Historically Implied Swaption Skews using Non-Parametric Methods

Evan Jackson

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

November 1, 2016

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy at the University of the Cape Town. It has not been submitted before for any degree or examination to any other University.

Evan Jackson

November 1, 2016

Abstract

This dissertation aims to derive historically realised volatilities for swaptions of a long-term nature within the South African market, which is illiquid and over-the-counter. To achieve this the dissertation adopts and constructs non-parametric methods which only make use of historical realised data of the underlying variable rather than any implied pricing history of the derivative itself. Stutzer's method of canonical valuation (1996) is adapted for use with interest rate derivatives of a long-term nature. However, under a simulation of swaption prices, canonical valuation is found to have a monotonic increase in pricing error for swaptions of maturities over 2 to 15 years. A new method is constructed, named the relative entropy approach, which is based on the work of Buchen and Kelly (1996) and is capable of pricing long-term interest rate derivatives using a smoothed continuous distribution of the historical realised data of the underlying variable only, while market implied pricing data can also be incorporated for calibration of the derivative to current market prices. Under simulation this method maintains consistent and bounded pricing error across swaption maturities of up to 15 years. This method is then used to obtain historical realised volatilities for swaptions of a long-term nature. The derived ten-year tenor swaption skews under the relative entropy approach observe smile characterisitcs similar to that of the market implied skew over short-term maturities and maintain a volatility smile, albeit diminishing, across moneyness for maturities up to 20 years. The skews are further tested for sensitivity to the input historical data as well as the precision of the skew under implementation of the relative entropy approach. Results show the derived swaption skews to be robust when using a historical data set greater than 1200 observations. The swaption skew is sensitive to the nature of the historical data used which is representative of particular market characteristics over certain historical periods. The relative entropy approach is concluded capable of pricing long-term swaptions in a market where little or no option pricing data exists and could be considered for use in practical applications.

Acknowledgements

I would like to thank my supervisors, Dylan Flint and Professor David Taylor for their assistance and input.

Thank you to Obeid Mahomed and Tom McWalter who gave up much of their time and provided me with invaluable insight over the course of my work. I am especially grateful to Ralph Rudd who has been an incredible support to me throughout my efforts.

A special thank you to Emily for her encouragement as well as help with editing. And finally to my parents who have been an absolute strength behind my every endeavor.

Contents

List of Figures

List of Tables

1. Introduction

Accurate pricing of contingent claims generally requires effective characterisation of the dynamics of the underlying under an equivalent martingale measure. Most familiar is the assumption that log returns of the underlying are Gaussian in nature within a complete market, amongst others. In reality, the assumption of Gaussian log returns is often an inadequate description, whilst markets are seldom complete.

Locally, the South African market is highly illiquid and this is combined with a lack of price transparency owing to the prevalence of over-the-counter trades. Hedging of market exposure is still a necessity for practitioners. How then does one adequately price unique instruments within such an incomplete and illiquid market? Moreover how does one effectively characterise the dynamics of the underlying?

Since the advent of the Black-Scholes-Merton pricing formula and the crash of 1987, many parametric methods have surfaced in an attempt to more accurately describe the underlying. Despite the complexity of these processes their parameters usually fail to completely capture the characteristics of the real-world (Garcia *et al.* (2010)). Parametric methods characterise the underlying dynamics through a set of estimated parameters and are therefore only able to capture as many characteristics of the underlying as specified by the model. Non-parametric methods have since developed in order to to reduce the necessity for restrictive assumptions and position themselves, largely, as model-free pricing techniques. Non-parametric methods essentially make use of historical implied option price data in order to price options while subsequent models have been derived which only require realised data of the underlying. These methods are therefore more likely to capture a greater set of characteristics influencing the underlying dynamics. Such techniques are advantageous in being very robust to model specification errors as well as being adaptive to the problem in question.

It is the objective of this dissertation to identify and evaluate a model which is able to price swaptions that are long-term in nature, extending up to ten-year terms for swaption tenor and up to twenty-year terms for swaption maturity. These instruments face unique challenges within the South African market as they are absent of much pricing data due to market illiquidity and non-transparency. Such instruments are valuable for the hedging of guaranteed annuity options by insurance companies and pension funds. To this end, the dissertation aims to construct implied long-term swaption skews using non-parametric means that only requires historical realised data of the underlying. The technique shall be evaluated based on the accuracy, sensitivity and precision of the resulting skews in consideration for practical applications such as hedging.

This research is compelling for at least three reasons: firstly, little work has been done on non-parametric modeling of interest rate derivatives. Secondly, analysis of instruments of a long-term nature within such a market is unique. Finally, there has been little practical implementation of non-parametric methods, especially within a market such as South Africa [1].

This dissertation extends the non-parametric method of Stutzer's canonical valuation (1996) for the pricing of interest rate derivatives before constructing a new method, named the relative entropy approach, based largely on the work of Buchen and Kelly (1996) which makes use of relative entropy principles under a continuous framework and is capable of pricing both short and long-term interest rate derivatives.

Chapter 2 discusses the principles of entropy with respect to non-parametric pricing theory. Chapter 3 provides a background for the pricing of swaptions within the context of non-parametric theory. Chapter 4 then provides general pricing methodologies for the pricing of interest rate derivatives using canonical valuation and the relative entropy approach. Chapter 5 evaluates the performance of the two methods under simulation within an idealised world using a one factor Vasicek model. Chapter 6 presents and discusses results of the derived swaption skew of ten-year tenor under the relative entropy approach using JSE swap curve data and compares this skew to the market implied swaption skew sourced from Bloomberg. Chapter 7 discusses the sensitivity of swaption skews derived under the relative entropy approach when using varying sets of the underlying historical realised data, as well as the precision of the skew when using the method itself. Chapter 8 finally provides conclusions to the study.

[1] A notable exception is the work done by De Araujo and Mare (2006) who examine the volatility skew in the South African market using the method of Duan (2002)

2. On Entropy and Non-Parametric Pricing Theory

The methods of canonical valuation and the relative entropy approach, both non-parametric pricing theories, rely on the principles of entropy theory. This chapter introduces the concept of entropy and the intuition behind the above pricing theories. Most of the technical detail regarding information theory can be found in Cover and Thomas (2012) and Buchen and Kelly (1996).

2.1 Information Theory and Entropy

The concept of information is too broad to be captured by a single definition. However we can define the quantity of entropy which has many intuitive notions as to the meaning of information and how it should be measured. We can initially view information as the amount of knowledge obtainable within a system of occurring events in the presence of noise. We can extend this notion to the amount of information one random variable contains about another. We note that as information is gained, so one's perception of the world changes. It is accepted that information is gained through the occurrence of an event x from set χ. This event is said to occur with probability $p(x),\ x \in \chi$.

Let $I(p(x))$ represent the information provided by the occurrence of an event from set χ with probability $p(x)$. We require a function that quantifies this gain in information. Furthermore the function must be non-negative for all probability $p(x)$ and decreasing with increasing probabilities $p(x)$. This is because firstly, we cannot gain negative information and secondly, because it is intuitive that the greater the likelihood of an event occurring, the less information its occurrence shall provide.

In information theory, the information obtained by the occurrence of event $x \in \chi$ with probability $p(x)$ is quantified as

$$I\left(p(x)\right) = -\ln\left(p(x)\right), \tag{2.1}$$

which is derived in Cover and Thomas (2012).

Let X be a χ-valued discrete random variable with probability mass function p. Entropy represents the expected quantity of information gained through the occurrence of an event from this distribution. In a discrete setting, Cover and Thomas (2012) define the measure of entropy to be

$$H(X) = -\sum_{x\in\chi} p(x)\ln\left(p(x)\right), \tag{2.2}$$

where x is a possible event of the discrete set χ containing N possible events. $H(X)$ is also referred to as the Shannon-Entropy (see Shannon, 2001).

By definition, entropy is a measure of uncertainty over the occurrence of a random event. It is a measure of 'missing information' in our system. We gain information through the occurrence of an event. Therefore the greater amount of information we expect to gain through the occurrence of an event within a state, the greater the entropy of our state prior to this event's occurrence, as we are more uncertain of which event will occur. For example observing the extremes, a minimum entropy of $H(X) = 0$ is obtained if $p(x) = 1$ for event $X,\ x \in \chi$ - where we are certain of one event occurring. A maximum entropy of $H(X) = \ln(N)$ is obtained if $p(x) = \frac{1}{N}\ \forall\ x = 1, ..., N;\ x \in \chi$ - where all events have an equal probability of occurring.

It should be noted here that we are measuring entropy based on the natural logarithm, in which case our unit of measure is the *nat*. One nat is the information gained by an event of probability $p(x) = \frac{1}{e}$ occurring, i.e. $I(p(x)) = -\ln(\frac{1}{e}) = 1$. Entropy is also commonly measured in base 2 logarithms, where the unit of measure is the *bit*.

2.2 Differential Entropy

Differential entropy is the entropy of a continuous random variable.

Let X be a continuous random variable with cumulative distribution function $F(x) = P[X \leq x],\ x \in \chi$. Should F be continuous then the random variable X is said to be continuous. This differs from Section 2.1 where X is discrete. Let $f = F'$ where the derivative exists for all $x \in \chi$. Then, f is the probability density function for the random variable X if $\int_{-\infty}^{\infty} f(x) = 1$.

Cover and Thomas (2012) define the differential entropy $h(x)$ of a continuous random variable X with density $f(x)$ as

$$h(X) = -\int_S f(x)\ln\left(f(x)\right)dx, \tag{2.3}$$

where S is the support set of X, defined as the set where $f(x) > 0$.

What is the relationship between discrete entropy and differential entropy? Cover and Thomas (2012) explains this relationship as follows.

Consider the aforementioned random variable X with density f. Let us divide the range of X into bins of width $\triangle$. We assume that the density is continuous within the bins. By the mean value theorem, there exists a value x_i within each bin such that

$$f(x_i) = \int_{i\triangle}^{(i+1)\triangle} f(x)dx.$$

Consider the quantized random variable $X^{\triangle}$ which is defined by

$$X^{\triangle} = x_i \qquad \text{if } \ i\triangle \leq X < (i+1)\triangle.$$

Therefore the probability that $X^{\triangle} = x_i$ is

$$p_i = \int_{i\triangle}^{(i+1)\triangle} f(x)dx = f(x_i)\triangle.$$

The entropy of the quantized version is

$$\begin{aligned} H(X^{\triangle}) &= -\sum_{-\infty}^{\infty} p_i \ln p_i \\ &= -\sum_{-\infty}^{\infty} f(x_i)\triangle \ln(f(x_i)\triangle) \\ &= -\sum \triangle f(x_i) \ln f(x_i) - \sum f(x_i)\triangle \ln \triangle \\ &= -\sum \triangle f(x_i) \ln f(x_i) - \ln(\triangle), \end{aligned} \tag{2.4}$$

since $\sum f(x_i)\triangle = \int f(x) = 1$. If $f(x)\ln f(x)$ is Riemann integrable the first term in 2.4 approaches the integral of $-f(x)\ln f(x)$ as $\triangle \to 0$ by definition of Riemann integrability. This proves the following theorem:

Theorem 2.1. *If the density $f(x)$ of the random variable X is Riemann integrable (ie. the limits of an integral are well defined), then*

$$H(X^{\triangle}) + \ln(\triangle) \to h(f) = h(X), \ \text{ as } \triangle \to 0. \tag{2.5}$$

Thus, the entropy of an n-nat quantisation of a continuous random variable X is approximately $h(X) + n$.

Note that the n-nat quantisation of a continuous random variable adheres to $\triangle = e^{-n}$.

This also shows that the entropy of a Riemann integrable discrete random variable will converge to the entropy of a continuous random variable as the discrete set x tends to infinity. This can be observed by substituting equation 2.4 in to equation 2.5 and using the definition of Riemann integrability.

2.3 Kullback-Leibler Distance and Relative Entropy

Define p as the probability mass function of the real-world measure P and q as the probability mass function of an equivalent martingale measure (EMM) Q, where $x \in \chi$. The pricing of a contingent claim requires a change of measure from P to Q. The relative entropy principle allows us to quantify the change in entropy, or 'entropic distance', between the two distributions, p and q.

Cover and Thomas (2012) define the relative entropy $D(q||p)$ of the probability mass function q with respect to the probability mass function p as

$$D(q||p) = E^Q\left[\ln\left(\frac{q(x)}{p(x)}\right)\right] = \sum_{x\in\chi} q(x)\ln\left(\frac{q(x)}{p(x)}\right). \tag{2.6}$$

This is also known as the Kullback-Leibler distance between two probability mass functions.

It holds that $D(q||p)$ is always greater than zero for all $q \neq p$ and zero for $q = p$. $D(q||p)$ does not, therefore, describe a gain or loss of entropy between p and q. It should be noted that this is not a true metric and is generally asymmetric; that is, the relative entropy between p and q is not always equal to the relative entropy between q and p.

Under a continuous framework, the relative entropy $D(f||g)$ between two probability densities f and g is defined as

$$D(f||g) = \int_S f(x)\ln\left(\frac{f(x)}{g(x)}\right), \tag{2.7}$$

where S is the support set of the continuous random variable X (where $f(x), g(x) > 0$). Note that Equation 2.7 is finite only if the support of f is contained within the support of g.

2.4 In the Context of the No-Arbitrage Market Constraints

Let S_t, $0 \leq t \leq T$, be a stochastic process that represents the path of the underlying for a contingent claim governed by b random processes. We define a contingent claim as a financial contract with a random payoff at maturity. The contingent claim payoff is realised dependent on which random outcome due to S_T has occurred at time T, that is, which state of nature has been realised at the time of the payoff. Let $A(t_i, T)$ represent the risk-free discount factor in the market at time t_i with maturity T and let s represent the number of securities.

A market is said to be complete when we can replicate any contingent claim with the existing securities. That is, the payoff of the contingent claim can be matched by a portfolio of securities for all possible states with probability one (Cvitanić and Zapatero (2004)).

For the market to be complete, we say that $s \geq b$. Under a complete market, the price of the contingent claim is unique. Moreover, there exists a unique Equivalent Martingale Measure (EMM). Brigo and Mercurio (2007) define an EMM as a probability measure such that the discounted asset price process is a martingale under expectation with respect to the measure numeraire.

A complete market is also one which is arbitrage free, namely that the probability of a positive return with no risk of loss is zero. Should a market be arbitrage free but incomplete, then many EMM's are said to exist - there are many possible no-arbitrage prices.

We find ourselves within an incomplete market as we are unable to uniquely replicate all contingent claims. Subsequently we want to choose the EMM that best describes the current information. We note the distribution q as the underlying distribution of our contingent claim under an EMM. We seek to find the optimal q distribution under the optimal EMM adhering to the no-arbitrage pricing constraints.

We choose to find the optimal q distribution (hence under the optimal EMM) using two relevant Entropy principles. The two methods available to us are the principle of maximum entropy and the relative entropy principle, which are discussed in sections 2.5 and 2.6 respectively.

2.5 Principle of Maximum Entropy

Literature in the field of thermodynamics has motivated that a closed thermodynamic system will evolve to obtain its maximum possible entropy under any set of given constraints. The theory was adopted within statistical mechanics by Jaynes (1957) who proposed that, under a set of known constraints, the distribution of maximum entropy will be the distribution holding the least possible information and, therefore, will be the distribution of minimum bias. It is the optimal distribution with which to represent missing or unknown information and any distribution of lower entropy would infer information that we do not know.

This leads us to the Principle of Maximum Entropy (PME). The PME states that the optimal distribution is the one with the largest remaining uncertainty, or entropy. In a financial context, as we move from the real-world distribution, p, to the EMM distribution, q, we imply a set of constraints under this EMM. In accordance with the PME, we seek to find the q distribution of maximum entropy.

2.5.1 Entropy Maximisation

Under implementation of the PME, we seek to maximise the Shannon-Entropy in the continuous case, adapted from Buchen and Kelly (1996).

Under the PME, we make no assumptions of the prior distribution. Empirically we assume that all observed events have an equal likelihood of occurrence; this will be discussed in section 2.7. We only consider the distribution q under Q, stating that the entropy of q would obtain a maximum under its given set of constraints. We therefore want to maximise the Shannon entropy of q:

$$-\int_{-\infty}^{\infty} q(x)\ln\left(q(x)\right) dx, \tag{2.8}$$

subject to the constraints

$$\int_{-\infty}^{\infty} q(x)dx = 1,$$

and

$$E^Q[r_j(x)] = \int_{-\infty}^{\infty} q(x)r_j(x)dx = c_j,$$

where $r_j(x)$ is the j^{th} function of the continuous set of observations $x \in \chi$ and c_j is the imposed j^{th} constraint, for $j = 1, ..., m$. i.e. we are imposing m constraints on our distribution where $r_j(x)$ is the function of x related to constraint c_j, such as a return of the underlying asset under observation. We require the function $r_j(x), x \in \chi$ to be 'well-behaved' - a Riemann integrable function.

We can solve the above problem using the method of Lagrange multipliers[1]. We seek to maximise

$$H(X) \equiv H(q) = \int_{-\infty}^{\infty} -q(x)\ln(q(x)) + (1+\lambda_0)q(x) + \sum_{j=1}^{m}\lambda_j q(x)r_j(x)\ dx, \tag{2.9}$$

where $\lambda_j,\ j = 0, ..., m$ are Lagrange multipliers. We find the maximum of $H(q)$ by taking the derivative with respect to $q(x)$ and equating to zero,

$$\frac{\delta H}{\delta q(x)} = \int_{-\infty}^{\infty} -\ln(q(x)) + \lambda_0 + \sum_{j=1}^{m}\lambda_j r_j(x)\ dx = 0, \tag{2.10}$$

[1] For a brief overview of using Lagrange multipliers see Stewart (2009)

which leads to the following solution:

$$q(x) = \frac{\exp\left(\sum_{j=1}^{m} \lambda_j r_j(x)\right)}{\int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^{m} \lambda_j r_j(x)\right) dx}, \tag{2.11}$$

where $\lambda_j,\ j = 1, ..., m$ are to be solved numerically in most cases by the following minimisation problem (see Ben-Tal (1985)):

$$F(\lambda_{j=1,...,m}) = \int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^{m} \lambda_j \int_{-\infty}^{\infty} q(x) r_j dx - c_j\right).$$

The above solution allows us to construct an EMM probability density function subject to its set of imposed no-arbitrage market constraints (discussed in Section 2.4).

2.6 The Relative Entropy Principle and its Minimisation

Consider the random variable X under P with distribution p, $x \in \chi$. We want to move to the optimal distribution q under Q. Under the relative entropy principle (REP), we assume that we have some prior knowledge of X under P. Moreover, we assume the prior distribution p under P. Conditional on us knowing this prior knowledge of X under P, the REP states that the optimal q distribution will be the distribution of minimum 'entropic distance' from p that adheres to the constraints under Q. The optimal q will be the distribution of least bias as we are assuming the least possible change in entropy from the known information under P whilst conforming to the known Q measure constraints. Relative entropy was discussed in section 2.3.

The minimisation of relative entropy (or Kullback-Leibler distance) seeks to minimise

$$D(q||p) = \int_{-\infty}^{\infty} q(x) \ln\left(\frac{q(x)}{p(x)}\right) dx, \tag{2.12}$$

subject to the constraints

$$\int_{-\infty}^{\infty} q(x) dx = 1$$

and

$$E^{Q}\left[r_j(x)\right] = \int_{-\infty}^{\infty} q(x) r_j(x) dx = c_j.$$

where $r_j(x)$ is the j^{th} function of the observation $x \in \chi$ and c_j is the imposed j^{th} constraint, for $j = 1, ..., m$.

As per the PME, the above problem can be solved using the method of Lagrange multipliers which is parallel to that of the entropy maximization derived in Buchen and Kelly (1996). We intend to find the minimum of

$$D(q||p) = \int_{-\infty}^{\infty} q(x) \ln\left(\frac{q(x)}{p(x)}\right) - (1+\lambda_0) q(x) - \sum_{i=1}^{m} \lambda_j q(x) r_j(x)\ dx, \tag{2.13}$$

where $\lambda_j,\ j = 0, ..., m$ are Lagrange multipliers. We achieve this through equating the derivative of $D(q||p)$ with respect to $q(x)$ to zero:

$$\frac{\delta D(q||p)}{\delta q} = \int_{-\infty}^{\infty} \ln\left(\frac{q(x)}{p(x)}\right) + p(x) - (1+\lambda_0) - \sum_{i=1}^{m} \lambda_j r_j(x)\ dx, \tag{2.14}$$

which leads us to the solution,

$$q(x) = \frac{p(x) \exp\left(\sum_{j=1}^{m} \lambda_j r_j(x)\right)}{\int_{-\infty}^{\infty} p(x) \exp\left(\sum_{j=1}^{m} \lambda_j r_j(x)\right)}\ dx. \tag{2.15}$$

Here $\lambda_j,\ j = 1, ..., m$ are solved for by finding the global minimum of the following unconstrained convex problem,

$$F(\lambda_1, \lambda_2) = \ln\left(\int_{-\infty}^{\infty} g(x) \exp\left(\lambda_1 c_1(x) + \lambda_2 c_2(x)\right) dx\right) - \left(\lambda_1 r(t_0, t_\alpha, t_\beta) + \lambda_2 S_{ATM}\right),$$

as per Buchen and Kelly (1996) which can be solved for numerically.

2.7 Discussion on Maximum Entropy, Minimum Relative Entropy and the Prior Distribution

Under a change of measure we move from the real-world measure, P, to an EMM, Q. While we have discussed methods that allow us to move to the optimal Q measure, we need to ask: from what distribution (or information) are we moving under P? The REP allows us to choose this distribution, while we shall learn that a uniform distribution is assumed implicitly under the PME. The PME is based on the Principle of Insufficient Reason (PIR), first considered within the realm of probability by Laplace (see marquis de Laplace (1840)), and is an extension of this generalisation.

2.7.1 The Principle of Insufficient Reason

Consider the discrete random occurrences of event x from a finite set χ. In the empirical sense, the values of x refer to historical observations of a random process, for example that of a spot interest rate, which occurs within a specific time period (set χ). We are required to assign probabilities to each of these occurrences. The PIR states that, for a set of mutually exclusive cases (events or outcomes), should we have no reason to believe that any one case is more possible than another then each case should be assigned an equal probability of occurring. For a subset of finite events $x \in \chi$,

$$p(x) = \frac{1}{n}, \qquad\qquad x = 1, ..., n. \tag{2.16}$$

Uffink (1995) states that the PIR is "Based on a symmetry in our belief and judgment in order to obtain numerical probabilities... and in this view the term probability should be understood as a degree of belief and hence, the uniform distribution represents exactly the situation where all events are equally credible." Recall the motivation behind Jaynes' PME, discussed in section 2.5.1, that the distribution of maximum entropy will be the distribution of least prejudice and which must therefore adopt an equivalent underlying judgment to that of the PIR: that in accordance with our degree of belief, we have no reason to favor the occurrence of any event over the occurrence of another beyond that of what is known. The uniform distribution is the distribution of maximum entropy.

The PIR previously received heavy criticism which led to an almost universal abandonment of the principle until it was later revived by Jaynes. Uffink notes that the objection made most frequently about the PIR is that, "one cannot simply derive empirical predictions from a lack of knowledge." Uffink also notes that, "the PIR is circular in that the only sensible meaning one can give to equally possible is equally probable."

As the PIR is a fundamental assumption within the PME, the PME carries scrutiny associated with the PIR albeit to a more generalised extent. Hence, while the PME is a widely acknowledged method for transformation of distributions, its application should be considered critically.

2.7.2 Examining the equivalence between the PME and the REP

Comparison of equations 2.11 to 2.15 leads to a clear relationship between the PME and the REP. Both equations have the same solution when the prior distribution p is assumed to be uniform. This result explains mathematically the concept of the PME. Under the PME we are moving from a measure P of maximum entropy (and therefore minimum information) and finding the least entropic distance required to move to a new measure Q. Rationally, moving from a state of highest possible entropy, the optimal Q distribution under the REP will be the Q distribution of maximum entropy.

Remark (On the equivalence between relative entropy minimisation and entropy maximisation)**.** *It should be noted that, while the PME and the Relative Entropy Principle (REP) are shown to be equivalent when assuming the PIR, a paper by Banavar and Maritan (2007) argues that a naive application of the PME can present an answer that depends on the level of initial information available and is not always equivalent to the REP. They suggest that the correct approach is, rather, the minimisation of the relative entropy with a suitable reference probability (the paper states maximisation of relative entropy but has viewed relative entropy as a negative function with a maximum value of zero when* $P = Q$*).*

2.7.3 Choosing the Prior Distribution under P

Of course relative entropy minimisation allows us to assume some prior knowledge of p. The question therefore is what initial distribution would be the most appropriate. As we assume knowledge of our distribution so its maximum possible entropy changes. Under no assumptions, the uniform distribution represents the distribution of maximum entropy. However, when considering a prior distribution where both the mean and variance are known the normal distribution represents the distribution of maximum entropy [2]. Our consideration regards what knowledge can be assumed that may be advantageous.

Buchen and Kelly (1996) note that assuming p to be a normal distribution of given mean and variance may be appropriate for use within a mathematical finance context where the normal distribution is often assumed to represent the log-return of the underlying asset (for example the famous Black-Scholes formula for equity derivatives). Their simulation results indeed find the assumption of p being the normal distribution to derive a more accurate representation of q than assuming p to be uniform.

Similarly, under the Relative Entropy Approach (REA) method, discussed in Section 4.2, we assume the empirical distribution of the underlying variable. It

[2] Proof of the above can be found in Cover and Thomas (2012)

would be naive to assume the empirical data as our population distribution under the P measure. However, as a sample distribution under P, it is an efficient distribution from which to move under the REP and, perhaps, our best estimate of the population distribution under P. Although not presented, testing of the REA under simulation showed less pricing error when assuming the empirical distribution over the uniform distribution under P and concurred with the findings of Buchen and Kelly (1996).

3. On the Non-Parametric Pricing of Swaptions

3.1 The $Q^{\alpha,\beta}$ Swap Measure

Proof and elaboration on all of the following theory and concepts can be found in Brigo and Mercurio (2007).

Let the current (t_0) price of a swaption of maturity t_α and tenor $t_\beta - t_\alpha$ be denoted $S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$, where $t_0 \leq t_\alpha < t_\beta$. The swaption's floating leg resets at times $t_\alpha, t_{\alpha+1}, ..., t_{\beta-1}$ and pays at times $\mathcal{T} = [t_{\alpha+1}, ..., t_{\beta-1}, t_\beta]$ against a fixed leg struck at r_k, while $\omega = 1(-1)$ for a payer (receiver) swaption. Denote the fair forward swap rate at time t_0 and of tenor $t_\beta - t_\alpha$, with payment times $\mathcal{T}$, as $r(t_0, t_\alpha, t_\beta)$ while the fair swap rate at time t_α is denoted by $r(t_\alpha, t_\alpha, t_\beta) = r(t_\alpha, t_\beta)$. The fair forward swap rate is given by

$$r(t_0, t_\alpha, t_\beta) = \frac{Z(t_0, t_\alpha) - Z(t_0, t_\beta)}{\sum_{i=\alpha+1}^{\beta} \tau_i Z(t_0, t_i)}, \qquad \text{where} \qquad \tau_i = t_{i+1} - t_i \tag{3.1}$$

where $Z(t_0, t_i)$ is the discount factor implied from the market over period (t_0, t_i), for $i \in \{\alpha, \alpha+1, ..., \beta\}$. Here we assume the discount factor as the t_0, t_i zero coupon bond.

The payoff of the swaption at time t_α can be constructed as

$$S_\omega(t_\alpha, t_\alpha, \mathcal{T}, t_\beta, r_k) = (\omega\,[r(t_\alpha, t_\beta) - r_k])^+ \sum_{i=\alpha+1}^{\beta} \tau_i Z(t_\alpha, t_i). \tag{3.2}$$

For computational convenience, we choose the numeraire to be a portfolio of zero coupon bonds $Z(t, t_i),\ i = \alpha+1, ..., \beta$ where $t_0 \leq t \leq t_\alpha$. We denote the numeraire as

$$C(t, t_\alpha, t_\beta) = \sum_{i=\alpha+1}^{\beta} \tau_i Z(t, t_i). \tag{3.3}$$

Brigo and Mercurio (2007) remark that the numeraire $C(t, t_\alpha, t_\beta)$ could be seen as the forward swap's "present value for [one] basis point."

By choosing the $C(t, t_\alpha, t_\beta)$ numeraire, the $r(t, t_\alpha, t_\beta), 0 \leq t \leq t_\alpha$, swap rate evolves according to a martingale under the $Q^{\alpha,\beta}$ measure, with $r(t_\alpha, t_\alpha, t_\beta) = r(t_\alpha, t_\beta)$ at maturity (the $r(t_\alpha, t_\beta)$ future fair swap rate is our swaption underlying). We refer to $Q^{\alpha,\beta}$ as the swap measure.

Hence, when pricing a swaption $S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$ we derive from the fundamental theorem of asset pricing:

$$\frac{S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)}{C(t_0, t_\alpha, t_\beta)} = E^{\alpha,\beta}\left[\frac{S_\omega(t_\alpha, t_\alpha, \mathcal{T}, t_\beta, r_k)}{C(t_\alpha, t_\alpha, t_\beta)}\right]$$

$$S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = C(t_0, t_\alpha, t_\beta) E^{\alpha,\beta}\left[\frac{(\omega\left[r(t_\alpha, t_\beta) - r_k\right])^+ \sum_{i=\alpha+1}^{\beta} \tau_i Z(t_\alpha, t_i)}{C(t_\alpha, t_\alpha, t_\beta)}\right]$$

$$S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = C(t_0, t_\alpha, t_\beta) E^{\alpha,\beta}\left[(\omega\left[r(t_\alpha, t_\beta) - r_k\right])^+\right]. \tag{3.4}$$

Clearly this is a powerful measure as under $Q^{\alpha,\beta}$ we are not required to simulate the yield paths driving $Z(t, t_i)$ in equation 3.3 up until $\mathcal{T}$ in order to discount back to t_0 as we would have to under the risk-neutral measure Q.

All subsequent non-parametric pricing theory of swaptions is derived under the $Q^{\alpha,\beta}$ swap measure. Moreover, all pricing of swaptions makes use of (3.4) by calculating the swaption payoff at time t_α through simulation of the $r(t_\alpha, t_\beta)$ swap rate (to be discussed in Section 3.3) and discounting to time t_0 using $C(t_0, t_\alpha, t_\beta)$.

3.2 Creating Daily Returns

Fundamental to our empirical simulation of the underlying is the creation of a set of daily returns. Given a set of H historical observations the variable $r_i,\ i = 1, ..., H$, we are able to create a set of N arithmetic returns, $x_i = r_i - r_{i-1}$, or a set of N daily performances, $x_i = \frac{r_i}{r_{i-1}}$. Both methods hold value as the arithmetic return functions over the space $(-\infty, \infty)$ while the performance functions over the space $(0, \infty)$. Choice of either return therefore depends on our assumptions of the observed process for simulation. For instance, it is natural to assume for equity derivatives that the stock price process exists only over the space $(0, \infty)$ and one would therefore use performances to model the stock process so as not to allow the existence of negative stock prices.

Regarding interest rates and interest rate derivatives, much of previous literature is skeptical of the allowance of negative rates due to their theoretical implausibility. However current market characteristics and the observation of negative market rates have led to a growing popularity in interest rate models that are able to incur negative rates. Hence, in the simulation of interest rates, one may assume the choice of arithmetic returns to allow for simulation over the space $(-\infty, \infty)$.

In order to allow for use of historical data incurring negative rates (and their subsequent simulation) for our non-parametric methods we make use of arithmetic daily returns in our empirical simulation. Choice of method is an engineering problem up to the discretion of the practitioner concerning assumptions around the underlying dynamics.

3.3 Empirical Analysis of Swaptions

Recall that the current time zero, the maturity of the swaption and the expiry of the underlying swap are denoted t_0, t_α and t_β respectively[1].

Similar to that of Monte Carlo analysis, these non-parametric pricing theories involve simulation of the underlying from time t_0 to t_α whereupon we can value the swaption under the $Q^{\alpha,\beta}$ swap measure in order to find a fair price for the swaption at t_0. What is of fundamental importance, therefore, is a critical understanding and reasoning behind the choice of variable for simulation.

It is clear that, for the swaption $S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$, the underlying asset is the $r(t_\alpha, t_\beta)$[2] future fair swap rate. Under the $Q^{\alpha,\beta}$ swap measure, at time t_0 we expect $r(t_\alpha, t_\beta)$ to evolve to the fair forward swap rate $r(t_0, t_\alpha, t_\beta)$. However, at time t_{0+1}, we can expect $r(t_\alpha, t_\beta)$ to evolve to the $r(t_{0+1}, t_\alpha, t_\beta)$ fair forward swap rate. By the time t_α our underlying and the fair forward swap rate $r(t_\alpha, t_\alpha, t_\beta)$ converge. We can therefore say that, under expectation (under the $Q^{\alpha,\beta}$ swap measure), our best description of the $r(t_\alpha, t_\beta)$ future fair swap rate at time t is the current fair forward swap rate $r(t, t_\alpha, t_\beta),\ t_0 \leq t \leq t_\alpha$.

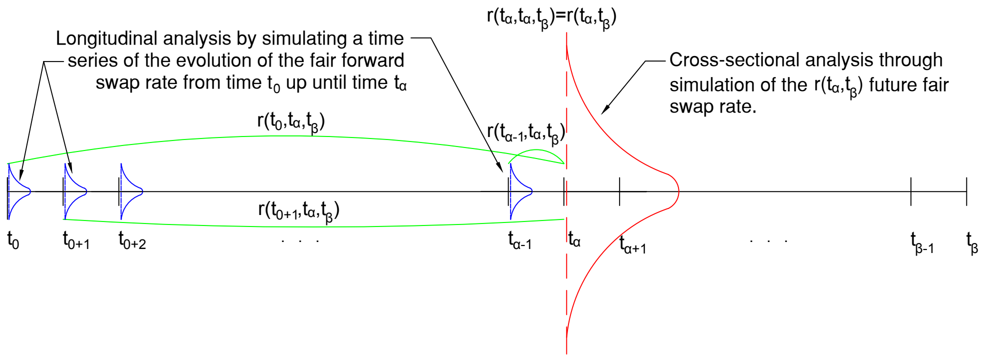
As non-parametric pricing theory requires empirical simulation of historical data, two methods are available to us. We can either model the underlying future fair swap rate $r(t_\alpha, t_\beta)$ cross-sectionally. Or we can simulate the fair forward swap rate, $r(t, t_\alpha, t_\beta)$, longitudinally, due to the fact that it is our best description of our underlying, $r(t_\alpha, t_\beta)$, at time t and converges to our underlying at time t_α. Figure 3.1 provides an illustration of the two methods of simulation. It is important that we find a method of analysis of the underlying and its evolution that is both tractable and plausible which we now discuss.

Cross-sectional simulation involves creating a distribution of the underlying variable relevant to the derivative itself, in this case the $r(t_\alpha, t_\beta)$ future fair swap rate at time t_α. The aim of this method is to simulate daily the current market fair swap rate $r(t_0, t_\beta - t_\alpha)$ up until time t_α to create a realisation of the future fair swap rate $r(t_\alpha, t_\beta)$. We repeat this simulation N times in order to create a set of N realisations

[1] For notational convenience we consider t_0, t_α, t_β in daily units.

[2] Recall that $r(t_\alpha, t_\beta) = r(t_\alpha, t_\alpha, t_\beta)$.

Fig. 3.1: Illustration showing simulation of the swaption underlying variable from time t_0 up until time t_α using a cross-sectional analysis and a longitudinal analysis.



and hence a cross-sectional distribution of $r(t_\alpha, t_\beta)$. To do this we implement the following algorithm.

Denote $r_{\alpha,\beta}(h)$ as the historical fair swap rate $r(t_0, t_\beta - t_\alpha)$ at calender time h, for $h \leq 0$. The method is implemented as follows:

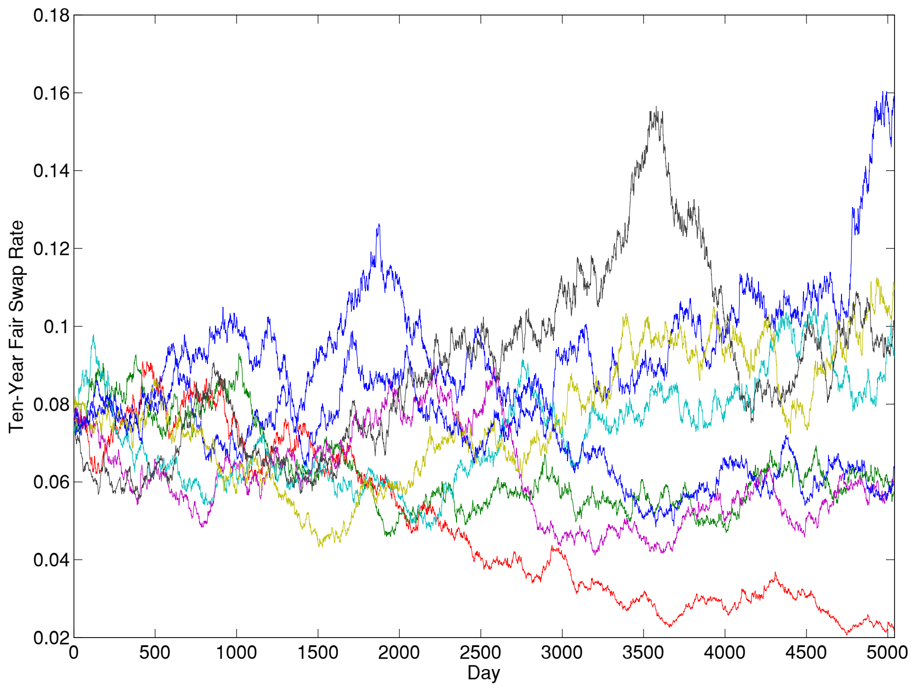
- From a set of H historical swap curves, construct a set of daily returns[3], $x_h = r_{\alpha,\beta}(h) - r_{\alpha,\beta}(h-1) \quad h = 0, ..., -H+1$ [4].
- Pick at random, and allowing for replacement, t_α returns from the set x_h, $h \in \{0, -1, ..., -H+1\}$ [5].
- Sum[6] the vector of drawn returns to create a single t_α period return X.
- Repeat this procedure N times to create a set of t_α period returns $X_i \quad i = 1, ..., N$.

Figure 3.2 shows samples of the simulated ten-year fair swap rate constructed from simulation using this algorithm. It is the set of realisations of the swap rate at time t_α that we use to create our distribution.

Longitudinal simulation involves creating a time-series of the expected evolution of the swaption underlying up until time t_α. Essentially our aim is to start with the market forward swap rate $r(t_0, t_\alpha, t_\beta)$ and simulate daily to the successive forward

[3] Should we opt to construct daily performances, $x_h = \frac{r_{\alpha,\beta}(h)}{r_{\alpha,\beta}(h-1)}$.
[4] Note h is negative as we are moving backwards in calender time
[5] h is essentially our random component for selection
[6] Should we be constructing performances the vector would be multiplied.

Fig. 3.2: Sample paths of the simulated daily ($\delta t = \frac{1}{252}$) ten-year fair swap rate over a ten year period under a cross-sectional analysis[7].



swap rate $r(t_j, t_\alpha, t_\beta),\ j = 1, ..., \alpha - 1$ up until time t_α. Hence we arrive at the future fair swap rate $r(t_\alpha, t_\alpha, t_\beta)$ at time t_α. Denote $r^j_{\alpha,\beta}(h)$ as the historical forward swap rate $r(t_0, t_j, t_\beta - t_\alpha + t_j)$ at calender time h, for $h \leq 0,\ j = 0, 1, ..., \alpha - 2, \alpha - 1$. The method is described as follows:

- From a set of H historical swap curves, create t_α sets of daily returns $x^j_h = r^j_{\alpha,\beta}(h) - r^j_{\alpha,\beta}(h-1) \quad h = 0, ..., -H+1,\ j = 0, 1, ..., \alpha - 2, \alpha - 1$.
- Pick at random, and allowing for replacement, a one day return from each set of x^j_h, for $j = 0, 1, ..., \alpha - 2, \alpha - 1,\ h \in \{0, -1, ..., -H+1\}$, in order to create a vector, of size t_α, of one day returns.
- Sum the created vector of daily returns in order to create a single t_α period return, X.
- Repeat this procedure N times in order to create a set of t_α period returns $X_i,\ i = 1, ..., N$.

It is interesting to note that equity derivatives could also be modeled longitudinally by simulating the $f(t_{0+i}, t_\alpha)$ equity forward contract for $i = 0, 1, ..., \alpha-2, \alpha-1$.

In fact, these two simulation methods can be seen as equivalent as their ultimate distributions will converge at t_α, just as a futures contract and a forward contract converge to the underlying spot value upon maturity.

It is clear however that, while both methods are plausible, the cross-sectional method is far more tractable than the longitudinal method under computational considerations, as the longitudinal method requires creation of the product of t_α and $H - 1$ daily returns whilst the cross-sectional method only requires creation of $H - 1$ daily returns. The cross-sectional method has therefore been used in all further non-parametric construction of swaption skews, as well as for non-parametric method simulation tests.

3.4 Sampling for Long-Term Derivatives

Previous non-parametric pricing theory has dealt with derivatives of short-term maturity, usually up to two years. However, the pricing of the proposed swaptions requires simulation of t_α period perfomances of up to 20 years. It should be noted that, under the method of canonical valuation, Stutzer (1996) derived a set of t_α period performances, X_i, $i = 1, ..., N$, through creating a set of rolling window period performances[8] using $X_i = \frac{S(-i)}{S(-i-t_\alpha)}$ for $i = 1, ..., H - t_\alpha$, where S represents the derivative underlying. This creates an obvious problem for long-term sampling as the maximum number of performances that can be created is $N = H - t_\alpha$. Hence when the gross period $t_\alpha \geq H$ there are zero possible performances available. In context, the data available within the South African market particular to the underlying is approximately 8 years currently. However, the t_α periods required reach a maximum of 20 years.

Our previous sampling methods, through allowing for random sampling of daily returns/performances while allowing for replacement, overcome these practical problems. For these sampling methods to hold however, we must assume daily returns/performances to be independently distributed. Tests for sample independence are conducted and reported in the results section in Chapter 6.

[7] Note that the figure represents the simulation of paths of the future fair swap rate up until time t_α. The set of realisations of this future fair swap rate at time t_α then become the 'cross-sectional' distribution.

[8] One may question the plausibility of Stutzer's method as it is clear that a rolling window period method will induce correlation between performances due to the significant number of identical daily performances present in rolling periods prior to and subsequent of each other (see Polakow *et al.* (2014)). This notion therefore violates the requirement of independence.

4. General Pricing Method for Interest Rate Derivatives

Due to the illiquid and non-transparent nature of the South African interest rate derivative market, little or no pricing data can be found for less popular contingent claims. Hence we are unable to make use of non-parametric methods that require only historical price data. We therefore require a method for the pricing of interest rate derivatives that requires historical swap curve data.

The pricing method of canonical valuation by Stutzer (1996) has been identified which allows us to price within a market where only sufficient underlying historical data can be found. This method makes use of the PME within a discrete setting.

A new method, named the relative entropy approach (REA), is then developed using relative entropy principles in a continuous setting and also allows for pricing using historical data of the underlying.

This chapter derives the generalised theory for the pricing of interest rate derivatives, and in particular swaptions, for the aforementioned methods.

4.1 Canonical Valuation for Interest Rate Derivatives

As we are now pricing interest rate derivatives, the method of Stutzer (1996)'s canonical valuation under the $Q^{\alpha,\beta}$ swap measure is presented[1]. The method makes use of the principles of maximum entropy.

While all previous literature on entropy has been presented in a continuous setting, canonical valuation is presented in a discrete setting as a discrete setting is fundamental to the basis of the method itself. The method seeks to take a discrete set of observations, each assumed to be of equal probability of occurrence, and maximise the entropy of that set, while adhering to a set of constraints under $Q^{\alpha,\beta}$, through modifying each observation's individual probability. The method does not make use of any continuous probability distribution for the set of observations.

[1] This is also similar to the method of Stutzer and Chowdhury (1999) used to price bond futures options.

Stutzer (1996) observes that use of a small amount of option pricing data can greatly increase the accuracy of results. Hence, in addition to the $Q^{\alpha\beta}$ swap measure constraints, we present the option of an additional market implied at-the-money swaption price constraint equivalent to that presented by Stutzer for equity derivatives.

4.1.1 Method

Consider the swaption $S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$ to be priced at time t_0 with expiry t_α and tenor length $t_\beta - t_\alpha$, where $t_\beta > t_\alpha \geq t_0$. The swaption's floating leg resets at times $t_\alpha, t_{\alpha+1}, ..., t_{\beta-1}$ and pays at times $\mathcal{T} = [t_{\alpha+1}, ..., t_{\beta-1}, t_\beta]$ against a fixed leg struck at r_k, while $\omega = 1(-1)$ for a payer (receiver) swaption.

Let $r(t_0, t_\beta - t_\alpha)$ represent the current fair swap rate with tenor $t_\beta - t_\alpha$ with payment times $\mathcal{T}$. The current fair forward swap rate is denoted by $r(t_0, t_\alpha, t_\beta)$ and has the tenor $t_\beta - t_\alpha$ with maturity at time t_α.

Let $r_x(t_\alpha, t_\beta)$, for $i \in \{1, ..., N\}$, be the x^{th} simulated fair swap rate at time t_α with tenor $t_\beta - t_\alpha$. Now consider a set of N single period (t_α) returns X, such that $r(t_0, t_\beta - t_\alpha) + X_x = r_x(t_\alpha, t_\beta), \quad x \in \{1, ..., N\}$[2]. The N possible simulated fair swap rates $r_x(t_\alpha, t_\beta)$ and returns X_x are constructed from a set of H historical daily observations of the fair swap rate with tenor $t_\beta - t_\alpha$. Their construction is discussed in Section 3.3.

Each observed performance X_x is assumed to have an equal probability of occurring over the single t_α period as per the PIR under the P measure. Each probability of occurrence is denoted $p(x) = \frac{1}{N}, \; x = 1, ..., N$. We seek to move to a $Q^{\alpha,\beta}$ swap measure[3], denoted by pmf q, using maximum entropy principles. Due to an incomplete market however, many candidate $Q^{\alpha,\beta}$ swap measures exist. We therefore move to the optimal $Q^{\alpha,\beta}$ swap measure $q^\star$ through maximisation of the Shannon Entropy (2.2),

$$-\sum_{i=1}^{N} q^\star(x) \ln(q^\star(x)), \tag{4.1}$$

[2] To reiterate, we use returns rather than performances in order to allow for existence of rates over the space $(-\infty, \infty)$ with positive probability.

[3] Here we define change of measure by Radon-Nikodym theorems. Please refer to Appendix B for further detail.

subject to the $Q^{\alpha,\beta}$ swap measure constraints

$$\sum_{x=1}^{N} q^{\star}(x) = 1,$$

$$E^{\alpha,\beta}\left[r(t_\alpha, t_\beta)\right] = \sum_{x=1}^{N} q^{\star}(x) r_x(t_\alpha, t_\beta) = r(t_0, t_\alpha, t_\beta) \tag{4.2}$$

and the market implied swaption price constraint

$$\boldsymbol{C}_x^{atm} q^{\star}(x) = C(t_0, t_\alpha, t_\beta) \sum_{x=1}^{N} (r_x(t_\alpha, t_\beta) - r(t_0, t_\alpha, t_\beta))^+ q^{\star}(x) = S_{ATM}$$

where

$$C(t_0, t_\alpha, t_\beta) = \sum_{j=1}^{n} \tau_j Z(t_0, t_j), \qquad \text{and} \qquad \tau_j = t_j - t_{j-1}.$$

$Z(t_0, t_j)$ is the discount factor which is bootstrapped off the swap spot rate curve over period $[t_0, t_j]$ and n is the number of payments in $\mathcal{T}$. S_{ATM} represents the current market price for an at-the-money swaption[4]of maturity t_α and tenor $t_\alpha - t_\beta$. As derived in Section 2.5, the solution to the above problem is

$$q^{\star}(x) = \frac{\exp(\lambda_1 r_x(t_\alpha, t_\beta) + \lambda_2 \boldsymbol{C}_x^{atm})}{\sum_{x=1}^{N} \exp(\lambda_1 r_x(t_\alpha, t_\beta) + \lambda_2 \boldsymbol{C}_x^{atm})},$$

for $x = 1, ..., N$, where λ_1, λ_2 can be solved by finding the global minimum of the convex function

$$F(\lambda_1, \lambda_2) = \sum_{x=1}^{N} \exp\left[\lambda_1 \left(r_x(t_\alpha, t_\beta) - r(t_0, t_\alpha, t_\beta)\right) + \lambda_2 \left(\boldsymbol{C}_x^{atm} - S_{ATM}\right)\right].$$

Further rigor of the above solutions can be found in Ben-Tal (1985).

Having solved for $q^{\star}(x)$ under $Q^{\alpha\beta}$ we calculate the swaption price using

$$S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = C(t_0, t_\alpha, t_\beta) \sum_{x=1}^{N} \left(\omega\left[r_x(t_\alpha, t_\beta) - r_k\right]\right)^+ q^{\star}(x). \tag{4.3}$$

[4] Note that S_{ATM} is equivalent for the payer and receiver swaption.

4.2 The Relative Entropy Approach for Interest Rate Derivatives

A new method, the relative entropy approach (REA), is now presented for the pricing of interest rate derivatives of a short and long-term nature. In an effort to price swaptions, the method utilises the $Q^{\alpha,\beta}$ swap measure. The method is of a continuous framework and makes use of historical data of the underlying only with the option to use market implied swaption prices as an additional constraint.

The REA is based, largely, on the work done by Buchen and Kelly (1996) using relative entropy principles (which they refer to as cross-entropy). However, it is extended to allow for pricing using historical data of the underlying variable rather than market data of derivative prices. The method makes use of smoothing techniques by Duan (2002) [5].

4.2.1 Method

Consider the swaption $S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$ to be priced at time t_0 with expiry t_α and tenor length $t_\beta - t_\alpha$, where $t_\beta > t_\alpha \geq t_0$. The swaption's floating leg resets at times $t_\alpha, t_{\alpha+1}, ..., t_{\beta-1}$ and pays at times $\mathcal{T} = [t_{\alpha+1}, ..., t_{\beta-1}, t_\beta]$ against a fixed leg struck at r_k, while $\omega = 1(-1)$ for a payer (receiver) swaption.

Let $r(t_0, t_\beta - t_\alpha)$ represent the current fair swap rate of tenor $t_\beta - t_\alpha$ with payment times $\mathcal{T}$. The current fair forward swap rate of tenor $t_\beta - t_\alpha$ and maturity t_α is $r(t_0, t_\alpha, t_\beta)$.

Let $r_i(t_\alpha, t_\beta)$, for $i \in \{1, ..., N\}$, be the i^{th} simulated fair swap rate at time t_α with tenor $t_\beta - t_\alpha$. Now consider a set of N t_α period returns X, such that $r(t_0, t_\beta - t_\alpha) + X_i = r_i(t_\alpha, t_\beta), \quad i \in \{1, ..., N\}$. Consider X_i as subset of N observations from the subset $\boldsymbol{X} \in \chi$ where χ is the set of all possible t_α returns. The N simulated future fair swap rates $r_i(t_\alpha, t_\beta)$ and returns X_i are constructed from a set of H historical daily observations of the fair swap rate with tenor $t_\beta - t_\alpha$. Their construction is discussed in Section 3.3.

Our first step is to smooth a function for the set of returns $X_i \in \chi$[6]. We define

[5] Note that Duan (2002) also adopts relative entropy principles. However the REA does not impose any assumptions on the empirical distribution of the underlying variable as required by Duan, who proposes his method for the pricing of path-dependent equity derivatives under the risk-neutral measure.

[6] Should we make use of t_α period performances rather than returns, it is suggested to smooth a function for the log of the performances (log-returns) as the following method smoothes a two-tailed distribution over the space $(-\infty, \infty)$.

the empirical cumulative distribution

$$\hat{G}(x, \boldsymbol{X}) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{X_i \leq x\}}. \tag{4.4}$$

We can now define G as the smoothed cumulative distribution function of $\hat{G}(x, \boldsymbol{X})$. Our variable x therefore represents the return of our future fair swap rate, $r(t_\alpha, t_\beta)$.

We further define g as a smoothed probability density function of the t_α period returns, $\boldsymbol{X}$, where $G(x) = \int_{-\infty}^{x} g(t)dt$.

We require smoothing of $\hat{G}(x, \boldsymbol{X})$ in order to obtain a function that is invertible. The function G further dampens sampling fluctuation and is integrable should it be well-behaved. G can be smoothed using a multinomial cumulative normal distribution to represent the function within 1.5 standard deviations of its mean and a Pareto maximal likelihood function to represent the tails of the distribution function. This procedure was suggested by Duan (2002) and is detailed in Appendix 5.1 of his paper.

In practice a kernel fitting function was found to fit the empirical cumulative distribution of the underlying data sufficiently within 1.5 standard deviations of its mean. A Pareto distribution was still used for the tails of the empirical distribution. The procedure was implemented using the 'paretotails' function within the MATLAB programming language. The function requires specification of a smoothing method within certain quantile bounds of the empirical distribution, while a pareto tails distribution function is fitted to the empirical distribution outside of these bounds. Here a kernel distribution function was fitted within the quantile bounds. The quantile bounds were estimated as equivalent to 1.5 standard deviations of the distribution mean. Quantile bounds were approximately 0.065 and 0.935 for the derived empirical cumulative distributions. Figure 4.1 shows the error between the empirical and fitted cumulative distributions, $\hat{G}(x, \boldsymbol{X})$ and G, when using the pareotails function.

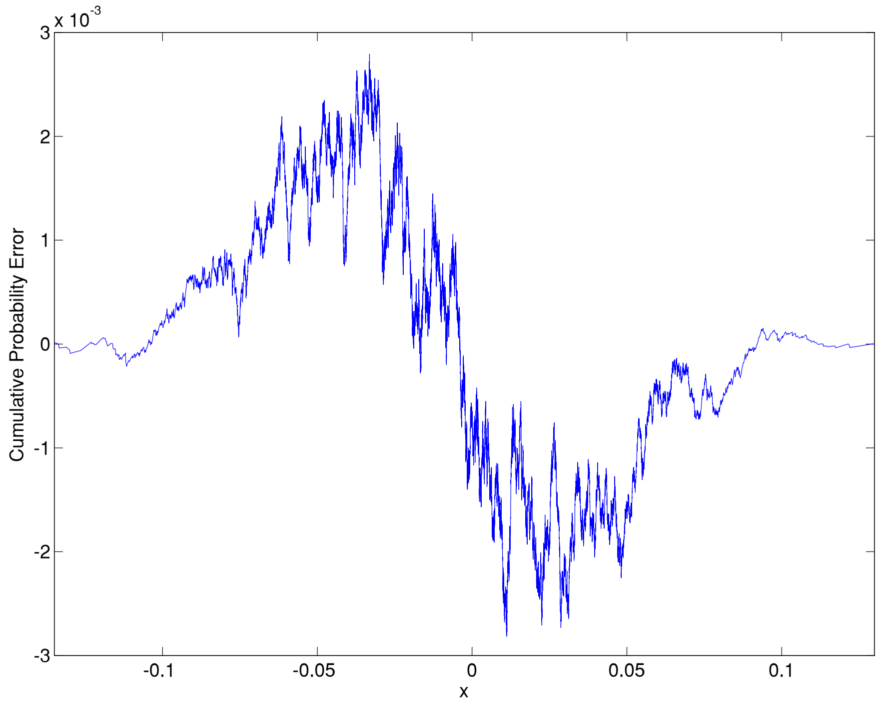
We seek to move to the optimal $Q^{\alpha,\beta}$ swap measure distribution, q, using relative entropy principles. We therefore seek to minimise (2.12). As g is our real world sample distribution we choose it as our best estimate of the real world population distribution p. We therefore substitute p for g. We are effectively moving from g to q via the principles of relative entropy as discussed in section 2.6. Hence we minimise

$$\int_{-\infty}^{\infty} q(x) \ln \left(\frac{q(x)}{g(x)} \right) dx \tag{4.5}$$

subject to the constraints

$$\int_{-\infty}^{\infty} q(x)dx = 1,$$

Fig. 4.1: Error between the empirical and fitted cumulative distributions, $\hat{G}(x, \boldsymbol{X})$ and G, for the future fair ten-year swap rate ten-year returns.



$$\int_{-\infty}^{\infty} c_1(x)q(x)dx = r(t_0, t_\alpha, t_\beta)$$

and

$$\int_{-\infty}^{\infty} c_2(x)q(x)dx = S_{ATM}$$

where

$$c_1(x) = r(t_0, t_\beta - t_\alpha) + x, \;\; c_2(x) = C(t_0, t_\alpha, t_\beta)\left(c_1(x) - r(t_0, t_\alpha, t_\beta)\right)^+$$

and

$$C(t_0, t_\alpha, t_\beta) = \sum_{j=1}^{n} \tau_j Z(t_0, t_j), \;\; \text{where} \;\; \tau_j = t_j - t_{j-1}.$$

$Z(t_0, t_j)$ is the discount factor which is bootstrapped off the swap spot rate curve over the period $[t_0, t_j]$ and n is the number of payments in $\mathcal{T}$. S_{ATM} represents the current market price for an at-the-money swaption of maturity t_α and tenor $t_\beta - t_\alpha$.

As derived in Section 2.6, the solution to the above problem is

$$q(x) = \frac{g(x)\exp\left(\lambda_1 c_1(x) + \lambda_2 c_2(x)\right)}{\int_{-\infty}^{\infty} g(x)\exp\left(\lambda_1 c_1(x) + \lambda_2 c_2(x)\right) dx}, \tag{4.6}$$

where $\lambda_1,\ \lambda_2$ are Lagrange multipliers.

It is now necessary to solve for the Lagrange parameters λ_1 and λ_2. Should the problem constraints be linearly independent, Buchen and Kelly (1996) derive a convenient solution to finding λ_1 and λ_2. A rigorous derivation of their solution can be found in Section IV of their paper. λ_1 and λ_2 can be solved by finding the global minimum of the scalar function $F(\lambda_1, \lambda_2)$ where

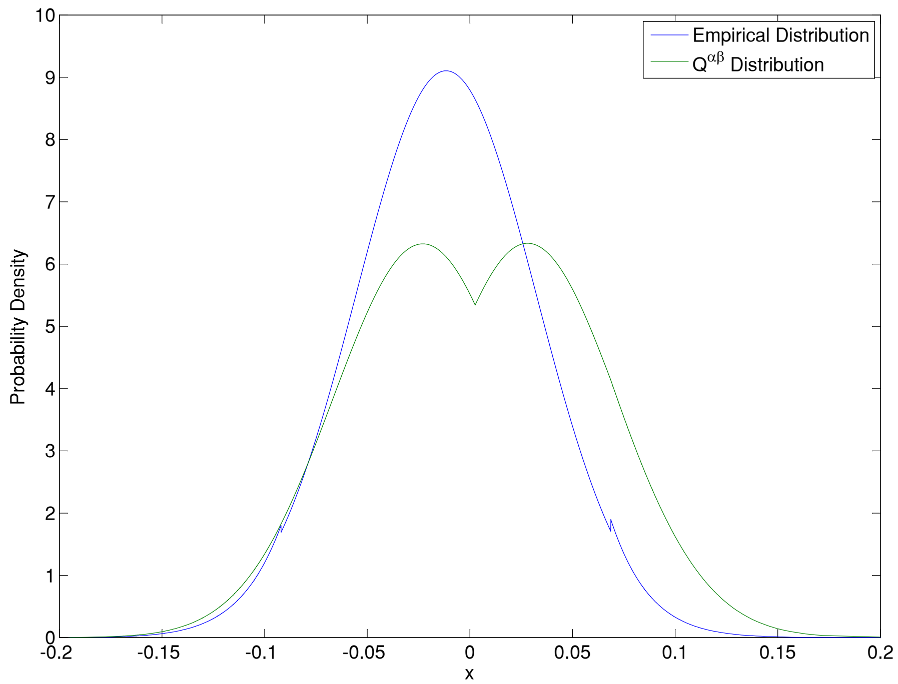
$$F(\lambda_1, \lambda_2) = \ln\left(\int_{-\infty}^{\infty} g(x)\exp\left(\lambda_1 c_1(x) + \lambda_2 c_2(x)\right) dx\right) - \left(\lambda_1 r(t_0, t_\alpha, t_\beta) + \lambda_2 S_{ATM}\right). \tag{4.7}$$

In practice the global minimum of $F(\lambda_1, \lambda_2)$ was solved for using the unconstrained minimisation function, 'fminunc', for two variables within the MATLAB programming software.

Figure 4.2 shows a comparison of the real-world distribution g and the optimal distribution q under the $Q^{\alpha,\beta}$ swap measure[7]. The double hump observed around $x = 0$ in the q distribution is reasoned due to the entropic manipulation of the original distribution (to remain as entropically close to g as possible) in compliance with the two main constraints - one to fit the expectation of historical swap rates to the market forward swap rate and the other to fit the expectation of swaption prices based on a historical swap rate data to the current ATM swaption price. Each distribution would be expected to be somewhat normal while the amalgamation of the two constraints leads to a more interesting shape. It is worth noting that this double hump is observed throughout distributions of swap rates of increasing tenor and swaption maturity, albeit to a diminishing extent.

[7] It is interesting to note that the $Q^{\alpha,\beta}$ distribution has gained entropy under its movement from the P distribution. The REP does not infer a decrease in entropy (Both $D(q||p) > 0$ and $D(p||q) > 0$ for $p \neq q$). It rather minimises the entropic distance between the two distributions while adhering to the imposed constraints. Only when assuming a uniform distribution under P, equivalent to the PME, can we assume a decrease in entropy in our movement between P and $Q^{\alpha\beta}$ as the uniform distribution holds maximum entropy.

Fig. 4.2: Comparison of the empirical real-world sample distribution, g, and the optimal $Q^{\alpha,\beta}$ distribution, q, obtained through relative entropy principles.



Finally, we can solve for the swaption price[8],

$$S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = C(t_0, t_\alpha, t_\beta) \int_{-\infty}^{\infty} \left(\omega \left[c_1(x) - r_k\right]\right)^+ q(x)dx. \tag{4.8}$$

4.3 Deriving the Swaption Skew

Having derived the relevant swaption pricing skew from either method of canonical valutaion or the relative entropy approach, we simply 'back out' the volatility skew using the market convention - we back out the implied volatility using the Black-76 model for swaptions. Under the Black-76 model[9],

$$S_\omega(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = \omega C(t_0, t_\alpha, t_\beta) \left[r(t_0, t_\alpha, t_\beta)\mathcal{N}(\omega d_1) - r_k \mathcal{N}(\omega d_2)\right] \tag{4.9}$$

[8] Equation 4.8 was calculated using the 'integral' function within the MATLAB programming language.

[9] A rigorous presentation of the pricing of forwards contracts can be found in the Black (1976) paper.

where $\omega = 1(-1)$ for a payer (receiver) swaption, and

$$d_1 = \frac{\log\left(\frac{r(t_0,t_\alpha,t_\beta)}{r_k}\right) + \frac{1}{2}\sigma^2 t_\alpha}{\sigma\sqrt{t_\alpha}}, \qquad d_2 = \frac{\log\left(\frac{r(t_0,t_\alpha,t_\beta)}{r_k}\right) - \frac{1}{2}\sigma^2 t_\alpha}{\sigma\sqrt{t_\alpha}}.$$

$S(\omega, t_0, t_\alpha, t_\beta), C(t_0, t_\alpha, t_\beta), r(t_0, t_\alpha, t_\beta)$ and r_k are defined as before while σ is the implied volatility of the swaption which is usually derived from market prices. However we derive this implied volatility from historical data using the proposed non-parametric methods. In order to imply the historical volatility from the relevant swaption price $S(\omega, t_0, t_\alpha, t_\beta)$[10] we make use of the 'fzero' function using the MATLAB programming language.

[10] Of course both the payer and receiver swaption of identical characteristics imply the same volatility.

5. Simulations in an Idealised World

In order to gain an understanding of the performance of the proposed methods to price long-term swaptions, the methods are tested in a idealised world across a range of moneyness and maturities. Testing is performed under a one factor Vasicek model framework. The Vasicek model is briefly introduced before the simulation procedure and results are presented.

5.1 A brief introduction to the Vasicek model

This section is taken largely from the presentation by Brigo and Mercurio (2007). The Vasicek model describes the dynamics of the short rate under the risk neutral measure using the Ornstein-Uhlenbeck process

$$dr_t = \kappa(\theta - r_t)dt + \sigma d\widetilde{W}_t \tag{5.1}$$

where r_t is the short rate at time t, $\widetilde{W}_t$ is a standard Brownian motion under Q, and constant parameters $\kappa, \theta, \sigma \in \mathbb{R}^+$ represent the rate of mean reversion, the level of mean reversion and the volatility respectively. For given times $t_2 > t_1$, the distribution of r_{t_2} given r_{t_1} is conditionally normal under Q with

$$\boldsymbol{E}^{\boldsymbol{Q}}[r_{t_2}|r_{t_1}] = e^{-\kappa(t_2-t_1)}r_{t_1} + \theta\left(1 - e^{-\kappa(t_2-t_1)}\right) \tag{5.2}$$

and

$$\boldsymbol{Var}^{\boldsymbol{Q}}[r_{t_2}|r_{t_1}] = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t_2-t_1)}\right). \tag{5.3}$$

Using equations 5.2, 5.3 we propose to simulate r_t for times $t_0 < t_1 < ... < t_n$ using

$$r_{t_{i+1}} = \boldsymbol{E}^{\boldsymbol{Q}}[r_{t_{i+1}}|r_{t_i}] + \boldsymbol{Var}^{\boldsymbol{Q}}[r_{t_{i+1}}|r_{t_i}]Z_{i+1} \tag{5.4}$$

for $i = 0, 1, ..., n$ where $Z_i \sim \mathcal{N}(0, 1)$.

The price of the zero coupon bond $Z(t_1, t_2)$ of yield $Y(t_1, t_2)$ can be derived under Vasicek dynamics as

$$Z(t_1, t_2) = \exp[Y(t_1, t_2)(t_2 - t_1)] \tag{5.5}$$

$$= \exp\left[(A(t_1, t_2) - B(t_1, t_2)r_{t_1})\right] \tag{5.6}$$

where

$$B(t_1, t_2) = \frac{1 - e^{-\kappa(t_2 - t_1)}}{\kappa} \tag{5.7}$$

and

$$A(t_1, t_2) = \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)(B(t_1, t_2) - (t_2 - t_1)) - \frac{\sigma^2}{4\kappa^2}B(t_1, t_2)^2. \tag{5.8}$$

Jamshidian (1989) derives an analytical solution for the price at time t of a European zero coupon bond option under the Q^T forward measure with strike X, option maturity T and bond maturity S:

$$\boldsymbol{ZBO}(t, T, S, X) = \omega\left[Z(t, S)\Phi(\omega h) - XZ(t, T)\Phi(\omega(h - \sigma_p))\right] \tag{5.9}$$

where

$$\sigma_p = \sigma\sqrt{\frac{1 - e^{-2\kappa(T-t)}}{2\kappa}}B(T, S)$$

and

$$h = \frac{1}{\sigma_p}\ln\frac{Z(t, S)}{Z(t, T)X} + \frac{\sigma_p}{2}.$$

$\omega = 1$ (-1) for a call (put) option while $\Phi(\cdot)$ is noted as the cumulative normal distribution function.

Making use of equation 5.9 we are now able to define the analytical price of a receiver swaption. Jamshidian (1989) presents a decomposition of a European coupon bearing bond option as the summation of European zero coupon bonds while a swaption can be viewed as an option on a coupon bearing bond. Consider a receiver swaption to be priced at time t_0 with maturity t_α and swap expiry t_β, where $t_0 < t_\alpha < t_\beta$. The swaption pays in arrears at times $\mathcal{T} = [t_{\alpha+1}, ..., t_{\alpha+n} = t_\beta]$ against a fixed leg struck at r_k. Let $\tau_i = t_i - t_{i-1},\ i = \alpha + 1, ..., \alpha + n$. Set $c_i := r_k\tau_i,\ i = \alpha + 1, ..., \alpha + n - 1$, and $c_n = 1 + r_k\tau_n$. We define r^* as the spot rate at time t_α where

$$\sum_{i=\alpha+1}^{\alpha+n} c_i \exp\left(A(t_\alpha, t_i) - B(t_\alpha, t_i)r^*\right) = 1.$$

Set $X_i := \exp\left(A(t_\alpha, t_i) - B(t_\alpha, t_i)r^*\right)$. Considering the above definitions, we derive the analytical price to the subsequent receiver swaption[1] as

$$S_R(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k) = \sum_{i=1}^{n} c_i\boldsymbol{ZBC}(t, T, t_i, X_i) \tag{5.10}$$

[1] Throughout the dissertation we assume a swaption notional of 1.

5.2 Simulation Procedure and Results

The test measures the pricing ability of the non-parametric methods within an idealised case where all dynamics are known[2]. The proposed methods are used to price receiver swaptions of ten-year tenor and of varying maturity and strike under a one factor Vasicek model framework.

Consider current time t_0 with $t_{-H} \leq t_0 \leq t_\alpha \leq t_\beta$. We simulate $H = 2000$ observations of the short rate using equation 5.4 with time steps of one day $\left(\delta t = \frac{1}{252}\right)$. Under simulation the same parameters as those of Duffee and Stanton (2012) are used which were estimated from the behaviour of Treasury data over a period of 30 years. Specifically, $\kappa = 0.0065$, $\theta = 0.052$[3] and $\sigma = 0.0175$ with $r_0 = 0.06$[4]. For each time step $t_i, i \in 0, -1, ..., -H$ we calculate the observed fair swap rate $r(t_i, t_\beta - t_\alpha)$ of tenor $t_\beta - t_\alpha = 10$ years, using equations 5.6 and 3.1.

Having created a set of N daily swap rates $r(t_i, t_\beta - t_\alpha)$, we implement the cross-sectional sampling method discussed in section 3.3 to obtain a set of $N = 10000$ future fair swap rates $r(t_0, t_\alpha, t_\beta)$. We can then implement the methods of canonical valuation and the relative entropy approach, discussed in sections 4.1 and 4.2 respectively, in order to price the receiver swaption $S_R(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$. Both methods make use of the ATM price constraint S_{ATM}.

A 'historical price' estimate is also implemented. This estimate allows us to compare the performance of the non-parametric methods relative to the performance of Vasicek estimates within its 'own world' while all methods are restricted to the data simulated[5]. Vasicek parameters κ, θ and σ are estimated from set of calculated yields on zero coupon bonds of terms 3 months, 6 months, 1 year, 2 years, 5 years and 10 years for each time step t_i using the simulated short rate. Parameters are

[2] This is a widely used testing method of the pricing ability of non-parametric methods and can be found in Duan (2002), Stutzer (1996), Gray and Newman (2005), Haley and Walker (2010) and Alcock and Gray (2005). Subsequent tests have measured the performance of non-parametric pricing of equity derivatives under a Black-Scholes model framework as opposed to the pricing of interest rate derivatives under a one factor Vasicek model framework, as in our case.

[3] While Duffee and Stanton (2012) estimate $\kappa\theta$ as a parameter, θ is taken as their estimated mean interest rate.

[4] Here we use the same parameters as Duffee and Stanton (2012) as we are then able to compare the results of our estimation technique using a linearised Kalman Filter with their results which followed a similar experiment using the same Kalman Filter. Similar paramater estimates were found when testing an albeit smaller set of South African swap rate data.

[5] Gray and Newman (2005) note that in practice it is not possible to verify a specific dynamic of the underlying which is a key motivation behind the use of non-parametric methods.

estimated through a log-likelihood optimisation routine using a Kalman Filter[6][7]. The estimated parameters are then used for the pricing of the receiver swaption under Vasicek using equation 5.10 where $M = 1$.

Pricing using the three methods is repeated for n=500 simulations for each receiver swaption $S_R(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$ over a range of maturities $t_\alpha = 0.5, 1, 2, 5, 10, 15$ years and moneyness $R/K = 0.85, 0.97, 1, 1.03, 1.125$. Each simulation is independent while all pricing methods make use of the same short rate or swap path per simulation. Denote the methods of the historical price estimate, canonical valuation and the relative entropy approach as HVM, CAN and REA respectively. Price estimates under HVM, CAN and REA are compared to the true Vasicek price (denoted VRS) for $S_R(t_0, t_\alpha, \mathcal{T}, t_\beta, r_k)$ using equation 5.10 with the known parameters. A pricing error is simply calculated per simulation for each method as

$$PE_{i,j} = \frac{Method_j - VRS}{VRS}, \quad i = 1, ..., n, \tag{5.11}$$

while j indicates the pricing method used. The absolute pricing error, $APE_{i,j}$, is calculated as the absolute of equation 5.11. A mean pricing error (MPE) per swaption can then be calculated per method from the n simulations as $MPE_j = \frac{1}{n}\sum_{i=1}^{n} PE_{i,j}$ while the mean absolute pricing error is similarly calculated as $MAPE_j = \frac{1}{n}\sum_{i=1}^{n} APE_{i,j}$.

It is important to note that our three prices estimates are calculated under different measures. We simulate the short rate and estimate the Vasicek parameters under Q[8] while the analytical swaption price under Vasicek is derived under the Q^T forward measure. Methods of canonical valuation and the relative entropy approach uses data under Q in this simulation to price the swaption under the $Q^{\alpha,\beta}$ forward swap measure through a change of measure. However the swaption price under Q is invariant by change of numeraire and prices under each measure are equivalent. We refer to proposition 2.2.1 and 'Fact Two' in Brigo and Mercurio (2007) which are presented in appendix B of this dissertation.

[6] Duffee and Stanton (2012) note the linearised Kalman Filter as a tractable and reasonably accurate estimation technique and recommend this technique where maximum likelihood is impractical. In our case maximum likelihood methods are noted to produce strongly biased parameter estimates.

[7] While not within the scope of this dissertation, a formal introduction of the linearised Kalman Filter can be found in Appendix C of this dissertation while further rigorous application can be found in Duffee and Stanton (2012). We implement the estimation technique under Vasicek directly from this paper. Moreover, the Vasicek simulation presented here is based largely on the Vasicek simulation of Duffee and Stanton (2012). Log-likelihood optimisation was accomplished using the constrained minimisation function 'fmincon' on the negative likelihood in MATLAB.

[8] We have not simulated the short rate under P due to the difficulty in estimation of the risk parameters as shown by Duffee and Stanton (2012) and as an effort to keep the simulation test succinct to its purpose. One could say that we have simulated the short rate under P with the market price of risk equal to zero.

Tab. 5.1: Parameter estimates of a one factor Vasicek model using the linearised Kalman Filter

Parameter	True Value	Mean	Std Dev
κ x10^2	6.50	6.51	0.20
θ x 10^2	5.20	5.22	0.72
σ x 10^3	17.50	17.55	0.32

Table 5.1 shows parameter estimate performance of the linearised Kalman Filter throughout the simulation for the one factor Vasicek model. The mean and standard deviation of the estimates for κ, θ, σ are similar to the results of Duffee and Stanton (2012). Moreover, performance is slightly better as we do not add noise to the simulated yield paths.

Table 5.2 displays the MAPE estimates for the HVM, CAN and REA methods. MAPE of HVM show a slight increase across moneyness and and maturity which is consistent with the results of Gray and Newman (2005). An increase in MAPE across maturity can be inferred through the use of equations 5.7, 5.8 and 5.9 when pricing the swaption under equation 5.10. CAN and REA MAPE is found to increase as options move further in or out the money which is reasoned to due their at-the-money implied pricing constraint. CAN MAPE is found to increase monotonically across maturity while REA MAPE maintains a slight decrease across maturity, albeit not always monotonic in nature.

HVM clearly outperforms the CAN MAPE across both moneyness and maturity despite the added at-the-money pricing constraint of CAN. REA MAPE however, while also possessing this constraint, is similar to HVM results and often outperforms the HVM method as maturity increases. The REA method clearly outperforms the method of CAN across moneyness and maturity with the exception of deep out-the-money options of short maturity. Moreover, out-performance by REA estimates is accentuated as maturity increases due to the inflating MAPE of CAN estimates.

What reasoning is there for the compounding increase in error between the CAN and REA estimates over an increase in maturity? Non-parametric methods re-weight the entire distribution of the underlying. It is noted by Haley and Walker (2010) that as this distribution becomes more dispersed so the precision of the non-parametric estimator decreases and hence error increases. It is well known that the greater the period of the option the more dispersed the distribution of the underlying[9]. Therefore the greater the maturity of the swaption the more dispersed

[9] In most Gaussian cases $d\widetilde{W}_t \sim \mathcal{N}(0, T)$.

Tab. 5.2: MAPE Estimates for a 10-Year Tenor Receiver Swaption:

	Time to Expiration (years)						
Moneyness (R/K)	$\frac{1}{2}$	**1**	**2**	**5**	**10**	**15**	**Method**
Deep Out-the-money	0.0056	0.0017	0.0034	0.0080	0.0048	0.0142	HVM
0.90	0.0233	0.0256	0.0356	0.0535	0.0662	0.0683	CAN
	0.0750	0.0384	0.0145	0.0012	0.0013	0.0012	REA
Out-the-money	0.0055	0.0030	0.0046	0.0060	0.0097	0.0116	HVM
0.97	0.0080	0.0082	0.0112	0.0162	0.0198	0.0203	CAN
	0.0063	0.0017	0.0008	0.0007	0.0006	0.0006	REA
At-the-money	0.0082	0.0042	0.0051	0.0058	0.0068	0.0123	HVM
1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	CAN
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	REA
In-the-money	0.0076	0.0060	0.0066	0.0059	0.0102	0.0114	HVM
1.03	0.0094	0.0089	0.0120	0.0165	0.0193	0.0224	CAN
	0.0014	0.0012	0.0011	0.0010	0.0009	0.0033	REA
Deep In-the-money	0.0090	0.0083	0.0087	0.0084	0.0093	0.0103	HVM
1.12	0.0493	0.0454	0.0511	0.0693	0.0796	0.0807	CAN
	0.0107	0.0084	0.0074	0.0061	0.0051	0.0045	REA

the returns of its underlying which, hence, incurs an increasing error in CAN. This does not effect REA to the same extent however. The reason for this is due to the smoothing of the discrete empirical distribution to a continuous setting. By fitting the tails of this distribution with a Pareto maximum likelihood function we have effectively constrained the distribution extremes to behave as per the calibrated Pareto function. This controls the distribution extremes as the distribution moves from the Q to $Q^{\alpha\beta}$ measure[10]. For CAN however, estimation of extreme events are discrete and unconstrained as the distribution moves from Q to $Q^{\alpha\beta}$. Hence an already dispersed distribution under the Q measure for a long-dated maturity often 'explodes' under a change of measure, leading to poor estimates of the distribution under $Q^{\alpha\beta}$ when using CAN and therefore exhibits poor pricing estimates.

MPE, while not giving an indication as to the accuracy of the pricing method, allows for analysis of the direction of error of pricing estimates under each method as well as an average pricing effectiveness. MPE results for the HVM, CAN and REA methods under simulation are found in Table 5.3. HVM is found to generally,

[10] Note that we only move from Q under simulation. In practice the methods CAN and REA would move from the real world measure P to the swap measure $Q^{\alpha\beta}$.

Tab. 5.3: MPE Estimates for a 10-Year Tenor Receiver Swaption:

	Time to Expiration (years)						
Moneyness (R/K)	$\frac{1}{2}$	**1**	**2**	**5**	**10**	**15**	**Method**
Deep Out-the-money	-0.0006	0.0007	0.0007	0.0001	-0.0005	-0.0015	HVM
0.90	-0.0128	-0.0233	-0.0355	-0.0535	-0.0662	-0.0680	CAN
	-0.0750	-0.0384	-0.0145	-0.0009	-0.0009	-0.0006	REA
Out-the-money	0.0009	0.0000	0.0011	0.0009	0.0007	0.0011	HVM
0.97	-0.0046	-0.0075	-0.0111	-0.0162	-0.0198	-0.0202	CAN
	-0.0063	-0.0017	-0.0007	-0.0007	-0.0005	-0.0005	REA
At-the-money	0.0039	0.0009	0.0003	0.0017	-0.0008	0.0039	HVM
1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	CAN
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	REA
In-the-money	0.0019	0.0040	0.0007	0.0031	0.0006	0.0022	HVM
1.03	0.0060	0.0082	0.0120	0.0165	0.0192	0.0171	CAN
	0.0013	0.0012	0.0011	0.0010	0.0008	-0.0018	REA
Deep In-the-money	0.0025	0.0051	0.0047	0.0033	0.0045	0.0056	HVM
1.12	0.0344	0.0416	0.0510	0.0693	0.0790	0.0799	CAN
	0.0102	0.0083	0.0074	0.0060	0.0047	0.0040	REA

but not always, overprice the receiver swaption while CAN and REA price estimates are found to be underpriced (overpriced) for out-the-money (in-the-money) receiver swaptions. The similar results of the absolute of MPE estimates in comparison to MAPE estimates shows that this is the case for almost all n swaption price estimates under CAN and REA. It can be reasonably assumed from equation 3.4 that the opposite would be observed for payer swaptions of equivalent moneyness. It is clear that, regarding average pricing effectiveness, REA is found to clearly outperform CAN across moneyness and maturity (bar deep out-the-money short maturity swaptions) while also performing admirably relative to and often outperforming HVM.

6. Results of JSE Swap Curve Volatility Skew Analysis

This chapter presents results of the constructed pricing surfaces and the swaption skew for swaptions of 10-year tenor using JSE noise-reduced swap curve data spanning from 16-Jan-2006 to 06-Jan-2015. This allowed for a total of approximately 2200 historical swap curve daily observations. All data was presented as either NACQ or NACS spot rates. The spot rates were therefore converted to NACC spot rates which could then be used to calculate a historical time series for the 10-year fair swap rate. Daily returns regarding the 10-year swap rate could then be calculated. Subsequently the sampling method in Section 3.4 could be used to derive t_α period returns, for $\alpha = 0.25, 0.5, ..., 30 - t_{\text{ten}}$ where t_{ten} is the swaption tenor.

The relative entropy approach was opted for use to construct the subsequent pricing surface under the method described in Section 4.2 while the respective swaption skew was derived using equation 4.9. Results concerning pricing accuracy under simulation are discussed in Chapter 5.

6.1 Sample Distributions

The sampling procedure as described in Section 3.4 was followed for construction of the JSE swaption skews. As mentioned, the procedure depends on the assumption that the daily returns data is independently distributed. This assumption was therefore verified for the JSE swap curve data used for the construction of the swaption skews. As swaption skews of 10-year tenor are to be constructed, historical data of the 10-year fair swap rate are to be used respectively. Each set of data was checked for auto-correlation as well as correlation between itself and all quarterly fair swap rates along the swap curve. Using a Ljung-Box Q-Test, auto-correlation of data could be tested to accept or reject the H_0 hypothesis that the data is free of auto-correlation at a 95% level. This test was repeated for correlation between different data sets.

Results of the above tests showed auto-correlation for the 10-year fair swap rate to be 13.6%. Tests on the 10-year fair swap rate data were found to accept the H_0

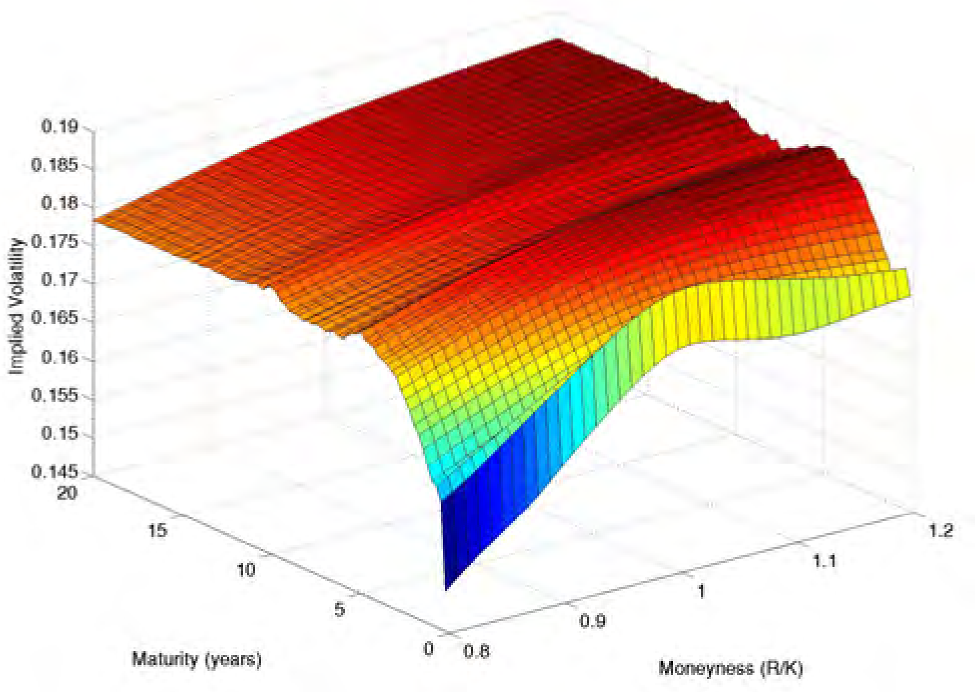
hypothesis at a 95% significance level.

Correlations between the 10-year fair swap rate data and quarterly fair swap rate data for the JSE swap curve were found insignificant while similar tests for significance were found to accept that the data was free of correlation with other data sets at a 95% significance level. The data required for the cross-sectional sampling method in section 3.4 was therefore assumed independent.

6.2 Results

Figure 6.1 shows the swaption volatility skew for swaptions of a ten-year tenor, of varying maturity and moneyness. These skews are derived under South African market conditions on the 06-Jan-2015. The respective constructed payer and receiver swaption price surfaces can be found in Appendix A.2.

Fig. 6.1: 10-Year swaption skew evaluated on 06-Jan-2015



A clear volatility smile exists across moneyness for 10-year swaptions of all presented maturities. The swaption smile over short-term maturities is significantly more pronounced than the smile over long-term maturity which also tends to be monotonic in nature.

The volatility level of each smile for a given maturity is, of course, very much determined by the current ATM 10-year swaption market implied volatilities due to the imposed ATM market price constraint under the relative entropy approach. Such a constraint will affect the first moment of the skew. However, higher moments implied by the historical data used within the pricing technique are still present within the skew - which is the fundamental interest in producing such a skew. It is clear through observation of the smiles across moneyness that the higher moments within the skew also stabilise over increasing maturity.

The diminishing variation of the skew, in conjunction with a diminishing smile across moneyness, over increasing maturity can be reasoned due to the effects of aggregation gaussianity (AG). It is reiterated here that the sampling method required for construction of long-term skews (discussed in Section 3.4) makes use of repeated re-sampling of a set of daily returns of the underlying in order to create a t_α period return through summation of the sampled daily returns. Assuming these daily returns to be independent, it is clear that the sampling procedure is vulnerable to the Central Limit Theorem. This is one of the stylized facts assumed within financial assets.

What is interesting to note is that much documented international literature has presented evidence of this stylized fact and has proposed financial asset log-returns to be overtly gaussian for terms as low as four weeks (Polakow *et al.* (2014)). However there is still a notable smile across moneyness for swaptions of maturities up to 20 years (should respective underlying distributions be overtly gaussian we would expect a flat skew as per the Black-Scholes model). Section 6.3 presents an analysis of the non-parametric swaption underlying distributions regarding their convergence to normality over increasing maturity t_α.

6.3 Testing for Aggregational Gaussianity

This section analyses the effects of aggregational gaussianity (AG) on the long-term swaption skews through conducting quantitative tests on the underlying distributions used in the construction of these skews. It is clear that, should AG have an effect on the underlying distribution, it will have a direct effect on the nature of the swaption skew itself as it is this distribution that fundamentally influences the swaption price, and therefore implied volatility. The testing methods are adopted from the work done by Polakow *et al.* (2014) who test for AG within the South African equity market.

Testing makes use of the Shapiro-Wilk (SW) and Anderson-Darling (AD) tests for goodness-of-fit to normality. Research suggests that the SW test has the greatest power when testing large sample sizes, followed closely by the AD test. A brief testing method for AG was performed on the swaption underlying distributions of specific tenor and maturity as follows:

- Construct a trial underlying distribution for a swaption of specific tenor and maturity, made up of 1000 sample returns, using the sampling method described in Section 3.4.
- Test the trial distribution for goodness-of-fit to normality using both the SW and AD tests. Record the test to either accept or reject the hypothesis that the sample distribution follows a normal distribution.
- Repeat the above steps to record a set of 1000 trials for each of the SW and AD tests.
- Calculate the percentage of trials rejected (ie. not following a normal distribution) for each of the SW and AD tests.

Results for the percentage of rejected tests for swaption underlying distributions of varying tenor and maturity are presented in Table 6.1 and Table 6.2 using the SW and AD tests respectively.

Similar to the results of Polakow *et al.* (2014) for short-term derivatives, the percentage of rejected tests diminish rapidly as maturity increases up to 1 year. This indicates, as expected, a convergence toward normality over increasing swaption maturity t_α because of the repeated re-sampling of daily returns in order to create a t_α period return (and hence distribution). What is interesting to note, however, is that a convergence toward normality is slower than what is projected by international literature. Moreover, the test percentage rejection slows and tends to oscillate across maturities exceeding 5 years. Hence, while there is a convergence toward normality, normality itself is not obtained by the distributions despite their increasing maturities.

This reflects the conclusions of Polakow *et al.* (2014) that AG is not a stable property within the SA market, albeit for interest rate rather than equity derivatives, and cannot be taken for granted as a stylised fact. With regards to historically implied swaption skews, the observation of a diminished but stable volatility smile across moneyness for swaptions of maturities exceeding 5 years is explained by the above results. A convergence to normality is observed, rapidly over short-term maturities, but is never obtained, hence the continued existence of the smile.

It is also important to note that the convergence to normality is faster for swaptions of greater tenor, although all tenors are found to oscillate around similar rejection percentages for maturities above 5 years. We therefore expect to see that swaption skews of greater tenors will be generally flatter, especially over short term maturities, but will all retain a smile across moneyness as maturity increases.

Tab. 6.1: Percentage of Trials failing Shapiro-Wilk Goodness-of-Fit Test for Normality

	Time to Expiration (years)							
Tenor (years)	$\frac{1}{12}$	$\frac{1}{2}$	**1**	**2**	**5**	**10**	**15**	**20**
1	64.6	38.0	18.4	11.4	8.3	5.5	7.3	5.6
2	67.9	21.8	12.3	7.9	6.8	6.5	6.0	6.8
5	37.3	15.4	10.9	8.2	7.5	8.3	4.9	6.5
10	30.2	12.9	10.2	7.6	6.1	5.7	6.7	6.1
15	32.4	12.8	8.5	6.6	8.3	6.2	6.1	5.6

Tab. 6.2: Percentage of Trials failing Anderson-Darling Goodness-of-Fit Test for Normality

	Time to Expiration (years)							
Tenor (years)	$\frac{1}{12}$	$\frac{1}{2}$	**1**	**2**	**5**	**10**	**15**	**20**
1	87.4	22.0	9.8	4.9	5.6	4.5	6.5	5.1
2	52.9	13.7	8.3	6.2	5.6	5.8	4.6	4.3
5	26.1	9.9	9.1	5.6	6.2	6.3	4.2	4.9
10	19.0	8.8	8.3	6.6	4.7	4.9	5.1	5.3
15	21.5	9.4	6.3	6.1	5.9	5.4	4.1	4.5

6.4 Comparison of Historically Implied vs. Market Implied Skews

Figure A.3 in Appendix A.2 shows the market implied volatility skew for the 06-Jan-2015, sourced from Bloomberg. Note that, due to a lack of pricing information within the SA market, the skew is constructed from a set of available 10-year swaption quotes as well as pricing information from caps and floors found within the market[1].

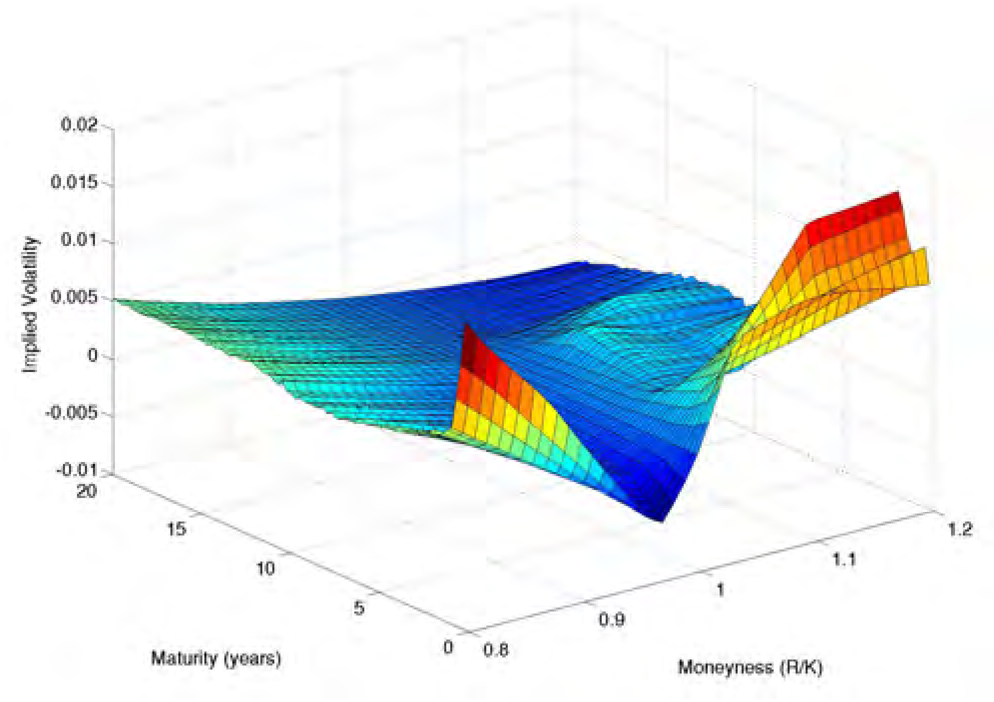
It is reiterated that the market-implied skew represents the swaption market dynamics based on supply and demand factors, liquidity and transparency. The

[1] For information regarding exact construction of the Bloomberg volatility cube see Levin and Zhang (2012). Data is sourced off the volatility cubes by Bloomberg to create the volatility skew.

skew derived under the relative entropy approach represents a historically-implied skew based on an entropic manipulation of the observed historical distribution of the underlying. The skews are not expected to be equivalent and the difference between them gives an indication of the swaption market dynamics, with the historical skew being an indication of the 'fair price'.

Figure 6.2 presents a skew showing the difference in volatility between the market-implied skew sourced from Bloomberg and the skew derived under the relative entropy approach (ie. the difference in volatility between Figure A.3 and Figure 6.1).

Fig. 6.2: Variation between the market implied and the historically-realised skews for a ten-year swaption evaluated on 06-Jan-2015



Comparison of market implied against historically implied skews, using Figure 6.2, reveals both in-the-money and out-the-money swaptions to be generally overpriced when compared to the historically-implied fair price. This overpricing is seen to increase as swaption moneyness moves further from unitary moneyness. Overpricing is also found to decrease across an increase in maturity.

General overpricing of swaptions within the SA market shows evidence of the lack of market price transparency and liquidity.

It is noted that the market-implied skew is close to flat from a maturity of 15 years and may indicate a lack of any pricing knowledge whatsoever that could be used for skew construction of swaptions of such a long-term nature (market quotes for 10-year swaptions on Bloomberg only extended to a maximum maturity of 10 years). The historical evidence of a smile beyond a maturity of 15 years is important for the trading of such long-term natured derivatives where no initial basis for a market quote exists.

Minor underpricing is also found for swaptions of 0.95 to 1 unit moneyness for maturities up to 10 years. This is the only range of swaptions which can be acknowledged to have more liquid and transparent trading within the SA market - where underpricing could indicate such market availability.

7. Skew Sensitivity Analysis

This chapter tests skews derived under the relative entropy approach with regards to their sensitivity to input sample data as well as skew precision. An understanding of the stability of the derived skews with respect to varying input data is important when considering the relative entropy approach for use in practical applications such as hedging.

Derived skews are tested for variation according to the size of the historical data sample used and according to the window period from which the historical data sample has been selected. The precision of the skew is tested when constructing random t_α period returns from the same sample set.

The JSE swap spot curve and at-the-money 10-year tenor swaption market prices presented for the 6-Jan-2015 (t_0) have been used for evaluation of all swaption skews in the subsequent tests using the relative entropy approach.

7.1 Results

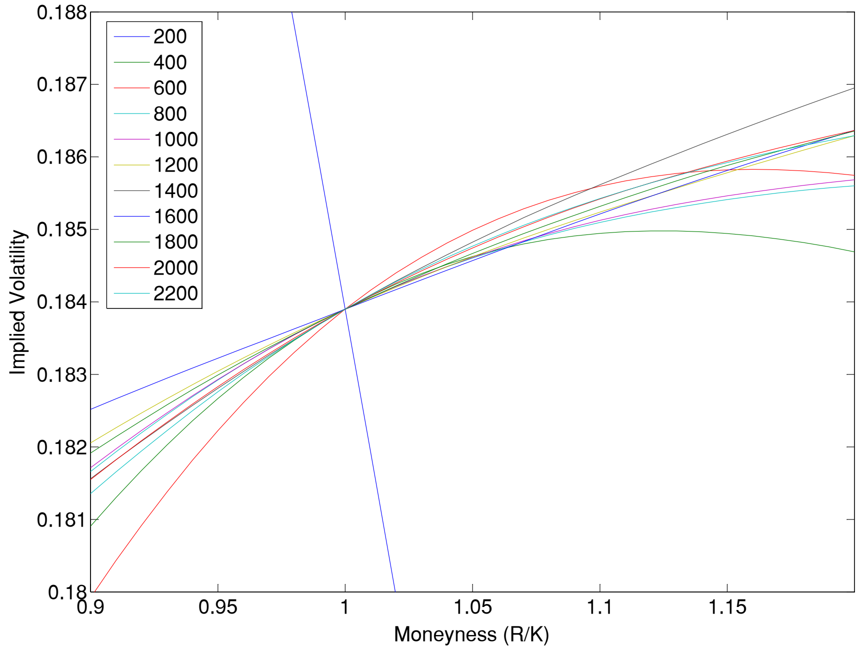
7.1.1 Input Sensitivity due to Sample Data Size

This section presents results of the sensitivity of the swaption skews to the size of the historical input data sample used. Given historical swap curve data from 16-Jan-2006 to 5-Jan-2015, daily returns were calculated for the 10-year fair swap rate. This gives a total of 2253 observed returns, denoted $X(i)$, for $i = -1, -2, ..., -2253$, where i refers to moving backwards in time from t_0. Selecting the initial sample size $h = 200$, t_α sample returns are created under the method described in Section 3.4 using the observed daily returns X(i), for $i = -1, -2, ..., -h$. This procedure was repeated for samples sizes $h = 200, 400, ..., 2200$.

Figure 7.1 shows increased sample size to have a significant effect on the 10Y10Y swaption volatility smile. Analysis of the 10Y1Y[1], 10Y2Y, 10Y5Y and 10Y15Y swaptions showed similar results. The non-parametric nature of the method evidently displays its assumption that the underlying probability distribution is wholly described by past market information. Questions, therefore, must be asked to as what set of past market information best describes the current market environment.

[1] Note that '10Y1Y' describes a swaption of ten-year tenor and five-year maturity.

Fig. 7.1: 10Y10Y swaption volatility smiles according to sample size using the relative entropy approach.



It can be reasoned that a sufficient sample size is required under the relative entropy approach in order to obtain a reliable swaption skew. It is proposed that a minimum sample size of 1200 returns be required in order to obtain a reliable skew, while the skew is observed to stabilise from a minimum of 600 sample returns.

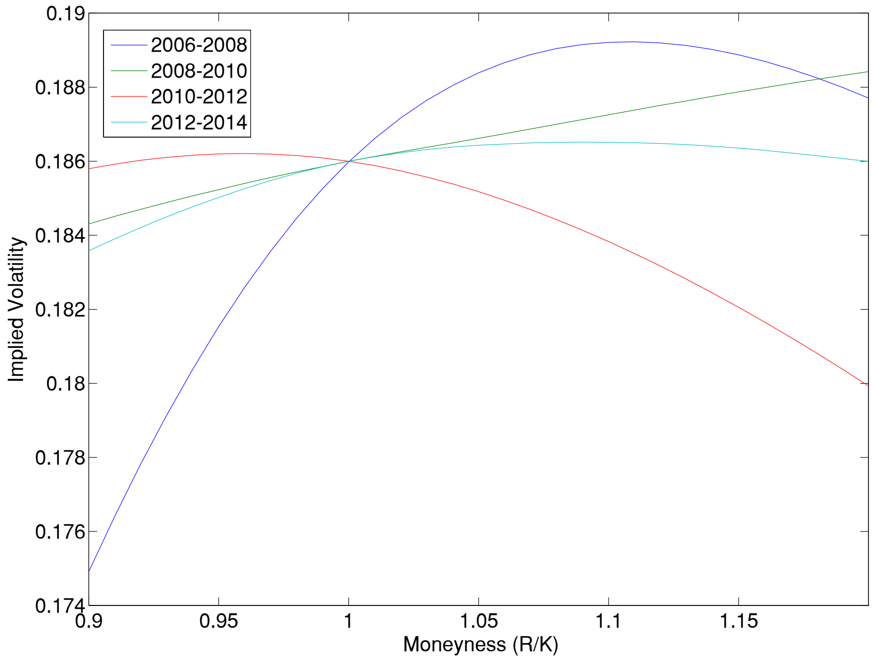
7.1.2 Input Sensitivity due to Sample Data Window Period

This section analyses the sensitivity of the swaption skews to the period of time in which the historical sample data is selected. Plots of the 10Y5Y swaption implied volatility smiles per bi-annual sample window period are shown in Figure 7.2.

Analysis of the 10Y1Y, 10Y2Y, 10Y10Y and 10Y15Y swaptions showed similiar results. However the 10Y5Y swaption smile was noted to have the most observable effects and is therefore the swaption presented.

It is evident from Figure 7.2 that the choice of sample data significantly influences the estimated volatility, due to the dynamics of the market present within the sample window period. To further understand the specific estimates of volatility given by Figure 7.2, we look at the return distribution moments of the sample window period data provided by Table 7.1.

Fig. 7.2: 10Y5Y swaption volatility smiles according to sample window period



Tab. 7.1: Moments of Sample Window Period Daily Returns for 10-Year Swap Rate

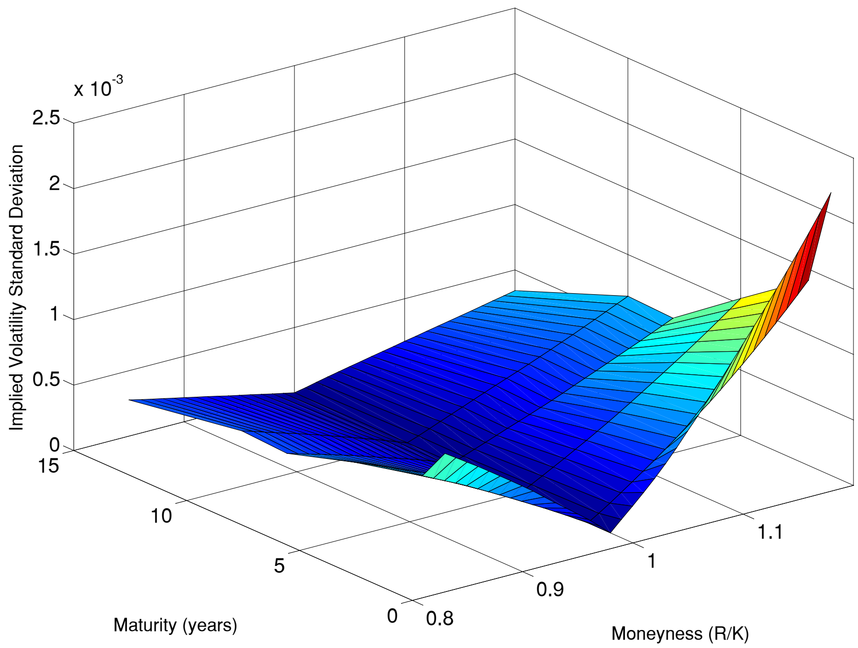
Moment	2006-2008	2008-2010	2010-2012	2012-2014
Mean	0.000204	-0.000018	-0.000290	-0.000042
Variance	0.000040	0.000084	0.000061	0.000080
Skewness	-0.423359	0.458689	0.424834	0.624602
Kurtosis	6.261931	6.633987	4.823127	6.925726

It is clear from Table 7.1 that window periods 2008-2010 and 2012-2014 show similar characteristics in their distributions and subsequently explain the similar volatility smile estimates for the two periods in Figure 7.2. Window period 2006-2008 is noted to be the only distribution with negative skewness and a positive mean which explains the higher gradient and curvature of the smile across moneyness. Such characterisctics could be due to the inclusion of the Credit Crisis of 2007-2008 within this window period. The other three sample periods are noted to have more moderate volatility smiles across maturity and incur similar characteristics in their first three moments. The interesting behavior of the smile produced using the window period 2010-2012, concerning its negative gradient, can be explained by a significantly negative mean in conjunction with a lower variance and kurtosis in comparison to the 2008-2010 and 2012-2014 window periods.

7.1.3 Precision of the Skew derived using the Relative Entropy Approach

This section seeks to understand the extent to which the random sampling method (discussed in Section 3.4) induces error within the swaption skew. Effects are considered to be due to, most notably, the precision of the random sampling method in order to create a consistent distribution as well as the ability of the distribution smoothing function to accurately fit the sample distribution. We seek to recreate the same swaption skew 100 times using the same sampling data with all parameters constant and observe the extent to which the 'identical' skews deviate from one another, hence understanding the induced error within the skew due to the method itself.

Fig. 7.3: Ten-year tenor swaption skew standard deviation due to sampling precision



Presented are results of the analysis showing the standard deviation skew for the 10-year tenor swaption, found in Figure 7.3. Results are derived through recreating the 10-year tenor swaption skew 100 times using of sample data from 16-Jan-2006 to 25-July-2014 to create a set of 10 000 t_α period samples . Using the 100 derived skews, a skew standard deviation could then be found. Due to the computationally expensive nature of the test, the skew has been derived using maturities of 1, 2, 5, 8, 10 and 15 years only.

Analysis of the skew shows a mean volatility standard deviation of $4.4107\text{x}10^{-4}$. Deviation of the skew is found to be significantly worse for short-term 10-year tenor swaptions that are deep ITM or OTM. This could be due to the greater smile found over short-term maturity compared to long-term maturity, as observed in figure 6.1. Consideration of the smile range per maturity and its respective standard deviation shows the swaption skew to be reasonably precise to the third decimal place (volatility measured here as a quantile).

8. Discussion and Conclusions

Stutzer (1996) made use of maximum entropy principles to derive a canonical non-parametric pricing method for options which only requires use of historical underlying data. Based largely on the work of Buchen and Kelly (1996), this dissertation has used the principles of relative entropy under a continuous approach in order to create a method for the pricing of interest rate derivatives of a long-term nature using only historical underlying data. The method, denoted the relative entropy approach, is used to price long-dated swaptions under the $Q^{\alpha,\beta}$ swap measure.

Initially both an extension of canonical valuation as well as the relative entropy approach are used to price swaptions of a long-term nature under simulation. However, simulation results show that the discrete method under Stutzer maintains a monotonic increase in pricing error over maturities of 2 to 20 years. Conversely, the continuous method of the relative entropy approach is found to be stable with regards to pricing error across maturity. This is reasoned due to the continuous method's ability to constrain the tails of the sample distributions while also using a more accurate prior distribution. The discrete method distribution however is found to 'explode' under change of measure when the initial distribution is too dispersed - a typical characteristic for distributions of long-dated maturity.

A swaption skew has subsequently been derived for swaptions of 10-year tenor within the SA market under the relative entropy approach. The derived historically implied skews show significant volatility smiles across moneyness for swaptions of short-dated maturities, while the smile is observed to flatten, but never completely, across moneyness over an increase in maturity. Subsequent tests for aggregational gaussianity on the swaption underlying distributions over an increase in maturity reveal these distributions to have gradual convergence to normality as maturity is increased. However full normality of the distributions is not reached for swaption maturities of up to 20 years which hence explains the maintained existence of a long-term volatility smile across moneyness. Furthermore, these results reflect those of Polakow *et al.* (2014) by questioning the existence of aggregational gaussianity as a stylised fact within the SA market, albeit for interest-rate rather than equity derivatives.

Comparison of the historically-implied 10-year swaption skew against the market-implied 10-year skew (derived by Bloomberg) for the 06-Jan-2015 show both skews to have similarities in their smile characteristics while general market overpricing is found for ITM and OTM swaptions of maturities less than 15 years. This is reasoned to the OTC and illiquid nature of the SA swaption market. Underpricing of swaptions of moneyness between 0.95 and 1 (R/K) for maturities less than 10 years is the only indication of more liquid and transparent swaptions traded within the market.

Testing the sensitivity of the skew to historical data inputs revealed that the choice of the window period in which data is sampled has a significant effect on the estimation of swaption volatility. It is concluded that, unless specific market characteristics are sought after, window periods of greater size (and therefore with a larger possible set of market dynamics or events as recorded historically) are generally more reliable for use for derivation of swaption skews, as one is more likely to have captured less probable events previously observed within the market.

Testing the sensitivity of the skew to the size of the data sample used revealed swaption volatility to stabilise when using a sample size greater than 600 daily returns while reliable estimates are found when using sample sizes greater than 1200.

A final test for precision of the skew under the relative entropy approach suggests the volatility skews to be reasonably precise to the third decimal place and could be extended for use in practical applications.

Bibliography

Alcock, J. and Gray, P. (2005). Dynamic, nonparametric hedging of european style contingent claims using canonical valuation, *Finance Research Letters* **2**(1): 41–50.

Banavar, J. and Maritan, A. (2007). The maximum relative entropy principle, *arXiv preprint cond-mat/0703622* .

Ben-Tal, A. (1985). The entropic penalty approach to stochastic programming, *Mathematics of Operations Research* **10**(2): 263–279.

Black, F. (1976). The pricing of commodity contracts, *Journal of financial economics* **3**(1): 167–179.

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

Buchen, P. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices., *Journal of Financial and Quantitative Finance* .

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*, John Wiley & Sons.

Cvitanić, J. and Zapatero, F. (2004). *Introduction to the economics and mathematics of financial markets*, MIT press.

De Araujo, M. and Mare, E. (2006). Examining the volatility skew in the south african equity market using risk-neutral historical distributions, *Investment Analysts Journal* (Issue 64): p.15–19.

Duan, J.-C. (2002). Nonparametric option pricing by transformation, *Rotman School of Management, University of Toronto* .

Duffee, G. R. and Stanton, R. H. (2012). Estimation of dynamic term structure models, *The Quarterly Journal of Finance* **2**(02): 1250008.

Garcia, R., Ghysels, E. and Renault, E. (2010). The econometrics of option pricing, *Handbook of Financial Econometrics* **1**: 479–552.

Gray, P. and Newman, S. (2005). Canonical valuation of options in the presence of stochastic volatility, *Journal of Futures Markets* **25**(1): 1–19.

Haley, M. R. and Walker, T. B. (2010). Alternative tilts for nonparametric option pricing, *Journal of Futures Markets* **30**(10): 983–1006.

Jamshidian, F. (1989). An exact bond option formula, *Journal of finance* pp. 205–209.

Jaynes, E. T. (1957). Information theory and statistical mechanics, *Physical review* **106**(4): 620.

Levin, K. and Zhang, J. (2012). The bloomberg volatility cube, *Documentation Bloomberg LP* .

marquis de Laplace, P. S. (1840). *Essai philosophique sur les probabilités*, Bachelier.

Polakow, D., Taylor, D. and Mahomed, O. (2014). Aggregational gaussianity in the south african equity market, *International Business & Economics Research Journal (IBER)* **13**(5): 1091–1106.

Shannon, C. E. (2001). A mathematical theory of communication, *ACM SIGMOBILE Mobile Computing and Communications Review* **5**(1): 3–55.

Stewart, J. (2009). *Multivariable calculus: concepts and contexts*, Cengage Learning.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation, *The Journal of Finance* **51**(5): 1633–1652.

Stutzer, M. and Chowdhury, M. (1999). A simple non-parametric approach to bond futures option pricing, *The Journal of Fixed Income* **8**(4): 67–76.

Uffink, J. (1995). Can the maximum entropy principle be explained as a consistency requirement?, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **26**(3): 223–261.

Appendix

A. Swaption Surfaces and Skews

A.1 Swaption Price Surfaces under the Relative Entropy Approach

Fig. A.1: Ten-year payer swaption price surface for 06-Jan-2015

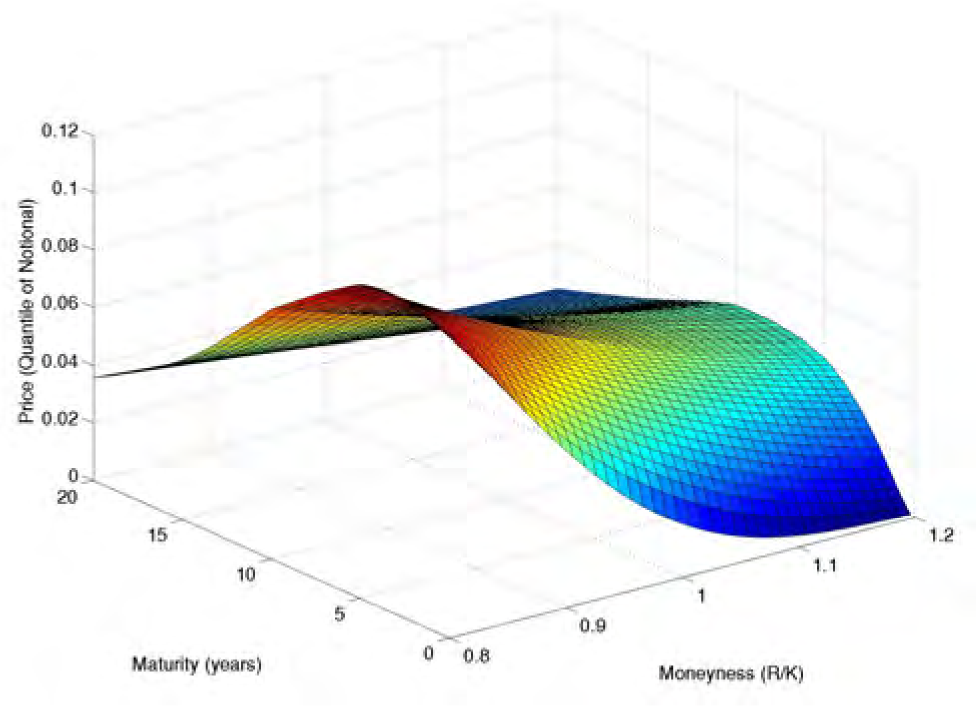
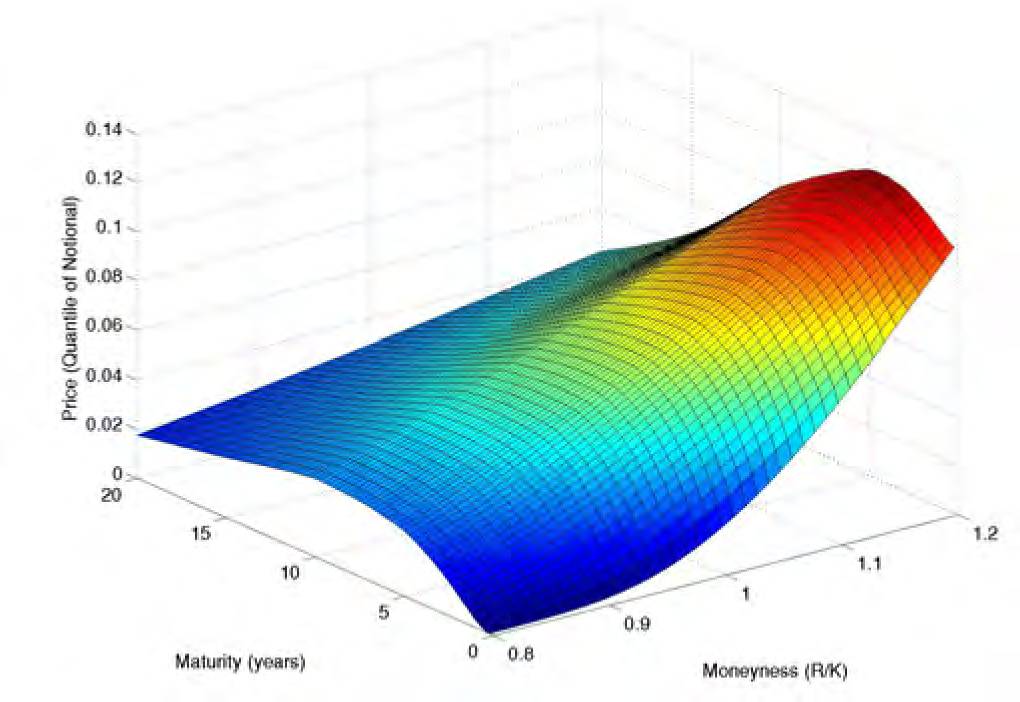
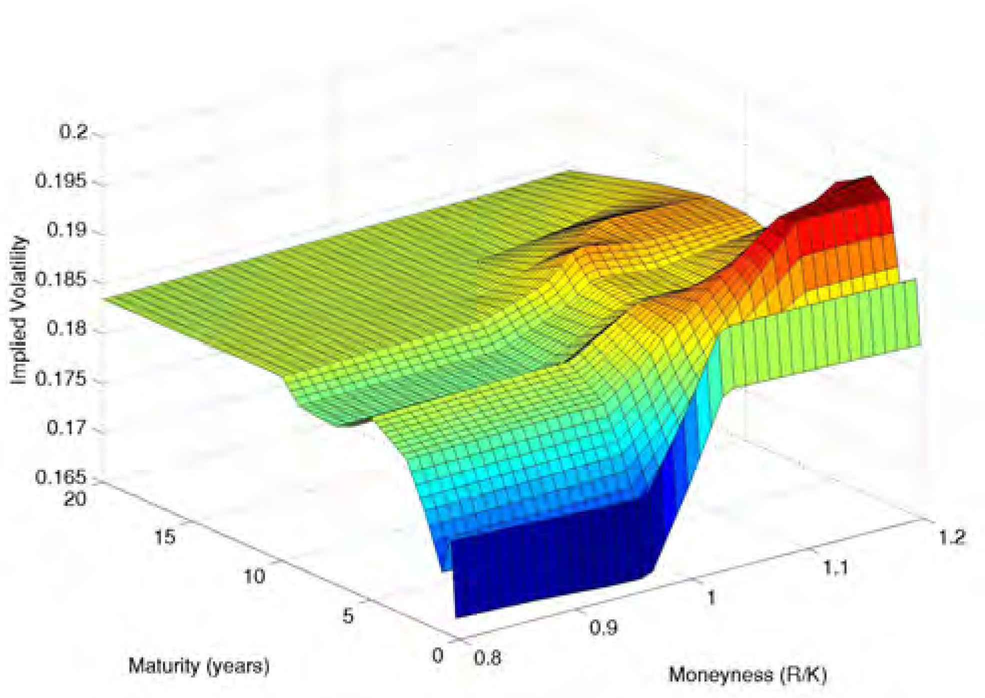


Fig. A.2: Ten-year receiver swaption price surface for 06-Jan-2015



A.2 Market Implied Volatility Skews

Fig. A.3: Ten-year market implied volatility skew for 06-Jan-2015 (Bloomberg)



B. Price invariance between measures

The following presentation is taken directly from Brigo and Mercurio (2007) in order to show the invariance in price of a tradable asset between change of measures.

FACT TWO. *The time-t risk neutral price*

$$Price_t = E_t^B \left[B(t) \frac{Payoff(T)}{B(T)} \right]$$

is invariant by change of numeraire: If S is any other numeraire, we have

$$Price_t = E_t^S \left[S(t) \frac{Payoff(T)}{S(T)} \right]$$

In other terms, if we substitute the three occurrences inside the boxes of the original numeraire with a new numeraire the price does not change. This second fact is a rephrasing of formula B.4 presented here under Proposition 2.2.1:

Proposition 2.2.1 *Assume there exists a numeraire N and a probability measure Q^N, equivalent to the intial Q_0, such that the price of any traded asset X (without intermediate payments) relative to N is a martingale under Q^N, ie.*

$$\frac{X_t}{N_t} = E^N \left\{ \frac{X_T}{N_T} | \mathcal{F}_t \right\}, \quad 0 \leq t \leq T \tag{B.1}$$

Let U be an arbitrary numeraire. Then there exists a probability measure Q^U, equivalent to the intial Q_0, such that the price of any attainable claim Y normalized by U is a martingale under Q^U, i.e.

$$\frac{Y_t}{U_t} = E^U \left\{ \frac{Y_T}{U_T} | \mathcal{F}_t \right\}, \quad 0 \leq t \leq T. \tag{B.2}$$

Moreover, the Radon-Nikodym derivative defining the measure Q^U is given by

$$\frac{dQ^U}{dQ^N} = \frac{U_T N_0}{U_0 N_T}. \tag{B.3}$$

The derivation of above is outlined as follows. By definition of Q^N, we know that for any tradable asset price Z,

$$E^N \left[\frac{Z_T}{N_T} \right] = E^U \left[\frac{U_0}{N_0} \frac{Z_T}{U_T} \right] \tag{B.4}$$

(both being equal to Z_0/N_0). By definition of the Radon-Nikodym derivative, we know also that for all Z

$$E^N\left[\frac{Z_T}{N_T}\right] = E^U\left[\frac{Z_T}{N_T}\frac{dQ^N}{dQ^U}\right]. \tag{B.5}$$

By comparing the right-hand sides of the last two equalities, from the arbitrariness of Z we obtain B.3. The general formula follows from immediate application of the Bayes rule for conditional expectations.

C. The Kalman Filter

The following presentation of the linearized Kalman Filter under the one factor Vasicek model under the Q measure follows a direct or similar presentation to Duffee and Stanton (2012). A more rigorous discussion and application of the Kalman Filter can be found in their paper. All parameters and dynamics used here are consistent with chapter 5.

Filtering is a natural approach when the underlying state is unobserved. Denote the parameters κ, θ, σ of the Vasicek model as ρ. The observation equation expresses the observed state, y_t, as a linear function of the unobservable state, x_t, plus a measurement error ϵ_t. The transition equation expresses the discrete time evolution of x_t as linear in x_t. These equations are determined by the parameters of the term structure model ρ. The structure is

$$y_t = H_0(\rho) + H_1(\rho)' x_t + \epsilon_t,$$
$$x_{t+1} = F_0(x_t, \rho) + F_1(x_t, \rho) x_t + \nu_{t+1}$$

The observation process noise ϵ_t is assumed as Gaussian white noise with variance covariance matrix $R(\rho)$. ν_{t+1} of the unobserved state process is denoted by the variance covariance matrix $Q(\rho)$.

Consider a set of j yields of tenor $[\tau_1, ..., \tau_j]$. The vectors of the observed state process y_t are

$$H_0(\rho) = \begin{bmatrix} \frac{-A(\tau_1)}{\tau_1} \\ \vdots \\ \frac{-A(\tau_j)}{\tau_j} \end{bmatrix}, \qquad H_1(\rho) = \begin{bmatrix} \frac{B(\tau_1)}{\tau_1} \\ \vdots \\ \frac{B(\tau_j)}{\tau_j} \end{bmatrix}.$$

Scalars of the unobservable state process x_t are

$$F_0(x_t, \rho) = \theta \left(1 - e^{-\kappa \delta t}\right),$$
$$F_1(x_t, \rho) = 1 - e^{-\kappa \delta t},$$
$$Q(x_t, \rho) = \frac{\sigma^2}{2\kappa} \left(1 - e^{-2\kappa \delta t}\right).$$

$x_{t|t}$ is denoted as the estimate of the state vector with a variance covariance matrix $P_{t|t}$. One-step ahead forecasts of the state vector and observable vector is denoted $x_{t+1|t}$ and $y_{t+1|t}$ respectively. The variance-covariance matrices of these forecasts are denoted $P_{t+1|t}$ and $V_{t+1|t}$ respectively.

The Kalman Filter follows the following recursion which begins with vector ρ. This vector is used to calculate the unconditional expectation and variance-covariance matrix for x_1, which we denote $x_{0|0}$ and $P_{0|0}$. It then takes the following steps:

- Use $x_{t|t}$ and ρ to evaluate $F_0(x_t, \rho), F_1(x_t, \rho)$ and $Q(x_t, \rho)$. Denote these values as F_{0t}, F_{1t} and Q_t.
- Compute the one-period-ahead prediction and variance of x_{t+1}, $x_{t+1|t} = F_{0t} + F_{1t}x_{t|t}$ and $P_{t+1|t} = F_{1t}P_{t|t}F'_{1t} + Q_t$.
- Compute the one-period-ahead prediction and variance of y_{t+1}, $y_{t+1|t} = H_0 + H'_1 x_{t+1|t}$ and $V_{t+1|t} = H'_1 P_{t+1|t} H_1 + R$.
- Compute the forecast error in y_{t+1}, $e_{t+1} = y_{t+1} - y_{t+1|t}$.
- Update the prediction of x_{t+1}, $x_{t+1|t+1} = x_{t+1|t} + P_{t+1|t}H_1V^{-1}_{t+1|t}e_{t+1}$ and $P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}H_1V^{-1}_{t+1|t}H'_1P_{t+1|t}$.

The finding of the true parameters ρ requires an appropriate objective function. Here we use the maximisation of the log-likelihood of observations which in turn minimises the observation prediction error and provides the maximum likelihood parameter estimates. Hence we maximise the log-likelihood:

$$l(\rho; y) = -\frac{1}{2}\sum_{t=1}^{T}\left[d\log(2\pi) + \log|V_{t|t-1}| + e'_t V^{-1}_{t|t-1} e_t\right].$$

Hence our parameter estimates are given by:

$$\hat{\rho}(y) = \max_{\rho}[l(\rho; y)].$$