag{6}$$

are sufficient to guarantee convergence of the SGD method under this setting.

Algorithm 1 The mini-batch SGD parameter update method:

1: Inputs: $\eta_1, \eta_2 \ldots, \boldsymbol{p}$
2: **while** stopping condition not met **do**
3: Sample a minibatch of m samples from the training set $\{x_1, x_2, \ldots x_m\}$
4: Compute $\hat{\boldsymbol{g}} \leftarrow \frac{1}{m} \sum_{i=1}^{m} \nabla C_{x_i}(\boldsymbol{p})$
5: $\boldsymbol{p} \leftarrow \boldsymbol{p} - \eta_k \hat{\boldsymbol{g}}$
6: $k \leftarrow k + 1$
7: **end while**

2.2 Backpropagation

As discussed before, a feedforward NN propagates information from the input nodes forward through the network to the output layer. In doing so, information about the error associated to each parameter update is also *backpropagated* through the network in order to adjust the parameters accordingly, i.e. in the direction of 'less error', where this direction is determined by the SGD method discussed in the previous section. Firstly, to implement the SGD method to train the network, we need to compute the partial derivatives of the cost function with respect to the parameters, that is to obtain $\nabla C(\boldsymbol{p})$. This technique exploits the fact that the cost function is a linear combination of terms involving each of the training points, and hence the individual partial derivatives with respect to each of the parameters will also be. To obtain expressions for these partial derivatives, it therefore suffices to consider the cost function at one fixed training point, and then compute the partial derivatives which will hold true for all other training points. Note that the only dependence of the cost function on the weight and bias parameters will be in the term involving the output of the ANN,

$$\boldsymbol{a}^{[L]} = \sigma_{L-1}(\boldsymbol{W}^{[L]}\boldsymbol{a}^{[L-1]} + \boldsymbol{b}^{[L]}) = \sigma_{L-1}(\boldsymbol{z}^{[L]}), \tag{7}$$

where we define $\boldsymbol{z}^{[l]} := \boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]} + \boldsymbol{b}^{[l]}$ for $l = 2, 3, \ldots, L$. Since the cost function can only be at a minimum if all partial derivatives are zero, this motivates referring to the expression given by

$$\delta_j^{[l]} = \frac{\partial C}{\partial z_j^{[l]}}, \tag{8}$$

for $1 \leq j \leq n_l$ and $2 \leq l \leq L$ as the error of the j^{th} neuron at layer l, i.e. the term we want to target a zero value for. Using the chain rule, we can summarise the errors and partial derivatives of the cost function with respect to the parameters as follows:

$$\boldsymbol{\delta}^{[L]} = \sigma'_{L-1}\left(\boldsymbol{z}^{[L]}\right) \circ \left(\boldsymbol{a}^{[L]} - \boldsymbol{y}\right), \tag{9}$$

$$\boldsymbol{\delta}^{[l]} = \sigma'_{l-1}\left(\boldsymbol{z}^{[l]}\right) \circ \left(\boldsymbol{W}^{[l+1]}\right)^{\top} \boldsymbol{\delta}^{[l+1]}, \qquad 2 \leq l \leq L-1, \tag{10}$$

$$\frac{\partial C}{\partial b_j^{[l]}} = \delta_j^{[l]}, \qquad 2 \leq l \leq L, \tag{11}$$

$$\frac{\partial C}{\partial w_{jk}^{[l]}} = \delta_j^{[l]} a_k^{[l-1]}, \qquad 2 \leq l \leq L. \tag{12}$$

The proof is given by Higham and Higham (2018). The gradients in Eqs. (9)-(12) are then computed using backpropogation as follows:

1. $\boldsymbol{a}^{[L]}$ is evaluated from a forward pass through the network, i.e. by computing $\boldsymbol{a}^{[1]}, \boldsymbol{z}^{[2]}, \boldsymbol{a}^{[2]}, \boldsymbol{z}^{[3]}, \ldots, \boldsymbol{a}^{[L]}$ in order;
2. $\boldsymbol{\delta}^{[L]}$ is then obtained from Eq. (9);
3. $\boldsymbol{\delta}^{[L-1]}, \boldsymbol{\delta}^{[L-2]}, \ldots, \boldsymbol{\delta}^{[2]}$ are then computed from Eq. (10) using a backward pass;
4. The partial derivatives of the cost function with respect to the parameters in Eqs (11) and (12) follow once we have the $\boldsymbol{\delta}^{[i]}$ and $\boldsymbol{a}^{[i]}$ for $1 \leq i \leq L-1$.

The pseudocode for this method is given in Algorithm 2.

Algorithm 2 The backpropagation method to train a neural network:

$N \leftarrow$ number of iterations
for $n = 1$ to N **do**
    Choose k randomly from $\{1, 2, \ldots, N\}$
    Draw a data point $\{x_k\}$ from the training data

    for $l = 2$ up to L **do**
        $\boldsymbol{z}^{[l]} - \boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]} + \boldsymbol{b}^{[l]}$
        $\boldsymbol{a}^{[l]} = \sigma(\boldsymbol{z}^{[l]})$
        $\boldsymbol{D}^{[l]} = \text{diag}(\sigma'(\boldsymbol{z}^{[l]}))$
    end for

    $\boldsymbol{\delta}^{[L]} = \boldsymbol{D}^{[L]}\left(\boldsymbol{a}^{[L]} - y(x_k)\right)$

    for $l = L - 1$ down to 2 **do**
        $\boldsymbol{\delta}^{[l]} = \boldsymbol{D}^{[l]}\left(\boldsymbol{W}^{[l+1]}\right)^{\top}\delta^{[l+1]}$
    end for

    for $l = L$ down to 2 **do**
        $\boldsymbol{W}^{[l]} \leftarrow \boldsymbol{W}^{[l]} - \eta\boldsymbol{\delta}^{[l]}\boldsymbol{a}^{[l-1]\top}$
        $\boldsymbol{b}^{[l]\leftarrow\boldsymbol{b}^{[l]}-\eta\boldsymbol{\delta}^{[l]}}$
    end for

end for

Before concluding this section, we make a note on the fact that when using SGD and backpropagation, it is not guaranteed that the minimum of the cost function will be a global minimum, but possibly only a local minimum, due to the non-convexity of error functions used in neural networks. Whilst this may appear to be a significant drawback to the technique, it is shown by LeCun et al. (2015) not to be the case.

2.3 The Keras library

As mentioned previously, Google have been key in developing machine learning technologies, one of these being the open-source NN library written in Python called Keras – developed by Chollet et al. (2015). This library is capable of running over TensorFlow, Theano or PlaidML, and was built with the intent of making deep learning an accessible area to explore. The idea here was to break down the barriers, that arise due to the complexity of NN and deep learning frameworks, that may act as a deterrent to researching in this area to some.

Due to the Keras library, creating deep learning networks that can be applied to a variety of applications has been made extremely accessible and fast. The library can be used to construct both convolutional and recurrent networks, as well as combinations of the two, and can run on CPU and GPU.

The core data structure in this library is a *model* which gives the way in which the layers are stacked, The simplest type of the model is the *Sequential model*, where the layers are stacked linearly on top of each other. If one wants to build a model that does not adhere to this structure, i.e. the architecture is more complex and gives rise to more flexibility, the *Functional API* can be used instead of the Sequential one.

We highlight the differences between the Functional API and the Sequential API in what follows.

Sequential API: model layers are stacked linearly on top of each other, making this a simple way for a user to construct straightforward neural networks. An ordered list of the layers is passed to the model, where the user specifies the dimension of each layer, as well as the activation function used as the data passes through the nodes at that layer.

Functional API: this allows for more complex models to be built, such as those that use multi-inputs, multi-outputs or shared layers. Much like layers, a functional model is callable on a tensor (hence why this setting allows for multiple inputs), and both the architecture and weights and bias parameters are used when the model is called. As a result, these models can process sequences of inputs (e.g. an image classification model can be developed to a video classification model), as well as share layers across different inputs to create a graph-like model.

Once a model and the architecture of its layers has been specified the model must be compiled. To do so, an optimiser and a loss function need to be chosen and fed as inputs to the **model.compile()** line of code. The model is then fitted to the input data using **model.fit()**. Testing data can then be input into **model.evaluate()** to evaluate the performance of the model on new data that it has not been trained on.

More detail and examples on these types of models in Keras is given in the extensive Keras library documentation by Chollet et al. (2015).

3 Deep Hedging

The first problem we address in this project is that of using a NN to obtain optimal delta hedging strategies. We begin by giving a brief summary of traditional delta hedging under the Black-Scholes framework, as in this setting we will use these 'theoretical' delta hedges as a benchmark for validating the delta hedges obtained from our NN. Deep hedging was introduced by Buehler et al. (2019) as a way to hedge derivatives exclusively with deep learning techniques, instead of using classic replication techniques. One of the main motivations for doing so lies in the fact that ML technologies can be used as a platform to hedge in incomplete markets (i.e. in a more realistic setting), which is something that has previously been viewed as inaccessible to do. Additionally, as traditional hedging techniques were developed when data was very limited, the vast relative amount of data available today motivates the exploration into alternative hedging techniques that can exploit the availability and scope of this data. Buehler (in a podcast by Quantcast (2019)) explains that under traditional hedging techniques, traders often overwrite or make modifications to the strategies given to them by these classical models to fix the small, but relevant, issues that aren't accounted for - such as transaction costs, external market factors or constraints specific to the trading desk, whether these be liquidity constraints or risk limits on the book. By implementing a deep hedging strategy, the machine can figure out these 'overwrites' for the trader based off of the input data, making the hedging process independent of human inputs and hence more efficient. Additionally, since the process is very data driven, there is no need to define an underlying model which can allow it to be automated quicker.

Since deep hedging is still in the relatively early stages of being developed and implemented in industry, however, one of its main limitations is that it can only effectively be applied to hedge products that are relatively liquid. Additionally, the amount of instruments trading in the market relative to the amount of times series data available is large, and as this is a data driven approach, practitioners implementing these techniques may need to find ways of simulating distributions that account for the sparsity, whilst also taking tail events into consideration. The data used to train models must also be clean, as small errors in the data may cause the machine to become unstable, however also have the ability to signal when there is an event coming up that might have a big impact on the market, for example, to ensure that the machine can be controlled to pick up these events and account for them in the output hedging strategy. Buehler also states that going forward, as more trading desks begin to adopt this strategy to hedge their positions, the trading teams must be required to have a minimum level of skill to handle and understand the machines, instead of just taking the outputs at face value. For instance, if the machine produces an answer, the trader needs to know roughly how and why it produced that output and whether they agree with it, and if not how they can alter

the objective function accordingly to both improve the machine, and ensure that there is no large financial implication from implementing a 'disagreeable' hedging strategy.

3.1 Classical delta hedging

The well-known Black-Scholes-Merton (BSM) model of Black and Scholes (1973), that is widely implemented in industry, provides a framework to price and hedge options using a risk-neutral argument under the assumptions of no arbitrage and thus market completeness. One of the key ideas in this model is that by buying and selling the underlying asset continuously over the life of an option in some predetermined quantity will allow an agent to hedge the option or, in other words, eliminate risk exposure associated with price movements in the underlying asset. The delta therefore represents the sensitivity of the value in the option to a unit change in the market price of the underlying asset. For example, if a trader has a long position in a European call option with $\Delta = 0.5$, they must short 0.5 units of the underlying in order to have a delta neutral position.

Let $(\Omega, \mathcal{F}, \mathbb{Q}, (\mathcal{F}_t)_{t\geq 0})$ be a filtered probability space where $\mathbb{Q}$ is the martingale measure and $(W_t)_{t\geq 0}$ is a Brownian motion on the probability space. Assume there exists some asset with price process dynamics $(S_t)_{t\geq 0}$ governed by the SDE

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t,$$

for $r \in \mathbb{R}$ the risk-free rate and $\sigma \in \mathbb{R}^+$ the volatility of the asset. Let $(C_t)_{t\geq 0}$ denote the price process of a vanilla European option written on the underlying asset (S_t) with strike prices $K > 0$ and maturity $0 < T < \infty$. The delta of this option can be expressed in closed form as $\delta_t = \Phi(d_1)$ if the option is a call option, and $\delta_t = \Phi(d_1) - 1$ if the option is a put option, where $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution, and

$$d_1 = (\log(S_t/K) + (r - 0.5\sigma^2)(T-t))/\sigma\sqrt{T-t}.$$

Although the assumptions in the BSM setting that are required to obtain these delta hedges are restrictive, and largely what the deep hedging techniques are trying to overcome, this framework is still widely used in financial institutions (perhaps due to the lack of alternative, yet still as easily implemented, hedging techniques available) and so we will use these 'theoretical' deltas as a benchmark to compare our results to in the following sections.

3.2 The problem: part 1

In this section, we outline our approach in tackling the first part of problem that this report addresses:

Implement and train a deep neural network to optimally hedge European options.

As discussed previously, this hedging technique has the advantage of being model-free, and trained purely using market data. However, since the data we will use throughout will be synthetic, we begin by simulating data under the BSM framework, and then extend this to the Heston Stochastic Volatility model, in order to compare our results with those from these classic models. All numerical results are presented in Section 5.

We consider a European call option with payoff at maturity $C_T = (S_T - K)^+$, and want to take observed prices of the underlying over the life of the option, $\{S_0, S_1, \ldots, S_t, \ldots S_T\}$ as inputs to the neural network to produce outputs that give us the optimal hedging strategy for the option at each instance in time, that is $\{\delta_0, \delta_1, \ldots, \delta_t, \ldots, \delta_{T-1}\}$ where δ_i denotes the amount of stock that must be held at time $0 \leq i < T$ to be delta neutral. Referring to Section 2, the aim of a NN is to optimally estimate some nonlinear function of the input data which here would be the function f_k that gives an optimal hedge $\delta_k = f_k(S_k, S_{k-1}, \ldots, S_0)$ at time $0 \leq k < T - 1$. However, since we have Markovianity in the BSM setting, we can consider the more efficient approach of finding the NN that gives the following delta values as the optimal hedging strategy:

$$\delta_0 = f(0, S_0) \tag{13}$$
$$\delta_1 = f(1, S_1) \tag{14}$$
$$\vdots \tag{15}$$
$$\delta_{T-1} = f(T-1, S_{T-1}) \tag{16}$$

i.e. the inputs used to produce the delta hedge at each instance in time include the time-step that we are at, and the stock price at that time.

Making use of the Keras functional API, discussed in Section 2.3, allows us to construct a multi-input and multi-output model to treat the model as a layer, calling it on each of the input tensors at each time point in order to obtain the output set of delta hedges over all times, This ensures that each time we call the model, both the architecture and its weights are reused, as the function f in Eqs. (13)-(16) is the same for each of the input tensors $\{[0, S_0], \ldots [T-1, S_{T-1}]\}$.

Since the idea of deep hedging is to find the optimal hedge accounting for P&L uncertainty, and the P&L of our position at maturity (sell the option for C_0, and trade in the market to keep a delta neutral position over the life of the option) is given by

$$V_T + C_0 - (S_T - K)^+ \tag{17}$$

where

$$V_T = \sum_{k=1}^{T} \delta_{k-1} \left(S_k - S_{k-1} \right), \tag{18}$$

we use a custom loss function given by

$$L(x, y) = \left(\sum_{k=1}^{T} y_{k-1} x_{k-1} + C_0 - (x_T - K)^+ \right)^2. \tag{19}$$

3.2.1 Architectural structure of the network

As mentioned in the above section, our network will not use the more commonly used Sequential model available in the Keras library, but rather the functional API. To reiterate, this allows us to maintain the structure, weights and biases applied to each input tensor $[i, S_i]$ to find the optimal f, rather than modelling multiple functions f_i for $i = 0, 1, \ldots, T-1$, each by a sequential model, to obtain each δ_i separately. This approach not only provides more intuitive delta hedges, but also significantly reduces the number of parameters that the model must estimate, thus improving the computational speed at which the model will run.

We propose a multi-input model where each input array has dimension 2, namely the time-step and the price of the underlying at that time. Within the model there will be 3 hidden layers, each containing 100 neurons. Finally, the last layer will consist of one neuron, which will output the optimal delta hedge for each time period. The structure is illustrated in Figure 2, however for ease of illustration we have shown 5 nodes per hidden layer instead of 100. The activation function we chose at layers 1,2 and 3 is the 'ReLU' activation function, and for layer 4 the 'sigmoid' activation function – this ensures our output delta will lie between 0 and 1. We concluded on this network structure after trying a few combinations of number of layers and nodes per layer and comparing the results, however if time constraints were not under consideration, we could run a more extensive study into the most effective structure for the neural network.

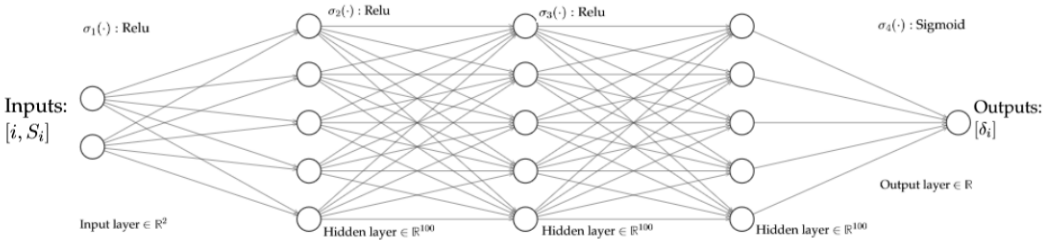


Figure 2: The structure of the neural network that we will train to hedge European call options.

3.2.2 The main challenges

The main challenges in approaching this problem lie in constructing a non-standard NN architecture using the Keras functional API instead of the more straightforward and frequently used Sequential model, as well as the need to code a custom loss function. Often in deep learning regression problems, built-in loss functions such as the mean squared error (MSE) loss are used. We also chose to run an unsupervised learning model where, although we know the theoretical deltas under the BSM framework, we did not feed these to the model as targets, instead we let the network learn the optimal deltas only using the input stock data.

We also note that deep hedging is a relatively novel concept, and hence the scope of available literature that can be used to guide this research, as well as to compare our results to, is small. Finally, as with any computational project, one must bear in mind the limits of the computational resources at hand.

4 Indifference Pricing using Deep Hedging

We now introduce indifference pricing theory in order to address the second half of the problem this report focuses on – that is, pricing options via deep hedging. In what follows, we refer to Buehler et al. (2019) throughout, and begin by giving some definitions that will play a role in defining the indifference price of a financial instrument.

Definition 4.1. Let $X_1, X_2 \in \mathcal{X}$ be asset positions (i.e. $-X_i$ is a liability for $i = 1, 2$). Then we define $\rho : \mathcal{X} \to \mathbb{R}$ to be a convex risk measure if it is:

1. Monotone decreasing: if $X_1 \leq X_2$ then $\rho(X_1) \geq \rho(X_2)$;
2. Convex: $\rho(\alpha X_1 + (1-\alpha)X_2) \leq \alpha\rho(X_1) + (1-\alpha)\rho(X_2)$ for $\alpha \in [0, 1]$;
3. Cash-Invariant: $\rho(X + c) = \rho(X) - c$ for $c \in \mathbb{R}$.

Intuitively, we can interpret the convex risk measure $\rho(X)$ as the minimum amount of cash required to be added to taking the asset position X in order for that position to be made 'acceptable' to an investor.

So the first part of Definition 4.1 implies that the greater the asset position taken by an investor, the less the amount of capital that must be added to make the position acceptable will be. Cash-invariance in this setting can be understood as follows: if an investor has an asset position X, then they will require a greater amount of capital added to this position for it to be 'acceptable' than if they had an asset position $X + c$. The difference in the amount of capital required is simply the difference in the asset positions, $c \in \mathbb{R}$.

Assume for some convex risk measure ρ, we define

$$\pi(X) := \inf_{\delta \in \mathcal{H}} \rho\left(X + (\delta \cdot S)_T\right) \tag{20}$$

where δ is a trading strategy in an asset with price process $(S_t)_{t \geq 0}$ and $\mathcal{H}$ is the set of admissible trading strategies. Here, $\pi(X)$ can be interpreted as the minimum additional amount of capital required to make the terminal position when taking an asset position X and adopting a trading strategy δ, i,e, $X + (\delta \cdot S)_T$, acceptable to an investor, i.e. the 'certainty equivalent' of that asset position.

To proceed in introducing the idea of an indifference price, consider the following two scenarios:

1. An agent sells a call option with payoff Z for initial price p_0, and has the position $-Z + p_0$ (assuming interest rates are 0 for simplicity) at maturity of the option;

2. The agent does not sell the call, and has no terminal position.

The indifference price of the call option is the value of p_0 such that we have $\pi(-Z + p_0) = \pi(0)$ where $\pi(x)$ is given by Eq. (20). Buehler et al. (2019) prove that $\pi(x)$ satisfied the cash-invariant property in Definition 4.1, and so it holds that

$$\pi(-Z + p_0) = \pi(0) \Rightarrow \pi(-Z) - p_0 = \pi(0) \tag{21}$$

and so the indifference price of the option is given by $p_0 = \pi(-Z) - \pi(0)$.
We now want to incorporate the risk preferences of an agent into our deep hedging problem, where the option we are hedging has payoff Z, and begin by assuming that an agent has utility represented by the exponential utility function $U(x) = -\exp(-\lambda x)$. To obtain the indifference price of this option using Eq. (21), we need to estimate $\pi(Z)$ using the NN. The paper demonstrates that under the assumption of an exponential utility function with risk aversion parameter λ for the agent trading in this market, the network will approximate $\pi(-Z)$ by

$$\pi^M(-Z) = \frac{1}{\lambda} \log \inf_{\theta \in \Theta} J(\theta), \tag{22}$$

where

$$J(\theta) = \mathbb{E}[e^{-\lambda(-Z + (\delta^\theta \cdot S)_T)}]. \tag{23}$$

It is shown that Eq. (22) converges to $\pi(-Z)$ as we increase the number of layers in the network.
To solve for the indifference price, we need an expression for $\pi(0)$.

Proposition 4.2. In the above setting, it holds that $\pi(0) = 0$.

Proof. When $\pi(0) = 0$, Eq. (23) becomes

$$J(\theta) = \mathbb{E}[e^{-\lambda((\delta^\theta \cdot S)_T)}]. \tag{24}$$

The trading strategy $(\delta^\theta \cdot S)_T$ can be approximated by $\int_0^T \delta_u dS_u$. Define

$$\begin{aligned} X_T &= -\lambda \int_0^T \delta_u dS_u, \text{ and the underlying asset} \\ \text{dynamics by } S_t &= S_0 \, \mathrm{e}^{\sigma W_t + (\mu - \frac{\sigma^2}{2})t}, \\ \text{then let } f(X_t, t) : &= \exp\left(-\lambda \int_0^t \delta_u dS_u\right). \end{aligned}$$

By application of Ito's Lemma:

$$\begin{aligned} \mathrm{d}(f(X_T, T)) &= \mathrm{e}^{X_T} \mathrm{d}X_T + \frac{1}{2} \mathrm{e}^{X_T} (\mathrm{d}X_T)^2, \\ \mathrm{e}^{X_T} &= \mathrm{e}^{X_0} + \int_0^T \mathrm{e}^{X_u} \mathrm{d}X_u + \frac{1}{2} \int_0^T \mathrm{e}^{X_u} \mathrm{d}\langle X \rangle_u \end{aligned}$$

but

$$\begin{aligned} \mathrm{d}\langle X\rangle_u &= \lambda^2\delta_u^2\mathrm{d}\langle S\rangle_u, \\ \mathrm{d}\langle S\rangle_u &= \sigma^2 S_u^2 \mathrm{d}u, \end{aligned}$$

and hence

$$\mathrm{e}^{X_T} = \mathrm{e}^{X_0} + \int_0^T \mathrm{e}^{X_u}\mathrm{d}X_u + \frac{1}{2}\int_0^T \lambda^2\delta_u^2\sigma_u^2 S_u^2 \mathrm{e}^{X_u}\mathrm{d}u.$$

However, since X_T is a martingale then:

$$\begin{aligned} \mathbf{E}[\int_0^T \mathrm{e}^{X_u}\mathrm{d}X_u] &= 0, \\ \implies \mathbf{E}[\mathrm{e}^{X_T}] &= \mathbf{E}[\mathrm{e}^{X_0}] + \mathbf{E}[\int_0^T \mathrm{e}^{X_0}\mathrm{d}X_u] + \mathbf{E}[\frac{1}{2}\int_0^T \mathrm{e}^{X_u}\lambda^2\delta_u^2\sigma_u^2 S_u^2\mathrm{d}u], \\ &= 1. \end{aligned}$$

The last term is zero since we are choosing the δ that minimises J, which in this case is zero. It follows that $\pi(0) = 0$. □

4.1 The problem: part 2

In this section, we outline the second part of the problem that this report addresses:

Use the indifference pricing framework to obtain prices for European call options via deep hedging.

The optimisation problem that we want to train the network to solve now becomes that in Eq. (20), i.e. finding the hedging strategy $\delta \in \mathcal{H}$ that minimizes $J(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the vector of weight and bias parameters in the network. This can again be solved using the stochastic gradient descent method and backpropagation algorithm, and we keep the structure of our network the same as in Section 3.2 however change the loss function so that our network now minimises over $J(\boldsymbol{\theta})$. As mentioned previously, under the exponential utility setting, this reduces to finding a minimum of Eq. (23), and so we replace our custom loss function by

$$L(x,y) = \exp\left(-\lambda\left(-(x_T - K)^+ + \sum_{k=1}^{T} y_{k-1}x_{k-1}\right)\right) \tag{25}$$

Since we are dealing with the exponential function here, the network does not encounter any problems differentiating this loss function, however under alternative assumptions one must ensure that $\triangledown J(\boldsymbol{\theta})$, i.e. the partial derivatives of J with respect to each of the weight and bias parameters, can be calculated efficiently.

Once the network solves the optimisation problem and obtains the output hedging strategy

$$\boldsymbol{\delta} = \{[\delta_0], [\delta_1], \ldots, [\delta_{T-1}]\},$$

we can calculate the indifference price of the option using Eq. (22) and Proposition 4.2.

5 Numerical Results

In this section, we present the results obtained from using our deep NN to approach the problems discussed in Sections 3.2 and 4.1

5.1 Data simulation

In order to train our network we needed a large amount of data. For simplicity, and as the main challenge was to build and train a neural network with no specific relevance to financial products trading in some specific market, we decided to use synthetic data throughout. This also allows us to check the results produced from our NN with theoretical benchmarks based off this simulated data. We generated Monte Carlo stock price observations first under the BSM model, and then under the Heston Stochastic Volatility model. The stochastic differential equations (SDEs) for these two models are given by

$$\mathrm{d}S_t = rS_t\mathrm{d}_t + \sigma S_t \mathrm{d}W_t \tag{26}$$

and

$$\begin{aligned} \mathrm{d}S_t &= rS_t\mathrm{d}_t + \sqrt{V_t}S_t\mathrm{d}W_t^s \\ \mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \epsilon\sqrt{(V_t)}\mathrm{d}W_t^v, \end{aligned} \tag{27}$$

respectively, where $(W_t)_{t\geq 0}, (W_t^s)_{t\geq 0}$ and $(W_t^v)_{t\geq 0}$ are Brownian motions, with

$$\mathrm{d}W_t^s \mathrm{d}W_t^v = \rho \mathrm{d}t$$

for $\rho \in [-1, 1]$. We did not calibrate our models to data as this was not a focus of the problem, so instead we just chose intuitive values for the parameters. In further work, how the output of the NN changes for different parameters could also be explored. The parameters we chose are as follows:

Parameter	Meaning	Value
r	interest rate	0
σ	BSM volatility	0.1
κ	rate of mean reversion	1.5
θ	the long run variance i.e. mean reversion variance	0.1
ϵ	the "vol of vol"	0.2
ρ	the correlation between the Brownian motions that drive the stock process and the stochastic volatility process	-0.2

We note that throughout, we simulate the data under the martingale measure to ensure that the network does not learn the drift trend, that is, calculate the hedging strategy that trades on this trend. Additionally, since the focus is not on finding the most efficient discretisation schemes for an SDE, for simplicity we used the Euler-Maruyama scheme to simulate from the Heston model, instead of a more complicated higher order scheme, however making use of the 'truncation scheme' shown by Bégin et al. (2015) here to ensure that the stochastic volatility process does not become negative for any $0 \leq t < T$.

To introduce the Euler scheme for an Ito process (Y_t) that satisfies the SDE

$$\mathrm{d}Y_t = \mu\left(t, Y_t\right)\mathrm{d}t + \sigma\left(t, Y_t\right)\mathrm{d}W_t \tag{28}$$

on some time interval $t \in [t_0, T]$ for $0 \leq t_0 < T < \infty$ we refer to Kloeden and Platen (2013).

Definition 5.1. For a given discretisation $t_0 = \tau_0 < \tau_1 < \ldots < \tau_n < \ldots < \tau_N = T$ of the time interval $[0, T]$, and some Ito process (Y_t) that satisfies the SDE (28), a Euler approximation is a continuous time stochastic process $X = \{X(t),\ t_0 \leq t \leq T\}$ satisfying the iterative scheme

$$X_{n+1} = X_n + \mu\left(\tau_n, X_n\right)\left(\tau_{n+1} - \tau_n\right) + \sigma\left(\tau_n, X_n\right)\left(W_{\tau_{n+1}} - W_{\tau_n}\right), \tag{29}$$

for $n = 0, 1, \ldots, N-1$, with initial value $X_0 = Y_0$ and where $X_n = X(\tau_n)$ denotes the value of the approximation at the discretisation time τ_n.

We refer to the time discrete approximation in Eq. (29) as X^{δ} where $\delta = \max_n\left(\tau_{n+1} - \tau_n\right)$. Further details of the Euler scheme as well as the conditions for strong and weak convergence and consistency are discussed in Kloeden and Platen (2013).
It follows that the Euler scheme for this model is given by

$$\begin{aligned} V_{n+1} &= V_n + \kappa(\theta - V_n)\left(\tau_{n+1} - \tau_n\right)\epsilon\sqrt{V_n}\left(W^v_{\tau_{n+1}} - W^v_{\tau_n}\right), \\ S_{n+1} &= rS_n\left(\tau_{n+1} - \tau_n\right) + \sqrt{V_n}S_n\left(W^s_{\tau_{n+1}} - W^s_{\tau_n}\right), \end{aligned} \tag{30}$$

where the notation follows that given in Definition 5.1. Once the Heston stock prices are generated with the required parameter inputs, a useful way to check the simulated values and indeed the correctness of how the algorithm has been implemented is to generate BSM stock prices under the same parameters, however taking the variance parameter here to be the long run variance in the second SDE shown in Eq. (30). Intuitively, we would expect the distribution of simulated stock prices under these two models to be somewhat similar, with the Heston

stock prices simply having a wider distribution around the mean value of the process at any given terminal time. This could be represented by a histogram, and we would expect the simulated Heston stock prices to show fatter tails and a lower peak when compared to the BS prices. The intuition here is that the stochastic volatility element to the Heston model adds further volatility to our stock process and thus generates a wider dispersion of stock prices at all times when compared to the BSM stock prices.

5.2 Black-Scholes framework

5.2.1 Deep hedging

To obtain the optimal delta hedging strategy under the BSM framework, we use the Monte Carlo BSM synthetic stock prices (with the parameters given in Section 5.1) as input to the neural network discussed in Section 3.2. Since the structure of our NN (shown in Figure 2) gives 20,601 parameters, we train the model using 100,000 sample paths, that is 100,000 realisations of each input training data array, This is because if the number of samples in the input data is low relative to the number of model parameters being estimated, the model runs the risk of overfitting. We take 100 normalised time-steps of the input stock data, which will produce 100 delta values representing the optimal hedge at each of those times. For the input layer and hidden layers, we chose 'ReLU' (rectified linear units) for the activation function, which is given by

$$\sigma(x) = \max(x, 0) \tag{31}$$

and is seen to be one of the simplest nonlinear activation functions that results in fast training for larger networks. On the output layer, we chose a sigmoid activation function, given by

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{32}$$

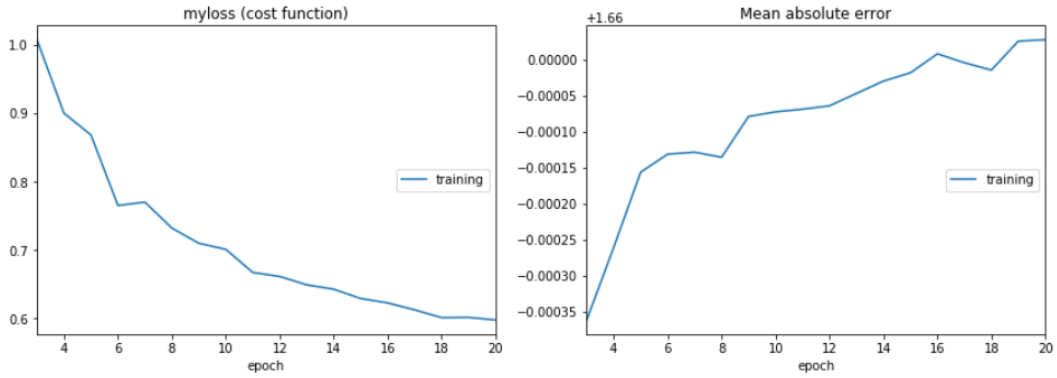
so to produce our delta output between 0 and 1.

Since we want to find the optimal delta hedge in terms of P&L uncertainty, we wrote a custom loss function to account for this – the loss function we used is shown in Eq. (33). Typically, the loss function will be a function of two inputs, 'yTrue' and 'yPred', where these will be the 'targets' and 'outputs' that are driven to those targets, respectively. If the 'yTrue' input to the loss function when training the model involved a target array of data separate to the input data, this will be an example of supervised learning, as the model is guided toward the correct values that it must drive 'yPred' to by altering the parameters. In our case, we take the increments in the input data stock prices as the 'yTrue' vector, and hence our model does not entirely fall under the category of supervised learning as there is no data besides the input data being used to train it. It is important to note that the loss

function must be written using the Keras backend tensor manipulation library to ensure that the model can obtain derivatives of the functions involved, as this loss function is the function that will be minimised by the model using the SGD and backpropagation techniques, and the output deltas will be the ones that give this minimum.

We chose to use the 'Adam' optimizer, developed by Kingma and Ba (2014), which is similar to SGD in the sense that is is an algorithm for gradient-based optimisation of stochastic objective functions, however this method computes adaptive learning rates for each parameter estimate instead of keeping the learning rate constant throughout and so one of the advantages is that the magnitude of the parameter updates at each iteration are invariant to the re-scaling of the gradient. This can help prevent the changes in parameter estimates as we approach the minimum of the loss function being large enough to result in the estimate 'jumping over' or missing the minimum. Pseudocode for the algorithm is given in Algorithm 1 in the paper, and its convergence is discussed by Reddi et al. (2019).

The code used to build and train the network to obtain an optimal delta hedging strategy in this setting is given in Appendix 8.1. After fitting the network using the input data in the 'stockdata' array, as shown in the code, we simulated 10000 new BSM sample paths using the same parameters and over the same number of time-steps to test the model, and compared the outputs to the theoretical deltas at each time-step. We simulated 10000 paths of each stock starting at values ranging between 10 and 220, in increments of 10, keeping the strike that the option written on these stocks has at 100, so that we can test how well our model predicts delta hedging strategies for out-the-money, at-the-money and in-the-money call options. We monitor the value of the loss function and the mean absolute error (MAE) at each epoch whilst testing the model, as shown in Figure 3.



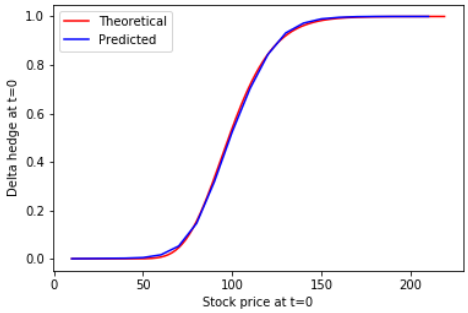

```
myloss (cost function):
training   (min:    0.597, max:   13.007, cur:    0.597)

Mean absolute error:
training   (min:    1.659, max:    1.664, cur:    1.660)
<keras.callbacks.History at 0x1aa30a10a58>
```

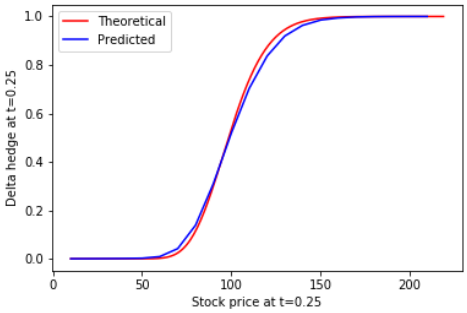
Figure 3: The value of the loss function and the mean absolute error (MAE) metric at each epoch whilst training the model.

We also compute the theoretical Black-Scholes delta at each time-step and use these values as a benchmark to assess the performance of our model. Since we are using synthetic data simulated under the assumptions of the Black-Scholes model, and have not included external factors such as transaction costs in our set-up, we would expect the theoretical and model output deltas to be close. This is also due to the fact that in the BSM framework, as discussed in Section 3.1, the delta is given by $\Phi(d_1)$, i.e, the function that the model is trying to learn before including any external factors is fairly simple. Figure 4 shows plots of the output delta hedges from the network against the BSM theoretical delta hedges, which are functions of the stock price and underlying parameters, at four different time steps $t = 0, 0.25, 0.5, 0.99$ where maturity is taken as $T = 1$ and $dt = 0.01$. We see that the network predicts the BSM delta hedge very accurately at early time-steps, however as $t \to T$ there is much less agreement between these two delta curves. As the option approaches maturity, in the BSM setting the delta hedge as a function of the underlying stock price approaches a step function with the step at the strike price, as shown by the red curve in Figure 4d. The reason for the NN producing a less agreeable delta hedge to the theoretical one here may be that as the function the network is trying to learn approaches a discontinuous function, the network is less capable of extracting this pattern from the input data. Additionally, in Figure 5 we plot the distribution of the P&L of the trading strategy that uses the delta hedges produced by the NN against the BSM delta hedges but based on the simulated stock data, i.e. the Monte Carlo estimates for the delta hedges in this setting. Under the theoretical BSM framework, the P&L would be a point mass at 0, as the 'perfect hedge' would be implemented. Whilst we see distributions in Figure 5 with a mode at 0 (indicat-

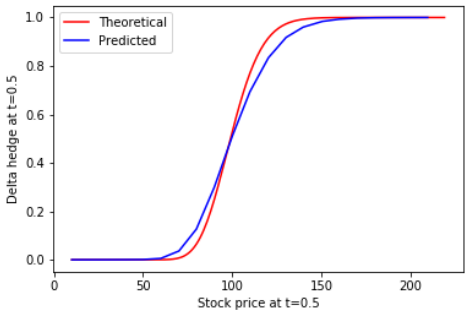
ing satisfactory performance of the network), we see deviation from a point mass at 0 largely due to the fact that we are no longer trading continuously, as per one of the assumptions under the BSM framework. If we reduced the step size dt in our model, we would expect to see less deviation in the P&L distribution around 0, i.e. the hedge approaches that in the theoretical setting.



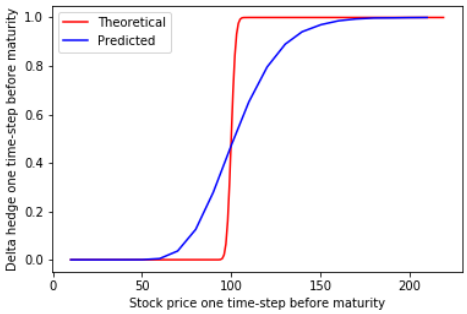
(a) Delta hedges at $t = 0$



(b) Delta hedges at $t = 0.25$



(c) Delta hedges at $t = 0.5$



(d) Delta hedges at $t = 0.99$ (one time-step before maturity)

Figure 4: Theoretical BSM delta hedges against the optimal delta hedges obtained from the NN as a function of stock price at time-steps $t = 0, 0.25, 0.5, 0.99$ (where $T = 1$ and $dt = 0.01$).

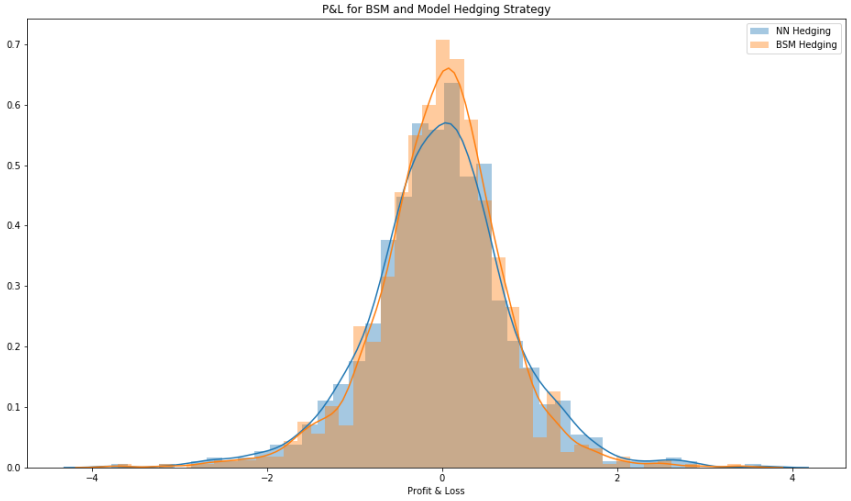


Figure 5: P&L distribution of the NN strategy vs. the Monte Carlo/sample BSM hedging strategy.

5.2.2 Indifference pricing

As discussed in Section 4.1, to obtain the indifference price of an option via deep hedging, we can use the same NN architecture, just changing the loss function to that given by Eq. (25). The indifference price of the option we are hedging can then be obtained using Eq. (22). Since we have an analytical, unique (due to no-arbitrage) price for options under the BSM framework, we use these prices to validate our indifference prices when training the network on the BSM simulated data. Table 1 shows the indifference prices relative to the BSM analytical price for different numbers of simulations, with the risk aversion parameter fixed at $\lambda = 0.025$.

Number of simulations	10 000	50 000	100 000
BSM Analytical price	7.96557	7.96557	7.96557
Indifference price	10.750114	8.26364	8.26176

Table 1: Indifference prices with $\lambda = 0.025$ vs. the BSM analytical price for the European call option.

Since trading is only possible at discrete times in our set-up, compared to the BSM setting where delta-hedging is assumed to be done continuously, the two prices won't completely coincide regardless of the number of samples of the underlying used to train the network. However, we still see that these prices are close, and as

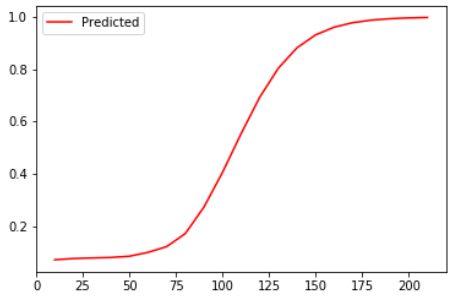
the number of simulations in the training data set increases, the indifference price approaches the theoretical BSM price, validating the performance of our NN here.

5.3 Heston stochastic volatility model

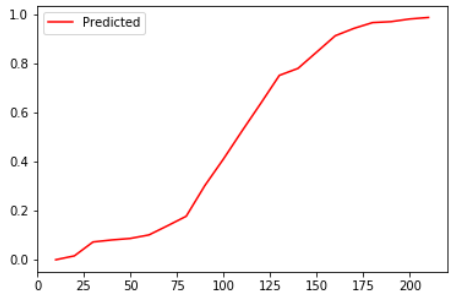
5.3.1 Deep hedging

After training our deep hedging NN on data simulated from under the BSM model, we now repeat this using training data simulated from the Heston model. We used the same number of simulated paths (100,000) over the same number of normalised time-steps (100). Since, unlike the BSM model, there is no analytic formula for Heston call option values, and by extension no closed form solution for the deltas of a Heston call option, we have no theoretical benchmark to compare our delta hedging strategies that are produced by the network. However, if we vary the inputs into our Heston model in a way such that the stock price path differs only slightly from the Black Scholes sample paths, then we may gradually observe how the deltas our model is computing differs from the Black Scholes deltas as we move further away from constant volatility and towards stochastic volatility.

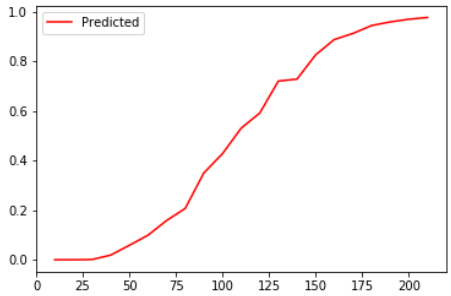
We use the loss function in the exponential utility setting and compute the deltas for varying risk parameters, all representing a risk-averse agent, i.e. $\lambda > 0$. The smaller the λ parameter, the less risk averse the investor. These results are demonstrated in Figures 6 – 8 for time-steps $t \in \{0, 0.25, 0.5, 1\}$ and $\lambda \in \{0.025, 0.05, 0.075\}$.



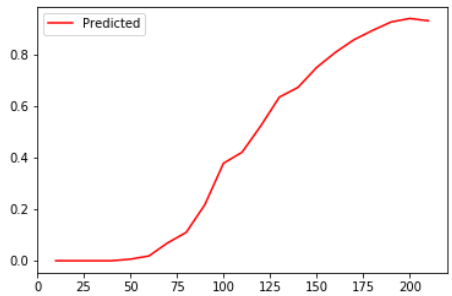
(a) Delta hedges at $t = 0$



(b) Delta hedges at $t = 0.25$

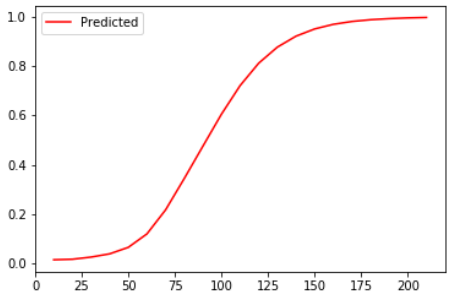


(c) Delta hedges at $t = 0.5$

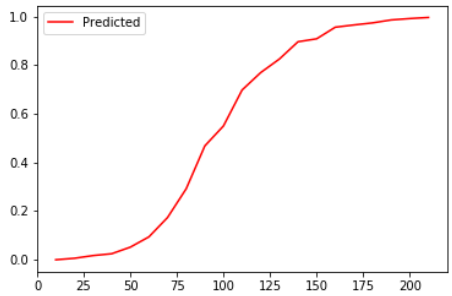


(d) Delta hedges one time-step before maturity

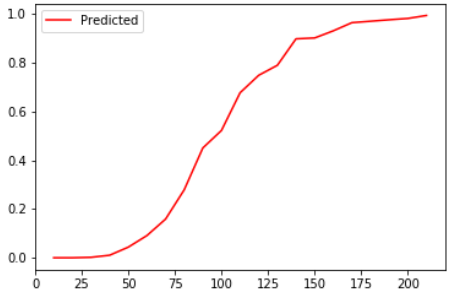
Figure 6: Computed Heston delta hedges obtained from the NN as a function of stock price at time-steps $t = 0, 0.25, 0.5, 1$ at a $\lambda = 0.025$.



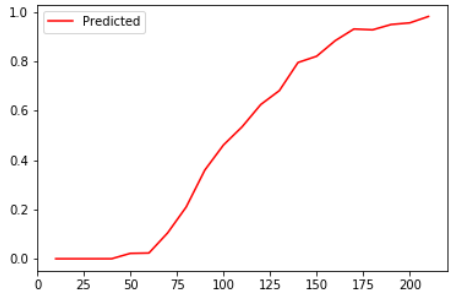
(a) Delta hedges at $t = 0$



(b) Delta hedges at $t = 0.25$

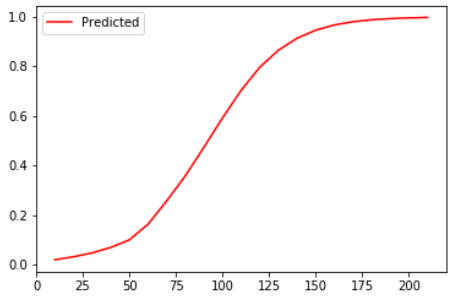


(c) Delta hedges at $t = 0.5$

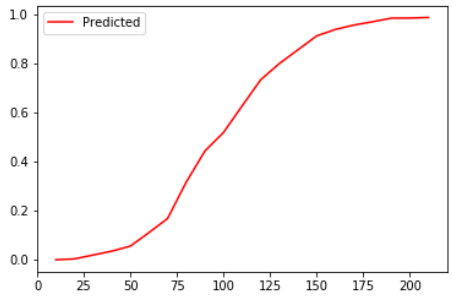


(d) Delta hedges one time-step before maturity

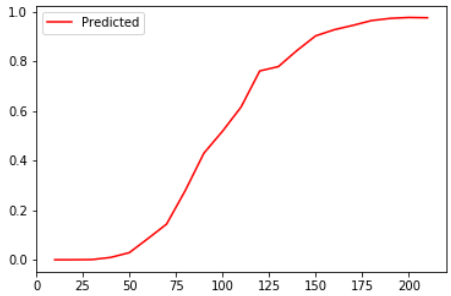
Figure 7: Computed Heston delta hedges obtained from the NN as a function of stock price at time-steps $t = 0, 0.25, 0.5, 1$ at a $\lambda = 0.05$.



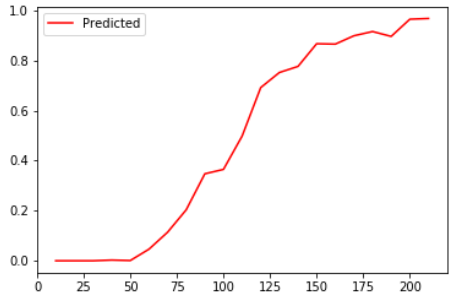
(a) Delta hedges at $t = 0$



(b) Delta hedges at $t = 0.25$



(c) Delta hedges at $t = 0.5$



(d) Delta hedges one time-step before maturity

Figure 8: Computed Heston delta hedges obtained from the NN as a function of stock price at time-steps $t = 0, 0.25, 0.5, 1$ at a $\lambda = 0.075$.

5.3.2 Indifference pricing

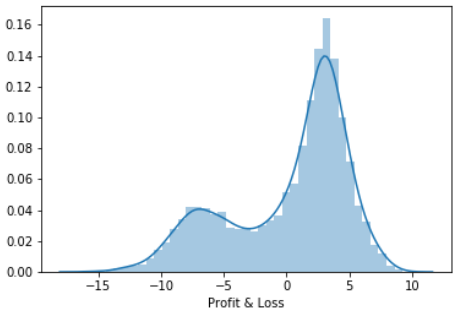
Using the delta hedges produced by our NN for the Heston training data under the exponential utility setting, we can apply the indifference pricing theory to obtain prices for the options written on this underlying. We do this similarly to the BSM setting, however we now compare our prices to the Monte Carlo prices of an option based off the same 100,000 sample paths that we used to train our network. The results are presented in Table 2.

λ	MCE	Indifference Price	Difference:
0.025	12.52437338910	12.612958175370	0.0885847862653
0.05	12.52437338910	12.648655879237	0.124282490131
0.075	12.52437338910	12.713756120590	0.18938273148

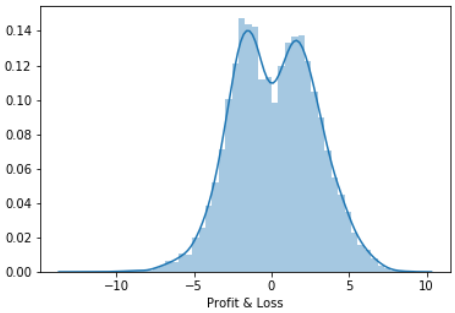
Table 2: Indifference prices vs. the Monte Carlo prices for European call options in the Heston model, for varying risk aversion parameters λ.

As we no longer have a complete market, as we do under the assumptions of the classical BSM model, the price of the option written on an underlying from the Heston model will be non-unique. Instead, an agent must specify some criteria, such as optimising their expected utility for a specified utility function, to define an acceptable minimum price for taking the position (which, here, is selling the option and trading in the market to hedge that option) – as discussed previously, this is the indifference price of the option. Agents with different risk preferences will accept different prices for the position to become acceptable, which is why we see discrepancies in the prices in the table as we change λ. Additionally, while we do not expect the indifference prices to compare exactly with the Monte-Carlo price, this can still be used as a benchmark when assessing the performance of the network – the closeness in prices show in the table suggests acceptable performance of the NN in producing the optimal delta hedges here.

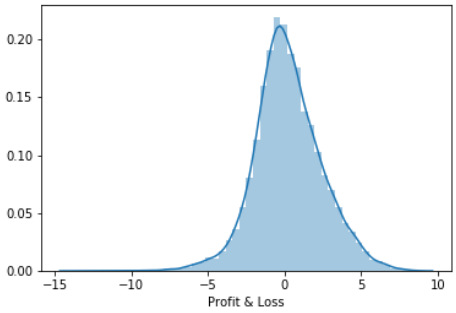
Finally, we plot the P&L distribution of the investor with risk aversion parameter $\lambda \in \{0.025, 0.05, 0.075\}$ who takes a position in the option and adopts the hedging strategy indicated by the NN – see Figure 9. Note that a P&L of 0 indicates a perfectly hedged position in the call option – this would be a desirable strategy for a largely risk averse investor. Figure 9c confirms this result, as we see the P&L start to condense around 0 at the higher risk aversion parameter $\lambda = 0.075$. If we were to increase λ further, we may expect this histogram to condense toward, but not quite reach (due to market incompleteness in the Heston setting), a point mass at 0. The lower the risk aversion parameter, the more desire the investor will have to see positive profit on their strategy (as they are willing to take on more risk to compensate for this). We see that when $\lambda = 0.025$ in Figure 9a, i.e. the investor is less risk averse than the other two cases, we observe a P&L distribution that is shifted to the right, with a higher mode than the other two cases. The result shown in Figure 9b for $\lambda = 0.05$ lies between these two cases as expected.



(a) P&L at $\lambda = 0.025$



(b) P&L at $\lambda = 0.05$



(c) P&L at $\lambda = 0.075$

Figure 9: P&L distribution of the NN strategy under exponential utility for $\lambda = 0.025, 0.05, 0.075$.

6 Extensions

In this section, we discuss the areas in which the content of this report may be extended in future work. The scope of the areas to which deep hedging can be applied to is vast, however we highlight the main 5 areas that would be of particular interest to us to explore going forward.

Transaction costs: One of the main advantages of deep hedging is that it allows for the hedge to account for factors that are overwritten by the assumptions of classical hedging techniques, such as delta hedging under the BSM framework. Something that plays a large role in the P&L of a traders book is transaction costs. By incorporating an extra term in our loss function, we can account for these transaction costs, and see how the optimal hedges produced by the network behave. If, for instance, we consider proportional transaction costs, that is where the cost is proportional to how much the hedging position changes between time steps, the change to the loss function will be as follows:

$$L(x,y) = \left(\sum_{k=1}^{T} y_{k-1} x_{k-1} + C_0 - (x_T - K)^+ + \sum_{k=1}^{T} c_{k-1} |y_{k-1}| \right)^2, \tag{33}$$

where c_k are now the costs that are dependent on changing the amount of stock held in each period. This is then the cost that is applied to each incremental absolute change in the delta across each period for $k = 1, \ldots, T$. The reasoning is that transaction costs are applied to buying or selling units of the underlying, the delta position, and is represented by the absolute size of the change in that position. It is this extra term that now allows us to compute an optimal delta strategy for transaction costs applied to the delta position. The following are some alternative types of transaction costs that could also be interesting to consider in further work:

- transaction costs proportional to the change in underlying price, not just size of the position, i.e. the trader must pay more to hedge when the underlying has jumped significantly between hedging times;
- transaction costs that depend on the size and liquidity of the market in which you are trading;
- trading costs that account for external factors such as levels of risk/ uncertainty in the market (perhaps due to upcoming economic events that are likely to move markets in an unpredictable way as traders are likely to want to scale back trading activity in times like these).

External market factors: Similar to transaction costs, we can include a variety of other external market factors that are likely to impact the hedging strategy adopted by a trader as inputs to the model. Under a classical model, a trader is likely to

compute the hedge, then make modifications based on how they think it should change in regards to these factors, whether they be upcoming geopolitical events that are likely to move markets, liquidity constraints, risk constraints of the trading desk, or others of this sort. These could be incorporated by introducing Lagrange-multipliers into the loss function or, as mentioned by Buehler et al. (2019) by the choice of activation function, for example a non-negative activation function could be chosen to impose a no short-selling restriction.

Non-zero interest rates: For simplicity throughout our study, we chose interest rates to be zero. Going forward, it could be explored how introducing interest rate dynamics, in particular the case where a stochastic interest rate model is considered, affect the optimal hedging strategies produced by the network. We assume that this could be incorporated through a new custom loss function, however we have not explored this in any depth.

Training with real data: Our network was trained on synthetic data, as this allowed us to obtain a benchmark for our results easily and quickly in order to evaluate the performance of the NN. As deep hedging techniques are purely data-driven and thus have the advantage of being model-free, one of the next steps would be to train the neural network on real stock data from a liquid market, and then back-test the model to assess its performance under a model-free setting.

Deep optimal stopping: Finally, the application of deep neural networks to optimal stopping policies is an area that could be explored in any follow-up work from this report. This could also be brought together with the deep hedging methodology to price American options, where results can be compared to analytical results to assess the performance of the neural network when used to solve problems of this type.

7 Conclusion

To conclude, we implemented a deep neural network using the functional API in the Keras library. We trained this network on data simulated from the Black-Scholes model and Heston stochastic volatility model, where these models were chosen so that we had benchmarks to assess the performance of our results, however deep hedging techniques can be purely model-free. In the Black-Scholes setting we compared our network output delta hedges to the theoretical delta hedges (which are available in closed form) and saw strong performance of the NN relative to this benchmark when the option was not approaching maturity. This stems from the fact that without accounting for external factors (such as transaction costs, liquidity constraints etc.), the function that the network is trying to learn is relatively simple. Whilst such a clear benchmark is not available for the setting where the data we simulated was from the Heston model, our results did not show anything that seemed out of the ordinary.

We then extended this to hedge derivatives under an exponential utility function to incorporate the risk preferences of an investor, which can be controlled through a risk aversion parameter. This also allows us to price options via deep hedging using the indifference pricing framework. The benchmark used to assess our results here was the theoretical BSM price of the option for the data simulated from the Black-Scholes model, and a Monte Carlo estimate of the price when the input data was simulated from the Heston model. Whilst we do not expect the prices obtained in these two different ways to be exactly the same, for reasons mentioned above, the results for the indifference prices of the financial contract did not show any large deviation from the benchmark prices that would suggest the deep neural network was producing nonsensical results.

Whilst the deep hedging neural networks discussed in this report have a large scope for both refinement and extension, they nonetheless illustrate an interesting starting point to exploring this new area of research in the overlap between mathematical finance and artificial intelligence. Additionally, it is clear the large amount of potential these concepts have in exploiting the large amounts of financial data that is readily available today, and continues to grow, to transform and automate high-dimensional hedging problems.

Bibliography

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

S. Aggarwal and S. Aggarwal. Deep investment in financial markets using deep learning models. *International Journal of Computer Applications*, 162(2):40–43, 2017.

J.-F. Bégin, M. Bédard, and P. Gaillardetz. Simulating from the heston model: A gamma approximation scheme. *Monte Carlo Methods and Applications*, 21(3):205–231, 2015.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, pages 1–21, 2019.

F. Chollet et al. Keras. `https://keras.io`, 2015.

D. Doyle and C. Groendyke. Using neural networks to price and hedge variable annuity guarantees. *Risks*, 7(1):1, 2019.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

C. F. Higham and D. J. Higham. Deep learning: An introduction for applied mathematicians. *arXiv preprint arXiv:1801.05894*, 2018.

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

P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.

Quantcast. Quantcast - a risk.net cutting edge podcast: Hans buehler. https://soundcloud.com/user-393208363, 2019.

S. J. Reddi, S. Kale, and S. Kumar. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.

H. J. Shin and J. Ryu. A dynamic hedging strategy for option transaction using artificial neural networks. *International Journal of Software Engineering and Its Applications*, 6(4):111–116, 2012.

8 Appendix

8.1 Deep hedging neural network code

The following is the code we wrote to build and train a deep neural network with loss function given by Eq. (33). The input data is given by the array 'stockdata'.

```
import numpy as np
import numpy.matlib as npm
import tensorflow as tf
import math
from tensorflow import keras
from keras.layers import Input, Dense, Concatenate
from keras.models import Model
from keras import backend as bk

# Setting up architecture of Neural Network
myinput_list = []
prediction_delta = []

# stack 4 layers
# important to have this before the loop as this ensures
#the same weights for each branch of Neural Network
layer1 = Dense(100, activation='relu')
layer2 = Dense(100, activation='relu')
layer3 = Dense(100, activation='relu')
layer4 = Dense(1, activation='sigmoid')

for i in range(nsteps):
    myinput = Input(shape=(2,))
    x = layer1(myinput)
    x = layer2(x)
    x = layer3(x)
    predictions=layer4(x)
    myinput_list.append(myinput)
    prediction_delta.append(predictions)

prediction_delta = Concatenate(axis=-1)(prediction_delta)

model = Model(inputs=myinput_list, outputs=prediction_delta)

# Defining custom loss function to be used in NN
```

```
#ypred=delta_pred , increments=y_true

def myloss(increments,delta_pred):
    temp=bk.sum(delta_pred*increments,axis=1)
    temp2=(temp+price-bk.maximum(float(0),
            bk.sum(increments,axis=1)+S0-K))**2
    #ensure 0 is a float
    return temp2

# Compiling and fitting the model
es=EarlyStopping(monitor='loss',patience=3,min_delta=0.0001)
model.compile(loss=myloss,optimizer='adam',metrics=['mae'])
model.fit(stockdata,np.diff(St),epochs=5,
                callbacks=[PlotLossesKeras()])
```

8.2 BSM data simulation code

The following code is used to simulate the sample paths of the underlying stock under the BSM framework in order to train the neural network.

```
# Parameters
nsim=100000
nsteps=100

K=100; S0=100; r=0; sigma=0.2; T=1; dt=T/nsteps;

# Set the seed
np.random.seed(0)

# call price function
def call_option_price(S0,K,T,r,sigma): #T is time to maturity
    d1=(np.log(S0/K)+T*(r+(sigma**2)/2))/(sigma*np.sqrt(T))
    d2=d1-sigma*np.sqrt(T)

    call_price=S0*norm.cdf(d1)-K*np.exp(-r*T)*norm.cdf(d2)
    return call_price

#generate BSM paths function
def stock_price_path(nsim,nsteps,S0,r,sigma):
    dt=1/nsteps
    Z=np.random.normal(0,1,size=[nsim,nsteps])
    exponent_value=np.exp((r-0.5*sigma**2)*dt+sigma*math.sqrt(dt)*Z)
```

```
    #creates a vector of our inital stock prices
    S_0=S0*np.ones((nsim,1))
    S=np.hstack((S_0,exponent_value))
    St=np.cumprod(S,axis=1) #axis=1, means along each row

    return St

St=stock_price_path(nsim,nsteps,S0,r,sigma)

    # Function creates a stock price path matrix.
    #The rows represent each simulation.
    #The columns represent stock value at each time point
    #Realisations are taken at nsteps time steps

# Function for d1 where the delta of a call option will be
#norm.cdf(d1)

def d1(s,r,sigma,t,T,K):
    d1 = 1/(sigma*math.sqrt(T-t))*(np.log(s/K)+
                (r+0.5*sigma**2)*(T-t))
    return d1

# Need to ensure that the input data is in correct format
#before fitting.
# It should be a list of arrays.
# First array in the list is like a column vector.
# The column array should look like:
#[[0,S_0^(0)],[0,S_0^(1)],...[0,S_0^(nsim)]]
# The second column will be the same as above,
#but the first item would 1 and the second item would be
# S_1^(0),S_1^(1),..,S_1^(nsim)
# and so on...

def create_stockdata(nsteps,nsim,St):
    stockdata=[]
    for i in range (0,nsteps):
        a0 = np.array([i, 0])
        a0a = npm.repmat(a0, nsim, 1)

        for j in range(0,len(a0a)):
            a0a[j][1] = St[j][i]
```

```
        stockdata.append(a0a)

    return stockdata

stockdata=create_stockdata(nsteps,nsim,St)
```

8.3 Heston data simulation code

The following code is used to simulate the sample paths of the underlying stock under the Heston Stochastic volatility framework,

```
import numpy as np
import numpy.matlib as npm
import math
from scipy.stats import norm
import matplotlib.pyplot as plt
from livelossplot import PlotLossesKeras
from scipy.linalg import cholesky

# Parameters
nsim=100000 #increase to 1000000
nsteps=100

K=100; S0=100; r=0; sigma=0.1; T=1; dt=T/nsteps #try sigma 0.2
lambda_val=0.025 #any value above 1 will result in nan

#Heston specific parameters
kappa = 1.5 #rate of mean reversion
theta = 0.1 #mean reversion, long run variance
epsilon = 0.2 #vol of vol
rho = -0.2


# Set the seed
np.random.seed(0)

# Functions used to calculate
def call_option_price(S0,K,T,r,sigma): #T is time to maturity
    d1=(np.log(S0/K)+T*(r+(sigma**2)/2))/(sigma*np.sqrt(T))
    d2=d1-sigma*np.sqrt(T)

    call_price=S0*norm.cdf(d1)-K*np.exp(-r*T)*norm.cdf(d2)
    return call_price
```

```
def heston(nsim,nsteps,S0,r,sigma,rho,kappa,theta,epsilon):
    #nsim = 1000
    #nsteps=100
    corr = np.array([[1, rho],[rho, 1]])
    L = cholesky(corr, lower=True)
    #note: default is an upper cholesky decomoposition
    # we need a lower one

    dt=T/nsteps
    S_arr=[]
    for i in range(0,nsim):
        Z = np.random.normal(0,1,size=[2,nsteps])
        #generates both the stock rv and the variance rv
        X = np.matmul(L,Z)
        #imposes the correlation structure on the rv
        #Zv = np.random.normal(0,1,size=[nsim,nsteps])
        Xs = X[0,:]
        Xv = X[1,:]

        S = np.zeros(nsteps+1)
        V = np.zeros(nsteps+1)

        #we now want to calculate our sample paths using the
        # Euler discretisation scheme
        for j in range(0,nsteps+1):
            if j==0:
                S[j] = S0 #setting our initial stock value
                V[j] = sigma #setting our initial variance
            else:
                V[j] = V[j-1] + kappa*(theta-max(0,V[j-1]))*dt
                        +epsilon*math.sqrt(max(0,V[j-1]))*Xv[j-1]
                        *math.sqrt(dt)
                S[j] = S[j-1] + r*S[j-1]*dt
                + math.sqrt(max(0,V[j-1]))
                *S[j-1]*Xs[j-1]*math.sqrt(dt)
        S_arr.append(S)

    return np.array(S_arr)

# Functions creates a stock price path matrix.
#The rows represent each simulation.
```

```
#The columns represent stock value at each time point
# Assume that we are creating many realisations of stock
# between time 0 and 1 (nsteps many realisations)
# Note that the number of columns is nsteps+1

# It is wise to create a function for d1 since the
#delta of a call option will be norm.cdf(d1)

def d1(s,r,sigma,t,T,K):
    d1 = 1/(sigma*math.sqrt(T-t))*(np.log(s/K)+(r+0.5*sigma**2)*(T-t))
    return d1

def mce_call(S,nsim,K,r):
#S here is a stock price path
#we need to average over the terminal values
    payoff = S[:,nsteps]-K

    disc_payoff = []
    for i in range(len(payoff)):
        #print(i)
        if payoff[i] <= 0:
            payoff[i] = 0
    disc_payoff = payoff*math.exp(-r*T)
    #now the discount is back to the current time t

    c_mce = np.mean(disc_payoff)
    return c_mce

St=heston(nsim,nsteps,S0,r,sigma,rho,kappa,theta,epsilon)
# Monte carlo price
price=mce_call(St,nsim,K,0)
```

8.4 Deep hedging neural network and indifference pricing

The following is the code we wrote to build and train a deep neural network with loss function given by Eq. (25), so to incorporate risk preferences into the optimisation problem via the exponential utility function. In this code, we use the optimal hedging strategy obtained from the network to price the option via the indifference pricing framework. The input data is given by the array 'stockdata'.

```
import numpy as np
import numpy.matlib as npm
import tensorflow as tf
```

```
import math
from tensorflow import keras
from keras.layers import Input, Dense, Concatenate
from keras.models import Model
from keras import backend as bk
from scipy.stats import norm
import matplotlib.pyplot as plt
from livelossplot import PlotLossesKeras
from keras.callbacks import History, EarlyStopping

# Setting up architecture of Neural Network
myinput_list = []
prediction_delta = []

# stack 4 layers
layer1 = Dense(100, activation='relu')
layer2 = Dense(100, activation='relu')
layer3 = Dense(100, activation='relu')
layer4 = Dense(1, activation='sigmoid')

for i in range(nsteps):
    myinput = Input(shape=(2,))
    x = layer1(myinput)
    x = layer2(x)
    x = layer3(x)
    predictions=layer4(x)
    myinput_list.append(myinput)
    prediction_delta.append(predictions)

prediction_delta = Concatenate(axis=-1)(prediction_delta)

model = Model(inputs=myinput_list, outputs=prediction_delta)

print(model.summary())

# Defining custom loss function to be used in NN
def myloss(increments,delta_pred):
#ypred=delta_pred, increments=y_true

    temp=bk.sum(delta_pred*increments,axis=1)
    cost=bk.sum(bk.abs(delta_pred[:,1:nsteps-1]
                -delta_pred[:,0:nsteps-2]),axis=1)
```

```
    temp2=temp-bk.maximum(float(0),bk.sum(increments,axis=1)
                          +S0-K)-0.1*cost
    temp3=bk.exp(-lambda_val*temp2)

    return temp3

# Compiling and fitting the model
es=EarlyStopping(monitor='loss',patience=3,min_delta=0.0001)
model.compile(loss=myloss,optimizer='sgd',metrics=['mae'])
model.fit(stockdata,np.diff(St),epochs=6,
          callbacks=[PlotLossesKeras(),es])

# Adam optimiser will also work.
#SGD works about the same.
#Nadam seems to not do as well as the other.

losses=model.history.history['loss']

indiff_price=(1/lambda_val)*np.log(losses[-1])
          -(1/lambda_val)*np.log(1)
# in most cases, the loss when payoff is zero is one.

# We now aim to compute the Profit & Loss distribution

# First generate sample paths
nsim2=10000
St_2=stock_price_path(nsim2,nsteps,S0,r,sigma)
stockdata_2=create_stockdata(nsteps,nsim2,St_2)

# Calculate the trading profit
predicted_delta=model.predict(stockdata_2)
stock_increments=np.diff(St_2)

martingale_transform=np.sum(predicted_delta*stock_increments,
                            axis=1)

# Calculate the payoff for each payoff
Z_T=[]
for i in range(0,nsim2):
    Z_T.append(max(St_2[i,nsteps]-K,0))
Z_T = np.array(Z_T) #This is the payoff at the end
# Z_T.shape
```

```
# Calculate the P&L for each sample path
PnL=martingale_transform-Z_T+indiff_price
np.mean(PnL)
```

Inferring OIS Discount Factors in the South African Market

TEAM UTS

RYAN BROODRYK, University of Cape Town
DYSON CHIWESHE, Reserve Bank of Zimbabwe
JAUNDRÉ SCHELTEMA, Sanlam Investment Management
ANDREW SOANE, University of Cape Town

Supervisor:
ERIK SCHLÖGL, University Technology Sydney

African Collaboration for Quantitative Finance and Risk Research

Contents

Chapter 1

Introduction

Collateral posted against financial derivative transactions accrues interest at an overnight rate, it is therefore integral that in the presence of roll-over risk, these derivative contracts are discounted at the overnight rate rather than the market benchmark rate. The purpose of this report is to infer a term structure of overnight index swap (OIS) discount rates from the South African interest rate market. In this market there are no basis swaps or Overnight Index swaps and therefore a term structure of OIS discount factors cannot be bootstrapped from the available market prices. Using a JIBAR/OIS spread, one could infer the term structure of OIS discount factors with the term structure of benchmark rates, however in the South African market where no JIBAR/OIS spread is observable this is not possible.
Näively one could use the term-structure of discount rates bootstrapped from South African data to discount collateralised derivatives, however the reason that a basis spread occurs is due to roll-over risk. Therefore in order to infer a term-structure of OIS discount rates we shall assign a degree of structure to the roll-over risk in the form of a stochastic model for the dynamics of the over night interest rate and the components which effect roll-over risk. We shall then estimate this model in the South African interest rate market in order to infer a term structure of OIS discount rates.

The concept of roll-over risk has become particularly relevant since the Global Financial Crisis, where refinancing at the prevailing market benchmark rate is now no longer considered guaranteed. This challenges fundamental principles of financial mathematics as one would expect the presence of a basis premium to introduce an arbitrage strategy, indeed one could borrow at the shorter maturity and lend at the longer maturity and realise what would have been considered a "risk-free" profit. This classical arbitrage strategy, however; does not account for the possibility of not being able to refinance ones investment at the market benchmark rate, the so called "roll-over" risk. The presence of roll-over risk means that the entire term-structure of interest rates can no longer be modelled as one process evaluated

at different maturities (eg. the Vasicek Model). This has led to a need for every term structure associated with a different tenor to be modelled separately, the so called "multi-curve" modelling technique. We have outlined a motivation for the need to consider roll-over risk from a financial and mathematical point of view in the following section.

In the presence of roll-over risk, one would pay a premium to borrow for longer tenors compared with shorter tenors. This is because borrowing at longer tenors results in the borrower avoiding the roll-over risk associated with the shorter tenors. This luxury is then priced as a premium applied to longer tenor borrowing rates. This is observed in the market as a basis spread, that is, a spread applied to a basis swap which exchanges floating legs of two different tenors. In particular, the LIBOR/OIS spread. The OIS rate implies a discount factor, this discount factor implies a yield, the LIBOR/OIS spread is the difference between this yield and LIBOR.

1 Motivation

Consider a market benchmark rate, calculated as an average of all panel members' cost of borrowing. For the sake of exposition, let the benchmark rate under consideration be the London Interbank Offered Rate (LIBOR). For simplicity, assume the panel member can borrow at LIBOR today and there is a possibility that the member can get downgraded in the near future. Furthermore, there is a possibility that the uncertainty in the market may cause other panel members to not lend at LIBOR but instead at a higher rate in the future. Therefore if the panel member planned on borrowing today and "rolling-over" their loan (i.e. refinancing their loan) at some future time point then they would be exposed to roll-over risk, that is the risk of not being able to refinance their loan at the prevailing benchmark rate in the future.

To illustrate the concept more concretely, we will present a simple example:
Consider a panel of banks making up the LIBOR panel. Now suppose Bank A belongs to this group. Now by definition of LIBOR, it is the average rate at which any bank can get funding from other panel members. Assume that Bank A can borrow from the panel at LIBOR. Now suppose that Bank A can borrow either for six months at LIBOR (**option A**) or it can borrow for three months at LIBOR, roll-over their investment for another three months and simultaneously enter a forward rate agreement (see section 2.2) to hedge their interest rate risk and essentially "lock" in a fixed rate from three months to six months (**option B**). The forward rate agreement, by definition, will swap the quoted three month LIBOR rate in three months time, regardless of whether Bank A is able to access funding at LIBOR in three months time or not with a fixed rate. Because Bank A may not be able to refinance their loan at LIBOR, Bank A is exposed to roll-over risk if it chooses option

B. Therefore Bank A should be willing to pay a premium for option A to avoid roll-over risk. Näively the premium that Bank A pays should result in an arbitrage opportunity as they are fully hedged from interest rate risk and we are assuming there are no other market risks that Bank A faces. However; due to the possibility of Bank A being unable to refinance their borrowing at LIBOR in three months, this premium is the price of avoiding this possibility. It is by this reasoning that a further investigation in to the structure of the roll-over risk needs to be done. Furthermore, in a market where no basis spread is observable such as South Africa, it will be useful to infer a basis spread from the observable market data and an estimated model.

2 Background

The concept of a basis spread was first considered in a foreign exchange setup in Boenkost and Schmidt (2004) whereby it was observed that cross-currency basis swaps had a liquidity-premium being charged on top of the floating rate of one leg of the basis swap. The authors justify the spread as a result of cross-currency basis swaps being of high demand, they also acknowledge the challenge of this spread to traditional arbitrage opportunities. The paper derives a cross-currency basis swap valuation procedure taking in to account this spread. Using this methodology, Kijima et al. (2009) apply an FX analogy to the single currency basis swaps. This was the first consideration of the "multi-curve" modelling approach.

It has become well understood since the Global Financial Crisis that a basis spread should be present, however; from a practical perspective, consistent and financially justifiable models have yet to appear in industry. Practitioners typically model basis spreads in an *ad-hoc* fashion, modelling them deterministically or stochastically but without any economic justification for the presence of the spreads.

The literature that concerns itself with modelling the basis spread in a economically justifiable manner, typically entails decomposing the roll-over risk into a credit and liquidity component. The justification being that firstly, roll-over risk emerges from the risk of a member in the benchmark rate panel being downgraded, the member is now forced to pay a higher cost of borrowing and therefore unable to refinance their investment at the benchmark rate. Secondly roll-over risk manifests from a panel member being unable to refinance their borrowing at the benchmark rate because of a liquidity premium being added to the benchmark rate during times of market distress.

Alfeus et al. (2017) develop a consistent model for fitting all observed tenors in a cross-sectional manner. Here the authors make the observation that modelling the term structure for each tenor is superfluous as one could model each tenor by sim-

ply one term structure for all tenors and the roll-risk on their own. This way, each term structure can be adjusted by the roll-over risk appropriate to its tenor.

An important aspect of the roll-over risk discussion is that typically, swap contracts are collaterised, i.e. each party to the contract is required to post daily margin to a margin account on a mark-to-market basis. This margin account accrues interest at an overnight rate. In the absence of roll-over risk, a simple arbitrage argument would imply that the value of the collaterised swap can be derived by using the term-structure of benchmark discount rates. However, in the presence of roll-over risk, it is necessary that we discount with respect to the overnight discount rate. Therefore, knowing the term-structure of benchmark discount rates is not adequate.

A significant contribution to the modelling of roll-over risk is from Filipović and Trolle (2013), it was here that the authors define their so called "Interbank risk" as the *risk of direct or indirect loss resulting from lending in the inter-bank money market*. Here, the authors decompose the LIBOR/OIS spread in to a default and non-default component. They model the probability of a representative entity of the LIBOR panel being removed due to their credit deterioration with a jump diffusion process. Using this model they are able to price a variety of collateralised swap contracts including credit default swaps. Using their pricing formulae for each instrument and credit default swap data, they estimate the risk-neutral dynamics of the credit spread process, leaving the non-default component as a "residual" spread, unexplained by the credit spread. Looking forward to the model of Backwell et al. (2019) (see section 2), the non-default component can be considered a liquidity spread. Backwell et al. (2019) explicitly take the view that the LIBOR/OIS spread is fully explained by a credit and funding spread, giving dynamics to both. Explicitly estimating the dynamics of the liquidity spread is typically not possible as there are no market instruments which price an individual members' liquidity risk premium, hence just as in Filipović and Trolle (2013), the order of estimation is to first estimate the dynamics of the credit spread with credit default swap data and then to estimate the entire dynamics of the model knowing the dynamics of the credit spread process.

Risk-neutral expectations of future LIBOR/OIS spreads are priced in to the overnight index swaps. Our report focuses on a market where there are no traded overnight index or basis swaps. We can therefore infer a term-structure of overnight index swap rates by modelling the market's risk-neutral dynamics of roll-over risk and applying the implied spreads to the observed market deposit rates. We shall follow the methodology of Backwell et al. (2019) and attempt to estimate their model to South African interest rate data. In this market, there are no traded basis swaps or overnight index swaps, therefore provided the South African data allows

for adequate estimation, we can infer a term-structure of OIS discount rates. This will entail developing the appropriate pricing formulae applicable to the South African interest rate market, estimating the risk-neutral dynamics of the various processes using the pricing formulae and an appropriate set of data and then computing the implied term-structure of overnight discount rates.

The report is structured as follows; section 2 explains the model of Backwell et al. (2019) as well as giving the South African interpretation of their pricing formulae as well as an overview of the data that is used for estimation. Section 3 explains the technique used to estimate the dynamics of the model using the pricing formulae of section 2 and a South African data set. Section 4 presents the results of the estimation as well as the model-implied term-structure of overnight discount rates. Section 5 then concludes the report with a discussion on the implementation of the model and the results.

3 Setup

We work on probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ equipped with a filtration $\mathbb{F}$ satisfying the usual conditions of right-continuity and $\mathbb{Q}$ - completeness representing the information available to the market. We assume that all processes under consideration in this report are adapted to $\mathbb{F}$. Denote by $\mathbb{E}_{\mathbb{Q}}[\cdot]$ the expectation with respect to $\mathbb{Q}$.

Chapter 2

Model

1 Model Dynamics

The instantaneous rate at which a bank can borrow is influenced by the idiosyncratic roll over risk of that bank. Let $k = 1, \ldots, n$ denote the k^{th} LIBOR panel member. We define the instantaneous borrowing rate for member k as

$$r_t^k = r_t^{\text{OIS}} + \phi_t^k + \lambda_t^k, \tag{2.1}$$

where (ϕ_t^k) denotes the idiosyncratic funding spread and (λ_t^k) is the idiosyncratic credit spread associated with member k, and r_t^{OIS} is the instantaneous Overnight Indexed Swap (OIS) rate. The OIS rate can be decomposed as $r_t^{\text{OIS}} = r_t + c_t$ where (r_t) is the risk free rate associated with the price P_{tT} at time $t \geq 0$ of a zero-coupon bond with maturity $T > t$. The process (c_t) is a market wide credit spread that reflects the overall credit quality of the financial market. We model the dynamics of (ϕ_t^k) and (λ_t^k) as jump diffusions with stochastic intensity as follows:

$$d\phi_t^k = -\kappa^\phi \phi_t dt + dJ_t^\phi, \quad \phi_0 = 0. \tag{2.2}$$

κ^ϕ is the mean reversion rate of (ϕ_t^k) and is assumed constant, (J_t^ϕ) is a pure jump process of finite variation, having only positive jumps, with stochastic intensity (ξ_t^ϕ), having dynamics

$$d\xi_t^\phi = \alpha^\phi(\theta^\phi - \xi_t^\phi)dt + \sigma^\phi \sqrt{\xi_t^\phi} dW_t^\phi. \tag{2.3}$$

$\xi_0^\phi > 0$, $\alpha^\phi \in \mathbb{R}$, $\theta \geq 0$ and $\sigma^\phi > 0$ are constants and (W_t^ϕ) is a $\mathbb{Q}$ - Brownian motion. The jump sizes are independent and identically exponentially distributed with mean $(\mu^\phi)^{-1}$ i.e.

$$J_t^\phi = \sum_{t=1}^{N^\phi(t)} X_t^\phi, \qquad X \sim \text{Exp}(\mu_\phi), \qquad N^\phi(t)|\xi_t^\phi \sim \text{Pois}(\xi_t^\phi).$$

Similarly for (λ_t^k),

$$d\lambda_t^k = -\kappa^\lambda \lambda_t du + dJ_t^\lambda, \quad \phi_0 = 0. \tag{2.4}$$

κ^λ is a mean reversion rate of (λ_t^k) and is assumed constant, (J_t^λ) is a pure jump process of finite variation, having only positive jumps, with stochastic intensity (ξ_t^λ)

$$d\xi_t^\lambda = \alpha^\lambda(\theta^\lambda - \xi_t^\lambda)dt + \sigma^\lambda\sqrt{\xi_t^\lambda}dW_t^\lambda. \tag{2.5}$$

$\xi_0^\lambda > 0$,$\alpha^\lambda \in \mathbb{R}$, $\theta \geq 0$ and $\sigma^\lambda > 0$ are constants and (W_t^λ) is a $\mathbb{Q}$ - Brownian motion. We will assume that the jump sizes are independent and identically exponentially distributed with mean $(\mu^\lambda)^{-1}$ i.e.

$$J_t^\lambda = \sum_{t=1}^{N^\lambda(t)} X_t^\lambda, \qquad X \sim \text{Exp}(\mu_\lambda), \qquad N^\lambda(t)|\xi_t^\lambda \sim \text{Pois}(\xi_t^\lambda).$$

Denote by A_{tT}^k the pre-default roll-over account associated with panel member k, that is, the discounted deposit account with panel member k which accumulates at (r_t^k) and is discounted at (r_t) assuming zero recovery in the event of default. Let τ^k denote the default time of entity k, note that by assumption this default time will be affected by entity k's idiosyncratic credit spread (λ_t^k) and the market-wide credit spread (c_t).

$$\begin{aligned}
A_{tT} &:= \mathbb{E}_\mathbb{Q}\left[\exp\left(-\int_t^T r_u du\right)\exp\left(\int_t^T r_u^k du\right)\mathbb{I}_{\{\tau^k > T\}}\middle|\mathcal{F}_t\right], \\
&= \mathbb{E}_\mathbb{Q}\left[\exp\left(-\int_t^T r_u du\right)\exp\left(\int_t^T (r_u^k) du\right)\exp\left(-\int_t^T c_u + \lambda_u^k du\right)\middle|\mathcal{F}_t\right], \\
&= \mathbb{E}_\mathbb{Q}\left[\exp\left(\int_t^T r_u + \phi_u^k\right)\middle|\mathcal{F}_t\right].
\end{aligned} \tag{2.6}$$

The second equality follows from the projection property of conditional expectations, see Lando (2009) for a more detailed proof. The third equality follows from equation (2.1).
Now denote by (Q_{tT}^k) the pre-default value of a zero-coupon bond issued by entity k, that is

$$\begin{aligned}
Q_{tT}^k &:= \mathbb{E}_\mathbb{Q}\left[\exp\left(-\int_t^T r_u du\right)\mathbb{I}_{\{\tau_k > T\}}\middle|\mathcal{F}_t\right], \\
&= \mathbb{E}_\mathbb{Q}\left[\exp\left(-\int_t^T (r_u + c_u + \lambda_u^k) du\right)\middle|\mathcal{F}_t\right], \\
&= \mathbb{E}_\mathbb{Q}\left[\exp\left(-\int_t^T (r_u^{\text{OIS}} + \lambda_u^k) du\right)\middle|\mathcal{F}_t\right].
\end{aligned} \tag{2.7}$$

Individual panel members are exposed to a term structure of unsecured discrete borrowing rates, all of which will be aggregated to form the market benchmark rate. We can imply such borrowing rates by considering a forward rate agreement (FRA) with reset at time t and maturity at time T entered in to by entity k. This contract will have the following payoff at maturity

$$V_T^k = \left(\exp\left(\int_t^T r_t^k du \right) - (1 + \delta K) \right) \mathbf{1}_{\{\tau_k > T\}}. \tag{2.8}$$

Note that the above payoff is contingent on entity k having not defaulted by time T. The price of the FRA at time t is

$$V_t^k = \mathbb{E}_{\mathbb{Q}} \left[V_T^k \, | \mathcal{F}_t \right] = A_{tT}^k - (1 + \delta K) Q_{tT}^k. \tag{2.9}$$

By setting $V_t^k = 0$, for all $t \leq T$, the fair rate $K \equiv R^k(t,T)$ is

$$R^k(t,T) = \frac{1}{\delta} \left(\frac{A^k(t,T)}{Q_{tT}^k} - 1 \right), \quad 0 \leq t \leq T. \tag{2.10}$$

Let the indices of the panel member banks at time t be $\{k_1(t), k_2(t), \ldots, k_m(t)\} = \{k \in (1, \ldots, m) | \phi_t^k = 0\}$. The panel average fair rate, i.e. the market benchmark rate is then defined as,

$$L(t,T) := \Upsilon(R^{k_t(1)}(t,T), \ldots, R^{k_t(m)}(t,T)). \tag{2.11}$$

The function $\Upsilon(\cdot)$ is market specific and is known *a priori* but the input observations are not observable. Modelling the benchmark rate in such way would be cumbersome as one would need to know the specific dynamics of each member of the panel which is subject to change when the panel changes. A more tractable approach would be to model the dynamics of the panel as a whole and infer the market benchmark rate as above from these dynamics. That is,

$$L(t,T) = \frac{1}{\delta} \left(\frac{A_{tT}^L}{Q_{tT}^L} - 1 \right). \tag{2.12}$$

where

$$A_{tT}^L = \mathbb{E}_{\mathbb{Q}} \left[\exp\left(\int_t^T \phi_u du \right) \middle| \mathcal{F}_t, \phi_t = 0 \right] \tag{2.13}$$

$$Q_{tT}^L = \mathbb{E}_{\mathbb{Q}} \left[\exp\left(- \int_t^T r_u^{\text{OIS}} + \lambda_u du \right) \middle| \mathcal{F}_t \right]. \tag{2.14}$$

In this setup the instantaneous roll-over risk at time t is zero, that is, at time t by virtue of being in the panel an entity is free of roll-over risk immediately, their funding and credit quality is modelled to deteriorate from time t onwards. Given the above argument we now assign a model to (r_t^{OIS}) as

$$dr_t^{\text{OIS}} = (\alpha^{\text{OIS}}(\theta^{\text{OIS}} - r_t^{\text{OIS}}) + \alpha^{\text{OIS},\phi}(\theta^{\phi} - \xi_t^{\phi}) + \alpha^{\text{OIS},\lambda}(\theta^{\lambda} - \xi_t^{\lambda}))dt + \sigma^{\text{OIS}}\sqrt{r_t^{\text{OIS}}}dW_t^{\text{OIS}}. \tag{2.15}$$

$\alpha^{\text{OIS}}, \alpha^{\text{OIS},\phi}, \alpha^{\text{OIS},\lambda} \in \mathbb{R}$, $\theta^{\text{OIS}} > 0$ and $\sigma^{\text{OIS}} > 0$ are constants and (W_t^{OIS}) is a $\mathbb{Q}$-Brownian motion. The process (ϕ_t) denotes the instantaneous funding spread and will have the same dynamics stated in (2.2). Furthermore at the time when the panel is updated, we will assume that there is no instantaneous roll-over risk, i.e. $\phi_t = 0$. Qualitatively the process (ϕ_t) models the sudden liquidity shocks and will decay at the rate κ^{ϕ} with mean jump size μ^{ϕ} arriving at intensity (ξ_t^{ϕ}). The process (λ_t) will have the same dynamics as (2.4). It will decay at rate κ^{λ} with mean jump size μ^{λ} and arriving at intensity (ξ_t^{λ}). Note that in this setup $\{(W_t^{\phi}), (W_t^{\lambda}), (W_t^{\text{OIS}})\}$ are independent $\mathbb{Q}$ - Brownian Motions.

2 Pricing

In order to price SAFEX, JIBAR, Forward Rate Agreements (FRA) and Interest Rate Swaps (IRS) we will first need to define the zero coupon bond and discounted floating rate. The reason for these definitions will become apparent in later sections. The price of a zero coupon bond discounted using (r_t^{OIS}) under the risk neutral measure $\mathbb{Q}$ is

$$P_{tT}^{\text{OIS}} := \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u^{\text{OIS}}du\right)\middle|\mathcal{F}_t\right] \tag{2.16}$$

and the discounted floating rate is

$$L_s(t,T) := \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_s^T r_u^{\text{OIS}}du\right)L(t,T)\middle|\mathcal{F}_s\right]. \tag{2.17}$$

But then

$$L(t,T) = \frac{1}{T-t}\left(\frac{A_{tT}^L}{Q_{tT}^L} - 1\right),$$

with

$$A_{tT}^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^T \phi_u du\right)\middle|\mathcal{F}_t, \phi_t = 0\right],$$

$$Q_{tT}^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u^{\text{OIS}}\phi_u\right)du\middle|\mathcal{F}_t, \lambda_t = 0\right].$$

Therefore

$$\begin{aligned}
L_s(t,T) = & \frac{1}{T-t}\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_s^T r_u^{\text{OIS}}du\right)\left(\frac{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^T \phi_u du\right)|\mathcal{F}_t,\phi_t=0\right]}{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T (r_u^{\text{OIS}}+\lambda_u)du\right)|\mathcal{F}_t,\lambda_t=0\right]}-1\right)\middle|\mathcal{F}_s\right], \\
= & \frac{1}{T-t}\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_s^t r_u^{\text{OIS}}du\right)\left(\frac{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^T \phi_u du\right)|\mathcal{F}_t,\phi_t=0\right]}{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T (r_u^{\text{OIS}}+\lambda_u)du\right)|\mathcal{F}_t,\lambda_t=0\right]}-1\right)\right. \\
& \left.\exp\left(-\int_t^T r_u^{\text{OIS}}du\right)\middle|\mathcal{F}_s\right], \\
= & \frac{1}{T-t}\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_s^t r_u^{\text{OIS}}du\right)\left(\frac{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^T \phi_u du\right)|\mathcal{F}_t,\phi_t=0\right]}{\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T (r_u^{\text{OIS}}+\lambda_u)du\right)|\mathcal{F}_t,\lambda_t=0\right]}-1\right)\right. \\
& \left.\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u^{\text{OIS}}du\right)\middle|\mathcal{F}_t\right]\middle|\mathcal{F}_s\right].
\end{aligned} \tag{2.18}$$

2.1 SAFEX and JIBAR

The South African Futures Rate (SAFEX) is the weighted average daily rate on overnight deposits. Let $\Delta = \frac{1}{365}$, then the overnight rate can be calculated as

$$L(t,t+\Delta) = \frac{1}{\Delta}\left(\frac{A_\Delta^L}{Q_\Delta^L}-1\right). \tag{2.19}$$

where

$$A_\Delta^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^{t+\Delta}\phi_u du\right)\middle|\mathcal{F}_t,\phi_t=0\right],$$

$$Q_\Delta^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\Delta} r_u^{\text{OIS}}+\lambda_u du\right)\middle|\mathcal{F}_t,\lambda_t=0\right].$$

The Johannesburg Interbank Agreed Rate (JIBAR) is traded for tenors of $1,3,6,9$ and 12. The market benchmark rate is then defined as,

$$L(t,T) = \frac{1}{T-t}\left(\frac{A_{tT}^L}{Q_{tT}^L}-1\right). \tag{2.20}$$

where

$$A_{tT}^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^T\phi_u du\right)\middle|\mathcal{F}_t,\phi_t=0\right],$$

$$Q_{tT}^L = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u^{OIS}+\lambda_u du\right)\middle|\mathcal{F}_t,\lambda_t=0\right].$$

2.2 Forward Rate Agreements

A Forward Rate Agreement (FRA) is an over the counter agreement to exchange a floating payment for a fixed payment according to the benchmark rate at some future date. For $T_0 < T_{i-1} < T_i$ the floating leg of a FRA with reset date T_{i-1} pays $L(T_{i-1}, T_i)\delta$ at time T_i. The fixed leg pays $f(T_0; T_{i-1}, T_i)\delta$, where $f(T_0; T_{i-1}, T_i)$ is the fair FRA rate, agreed upon at time T_{i-1}. The fair FRA rate is calculated by equating the expected discounted value of the fixed leg with the expected discounted value of the floating leg, that is

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_{T_0}^{T_i} r_u^{\text{OIS}} du\right) f(T_0; T_{i-1}, T_i)\delta \middle| \mathcal{F}_{T_0}\right] \\ =&\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_{T_0}^{T_i} r_u^{\text{OIS}} du\right) L(T_{i-1}, T_i)\delta \middle| \mathcal{F}_{T_0}\right]\end{aligned}$$

which implies that

$$\delta f(T_0, T_{i-1}, T_i)\delta P_{T_0,T_i}^{\text{OIS}} = \delta L_{T_0}(T_{i-1}, T_i).$$

Therefore

$$f(T_0, T_{i-1}, T_i) = \frac{L_{T_0}(T_{i-1}, T_i)}{P_{T_0,T_i}^{\text{OIS}}}. \tag{2.21}$$

2.3 Interest Rate Swaps

Interest rate swaps are contracts between two parties to exchange a stream of fixed-rate payments for a stream of floating-rate payments indexed to JIBAR of a particular maturity. Consider the discrete tenor structure:

$$t = T_0 < T_1 < \cdots < T_N = T,$$

and let $\delta = T_i - T_{i-1}$ denote the length of the tenor dates. Then for some initiation time T_0, define

$$T_i = T_0 + i\delta,$$

for $i = 1, 2, \ldots, n$, where n is the number of interest payments. Let the fair swap rate be denoted by $S(T_0, n)$. The fair rate must ensure that the present value of the floating payments is equal to the present value of the fixed payments. This can be expressed as

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}}\left[\sum_{i=1}^{n/\delta} \exp\left(-\int_{T_0}^{T_i} r_u^{\text{OIS}} du\right) \delta L(T_{i-1}, T_i) \middle| \mathcal{F}_{T_0}\right] \\ =&\mathbb{E}_{\mathbb{Q}}\left[\sum_{i=1}^{n/\delta} \exp\left(-\int_{T_0}^{T_i} r_u^{\text{OIS}} du\right) \delta S(T_0, n) \middle| \mathcal{F}_{T_0}\right].\end{aligned}$$

Thus

$$\begin{aligned}&\delta\sum_{i=1}^{n/\delta}\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_{T_0}^{T_i}r_u^{\text{OIS}}du\right)L(T_{i-1},T_i)\middle|\mathcal{F}_{T_0}\right]\\=&\delta S(T_0,n)\sum_{i=1}^{n/\delta}\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_{T_0}^{T_i}r_u^{\text{OIS}}du\right)\middle|\mathcal{F}_{T_0}\right],\end{aligned}$$

which implies that

$$S(T_0,n)=\frac{\sum_{i=1}^{n/\delta}L_{T_0}(T_{i-1},T_i)}{\sum_{i=1}^{n/\delta}P^{\text{OIS}}_{T_{i-1},T_i}}. \tag{2.22}$$

3 Affine Transformation

The aim of this report is to estimate the above model using South African data. Ideally one would have access to data pertaining to the individual JIBAR panel members credit spread. This would allow one to estimate the dynamics of (λ_t) and then using those dynamics, estimate the dynamics of the entire model. This is not possible in the South African market as we do not have access to South African credit default swap data. This implies that we are not able to disentangle the effects of (ϕ_t) and (λ_t) on the pricing formulae above. In order to get around this problem, we shall set $\lambda_t \equiv 0$ for all $t \geq 0$. This implies that instead of modelling the JIBAR/OIS spread as two components, it is modelled as one component.

In order to solve the conditional expectations (2.16) and (2.18), we use the methodology of Duffie et al. (2000). This is formalised as follows:
Define the state vector process (X_t), as

$$X_t=\begin{bmatrix}r_t^{\text{OIS}}\\ \phi_t\\ \xi_t^{\phi}\end{bmatrix}. \tag{2.23}$$

The SDE for (X_t) is then

$$dX_t=\mu(X_t)dt+\sigma(X_t)dW_t+dJ_t \tag{2.24}$$

where $\mu(X_t)\in\mathbb{R}^3$, $\sigma(X_t)\in\mathbb{R}^3\times\mathbb{R}^3$ and W_t is a 3-dimensional $\mathbb{Q}$ - Brownian Motion.

$$J_t=\sum_{k=1}^{N(t)}Z_k$$

is a 3-dimensional jump process, where $\{Z_1, Z_2, ...\}$ are independent exponential random variables and $N(t)$ is a counting process with intensity $\lambda(X_t) \in \mathbb{R}^3$. Using equations (2.2), (2.3) and (2.15) we can write this SDE in affine form. That is,

$$\mu(x) = K_0 + K_1 x, \quad \sigma(x)\sigma(x)^\top = H_0 + H_1 x \quad \text{and} \quad \lambda(X_t) = L_0 + L_1 X_t.$$

(See appendix A for more detail).
It then follows for $\rho_0 \in \mathbb{R}^3$, $\rho_1 \in \mathbb{R}^3 \times \mathbb{R}^3$, $a \in \mathbb{R}^3$ and $\omega \in \mathbb{R}^3 \times \mathbb{R}^3$ that

$$\mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_t^T (\rho_0 + \rho_1 \cdot X(u))du} e^{a + w \cdot X_T} \,|\mathcal{F}_t\right] = e^{\varphi(t,T) + \psi(t,T)\cdot X_t} \tag{2.25}$$

where, for each fixed t, the functions $\varphi(\cdot, t)$ and $\psi(\cdot, t)$ solve generalised Riccatti equations

$$\begin{aligned} -\frac{\partial \psi_s(s,t)}{\partial s} &:= \psi(s,t)\cdot K_1 + \frac{1}{2}\varphi(s,t)^\top H_1 \varphi(s,t) + \ell_1 \mathbb{E}_{\mathbb{Q}}\left[e^{\psi(s,t)\cdot Z_t} - 1\right] - \rho_1, \\ -\frac{\partial \varphi_s(s,t)}{\partial s} &:= \psi(s,t)\cdot K_0 + \frac{1}{2}\varphi(s,t)^\top H_0 \varphi(s,t) + \ell_0 \mathbb{E}_{\mathbb{Q}}\left[e^{\psi(s,t)\cdot Z_t} - 1\right] - \rho_0, \end{aligned} \tag{2.26}$$

with the boundary conditions $\varphi(t,t) = a$ and $\psi(t,t) = w$.

4 Implementation on South African Data

We have derived pricing formulae in section 2 for various South African market instruments and in appendix B we will give the exact specifications of the market instruments we are using. The time period of data available to us is daily data from 4 January 2004 to 18 October 2018 with all instruments having full data available from 2005. In this section we qualitatively analyse these instruments. Figure 2.1 shows the evolution of the JIBAR 3-month rate over the period January 2005 to October 2019.

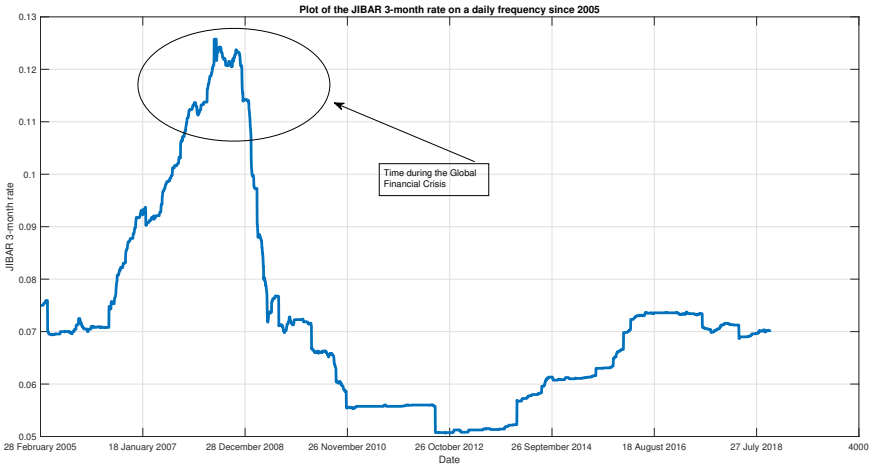


Figure 2.1: 3-month JIBAR from January 2005 to October 2019.

The increase in JIBAR around the time of the Global Financial Crisis shows the impact of the event on the South African market. Thereafter, the rates seem to stabilise around 5-7.5% and this is in part a function of the South African Reserve Bank's target to control inflation and stabilise the currency.

The following cross section of the market on 18 October 2018, shows each of the instruments plotted according to their term structure (shown as multiples of a year).

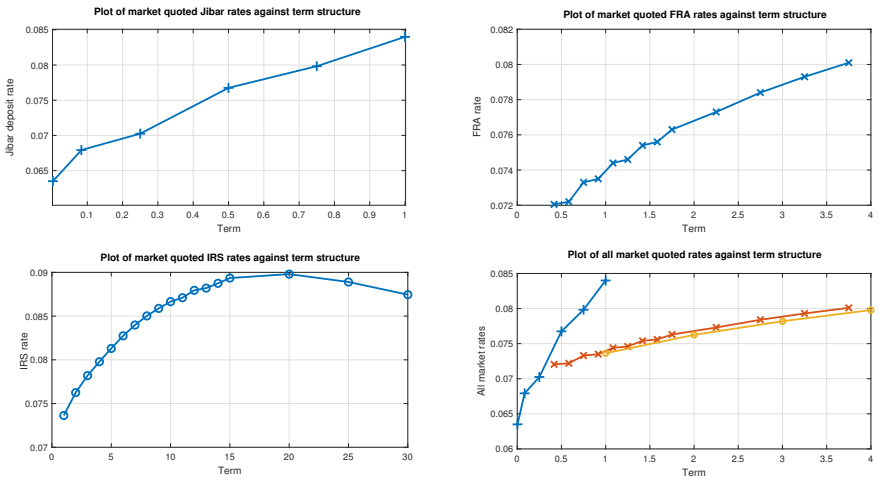


Figure 2.2: Cross-sectional look at the South African Interest Rate market on 18 October 2018.

From figure 2.2, we can see that all rates are increasing as a function of their maturity, except for the quoted swap rates which seem to decline after 20 years. The quoted forward rates are increasing as a function of the term and hence we would expect the quoted swap rates to increase as well. The long end decline of the swap rates could be a function of demand for these types of swaps as well as a less liquid market using derivative instruments that have this long a maturity.

If we look at the market quoted prices over the entire history of the dataset we find some anomalies. During the month of February 2005 the quoted market prices for a 10,12 and 15-year IRS were around 7.7% while the 11,13 and 14-year quotes were around 9.5%. There are various other market anomalies but nothing as severe as the disparate quotes for the IRS during 2005. Following the JIBAR curve, any inference or models that are to be fit longitudinally are better fit to the period after 2010 where rates are more stable.

Chapter 3

Estimation

1 Unscented Kalman Filter

We wish to estimate the parameters used in our model dynamics. Since we are not just working with cross-sectional data, but rather a time-series of instrument prices, it becomes necessary to use a filter. We follow the approach of Filipović and Trolle (2013) and use an unscented Kalman filter together with maximum likelihood estimation. For a given parameter set, we want to compute the likelihood of seeing our observed prices if those were the correct parameters. This value will then be fed to an optimiser to calculate the next estimate of the parameter set.

1.1 Market Price of Risk

When we price instruments for a given parameter set, we are working in the risk-neutral world under the measure $\mathbb{Q}$. When we simulate the state transition of the latent vector, we need to know its dynamics under the equivalent real world measure $\mathbb{P}$. These measures are equivalent, but may be different. It is necessary to model the market price of risk to allow for this when changing measures. We adjust the approach used in Filipović and Trolle (2013) to our model by assuming that the market price of risk process is given by

$$\Gamma(t) = \left(\Gamma_r \sqrt{r^{OIS}(t)}, \Gamma_\phi, \Gamma_\lambda, \Gamma_{\xi^\phi} \sqrt{\xi^\phi(t)}, \Gamma_{\xi^\lambda} \sqrt{\xi^\lambda(t)}\right)^T. \tag{3.1}$$

Γ_ϕ and Γ_λ are set to zero as by the definition of the model, the credit and funding deterioration processes ((λ_t) and (ϕ_t)) are measures of the entity-specific credit and funding deterioration. These processes therefore model an idiosyncratic risk and should not have an associated market risk premium.

The Brownian motions under $\mathbb{P}$ are as follows:

$$\begin{aligned}
dW_t^{OIS\prime} &= dW_t^{OIS} + \Gamma_r \sqrt{r^{OIS}(t)}dt, \\
dW_t^{\phi\prime} &= dW_t^{\phi} + \Gamma_{\xi^\phi} \sqrt{\xi^\phi(t)}dt, \text{ and} \\
dW_t^{\lambda\prime} &= dW_t^{\lambda} + \Gamma_{\xi^\lambda} \sqrt{\xi^\lambda(t)}dt,
\end{aligned}$$

where dW_t^{OIS}, dW_t^{ϕ} and dW_t^{λ} are the Brownian motions under $\mathbb{Q}$ introduced in 1. The new dynamics are then given by

$$\begin{aligned}
dr_t^{\text{OIS}} =&(\alpha^{\text{OIS}}(\theta^{\text{OIS}} - r_t^{\text{OIS}}) + \alpha^{\text{OIS},\phi}(\theta^{\phi} - \xi_t^{\phi}) + \alpha^{\text{OIS},\lambda}(\theta^{\lambda} - \xi_t^{\lambda}) + \Gamma_r \sigma^{OIS} r_t^{OIS})dt \\
&+ \sigma^{\text{OIS}} \sqrt{r_t^{\text{OIS}}} dW_t^{\text{OIS}}. \\
d\xi_t^{\phi} =&(\alpha^{\phi}(\theta^{\phi} - \xi_t^{\phi}) + \Gamma_{\xi^\phi}\sigma^{\phi}\xi_t^{\phi})dt + \sigma^{\phi}\sqrt{\xi_t^{\phi}} dW_t^{\phi}. \\
d\xi_t^{\lambda} =&(\alpha^{\lambda}(\theta^{\lambda} - \xi_t^{\lambda}) + \Gamma_{\xi^\lambda}\sigma^{\lambda}\xi_t^{\lambda})dt + \sigma^{\lambda}\sqrt{\xi_t^{\lambda}} dW_t^{\lambda}.
\end{aligned}$$

1.2 Algorithm Overview

A Kalman filter takes two processes - a hidden process (x_t) we call the latent state, and an observed process (y_t) whose value at each time step is known. We also supply functions f and h that model how (x_t) evolves in time and how (x_t) and (y_t) are related. In our simplified model we cannot distinguish between downgrade risk and liquidity risk, and it follows that $(x_t) = [r_t^{OIS}, \xi_t^{\phi}]^T$ and each y_t is the vector of prices of the thirty-two instruments[1] in our data-set at time t. We have $f : {\mathbb{R}_+}^2 \times {\mathbb{R}_+}^2 \longmapsto {\mathbb{R}_+}^2$ such that

$$x_{t+1} = f(x_t, w_t),$$

where (w_t) is a non-additive noise process corresponding to the Brownian motion in (2.15). Similarly, $h : {\mathbb{R}_+}^2 \longmapsto {\mathbb{R}_+}^{32}$ such that

$$y_t = h(x_t) + v_t,$$

where v_t is an additive measurement noise term. h maps x_t to the prices calculated in 2. We are then required to provide covariance matrices R and Q for the measurement and process noise respectively. A regular Kalman filter uses a linear h and an extended Kalman filter uses a first order linearization of a non-linear h. The unscented Kalman filter maintains the full non-linearity of h, and instead approximates the distribution by generating a set of "sigma points" spread about x_t. The full algorithm is specified in Appendix C.

[1] We are only estimating using the FRAs and IRSs. Looking at the bottom right of figure 2.2 it seems plausible that the market for deposits is out of sync of that for the other instruments. We use this as justification to drop the deposits from our estimation data set.

2 Optimisation

Parameter estimation involves finding a parameter set $\hat{\Theta}$ such that the log-likelihood, $\mathcal{L}(\Theta)$, is maximised, i.e:

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta).$$

The approach taken with the unscented Kalman Filter implies that the density function of our observed variables (Y_t) is approximated by a Normal distribution and hence the above estimation procedure is known as a quasi-maximum likelihood estimation procedure.
In our setup, the parameter set we aim to estimate is defined as

$$\Theta = \{\alpha^{\text{OIS}}, \alpha^{\phi}, \alpha^{\text{OIS},\phi}, \kappa^{\phi}, \Gamma^{\text{OIS}}, \Gamma^{\phi}, \mu^{\phi}, \sigma^{\text{OIS}}, \sigma^{\phi}, \theta^{\text{OIS}}, \theta^{\phi}\}.$$

In the optimisation routine, certain model restrictions are placed on the parameter set (eg. the Feller condition: $2\alpha^{\text{OIS}}\theta^{\text{OIS}} > (\sigma^{\text{OIS}})^2$). In the South African market, interest rates have historically and currently been above the 7% threshold, and as such we do not allow models where the interest rates can go to zero or be negative.
Assessing the quality of our parameter estimates is done by considering their standard errors. This is done by inverting the information matrix

$$\text{Var}(\Theta) = I(\Theta)^{-1}, \tag{3.2}$$

where

$$I(\Theta) = -\mathbb{E}[H(\Theta)],$$

and $H(\Theta)$ is the Hessian matrix defined as

$$H(\Theta) = \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \Theta \Theta^T}.$$

In essence, taking the square root of the diagonal entries in (3.2) gives us the standard errors associated with each of the estimated parameters. From there we can construct confidence intervals and perform hypothesis tests to determine the significance of each estimated parameter. An assumption we make is that the expectation $\mathbb{E}[H(\Theta)]$ is just equal to $H(\Theta)$ and this should not be a problem as our data set is not small (i.e. no small sample problems).
We will also consider first-order optimality conditions when assessing the output. Essentially if the value of this quantity is less than 10^{-7} then we can be sure that the gradient of the (negative) log-likelihood function is close to zero at the evaluation points.
Further practical implementation details are given in appendix D.

Chapter 4

Results

In this chapter we will present the results from the estimation procedure of section 3. Due to time constraints, the results below were run on a limited set of data. The model was estimated using data for 23 August 2018 to 31 October 2018. While we acknowledge that this limits the intepretability of our results quite significantly, the parameters in the model are not time dependent and so due to ever changing market conditions, one shouldn't estimate on too big a data set. It is for this reason that our results still hold some validity and allow for meaningful interpretation.

1 Implied OIS Discount Curve from Cross-Section Estimation

The following results show two sample cross-section estimations. Given the complexity of the model and the nature of the cross-section estimation, it is expected that the output of these estimations will lack robustness or intepretability. The aim of this step was to get a set of reasonable starting parameters to feed in to the longitudinal estimation. The dates used were 1st May 2008 and 18th September 2018. Table 4.1 shows the estimated parameters for the two estimations. Figures 4.1 and 4.2 show the estimated overnight, deposit, FRA and swap rates for the given estimation.
Figures 4.3, 4.4, show the OIS discount curves from these two estimations.

Parameter	1 May 2008	18 September 2018
α^{OIS}	1.70	0.32
θ^{OIS}	0.15	0.09
$\alpha^{OIS,\phi}$	4.18	6.08
σ^{OIS}	0.48	0.02
κ^{ϕ}	5.41	7.59
α^{ϕ}	8.17	7.81
θ^{ϕ}	5.63	7.64
σ^{ϕ}	2.61	3.88
μ^{ϕ}	4.94	7.24

Table 4.1: Parameters for cross-sectional estimations

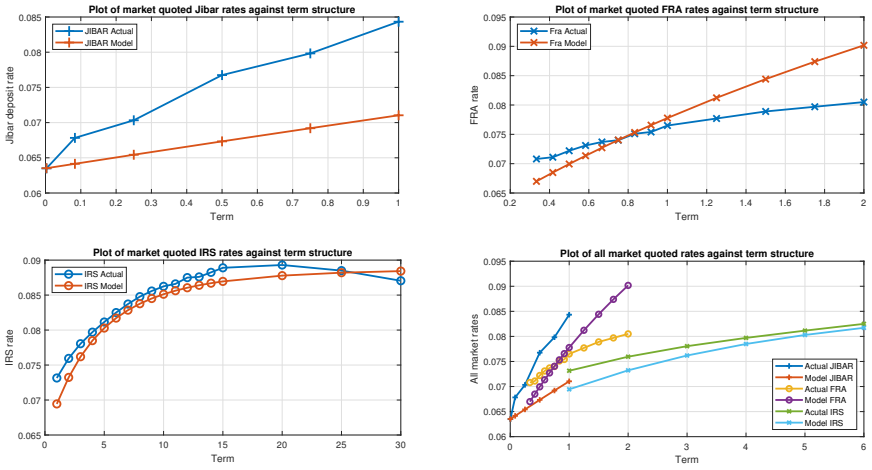


Figure 4.1: Cross-sectional estimation for the 18th September 2018 showing deposit rates (top left), FRA rates (top right), IRS rates (bottom left) and combined rates (bottom right)

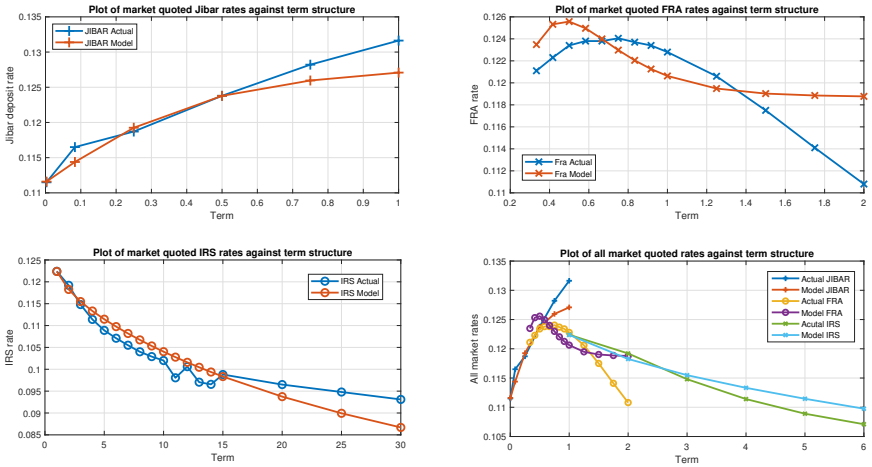


Figure 4.2: Cross-sectional estimation for the 1st May 2008 showing deposit rates (top left), FRA rates (top right), IRS rates (bottom left) and combined rates (bottom right)

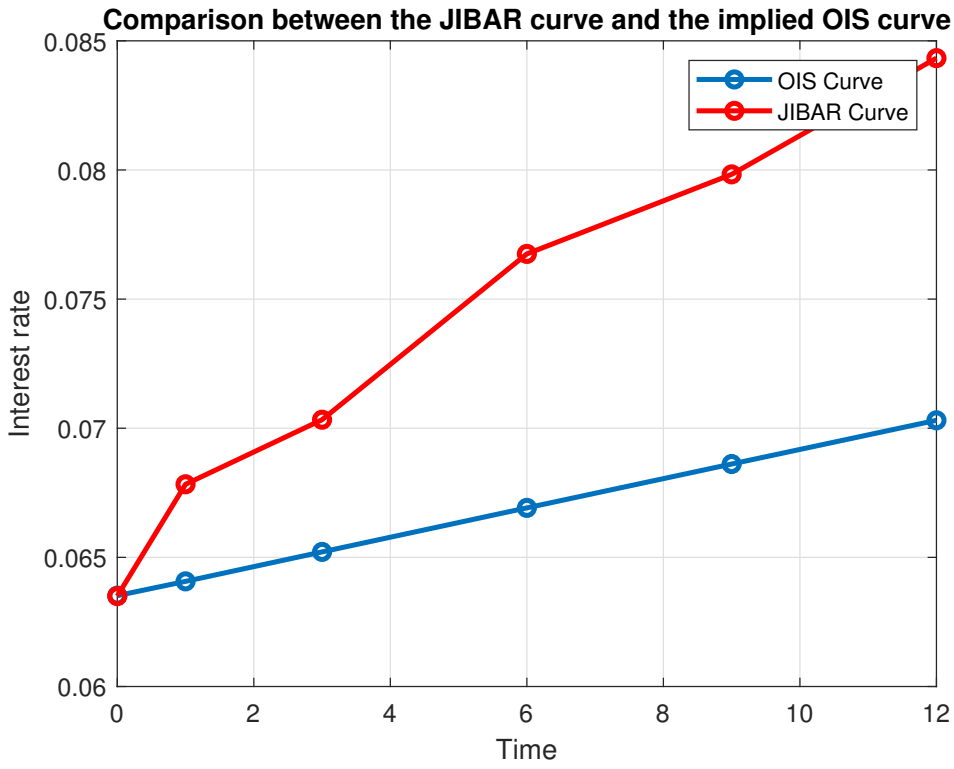


Figure 4.3: Cross-sectional JIBAR and OIS curves for 18th September 2018

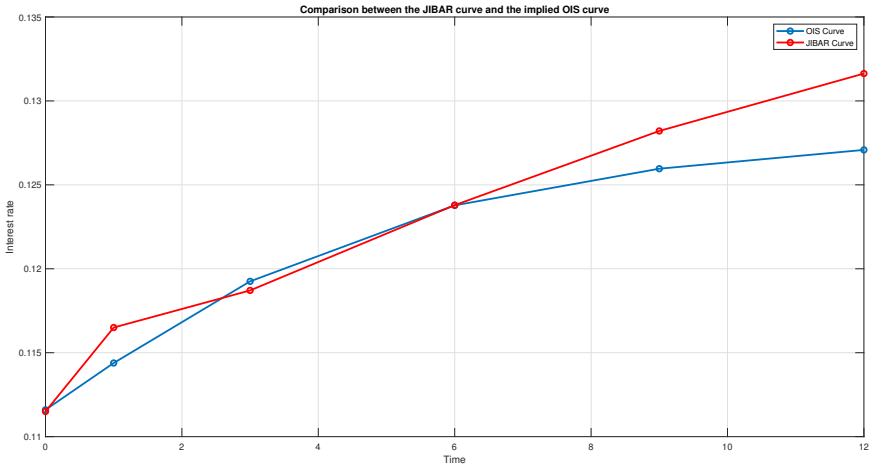


Figure 4.4: Cross-sectional JIBAR and OIS curves for 1st May 2008

2 Implied OIS Discount Curve from Longitudinal Estimation

The aim of this report is to fit the model of Backwell et al. (2019) to South African data in order to infer a term-structure of overnight index swap rates. The estimation procedure of section 3 was used to produce a point estimate of the model parameters from section 2. Below is a table of model parameter values:

Parameter	Estimate
α^{OIS}	9.62
α^{ϕ}	3.73
$\alpha^{\text{OIS},\phi}$	0.01
κ^{ϕ}	10.00
σ^{OIS}	0.01
σ^{ϕ}	0.001
θ^{OIS}	0.00
θ^{ϕ}	1.86
Γ_r	0.13
Γ_{ϕ}	0.04
μ^{ϕ}	0.50

Using these model parameter values, we test the goodness-of-fit from a cross-sectional point of view. That is we can use the model to price the deposit, FRA and swaps that are observable in our data. We can then compare this output with the actual prices observed. Figure 4.5 give this comparison on 31 October 2018.

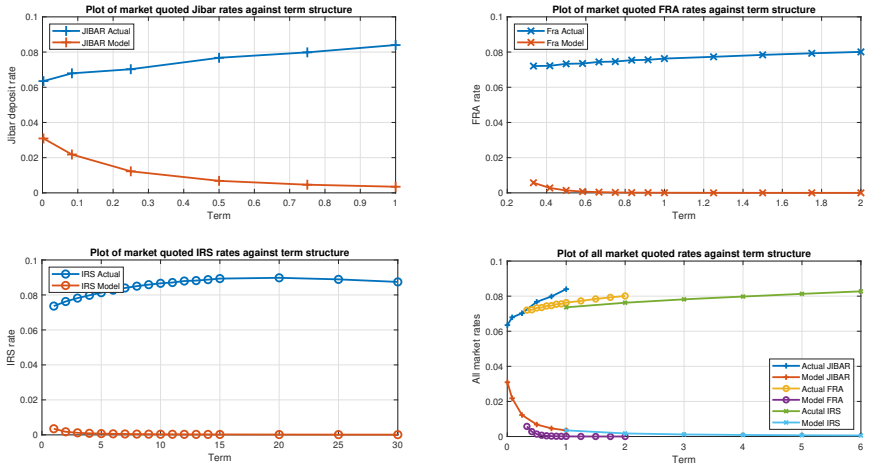


Figure 4.5: Cross-sectional look at model fit on 31 October 2018 showing deposit rates (top left), FRA rates (top right), IRS rates (bottom left) and combined rates (bottom right)

From figure 4.5 we can see that the model does not fit the South Africa data well on this day. This could be due to several reasons. Firstly, due to time constraints, the optimisation process was only able to run on a limited set of data, this meant that parameter estimation would be based on a small sample size and thus inaccurate. Secondly, from figure 4.5 we observed that the deposit rates, FRA rates and IRS rates have large inconsistencies. By assumption, the model we have estimated bases its pricing methodology on a market which prices these instruments in a consistent manner. The model estimation procedure then cannot distinguish between model error or inconsistencies in the data. Thirdly, because of the inconsistencies in the data through time, the Unscented Kalman Filter was unable to accurately estimate the state process, this meant that the likelihood calculation was unstable and unable to infer the correct parameter estimates.
Figure 4.6 shows the actual JIBAR observed rate for 31 October 2018 and the model implied OIS rate for that date.

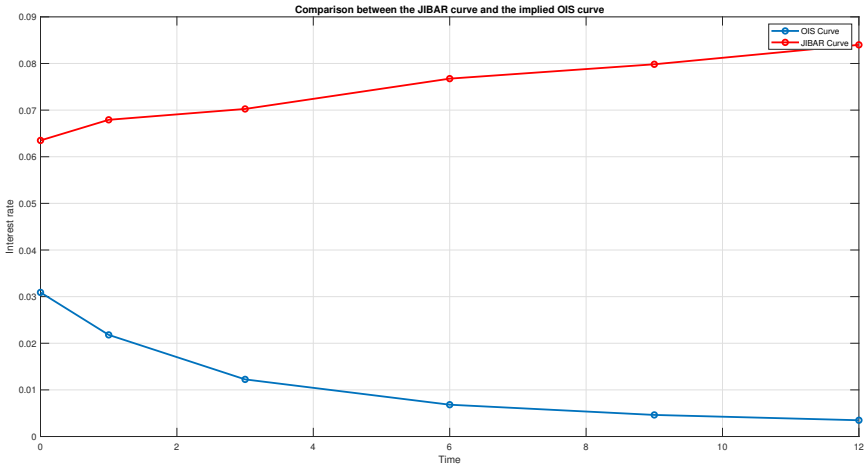


Figure 4.6: JIBAR and OIS curve on 31 October 2018.

Chapter 5

Conclusion and Further Research

The aim of this report was to use the roll-over risk model of Backwell et al. (2019) estimated on South African data to infer an overnight index swap discount curve. In this market there are no traded basis or overnight index swaps, we therefore model the JIBAR/OIS spread using roll-over risk, inferring a term structure of OIS rates from the observed JIBAR rates. The purpose of the exercise is to be able to price collateralised derivative contracts. These contracts have collateral posted to a margin account on a daily mark-to-market basis, therefore in order to price these instruments, it is required to discount at the appropriate OIS discount factor. Estimation on a South African data set required us to develop the appropriate pricing formulae for each South African instrument that we could observe. Using these pricing formulae we were able to estimate the dynamics of the roll-over risk model. This estimation was done using an Unscented Kalman Filter coupled with a likelihood maximisation routine. The results are consistent with our assertions that JIBAR should charge a premium over and above the OIS rate to avoid roll-over risk.

Further research could be done into a robustness study of our estimation procedure. This could be done by excluding instruments or observations from our data set and observing the effects of these exclusions on our OIS discount curve. It would also be beneficial to test the output of our estimation procedure on out-of-sample data. That is, perform the estimation on a certain set of data and then test the goodness-of-fit on a separate set of data. The estimation procedure developed in this report produces a point estimate of the model parameters and the OIS discount curve, it would be beneficial to use a particle filter in the parameter estimation step in order to output a posterior distribution of each model parameter. This would allow us to derive a confidence interval for the OIS discount curve which would be useful from a practical standpoint as it gives an indication of the uncertainty in the model parameters.

As stated in the results section, the data used to estimate the model had inconsistencies between the deposit, FRA and IRS rates, this is stated as a reason for the poor model fit of our model in section 2. Further research could be done to coerce the data in such a way to limit the effect that these inconsistencies have on the fit of the model.

Bibliography

Alfeus, M., Grasselli, M., Schlogl, E., 2017. A consistent stochastic model of the term structure of interest rates for multiple tenors.

Backwell, A., Macrina, A., Schlogl, E., Skovmand, D., 2019. Term rates, multicurve term structures and overnight rate benchmarks: a roll-over risk approach.
URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3399680

Boenkost, W., Schmidt, W. M., 2004. Cross currency swap valuation. HfB Business School of Finance & Management.

Duffie, D., 2008. Financial modelling with affine process.
URL http://www.nccr-finrisk.uzh.ch/media/pdf/affix.pdf

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68 (6), 1343–1376.

Filipović, D., Trolle, A. B., 2013. The term structure of interbank risk. Journal of Financial Economics 109 (3), 707–733.

Kijima, M., Tanaka, K., Wong, T., 2009. A multi-quality model of interest rates. Quantitative Finance 9 (2), 133–145.

Lando, D., 2009. Credit risk modeling: theory and applications. Princeton University Press.

MathWorks, 2016a. Matlab filter package documentation.
URL http://tiny.cc/fmtc2

MathWorks, 2016b. Unscented Kalman filter documentation.
URL http://tiny.cc/fmtc

Appendix A

Riccati Equations

In this section we develop pricing formula for each of the risk neutral expectations used in the pricing equations for the various interest rate derivatives. We give a theoretical overview of the techniques required to construct the relevant Ricatti Equations and then we proceed to give the relevant formulas for all the expectations.

1 Background

We consider a general jump-diffusion model, X_t, on $D = \mathbb{R}^m \times \mathbb{R}^n$ specified by:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t + dJ_t$$

where

1. $\mu : D \to \mathbb{R}^d, \;\; \sigma : D \to \mathbb{R}^{d\times d}$
2. W_t is a Brownian motion in $\mathbb{R}^d$
3. J_t is a right-continuous jump process, with jumps sizes following some distribution law $\nu(dx)$ with jump intensity given by the process $\lambda(X_{t-})$, for some $\lambda : D \to [0, \infty)$
4. We assume that the Brownian motion W, the jump sizes of Z and the jump times are all independent.

In order to develop a tractable model we require that drift, diffusion and jump intensity all be affine functions of X_t, i.e. we require that:

$$\mu(x) = b + \beta_1 x_1 + \cdots + b_d x_d$$

$$\sigma(x)\,\sigma(x)^{\mathsf{T}} = a + \alpha_1 x_1 + \cdots + \alpha_d x_d$$

$$\lambda(x) = m + \mu_1 x_1 + \cdots + \mu_d x_d$$

where $b,\ \beta_i \in \mathbb{R}^d; a,\ \alpha_i \in \mathbb{R}^{d\times d}$ and $m,\ \mu_i \in [0,\ \infty)$.
We are now able to specify the exponential affine form of the following expectation used to calculate the characteristic function of X_t:

$$\mathbb{E}\left[e^{u\cdot X_T} \mid \mathcal{F}_t\right] = \exp\left(\phi(T-t,\ u) + X_t\psi(T-t,\ u)\right),\ \ 0 \le t \le T$$

and ϕ & ψ solve the following system of Riccati equations:

$$\phi(t,\ u) = F(\psi(t,\ u)), \qquad \phi(0,\ u) = 0$$

$$\psi(t,\ u) = R(\psi(t,\ u)), \qquad \psi(0,\ u) = u$$

with the function F & R being defined by the following system of equations:

$$F(u) = b^\intercal u + \frac{1}{2}u^\intercal a\, u + mk(u)$$

$$R_1(u) = \beta_1^\intercal u + \frac{1}{2}u^\intercal \alpha_1 u + \mu_1\kappa(u)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$R_d(u) = \beta_d^\intercal u + \frac{1}{2}u^\intercal \alpha_d u + \mu_d\kappa(u)$$

Here $\kappa(u) = \int_{\mathbb{R}^d}(e^{u\cdot x} - 1)\,\nu(dx)$. We note that if $\nu(dx)$ follows an exponential distribution with parameter ϑ then elementary calculus gives $\kappa(u) = \frac{u}{\vartheta - u}$.
As a simple example of the above, consider the following diffusion process:

$$dX_t = -\lambda(X_t - \theta)\,dt + \sigma\sqrt{X_t}dW_t + dJ_t$$

where J_t is a pure jump process with constant intensity ε and jumps sizes exponentially distributed with parameter α. Consequently the functions F and R can be specified to construct the given Riccati equations:

$$F(u) = \lambda\theta u + \frac{\epsilon u}{\alpha - u}, \quad R(u) = -\lambda u + \frac{u^2}{2}\sigma^2$$

2 Implementation Methodology: Riccati Equation

The general time-inhomogeneous Riccati equation in Duffie (2008), page 29 implementation in Matlab requires the use of Matlab's *ode23tb* function . The given function can solve any (well specified) system of ODE's given a set of initial conditions. In our implementation we only consider terminal boundary value conditions and hence we have to solve the differential equations backwards in time, i.e. our

terminal conditions essentially become initial conditions and we move backwards in time until we get our solution at time s.
The coded solution is split into two parts: The initialisation of all the parameters together the wrapper around *ode23tb*and the output and the helper function *bvpfcn* containing a definition of the actual system of ODE's.
A problem in the implementation is that the given Riccati Equation are given in a Matrix form while *ode23tb* can only handle vector constraints. To this end we allow the computer to compute a single column vector of values and then these are passed to the helper functions. This input is reshaped into a matrix (of the appropriate dimensions) and calculations continue using this matrix form. At the end of the helper function the final matrix is then again converted into a single column by stacking all the columns underneath each other. This effectively allows us to solve the Matrix equations with relative ease.
A technical aspect of the Riccati equations involves the expectation:
$E\left[e^{\psi(s,\,t)Z_i}-1\right]=k\left(\psi\left(s,t\right)\right)$. This expectation is implemented on the assumption that the jump sizes under the jump process follow an i.i.d exponential distribution with some parameter μ. We can explicitly then calculate this expectation:

$$k\left(s\right)=\int_0^\infty\left(e^{sx}-1\right)ue^{-\mu x}dx=\frac{s}{\mu-s}$$

Since the equations involving $\psi\left(s,t\right)$ are matrix valued, we take the quotient to mean element-wise division. In cases where we encounter any $\frac{0}{0}$ division we set the corresponding values in output matrix (which will contain NaNs (not-a-number value)) to zero since by definition if $s=0$ in the above integral then the integral evaluates to 0.
Lastly, since the Riccati equations are matrix and vector valued, we allow the user to obtain an appropriate slice of these variables corresponding to the type of expectation they are trying to take. In essence we are computing the solution to 5 expectations simultaneously, 3 of which are only needed in all our pricing formulae and hence we can take any appropriate slice to work out the correct value for that expectation.

3 Process Evaluation

For the process

$$X_t=\left[\begin{array}{c}r_t^{\text{OIS}}\\ \phi_t\\ \xi_t^{\phi}\end{array}\right], \tag{A.1}$$

with SDE

$$dX_t=\mu(X_t)dt+\sigma(X_t)dW_t+dJ_t \tag{A.2}$$

where W_t is a 5-dimensional Wiener process and

$$J_t = \sum_{t=1}^{N(t)} X_t,$$

where $X_1, X_2, \ldots$ are independent identically distributed random variables in $\mathbb{R}^d$ which are independent of W and N is a counting process with intensity. Using the method outlined in Duffie et al. (2000) we have the following:

$$\begin{aligned}\mu(x) &= K_0 + K_1 x \\ &= \begin{bmatrix} \alpha\theta \\ 0 \\ \alpha^{\phi}\theta^{\phi} \end{bmatrix} + \begin{bmatrix} -\alpha^{\text{OIS}} & 0 & -\alpha^{\text{OIS},\phi} \\ 0 & -\kappa^{\phi} & 0 \\ 0 & 0 & -\alpha^{\phi} \end{bmatrix},\end{aligned} \tag{A.3}$$

where $\alpha\theta = \alpha^{\text{OIS}}\theta^{\text{OIS}} + \alpha^{\text{OIS},\phi}\theta^{\text{OIS},\phi}$.

$$\begin{aligned}\sigma(x)\sigma(x)^{\top} &= H_0 + H_1 x \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} (\sigma^{\text{OIS}})^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (\sigma^{\phi})^2 \end{bmatrix},\end{aligned} \tag{A.4}$$

with stochastic intensity of jumps given by

$$\begin{aligned}\Lambda^{(j)}(x) &= \ell_0 + \ell_1 \cdot x \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},\end{aligned} \tag{A.5}$$

Appendix B

Data

The following instruments (daily timeseries), with Bloomberg codes, were available to us to use from 4 January 2000 until 31 October 2018. DEP refers to a market reference rate indexed to JIBAR, FRA refers to a Forward Rate Agreement indexed to 3-month JIBAR with the quoted tenor indicating the start and end dates to which the quoted rate should apply, i.e. "1mx4m would indicate a FRA applying in one months time for the next 3 months. "IRS indicates an interest rate swap linked to 3-month JIBAR with both floating and fix legs having a 3-month frequency.

Instrument	Tenor	Unadj Term	BBG Code	Data Availability
DEP	**1d**	**0,00**	**RAONON**	4 January 2000
DEP	**1m**	**0,08**	**JIBA1M**	4 January 2000
DEP	**3m**	**0,25**	**JIBA3M**	4 January 2000
FRA	**1mx4m**	**0,33**	**SAFR0AD**	4 January 2000
FRA	**2mx5m**	**0,42**	**SAFR0BE**	4 January 2000
DEP	**6m**	**0,50**	**JIBA6M**	4 January 2000
FRA	**3mx6m**	**0,50**	**SAFR0CF**	4 January 2000
FRA	**4mx7m**	**0,58**	**SAFR0DG**	4 January 2000
FRA	**5mx8m**	**0,67**	**SAFR0EH**	4 January 2000
DEP	**9m**	**0,75**	**JIBA9M**	28 February 2005
FRA	**6mx9m**	**0,75**	**SAFR0FI**	4 January 2000
FRA	**7mx10m**	**0,83**	**SAFR0GJ**	4 January 2000
FRA	**8mx11m**	**0,92**	**SAFR0HK**	4 January 2000
DEP	**12m**	**1,00**	**JIBA12M**	4 January 2000
FRA	**9mx12m**	**1,00**	**SAFR0I1**	4 January 2000
IRS	**1y**	**1,00**	**SASW1**	4 January 2000
FRA	**12mx15m**	**1,25**	**SAFR011C**	1 March 2001
FRA	**15mx18m**	**1,50**	**SAFR1C1F**	2 May 2005
FRA	**18mx21m**	**1,75**	**SAFR1F1I**	2 May 2005
FRA	**21mx24m**	**2,00**	**SAFR1I2**	2 May 2005
IRS	**2y**	**2,00**	**SASW2**	4 January 2000
IRS	**3y**	**3,00**	**SASW3**	4 January 2000
IRS	**4y**	**4,00**	**SASW4**	4 January 2000
IRS	**5y**	**5,00**	**SASW5**	4 January 2000
IRS	**6y**	**6,00**	**SASW6**	4 January 2000
IRS	**7y**	**7,00**	**SASW7**	4 January 2000
IRS	**8y**	**8,00**	**SASW8**	4 January 2000
IRS	**9y**	**9,00**	**SASW9**	4 January 2000
IRS	**10y**	**10,00**	**SASW10**	4 January 2000
IRS	**11y**	**11,00**	**SASW11**	18 February 2003
IRS	**12y**	**12,00**	**SASW12**	1 February 2001
IRS	**13y**	**13,00**	**SASW13**	18 February 2003
IRS	**14y**	**14,00**	**SASW14**	18 February 2003
IRS	**15y**	**15,00**	**SASW15**	4 January 2000
IRS	**20y**	**20,00**	**SASW20**	4 January 2000
IRS	**25y**	**25,00**	**SASW25**	21 August 2000
IRS	**30y**	**30,00**	**SASW30**	21 August 2000

Appendix C

Unscented Kalman Filter

1 Algorithm

Here we present the algorithm used in the implementation below. The full documentation is available at MathWorks (2016b). Let (x_k) be the latent state process we wish to estimate, and (y_k) the observed process we have access to. The algorithm demands the following inputs:

- f, which describes the dynamics of (x_k),
- h, which relates the underlying x_k to the observed y_k,
- Q, the covariance of the process noise error,
- R, the covariance of the measurement noise error and
- $x[0|-1]$, the initial estimate of $x[0]$.
- $P[0|-1]$, the initial state covariance. A higher value should be used if one is not sure of $x[0|-1]$.
- α, β and κ which are scaling parameters.

These are used to model the state transition and measurement functions which are given by

$$\begin{aligned} x[k+1] &= f(x[k], w[k]) \text{ or } f(x[k]) + w[k], \\ y[k] &= h(x[k], v[k]) \text{ or } h(x[k]) + v[k], \\ w[k] &\sim \mathcal{N}(0, Q), \\ v[k] &\sim \mathcal{N}(0, R), \end{aligned}$$

where we call f (and similarly, h) additive if $x[k+1] = f(x[k]) + w[k]$. In what follows we will assume that f is non-additive and h is additive. The parameters α

and κ control the spread of the sigma points about the mean. This spread is directly proportional to α and $\sqrt{\kappa}$. Typically κ is set to 0 and the spread is controlled solely by α. β can be used if you know something about the distribution of (x_k), but defaults to 2 which is optimal for Gaussian distributions.
For each time step k, we will alternate between correction and prediction steps. For correction we update the estimates of the mean $x[k|k]$ and covariance $P[k|k]$ of the latent state, given prior estimates $x[k|k-1]$ and $P[k|k-1]$ by comparing the measure estimate $\hat{y}[k]$ to the observed $y[k]$. For prediction we compute the next time step's prior estimates $x[k+1|k]$ and $P[k+1|k]$.

1.1 Correction

First we choose a set of $2M+1$ sigma points, where M is the size of the covariance matrix $P[k|k-1]$ and the number of states. These are distributed about the mean $x[k]$ by adding and subtracting columns of the matrix square root of the scaled covariance matrix.

$$\begin{aligned}
&\hat{x^{(0)}}[k|k-1] = \hat{x}[k|k-1], \\
&\hat{x^{(i)}}[k|k-1] = \hat{x}[k|k-1] + \Delta x^{(i)} \qquad i = 1, 2, \ldots, 2M, \\
&\Delta x^{(i)} = (\sqrt{cP[k|k-1]})_i \qquad i = 1, 2, \ldots, M, \\
&\Delta x^{(M+i)} = -(\sqrt{cP[k|k-1]})_i \qquad i = 1, 2, \ldots, M.
\end{aligned}$$

Here $c = \alpha^2(M+\kappa)$ is a scaling factor and $\sqrt{cP}$ is the matrix square root such that $\sqrt{cP}\sqrt{cP}^T = cP$.
Each sigma point is then passed through the measurement function h to compute a predicted measurement,

$$\hat{y^{(i)}}[k|k-1] = h(\hat{x^{(i)}}[k|k-1]) \qquad i = 1, 2, \ldots, M.$$

These measurement are then combined to yield a predicted measurement $\hat{y}[k]$ for time k. We define a set of measurement weights W_M^i as

$$\begin{aligned}
W_M^0 &= 1 - \frac{M}{\alpha^2(M+\kappa)}, \\
W_M^i &= \frac{1}{2\alpha^2(M+\kappa)} \qquad i = 1, 2, \ldots, M. \\
\hat{y}[k] &= \sum_{i=0}^{2M} W_M^i \hat{y^{(i)}}[k|k-1]
\end{aligned}$$

Note that all the non-central sigma points are weighted equally. With a slightly different set of weights W_c^i (differing only in the central weight W_c^0) we estimate

the covariance of the predicted measurement as

$$
\begin{aligned}
W_c^0 &= (2 - \alpha^2 + \beta) - \frac{M}{\alpha^2(M+\kappa)}, \\
W_c^i &= \frac{1}{2\alpha^2(M+\kappa)} \qquad i = 1, 2, \ldots, M. \\
P_y &= \sum_{i=0}^{2M} W_c^i (\hat{y^{(i)}}[k|k-1] - \hat{y}[k])(\hat{y^{(i)}}[k|k-1] - \hat{y}[k])^T + R
\end{aligned}
$$

These values for P_y and $\hat{y}[k]$ can then be used in the maximum likelihood calculation. We then estimate the cross-covariance between $\hat{x}[k|k-1]$ and $\hat{y}[k]$ as

$$
P_{xy} = \frac{1}{2\alpha^2(M+\kappa)} \sum_{i=1}^{2M} W_c^i (\hat{x^{(i)}}[k|k-1] - \hat{x}[k|k-1])(\hat{y^{(i)}}[k|k-1] - \hat{y}[k])^T
$$

We can start the summation at 1, since the zeroth sigma point $\hat{x^{(0)}}[k|k-1]$ is just $\hat{x}[k|k-1]$ itself. We are now ready to calculate the Kalman gain $K = PxyPy^{-1}$ and finish updating the state and state error covariance estimates as

$$
\begin{aligned}
\hat{x}[k|k] &= \hat{x}[k|k-1] + K(y[k] - \hat{y}[k]) \\
P[k|k] &= P[k|k-1] - KP_yK^T
\end{aligned}
$$

1.2 Prediction

We now predict the state and state error covariance at the next time step. We start by propagating sigma points. This time, since we have a non-additive noise for f we will additionally compute $2L$ sigma points $\hat{w}[k|k]$ for the error terms about the mean of zero using the error covariance Q with size L. this time about $\hat{x}[k|k]$ using the covariance matrix $P[k|k]$. That is

$$
\begin{aligned}
\hat{x^{(0)}}[k|k] &= \hat{x}[k|k], \\
\hat{x^{(i)}}[k|k] &= \hat{x}[k|k] + \Delta x^{(i)} \qquad i = 1, 2, \ldots, 2M, \\
\Delta x^{(i)} &= (\sqrt{cP[k|k]})_i \qquad i = 1, 2, \ldots, M, \\
\Delta x^{(M+i)} &= -(\sqrt{cP[k|k]})_i \qquad i = 1, 2, \ldots, M. \\
\hat{w}^{(i)}[k|k] &= (\sqrt{cQ[k|k]})_i \qquad i = 1, 2, \ldots, L, \\
\hat{w}^{(M+i)}[k|k] &= -(\sqrt{cQ[k|k]})_i \qquad i = 1, 2, \ldots, L.
\end{aligned}
$$

Each sigma point is then passed through the state transition function. The $2L + 1$ sigma points for $x[k|k]$ are passed in with zero error component. The $2M$ sigma points for $w[k|k]$ are passed in as the error argument alongside the original mean

$x[k|k]$ used to generate the $x[k|k]$ sigma points. The whole set of $2(L+M)+1$ points are then combined with the same W_m weights as before to get $\hat{x}[k+1|k]$.

$$\begin{aligned}
&\hat{x}^{(i)}[k+1|k] = f(\hat{x}^{(i)}[k|k], 0) \qquad i = 0, 1, 2, \ldots, 2M, \\
&\hat{x}^{(1+2M+i)}[k+1|k] = f(\hat{x}^{(0)}[k|k], \hat{w}^{(i)}[k|k]) \qquad i = 1, 2, \ldots, 2L, \\
&W_M^0 = 1 - \frac{M}{\alpha^2(M+\kappa)}, \\
&W_M^i = \frac{1}{2\alpha^2(M+\kappa)} \qquad i = 1, 2, \ldots, 2(M+L). \\
&\hat{x}[k+1|k] = \sum_{i=0}^{2(M+L)} W_M^i \hat{x^{(i)}}[k+1|k]
\end{aligned}$$

The covariance is propagated in similar fashion, as

$$\begin{aligned}
&W_c^0 = (2 - \alpha^2 + \beta) - \frac{M}{\alpha^2(M+\kappa)}, \\
&W_c^i = \frac{1}{2\alpha^2(M+\kappa)} \qquad i = 1, 2, \ldots, M. \\
&P_x[k+1|k] = \sum_{i=0}^{2(M+L)} W_c^i (\hat{x^{(i)}}[k+1|k] - \hat{x}[k+1|k])(\hat{x^{(i)}}[k+1|k] - \hat{x}[k+1|k])^T
\end{aligned}$$

Note we do not add Q here as we did with P_y and R, because the effect of the noise has already been incorporated by the additional sigma points.

2 Implementation

We use a *Matlab*'s 'Control System Toolbox' and 'Signal Processing Toolbox' to implement the unscented Kalman filter specified in MathWorks (2016a). We need to provide it with $f, h, Q, R, x[0:-1], P[0|-1]$ as defined in the previous subsection. Following the dynamics of our simplified model, we specify $f : \mathbb{R}_+^2 \times \mathbb{R}_+^2 \longrightarrow \mathbb{R}_+^2$ as

$$\begin{aligned}
f\left(\begin{bmatrix} \hat{r}_k^{OIS} \\ \hat{\xi}_k \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) = \begin{bmatrix} \hat{r}_k^{OIS} \\ \hat{\xi}_k \end{bmatrix} + \begin{bmatrix} +\alpha^{OIS}(\theta^{OIS} - \hat{r}_k^{OIS}) + \alpha^{OIS,\phi}(\theta^{\phi} - \hat{\xi}_k) + \Gamma_r \sigma^{OIS} \hat{r}_k^{OIS} \\ \alpha^{\phi}(\theta^{\phi} - \hat{\xi}_k) + \Gamma_{\xi} \sigma^{\phi} \hat{\xi}_k \end{bmatrix} dt \\
+ \begin{bmatrix} w_1 \sigma^{OIS} \sqrt{\hat{r}_k^{OIS}} \\ w_2 \sigma^{\phi} \sqrt{\hat{\xi}_k} \end{bmatrix} \sqrt{dt}
\end{aligned}$$

Our measurement function h maps each pair $\hat{x}_k = (\hat{r}_k^{OIS}, \hat{\xi}_k)$ to the price vector $\hat{y}_k \in \mathbb{R}_+^{32}$ made up of the prices of the FRAs and IRSs outlined in the Data section B. Pricing formula are as in 2. The remaining inputs and parameters are specified

as follows:

$$
\begin{aligned}
&Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&R = 0.01 \\
&x[0|-1] = [0.07, 0.1]^T \\
&P[0|-1] = \begin{bmatrix} \sigma^{ois}\sqrt{0.07} & 0 \\ 0 & \sigma^{\phi}\sqrt{0.1} \end{bmatrix} \Delta t \\
&\alpha = 0.9 \\
&\kappa = 0 \\
&\beta = 2
\end{aligned}
$$

where $\Delta t = 1/365$

Appendix D

Optimisation

The estimation methodology uses a local nonlinear minimiser in *Matlab*, *fmincon*, to do the estimation where the optimisation is given in *Matlab's solver* form. The optimisation requires us to define a set of optimisation variables that are then fed into the objective function (which returns the negative log-likelihood function) and then an iterative process occurs. This iterative process can be broken down as follows:

1. Evaluate the objective function at the given set of parameters.
2. Take each parameter in turn and add a small value to it and evaluate the objective function. For N parameters in our parameters set, this involves N new objective function evaluations.
3. Depending on the results of step 2, work out the appropriate update rule to change all the parameters and evaluate the objective function again. Hopefully the objective function is smaller otherwise the update rule is revised.
4. Repeat steps $1-3$ until the norm of the change in the parameters in step 2 is less than 10^{-10} or until the first order optimality condition is less that 10^{-7}.

After step 4, the parameters obtained are considered our estimated model parameters.
In the estimation process we do limit the upper and lower bounds of each of the parameters as follows:

	Lower Bound	Upper Bound
α^{OIS}	0	10
α^{ϕ}	0	10
$\alpha^{\text{OIS},\phi}$	0	10
κ^{ϕ}	0	10
Γ^{OIS}	0	1
Γ^{ϕ}	0	1
μ^{ϕ}	0	10
σ^{OIS}	0	30
σ^{ϕ}	0	10
θ^{OIS}	0	10
θ^{ϕ}	0	10

Table D.1: Upper and Lower bounds for Θ.

Finally, since the optimisation process in *solver* form does not give the Hessian matirx, we need to approximate it numerically. This is done by numerically calculating the second order derivatives.