

FINANCIAL MATHEMATICS TEAM CHALLENGE

A collection of the four reports from the 2014 Financial
Mathematics Team Challenge.

ACQuFRR

AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH



Preamble

The Financial Mathematics Team Challenge (FMTC) was born in February 2013 out of one of the casual, contemplative conversations we often have. We were mulling over a few ideas around facilitating an event for South African postgraduate students in Financial and Insurance Mathematics where they could focus on a topical research project with industry relevance while simultaneously developing links to international students and academics in the field. We had been inspired in part by Nadim Sah drawing our attention to the MITACS Industrial Math Summer School. We also wanted the event to serve as an opportunity to bring international researchers to South Africa, and to give them a glimpse of the dynamic environment that is developing at the University of Cape Town in the African Institute for Financial Markets & Risk Management. A final, indispensable ingredient had to be that the participating students would work in teams and be exposed to a healthy dose of fair competition.

The first FMTC was held from the 3rd to the 13th of July 2014 at the University of Cape Town, South Africa. The Challenge brought together postgraduate students from France, South Africa and the UK to pursue intensive research in Financial Mathematics. Four teams of Masters and PhD students each worked on a different research problem during the ten days. Professional and academic experts from France, South Africa and the UK individually mentored the teams, fostering teamwork and providing guidance. The students applied themselves with incredible dedication and exemplary vigour.

The research included topical projects on *multi-curve interest rate modelling* and *counterparty risk valuation adjustments*, *pricing and hedging of commodity gap options*, and *portfolio diversification measures*. They were either proposed directly by our industry partners or chosen from areas of current relevance to the finance industry. In order to prepare the teams, guidance and preliminary reading was given to them a month before the meeting in Cape Town. During the final two days of the challenge, the teams presented their conclusions and solutions in extended seminar talks. The team whose research findings were elected as the best won a cash prize and was awarded a floating trophy.

The teams were asked to write a report containing a critical analysis of their research problems and the results that they had obtained. This volume contains these four reports, and will be available to future FMTC participants. It may also be of use and inspiration to Masters and PhD students in Financial and Insurance Mathematics.

The first Financial Mathematics Team Challenge was a wonderful opportunity for students to interact and collaborate on research in Risk, Investments, Insurance and Financial Mathematics. Motivated by its success, we are already planning its second edition.

David Taylor, University of Cape Town

Andrea Macrina, University College London & University of Cape Town

Contents

1. **D. Maxwell, Z. Pitsillis and R. Rudd**
Commodity Future Spread Options
2. **Y. Jiang, J. T. J. Mbongo Nkouna and T. M. Nguyen**
bA Multi-Curve Models with Counterparty Risk
3. **A. Backwell, M. Kateregga, F. Pasini and T. von Boetticher¹**
Pricing Kernels, Multi-Curve Models and Swaption Pricing
4. **O. Mahomed, W. Ushan and M. Khumalo**
Portfolio Diversification using Higher moment Measures - Entropy and the Diversification Delta

¹Winning team of the first Financial Mathematics Team Challenge

Commodity Future Spread Options

Team 1

D. MAXWELL, University of Cape Town
Z. PITSILLIS, University of Cape Town
R. RUDD, University of Cape Town

Supervisors:

T. MCWALTER, University of Cape Town
J. DUNNE, Standard Chartered

Contents

1	The Problem	4
1.1	Modelling Commodities	5
1.1.1	Reduced-form Versus Structural	6
1.1.2	Two Factor Models	7
2	The Model	8
2.1	The Schwartz97	8
2.1.1	Futures Term Structure	10
2.1.2	Volatility Term Structure	10
3	The Simulation	13
3.1	Pricing a European Call	14
3.1.1	Delta Hedging	16
3.1.2	Profit and Loss Simulations	17
3.2	Pricing the European Spread Option	18
3.2.1	Delta Hedging	19
3.2.2	Profit and Loss Simulations	21
4	The Calibration	24
4.1	The Kalman Filter	24
4.1.1	State Space Representation of the Schwartz 2 Factor Model	24
4.1.2	Parameter Estimation via the Kalman Filter	26
4.1.3	Difficulties Associated with Calibration	27
4.1.4	Calibration Process	27
4.2	Applying the Kalman Filter	28
4.2.1	The Data and Model Results	28
4.2.2	Model Assessment	32
5	Bonus: Seasonality	32
5.1	Modelling Seasonality	35
5.1.1	The Model	37
5.1.2	The Data	37
5.2	Applying the Model	37
5.2.1	Estimating the Seasonal Premia	37
5.2.2	Application of Seasonal Premia	41

6	Hedging	45
6.1	Historical Profit and Loss	45
6.1.1	The European Call	45
6.1.2	The Spread Option	46
6.2	Conclusion	48
	Bibliography	48

1 The Problem

In the FX market, the spot exchange rate and the future exchange rate possess an elastic relationship due to the behaviour of the future exchange rate being driven by current interest rate dynamics. Thus, when the spot exchange rate changes, provided interest rates remain unchanged, the futures curve will undergo a parallel shift.

This allows for a long position in the spot to be hedged by a short position in the future. When the spot falls, it is expected that the near-dated future position will also move downwards, as the overall shape of the futures curve remains unchanged.

Commodities prices exhibit mean reversion, essentially driven by the market's view that the price of a commodity will return to some long run mean. This property causes a unique behaviour in the relationship between the commodity spot price (often referred to as the 'cash price') and the commodity futures price. In the commodities market, when spot prices fall significantly, more often than not the futures curve will be upward sloping, as the market expects the price to return to some long-term mean. And vice versa, when spot prices increase significantly there is typically a downward sloping forward curve. The result of this is 'double exposure'.

Consider an investor who takes a long position in the spot commodity and attempts an offsetting hedge with a short position in the future. If the spot prices rise, the investor may show a gain on both positions, as there is not a corresponding rise in the future. However, should spot prices fall, the investor may lose on both positions, as the future value may increase. The relationship between the spot and the future in commodities is inelastic.

A commodity spread option would allow the investor to take a call or put on the difference between two points on the futures curve and, hence, it could be used as a form of protection against the losses in the situation described above.

The payoff of a call option on the spread would be given by

$$D_t = (G(t, T_1) - G(t, T_2) - K)^+,$$

where t is the expiry of the option, K is the strike and $G(t, T_1)$ and $G(t, T_2)$ are the prices of the futures expiring at $t < T_2 < T_1$ respectively.

This report will investigate implementing a model for the commodities market that will allow for the efficient pricing and hedging of the European spread option described above under the real-world evolution of the futures curve. The remainder of this section proceeds with an overview of the relevant literature and then highlights the choice of the model for implementation. Section 2 will investigate the analytical dynamics of the chosen model, whereas Section 3 will cover the implementation of simulation and pricing under the model dynamics. The calibration of the model parameters to real world LME Copper futures is then discussed in Sec-

tion 4, before a brief digression into a potential modelling approach of the seasonality effect present in natural gas prices. Finally, the report concludes with the results of the model when applied to pricing and hedging a European spread option on the LME Copper futures, instantiating at 11 December 2012. Here, it is found that the model captures enough complexity to realistically price and hedge the relevant option, under the constraint of a well-specified forward-looking volatility.

1.1 Modelling Commodities

There are four “stylised facts” regarding the behaviour of the commodity market and commodity term structure and it is possible to categorise models for commodity markets according to their ability to capture these features. These features are presented below.

1. Backwardation and Contango

The commodity futures curve is in contango when it is upward-sloping or in backwardation when it is downward-sloping. Often the shape of the curve is described in terms of the ‘basis’, which is the difference between the spot price and the futures price. Thus when in contango there is a negative basis as future prices are greater than spot prices, and when in backwardation the basis is positive as future prices are less than current prices. A humped futures curve is also common.

There are two main approaches to explaining the dynamics driving the changes in the futures curve shape. The ‘Hedging Pressures’ approach, dating back to [Key30] is based on the premise that “hedgers on average hold a short position in the futures market” and hence are “willing to pay a risk premium in order to hedge their exposure to spot price”, the net result of this is “futures prices are a downward biased estimator of future spot prices” [BP13]. The more widely accepted approach is called the ‘theory of storage’. This branch relates the spot and futures price by the idea of a ‘convenience yield’. Using the definition from [Sch97], the convenience yield can be viewed as the gain or “flow of services” that is experienced by the holder of the underlying spot but not by the holder of the future.

Typically, the convenience yield is modelled as net convenience yield, which is equal to the gross convenience yield less the cost associated with holding the underlying spot [BP13]. Thus the net convenience yield can be negative. According to the theory of storage, the spot-futures price relationship is

$$G(t_0, T) = S_{t_0} \exp(r - \delta)(T - t_0),$$

where δ is the net convenience yield and r is the risk-free rate (here both are assumed constant). This implies that the slope of the futures curve is positive if the risk-free rate is greater than the convenience yield and negative if it is

not. Relating this to the concepts of supply and demand, the inverse relationship between price and inventory levels means that the equation above implies an inverse relationship between convenience yield and inventory levels.

2. Mean-reversion

There is a strong mean-reverting behaviour in spot commodity markets [NS04], largely believed to be due to the dynamics of supply and demand interactions. Essentially, prices rise when shortages occur and this tends to raise the level of investments which will push up supply and bring price down again (see [BP13]). Whilst there are some studies which question the presence of mean reversion in some commodities (see [BLO97]), it is generally accepted that mean-reversion is an important aspect of the commodities market [Pin01].

3. The Samuelson Effect

[Gem05] describes the ‘Samuelson Effect’ as the observation that, all else being equal, the volatility of futures prices tends to increase as the time to maturity decreases. It is believed that this is due to the increased sensitivity of the futures price to current information as it nears its time to maturity as was originally proposed in [Sam65].

4. Seasonality

Seasonality is common in some commodities (see [Gem05]), and is driven by supply side factors (seasonal production cycles, such as agricultural commodities) or demand side factors (such as an increase in demand in the USA for natural gas during the winter months). Beyond the seasonality in the price level and convenience yield, there is evidence of seasonality in the volatility corresponding to that exhibited in the spot.

1.1.1 Reduced-form Versus Structural

There exist two distinct approaches used to model commodities and their futures prices. The ‘structural approach’ has an economic grounding and attempts to model commodity price behaviour by first analysing the underlying economic principles that drive the factors affecting the supply and demand, such as inventories. The ‘reduced-form’ approach is typically based on stochastic models and are widely used in the financial world as they often result in closed-form solutions for certain options. A criticism of this approach is that it often leads to very little economic insight [BP13]. [Pir11] and [RSS00] both raise concerns on the use of reduced-form models, however both acknowledge the usefulness of the reduced-form framework when applied to derivative pricing.

Within the reduced-form class of models, again, a distinction can be drawn between two broad types of model, namely ‘spot price models’ and ‘no-arbitrage

models'. Spot price models are primarily concerned with modeling the spot price, and the futures prices are derived from the modelled spot price dynamics using arbitrage arguments, however as a result one may find that the term structure of these implied futures prices do not conform to the term structure observed in the market [BP13]. No-arbitrage models include the current market futures prices as an input to the model, and the dynamics of the futures curve is inferred by enforcing no-arbitrage assumptions [BP13].

For the purposes of solving the research problem, it is only necessary to consider the prominent reduced-form models. The earliest models used were typically one factor models for the commodity spot price using geometric Brownian motion. One of the most popular of these models was the [BS85] model:

$$dS_t = \mu S_t dt + \sigma_S S_t dW_t$$

where S_t is the spot price, μ is the drift of the spot price, σ_S is the spot price volatility and dW_t is standard Brownian motion.

The pitfalls of this model were its inability to capture mean-reversion and the Samuelson effect particularly well. To overcome this [Sch97] proposed a one factor mean-reverting process to model the spot price;

$$dS_t = S_t \kappa (\mu - \ln S_t) dt + \sigma_{S_t} dW_t$$

where κ is the speed of spot price adjustment and μ is the long run mean log price.

There are other one factor models (see [Bre91]), but those outlined above largely contributed to the further development of term structure models. One factor models are unable to capture all the characteristics of the term structure of commodity futures prices, for example some models fail to capture the Samuelson effect, as well as the fact that basis behaves differently in contango and backwardation [Lau03].

1.1.2 Two Factor Models

The natural progression was to add another factor to improve model performance. The [Sch97] model (referred to hereafter as the Schwartz97 model), which was an enhancement of [GS90], became one of the "most famous term structure models of commodity prices" [Lau03]. In this model, the two factors are the spot price and the convenience yield. Some models have implemented the second state variable as the long-term price (see [Gab91]) or some quantity analogous to the convenience yield. The dynamics of the Schwartz97 model are examined in Section 2. The addition of the second state variable has in most cases allowed for "richer shapes of curves" [Lau03], however it increases the complexity of the modelling.

There has been some development of three factor models, mainly with the goal of allowing for stochastic interest rates. However, it appears that the improvement in the models performance does not outweigh the extra computational work that

is required to calibrate the extra state variable. The three factor model by [Sch97], which models the long-term forward price as the third factor, is compared to the Schwartz97 model in [Lau03] and the results appear “empirically very similar”.

In this report the Schwartz97 two-factor model is implemented. This model captures many aspects of forward curve term structures that one factor models can not, and performs as well as most three factor models. It allows for contango and backwardation of the futures curve, mean-reversion of the spot price as well as the Samuelson effect. Lastly, it appears to have been received very favourably in the literature [Lau05]. It does however, not address seasonality (see Section 5).

2 The Model

2.1 The Schwartz97

The typical “stylised facts” of commodity markets, as described in Section 1.1 are spot price mean reversion, contango or backwardation in the futures curve and the decreasing volatility of the futures curve in maturity. In order to capture these phenomena the Schwartz97 futures curve is used. In order to broadly capture the “stylised facts” Schwartz [Sch97] posits a mean reverting stochastic convenience yield in conjunction with the spot price process. The futures curve is then derived from deterministic interest rates, the prevailing (but often unobservable) commodity spot price and the stochastic convenience yield. The interplay of the interest rate, convenience yield and spot price as well as their long-term averages imply the shape of the futures curve. That is, the contango or backwardation as well as the curvature of the futures term structure depends on the relationships of all these variables. In addition the model allows for (but does not enforce) decreasing volatility of the futures curve in maturity.

The joint dynamics of the convenience yield and the commodity spot price serve to proxy the more fundamental dynamics of the spot price and inventories. Simplistically, when inventories are high, spot prices become depressed and there is a low marginal reward for physically holding the commodity and, hence, a low convenience yield. Schwartz [Sch97] captures the relation through the correlation of changes in the convenience yield and log spot.

In order to allow for mean reversion in the spot price, the drift of the spot price is linked to the level of the convenience yield. So, through their joint relationship and the mean reversion in the convenience yield there is mean reversion in the spot. The prevailing spot and convenience yield, their respective long-term averages and rates of mean reversion are combined with the interest rate to form the futures curve.

Schwartz’s two factor model [Sch97] is presented here in the notation of Erb *et al* [Erb14]. The commodity spot price S_t and the stochastic convenience yield δ_t are modelled as joint stochastic processes. Their stochastic differential equations

under the real world measure \mathbb{P} are given by

$$\begin{aligned}d\delta_t &= \kappa(\alpha - \delta_t)dt + \sigma_\epsilon dW_t^{(\epsilon)} \\dS_t &= (\mu - \delta_t)S_t dt + \sigma_S S_t dW_t^{(S)}\end{aligned}$$

where $W^{(S)}$ and $W^{(\epsilon)}$ are correlated Brownian motions, with instantaneous correlation $dW_t^{(\epsilon)}dW_t^{(S)} = \rho dt$.

The dynamics of the convenience yield are that of an Ornstein-Uhlenbeck process with long-term mean α , rate of mean reversion κ and convenience yield volatility σ_ϵ . The spot price can then be seen as an asset with return μ and volatility σ_S with an addition stochastic dividend yield in the form of the convenience yield. Note that a positive correlation between the returns on the spot and the convenience yield will induce mean reversion in the spot. Probabilistically, when S_t is large, δ_t will be large and thus the drift of the spot price will be reduced and may even become negative.

This model has two sources of risk and, thus, in order to construct a risk neutral measure \mathbb{Q} it is necessary to have both a market price of spot risk and a market price of convenience yield risk. The market price of spot risk is written as

$$\frac{\mu - r}{\sigma_S},$$

where r is the instantaneous risk-free rate. The market price of convenience yield risk is written as a single constant parameter λ . This implies the risk neutral dynamics of the factors will be

$$\begin{aligned}d\delta_t &= [\kappa(\alpha - \delta_t) - \lambda]dt + \sigma_\epsilon d\widetilde{W}_t^{(\epsilon)} \\dS_t &= (r - \delta_t)S_t dt + \sigma_S S_t d\widetilde{W}_t^{(S)}\end{aligned}$$

where $\widetilde{W}^{(\epsilon)}$ and $\widetilde{W}^{(S)}$ are \mathbb{Q} Brownian motions with instantaneous correlation $d\widetilde{W}^{(\epsilon)}d\widetilde{W}^{(S)} = \rho dt$. For simplicity, it is convenient to write the risk neutral dynamics of the convenience yield as

$$d\delta_t = [\kappa(\tilde{\alpha} - \delta_t)]dt + \sigma_\epsilon d\widetilde{W}_t^{(\epsilon)}$$

where $\tilde{\alpha} = \alpha - \lambda/\kappa$.

Since a risk neutral measure can be constructed, futures prices with maturity T at t_0 can be expressed as a risk neutral expectation of the spot price at maturity. That is,

$$G(S_0, \delta_0, t_0, T) = E^{\mathbb{Q}}[S_T | \mathcal{F}_{t_0}] \tag{1}$$

$$= S_{t_0} e^{A(T-t_0) + B(T-t_0)\delta_{t_0}} \tag{2}$$

where A and B are the functions given by

$$\begin{aligned}
A(T - t_0) &= \left(r - \tilde{\alpha} + \frac{\sigma_\epsilon^2}{2\kappa^2} - \frac{\sigma_S \sigma_\epsilon \rho}{\kappa} \right) (T - t_0) + \frac{1}{4} \sigma_\epsilon^2 \frac{1 - e^{-2\kappa(T-t_0)}}{\kappa^3} \\
&\quad + \left(\kappa \tilde{\alpha} + \sigma_S \sigma_\epsilon \rho - \frac{\sigma_\epsilon^2}{\kappa} \right) \frac{1 - e^{-\kappa(T-t_0)}}{\kappa^2}, \\
B(T - t_0) &= -\frac{1 - e^{-\kappa(T-t_0)}}{\kappa}.
\end{aligned}$$

2.1.1 Futures Term Structure

Schwartz's formulation allows for varied shapes in the futures curve, depending on the relationships between the parameters. In the same vein as Lautier [Lau05] the effect of various changes to the parameters on the futures term structure under Schwartz's two factor model is presented to illustrate its adaptability.

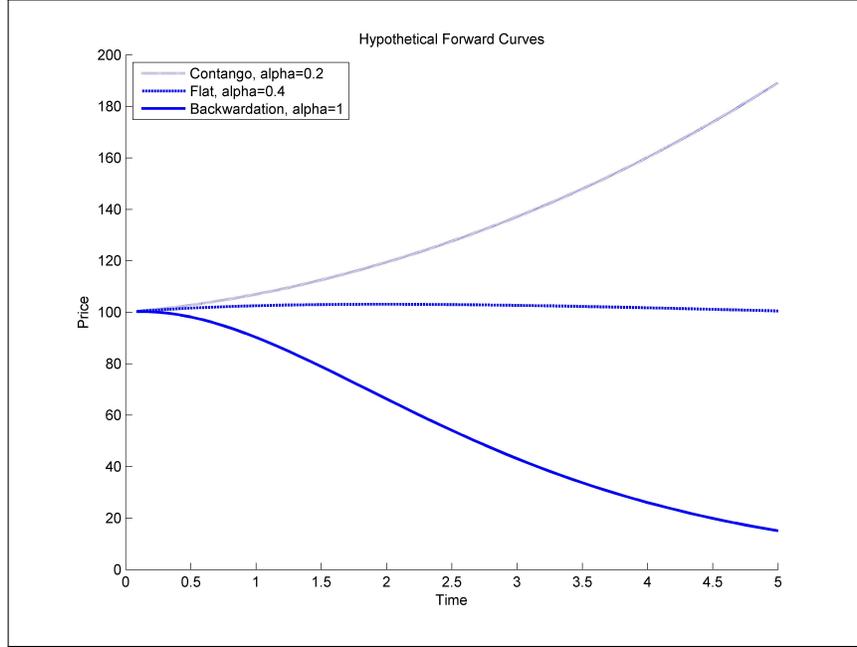
The first characteristic is that the curve can be either in backwardation or contango, based on the size of the long-term mean convenience yield α and its relation to the risk-free rate r . If the long-term mean convenience yield is comparatively low then the futures curve is upwards sloping. If the long-term mean average convenience yield is comparatively high then the futures curve is downward sloping. This corresponds with the economic intuition, as outlined by Geman [Gem05], that backwardation occurs when the benefits of holding the asset in physical form outstrip the financing (opportunity) costs associated with purchasing the commodity. Conversely, contango occurs when the benefits of holding the commodity in financial form, that is a futures contract and cash deposit, outstrip the benefits of physically holding the asset. The Schwartz97 model can capture a range of these phenomena. Hypothetical futures curves for different values of α are plotted in Figure 1 demonstrating a flat curve and curves in contango and backwardation. The risk-free rate and convenience yield used for these plots was 1%.

The convenience yield is modelled as a stochastic process and short term variations in the convenience yield lead to curvature in the futures term structure. Again, this has an economic interpretation, short term gains (falls) in the convenience yield imply a higher (lower) value associated with holding the commodity in physical form. This is observed in the effect of convenience yield variations on the short end of the futures term structure, where humps or saddles can occur. Hypothetical futures curves are plotted in Figure 2 showing the effect of the prevailing convenience yield on the same curves as before.

2.1.2 Volatility Term Structure

The Samuelson effect can be captured by the Schwartz97 model. However it will emerge in the futures curve model if and only if the volatility of the convenience yield is sufficiently small. This can be seen as follows.

Figure 1: Hypothetical futures curves as calculated in the Schwartz two factor model with different long-term mean convenience yields



The log futures price at t for a given maturity T can be written as

$$\ln G(S_t, \delta_t, t, T) = \ln S_t + A(T - t) + B(T - t)\delta_t.$$

Now, both S_t and δ_t are random variables. For a fixed t , the variance of the futures contract maturing at time T can be written as

$$\text{Var}[\ln G(S_t, \delta_t, t, T)] = \text{Var}[\ln S_t] + B(T - t)^2 \text{Var}[\delta_t] + B(T - t) \text{Cov}[\ln S_t, \delta_t].$$

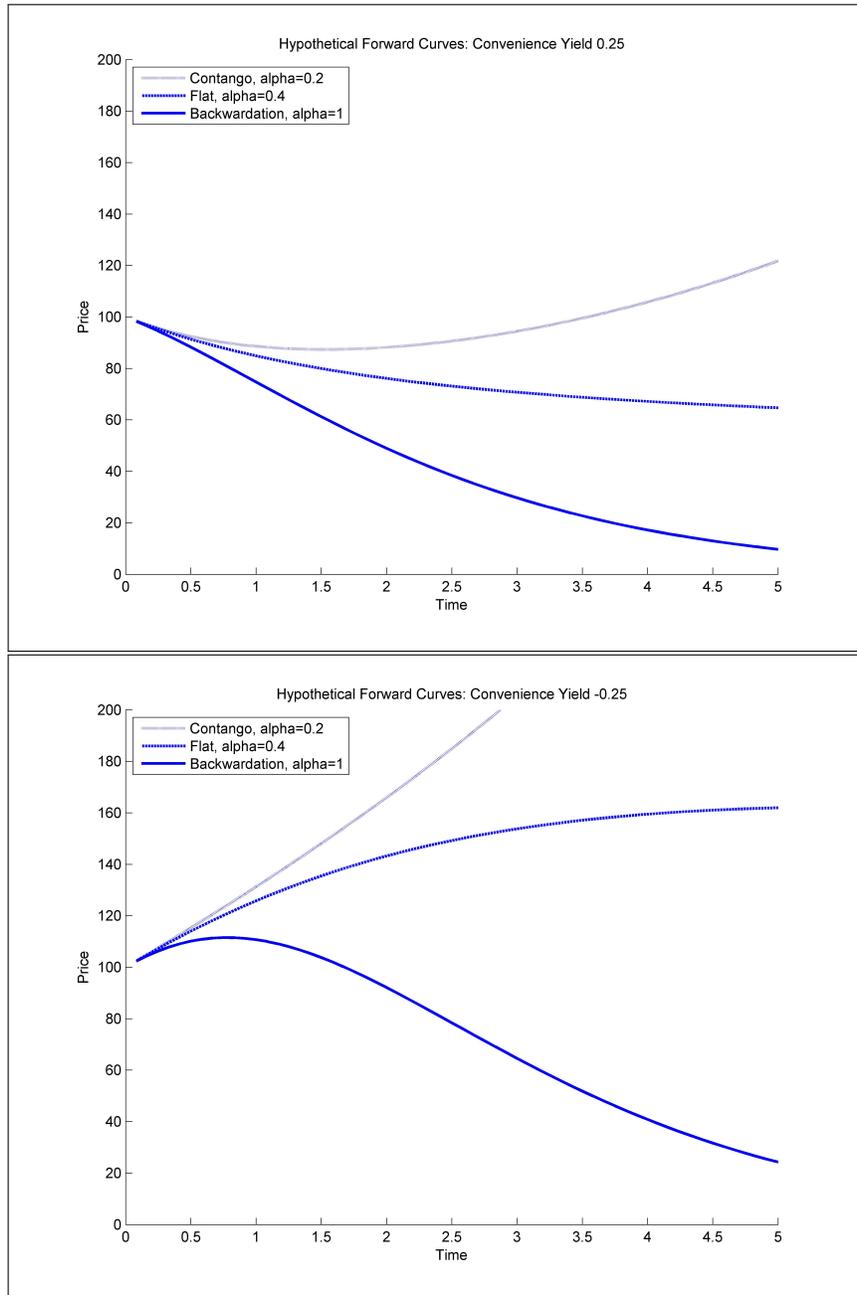
Since $\ln S_t$ and δ_t do not depend on the future's maturity T , only $B(T - t)$ needs to be analysed. It is clear that B decreases as T increases

$$\frac{dB(T - t)}{dT} = -\exp(-\kappa(T - t)) < 0.$$

However, since $B(T - t)$ is negative it does not follow that volatility of a futures contract always decreases as maturity increases. In fact

$$\begin{aligned} \frac{d\text{Var}[\ln G(S_t, \delta_t, t, T)]}{dT} &= 2B(T - t) \frac{dB(T - t)}{dT} \text{Var}[\delta_t] + \frac{dB(T - t)}{dT} \rho \text{SDev}[\ln S_t] \text{SDev}[\delta_t] \\ &= (2B(T - t) \text{SDev}[\delta_t] + \rho \text{SDev}[\ln S_t]) \frac{dB(T - t)}{dT} \text{SDev}[\delta_t] \end{aligned}$$

Figure 2: Hypothetical futures curves as calculated in the Schwartz two factor model with different long-term mean convenience yields and different convenience yields



is negative if and only if

$$\text{SIDev}[\delta_t] \leq -\frac{\rho \text{SIDev}[\ln S_t]}{2B(T-t)}.$$

This inequality has an intuitive interpretation. Since the level of the convenience yield affects the curvature of the futures curve as shown in Figure 2, if the convenience yield is excessively volatile relative to the spot, the entire futures curve will be volatile.

Regardless of the volatility of the convenience yield there is a limit to the volatility of the futures curve. As Lautier [Lau05] shows, the limit of the futures contract volatility is given by

$$\lim_{T \rightarrow \infty} \text{Var}[\ln G(S_t, \delta_0, t, T)] = \text{Var}[\ln S_t] + \frac{\text{Var}[\delta_t]}{\kappa^2} - \frac{\text{Cov}[\ln S_t, \delta_t]}{\kappa}$$

This is an important feature of the Schwartz model.

3 The Simulation

The Schwartz97 two-factor model is especially amenable to Monte Carlo simulation as it is possible to derive the joint transition density for $X_{t_0} = \ln(S_{t_0})$ and the convenience yield δ_{t_0} , see [Erb14]. Under \mathbb{Q} , it is given by

$$\begin{pmatrix} X_{t_0} \\ \delta_{t_0} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \bar{\mu}_X(X_{t_0}, \delta_{t_0}, \tau) \\ \bar{\mu}_\delta(\delta_{t_0}, \tau) \end{pmatrix}, \begin{pmatrix} \sigma_X^2(\tau) & \sigma_{X\delta}(\tau) \\ \sigma_{X\delta}(\tau) & \sigma_\delta^2(\tau) \end{pmatrix} \right), \quad (3)$$

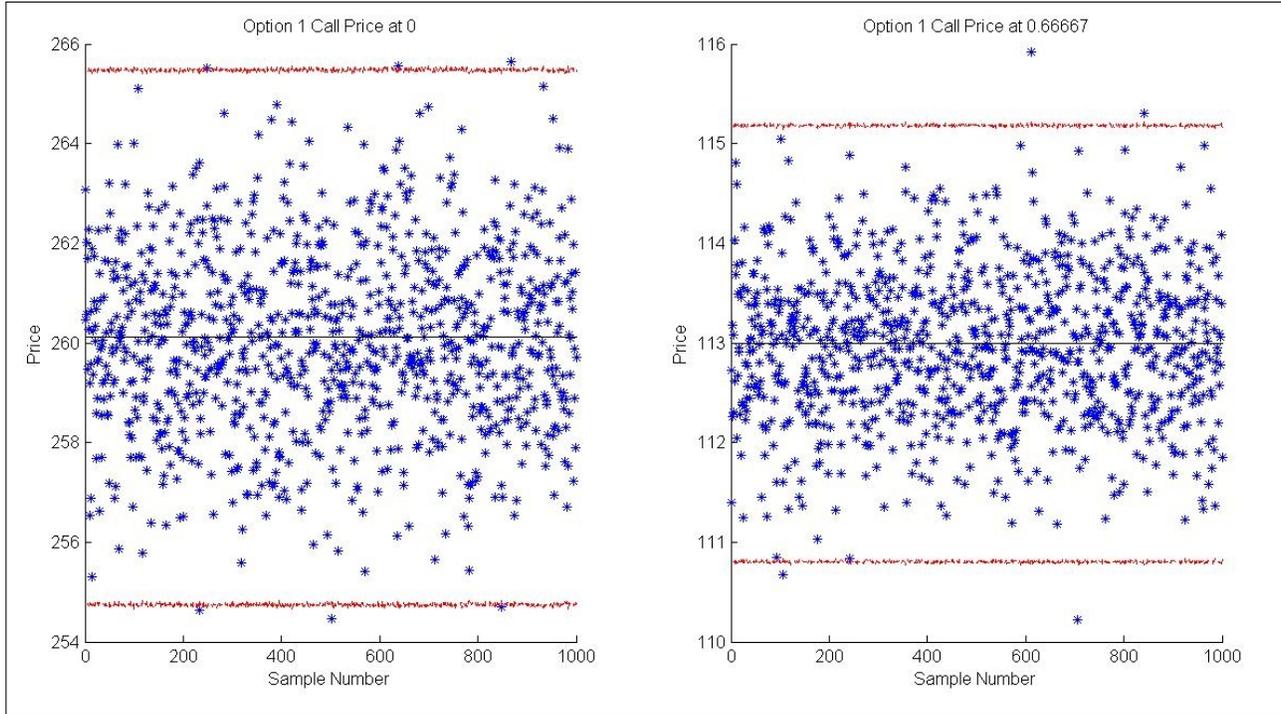
with parameters

$$\begin{aligned} \bar{\mu}_X(X_{t_0}, \delta_{t_0}, \tau) &= X_{t_0} + \left(r - \frac{1}{2}\sigma_S^2 - \bar{\alpha} \right) \tau + (\bar{\alpha} - \delta_{t_0}) \frac{1 - e^{-\kappa\tau}}{\kappa} \\ \bar{\mu}_\delta(\delta_{t_0}, \tau) &= e^{-\kappa\tau} \delta_{t_0} + \bar{\alpha}(1 - e^{-\kappa\tau}) \\ \sigma_X^2(\tau) &= \frac{\sigma_\epsilon^2}{\kappa^2} \left(\frac{1}{2\kappa}(1 - e^{-2\kappa\tau}) - \frac{2}{\kappa}(1 - e^{-\kappa\tau}) + \tau \right) \\ &\quad + 2 \frac{\sigma_S \sigma_\epsilon \rho}{\kappa} \left(\frac{1 - e^{-\kappa\tau}}{\kappa} - \tau \right) + \sigma_S^2 \tau \\ \sigma_\delta^2(\tau) &= \frac{\sigma_\epsilon^2}{2\kappa} (1 - e^{-2\kappa\tau}) \\ \sigma_{X\delta}(\tau) &= \frac{1}{\kappa} \left\{ \left(\sigma_S \sigma_\epsilon \rho - \frac{\sigma_\epsilon^2}{\kappa} \right) (1 - e^{-\kappa\tau}) + \frac{\sigma_\epsilon^2}{2\kappa} (1 - e^{-2\kappa\tau}) \right\} \end{aligned}$$

Thus, Monte Carlo samples can be generated for the log spot and convenience yield at any time, t_0 , through

$$\begin{pmatrix} X_{t_0} \\ \delta_{t_0} \end{pmatrix} = \begin{pmatrix} \bar{\mu}_X(X_0, \delta_0, \tau) \\ \bar{\mu}_\delta(\delta_0, \tau) \end{pmatrix} + LZ, \quad (4)$$

Figure 3: Monte Carlo call price for Option 1 with three standard deviation bounds. The first figure represents the initial price and the second the price one month before expiry. The black line is the analytical solution.



with X_0 and δ_0 given, Z a 2 by 1 vector of standard normal random numbers and L the Cholesky decomposition of the covariance matrix from Equation (3).

At t_0 , the entire futures curve, for all maturities T , is determined by S_{t_0} and δ_{t_0} via Equation 1.

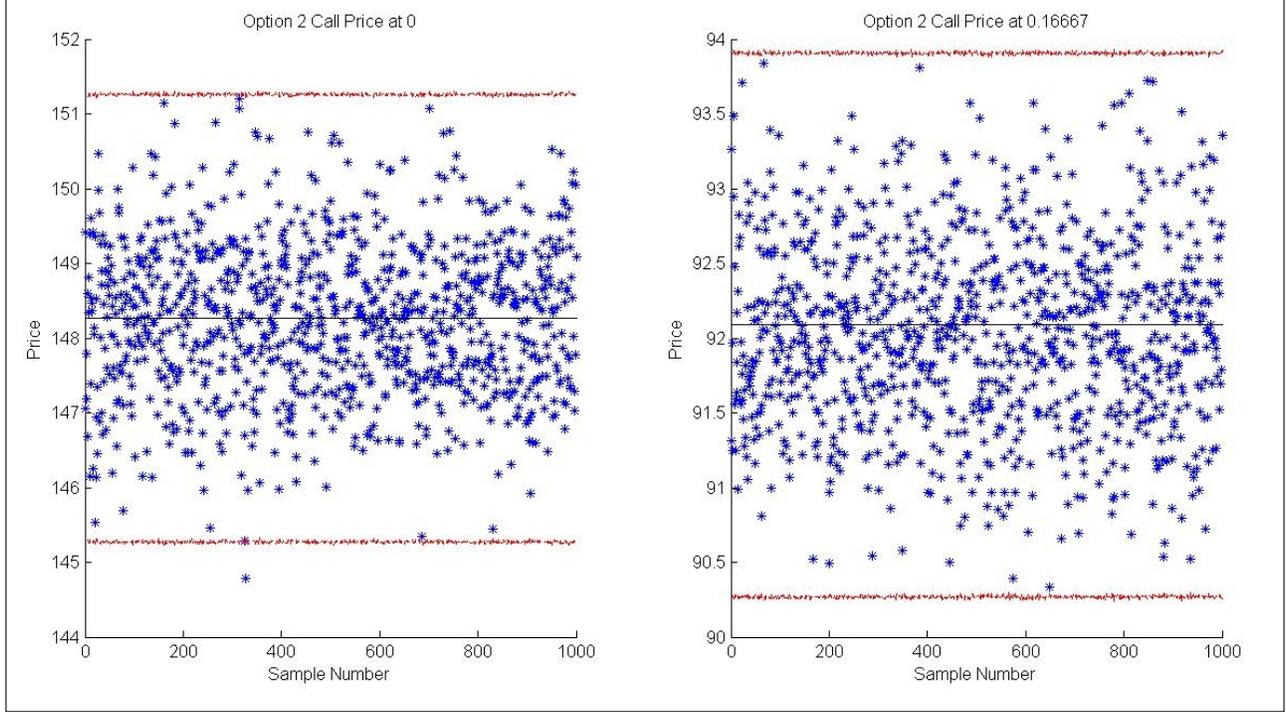
This allows for Monte Carlo simulation of the futures curve at any point and the extension to a simulation of the time-development of the futures curve is obvious.

3.1 Pricing a European Call

Under the Schwartz97 model, an analytical solution exists for the case of a European call written on a commodity future (see [MS98] and [HR98]). This serves as an useful check on the implementation of the Monte Carlo simulation.

Keeping with the notation of [Erb14], but expanding to the case when t_0 does not equal 0, the analytical price for a European call, C^G , expiring at t , on the com-

Figure 4: Monte Carlo call price for Option 2 with three standard deviation bounds. Initial price and one month before expiry. The black line is the analytical solution.



modity future G expiring at T with strike K is given by

$$\begin{aligned} C^G &= E^{\mathbb{Q}}[e^{-r(t-t_0)}(G(S_t, \delta_t, t, T) - K)^+] \\ &= e^{-r(t-t_0)}[G(S_{t_0}, \delta_{t_0}, t_0, T)\Phi(d_+) - K\Phi(d_-)], \end{aligned} \quad (5)$$

with

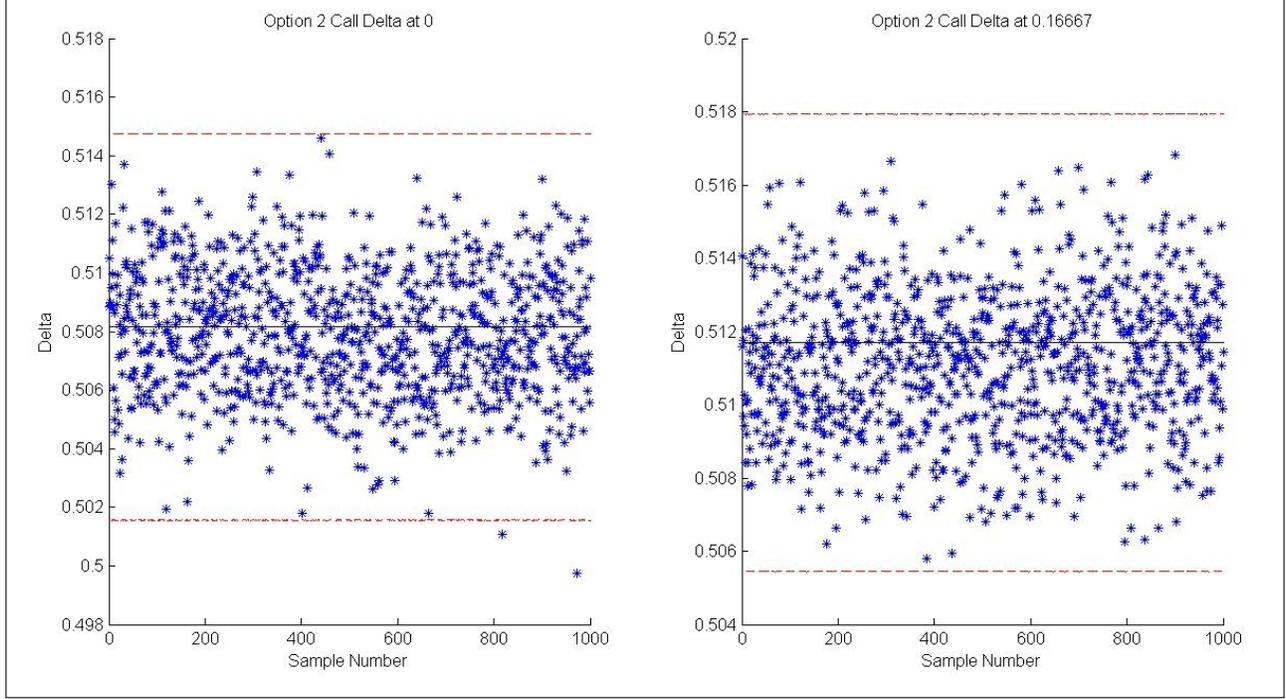
$$d_{\pm} = \frac{\ln \frac{G(S_{t_0}, \delta_{t_0}, t_0, T)}{K} \pm \frac{1}{2}\sigma^2}{\sigma},$$

and

$$\begin{aligned} \sigma^2 &= \sigma_S^2(t-t_0) + \frac{2\sigma_S\sigma_{\epsilon}\rho}{\kappa} \left(\frac{1}{\kappa}e^{-\kappa(T-t_0)}(e^{\kappa(t-t_0)} - 1) - (t-t_0) \right) \\ &+ \frac{\sigma_{\epsilon}^2}{\kappa^2} \left((t-t_0) + \frac{1}{2\kappa}e^{-2\kappa(T-t_0)}(e^{2\kappa(t-t_0)} - 1) - \frac{2}{\kappa}e^{-\kappa(T-t_0)}(e^{\kappa(t-t_0)} - 1) \right). \end{aligned}$$

Figures 3 and 4 illustrate the Monte Carlo pricing of a 9-month call on the 12-month future (Option 1) and a 3-month call on the 9-month future (Option 2). The necessity of checking the effectiveness at two different points during the lifetime of the options will become clear in the next section.

Figure 5: Monte Carlo call delta for Option 2 with three standard deviation bounds. Initial delta and one month before expiry. The black line is the analytical delta.



3.1.1 Delta Hedging

From Equation (5) it is possible to derive the analytical delta of the call option with respect to the underlying future,

$$\frac{\partial C^G}{\partial G} = e^{-r(t-t_0)} \Phi(d_+). \quad (6)$$

This can be used to test the effectiveness of the implementation of a Monte Carlo central-difference delta, which will be necessary for the spread option where a closed-form delta does not exist. It is important to note that the central-difference delta can only be computed with respect to the processes that are being simulated. The central-difference delta with respect to the spot commodity price is given by

$$\Delta^S = \frac{\hat{C}^G(S_{t_0} + \Delta S_{t_0}) - \hat{C}^G(S_{t_0} - \Delta S_{t_0})}{2\Delta S_{t_0}}, \quad (7)$$

where $\hat{C}^G(S_{t_0} + \Delta S_{t_0})$ is the Monte Carlo price of the call option generated with the initial spot price shifted upward by ΔS_{t_0} . This is an approximation to $\frac{\partial C^G}{\partial S}$ and an application of the Chain Rule is necessary to relate it to the delta with respect to

the underlying future,

$$\begin{aligned}\frac{\partial C^G}{\partial G} &= \frac{\partial C^G}{\partial S} \frac{\partial S}{\partial G} \\ &\approx \Delta^S e^{-A(T-t_0)-B(T-t_0)\delta_{t_0}}.\end{aligned}$$

It is possible to compute the central-difference delta with a shift in δ_{t_0} instead, but from a practical perspective it is more sensible to specify the sensitivity with respect to a move in the underlying future, ΔG_{t_0} , which is simple to relate to a move in the spot by

$$\Delta S_{t_0} = \Delta G_{t_0} e^{-A(T-t_0)-B(T-t_0)\delta_{t_0}}. \quad (8)$$

Figure 5 compares the central-difference delta approximation to the analytical delta for a 3-month call struck on the 9-month future at inception and one month before expiry. If the analytical adjustment owing to the Chain Rule wasn't applied, the right-hand figure would show a clear bias whereas the delta at inception would be unaffected.

3.1.2 Profit and Loss Simulations

After establishing the Monte Carlo pricing and central-difference delta, a pathwise profit and loss simulation can be constructed to test the coherence of the system.

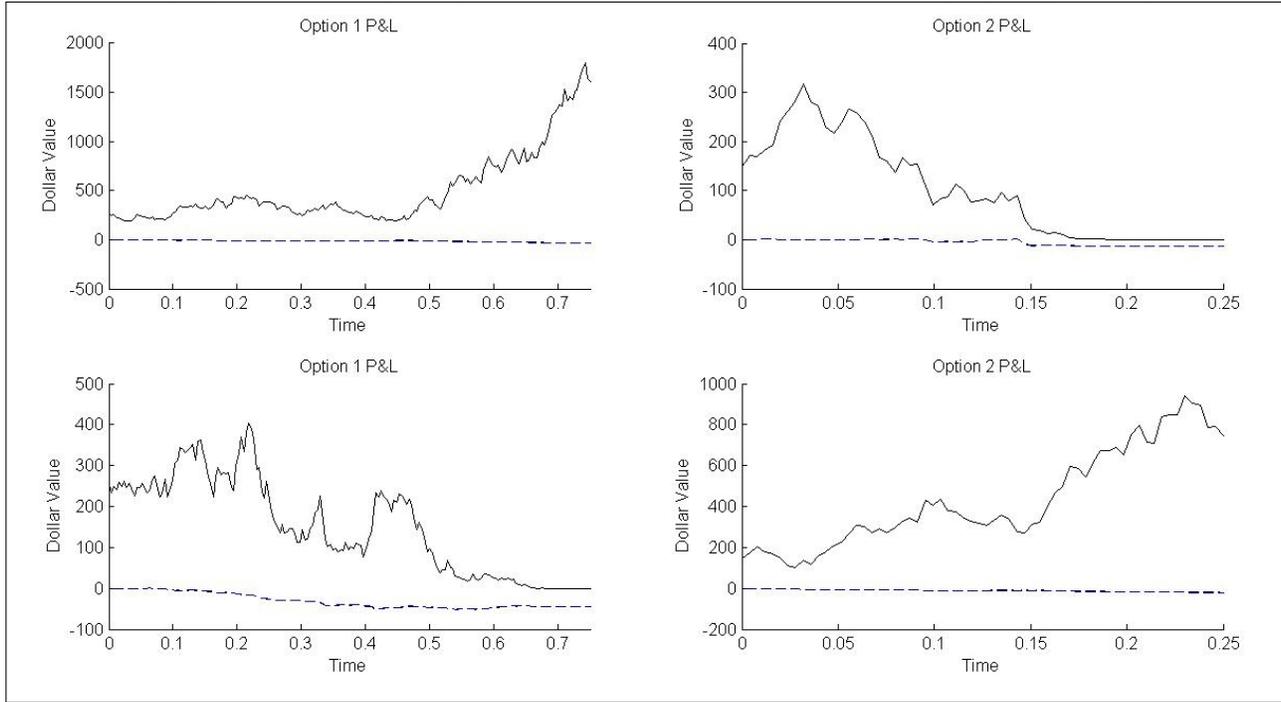
The profit and loss is computed from the perspective of a market participant selling the option, using the following framework:

1. At the initial time, $t_0 = 0$, the hedger sells the option and buys the delta of the underlying. He borrows capital if necessary, otherwise he invests the surplus at the risk-free rate.
2. At each time-step before the maturity, the hedger updates his position in the underlying to the new delta by buying the difference between the delta at this time-step and the previous time-step at the current price of the underlying. He retains his short position in the option and his cash position accumulates from the previous time-step.
3. At maturity, he converts his holding of the underlying into cash and pays out the payoff of the option (if in-the-money).

The hedger's total profit and loss at the end of the period is his final portfolio value.

Two example paths are presented for Option 1 and Option 2 in Figure 6.

Figure 6: Examples of pathwise profit and loss. The solid black line is the evolving option price, where the black and blue dashed lines are the profit and loss of the hedge portfolio using analytical and central-difference delta hedging respectively.



3.2 Pricing the European Spread Option

The price of the European spread option will be the expectation under the risk-neutral measure of its final payoff,

$$D_{t_0} = E^{\mathbb{Q}}[e^{-r(t-t_0)}(G(S_t, \delta_t, t, T_1) - G(S_t, \delta_t, t, T_2) - K)^+].$$

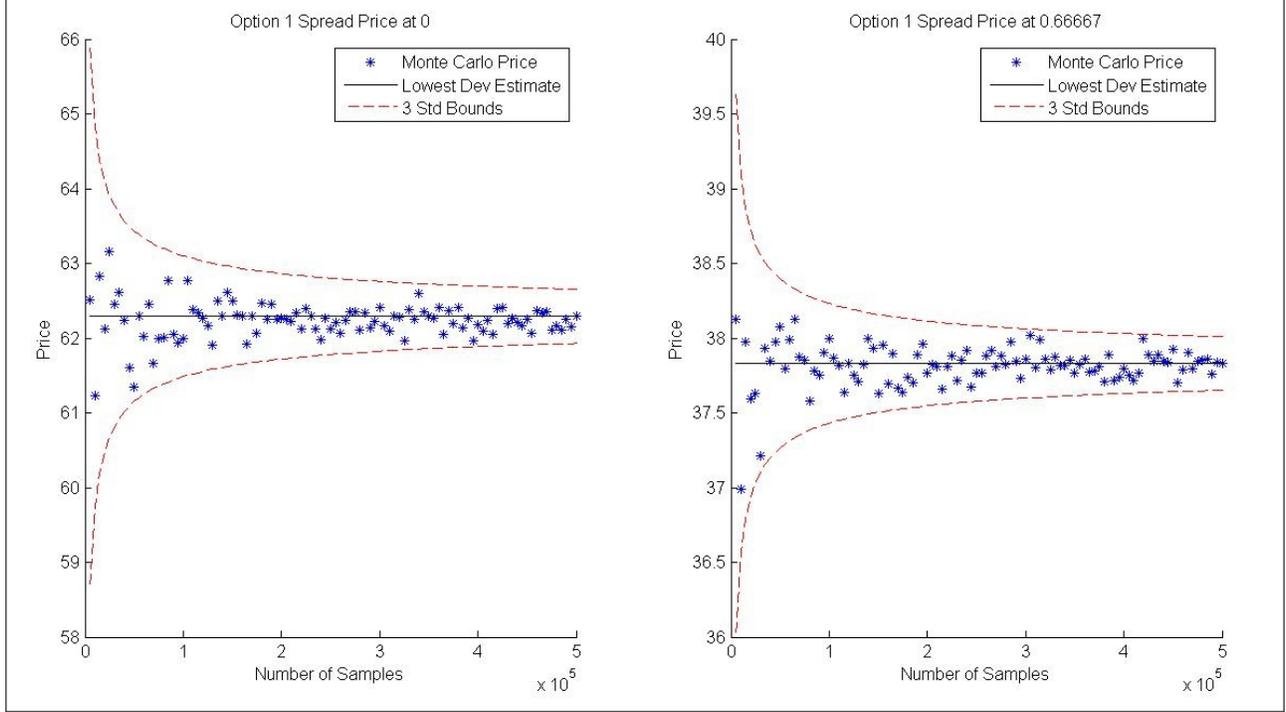
Although no closed-form analytical solution has been determined¹, this option can be priced via the Monte Carlo simulation of the joint underlying processes.

Without a closed-form solution, the resultant Monte Carlo price cannot be checked for accuracy. However, it is possible to examine the behaviour of the solution as the number of contributing paths is increased. This behaviour is illustrated in Figure 7. Qualitatively, it appears that the solution is converging.

For the rest of this section, consider Spread Option 1 as a 9-month European call on the difference between the 16-month and 12-month futures and Spread Option 2 as a 3-month European call on the difference between the 18-month and 9-month futures.

¹Although [AV07] calculate analytic approximations to spread options based on two-dimensional geometric Brownian motion, the results cannot be applied in this case.

Figure 7: Behaviour of the Monte Carlo spread option price at inception and one month before maturity as a function of increasing sample size.



3.2.1 Delta Hedging

To hedge the spread option, positions must be taken in each of the underlying futures. This requires the delta to be computed with care, using the Chain Rule for partial derivatives.

For ease of presentation, the dependence on the current time, t_0 , will be omitted from the notation. Let $D(G^{T_1}, G^{T_2})$ be the price of the spread option with $G^{T_1}(S, \delta)$ and $G^{T_2}(S, \delta)$ the underlying futures. The derivatives of interest are $\frac{\partial D}{\partial G^{T_1}}$ and $\frac{\partial D}{\partial G^{T_2}}$, which must be expressed in terms of $\frac{\partial D}{\partial S}$ and $\frac{\partial D}{\partial \delta}$ and deterministic expressions, as these are the only derivatives that can be computed via the central-difference Monte Carlo simulation.

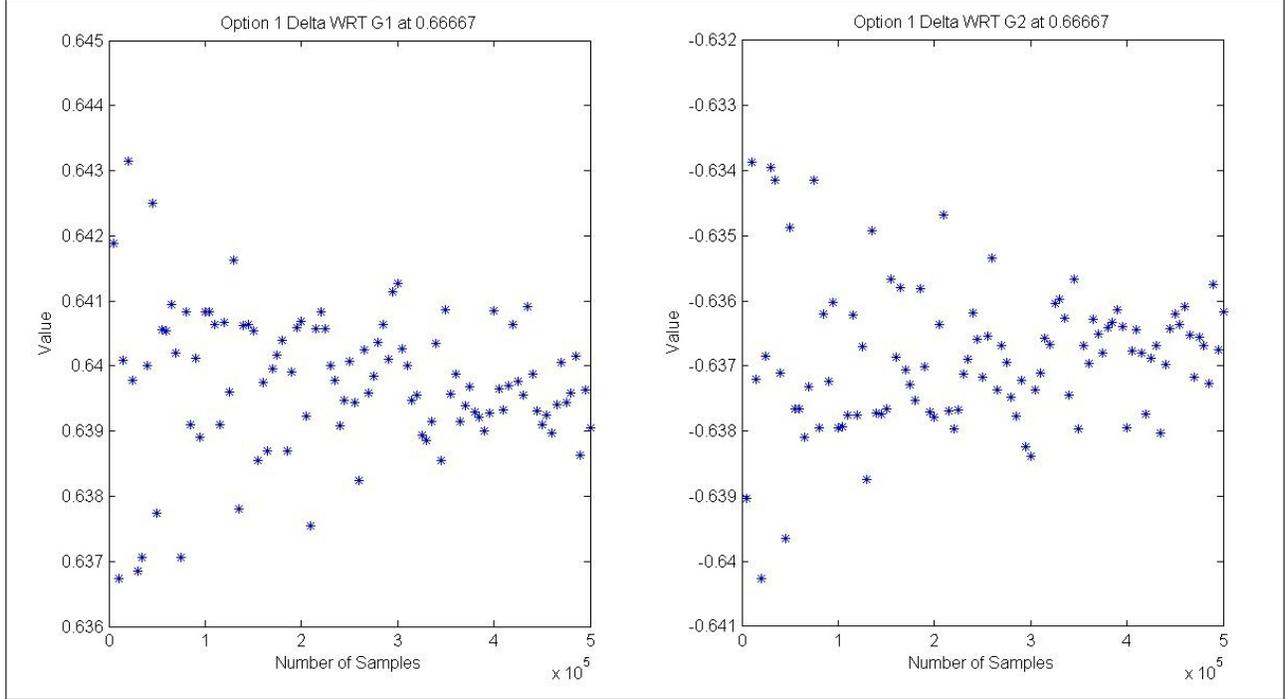
This can be achieved by simultaneously solving the equations,

$$\frac{\partial D}{\partial S} = \frac{\partial D}{\partial G^{T_1}} \frac{\partial G^{T_1}}{\partial S} + \frac{\partial D}{\partial G^{T_2}} \frac{\partial G^{T_2}}{\partial S}$$

and

$$\frac{\partial D}{\partial \delta} = \frac{\partial D}{\partial G^{T_1}} \frac{\partial G^{T_1}}{\partial \delta} + \frac{\partial D}{\partial G^{T_2}} \frac{\partial G^{T_2}}{\partial \delta}.$$

Figure 8: Behaviour of the Monte Carlo Spread Option 1 delta with respect to each underlying one month before maturity as a function of increasing sample size.



This results in

$$\frac{\partial D}{\partial G^{T_1}} = \frac{\partial D}{\partial \delta} \left[\frac{1}{G^{T_1} [B(T_1 - t_0) - B(T_2 - t_0)]} \right] + \frac{\partial D}{\partial S} \left[\frac{1}{\frac{G^{T_1}}{S} \left[1 - \frac{B(T_1 - t_0)}{B(T_2 - t_0)} \right]} \right],$$

with the obvious symmetric extension to $\frac{\partial D}{\partial G^{T_2}}$.

Now central-difference approximations can be used for $\frac{\partial D}{\partial S}$ and $\frac{\partial D}{\partial \delta}$ to completely determine the delta of the option with respect to each underlying future. Calculating the standard deviation of the Monte Carlo central-difference approximation is beyond the scope of this work, but the qualitative behaviour of the above expression is illustrated in Figure 8. Spread Option 1 is a 9-month European call on the spread between the 16-month and 12-futures (7- and 3-month futures at option maturity).

The deltas for the spread option possess an interesting feature: they approximately mirror each other. Standard Chartered's Jayson Dunne intuited that it should be possible to hedge the spread option by taking a position in the spread

itself (as opposed to different positions in each of the underlying futures), despite their imperfect correlation. This appears to be confirmed by the central-difference delta calculations.

3.2.2 Profit and Loss Simulations

As for the call option, having an expression for the delta with respect to each underlying allows for a hedging simulation. In the same framework as before, sample paths are simulated and a hedge portfolio is constructed. For illustrative purposes, two sample paths for Spread Option 1 and Spread Option 2 are displayed in Figure 9. The distribution of the simulated profit and loss is displayed in Figure 10.

The histogram was generated using 4500 samples of the profit and loss. Each profit and loss was computed using daily delta hedging with a 21-day month. Each daily price and corresponding delta was computed from independent 50 000 sample Monte Carlo simulations. The total simulation was spread across 6 computers and ran for approximately 2 hours.

The histogram is encouraging: the mean is slightly to the left of zero, which is to be expected when discretely hedging a continuous time model. It also alludes to the premium that should be charged to account for the fact that all risk, specifically convenience yield risk, is not hedgeable. This qualitatively validates the internal consistency of the implementation.

Figure 9: Examples of pathwise profit and loss for Spread Option 1 and Spread Option 2. The solid black line is the evolving option price, whereas the dashed black line is the profit and loss.

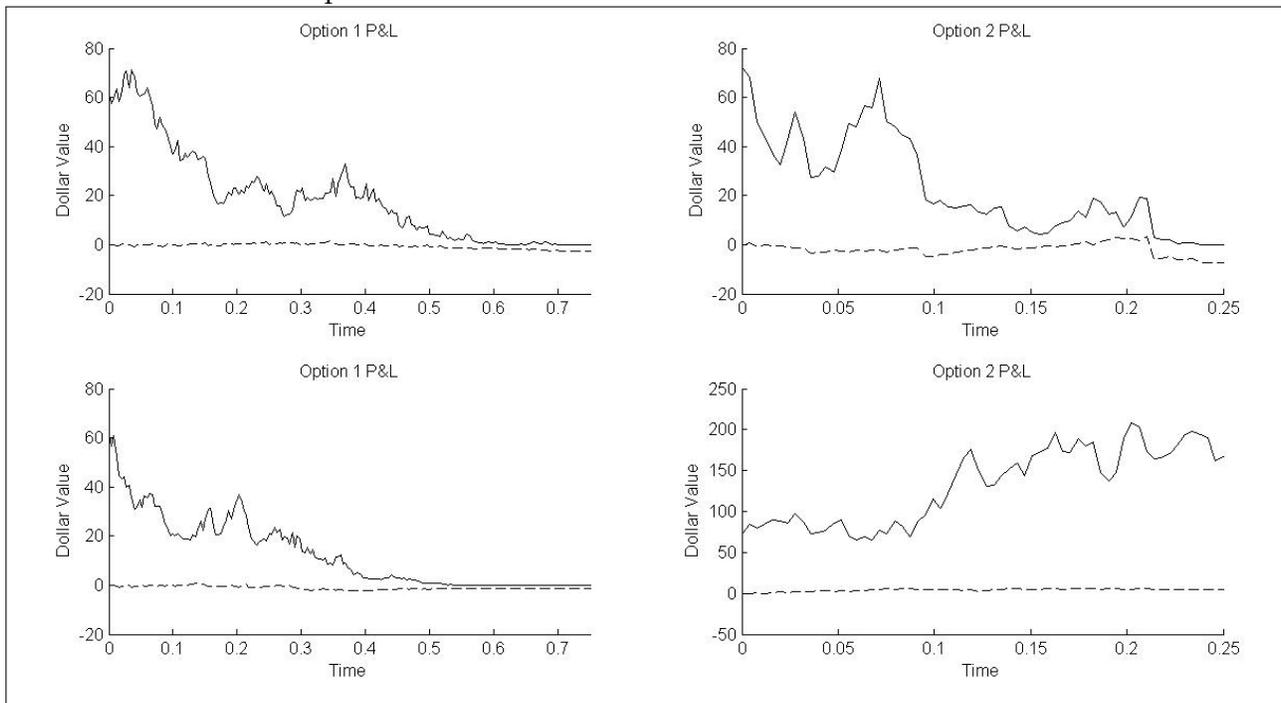
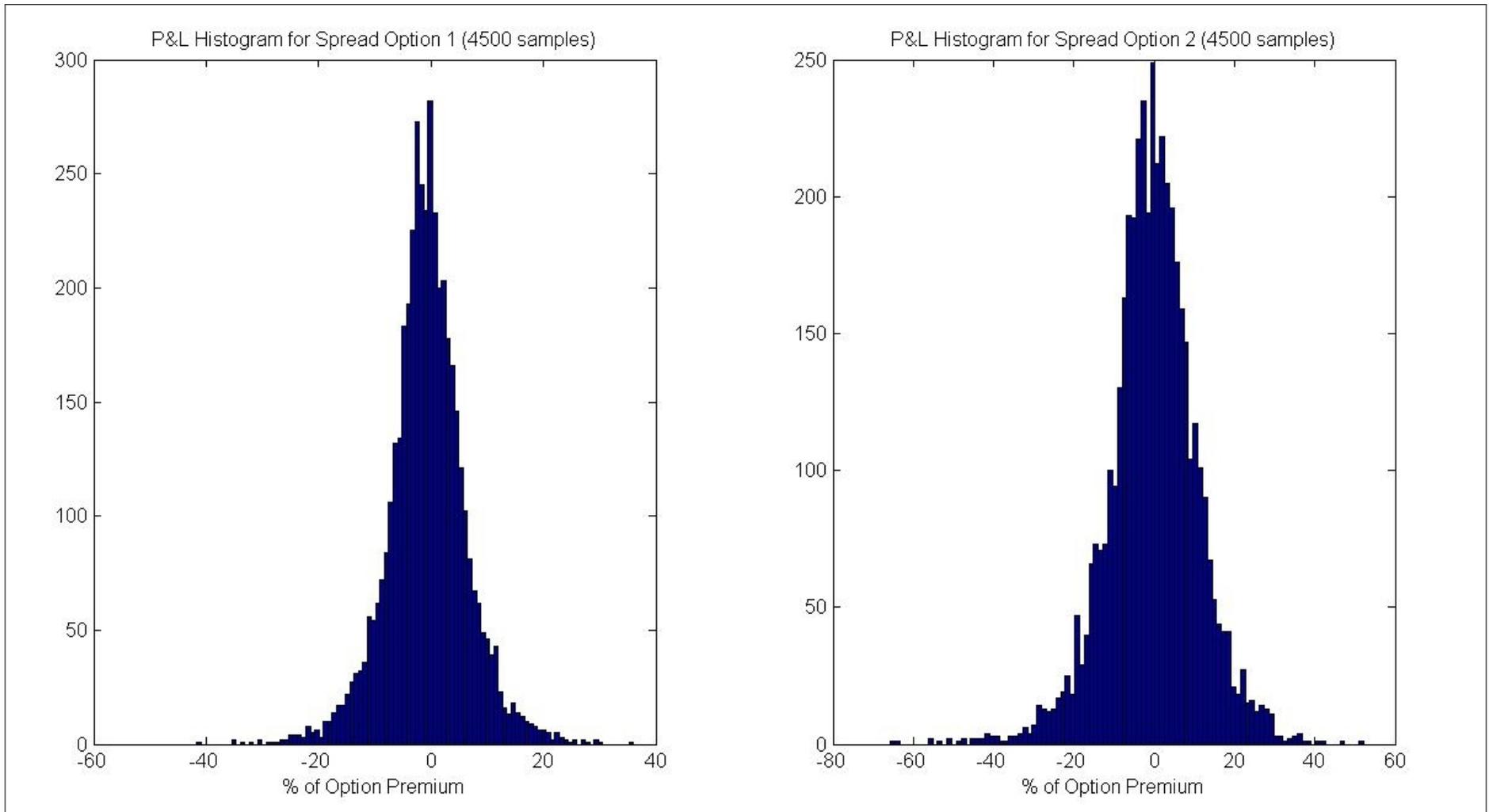


Figure 10: Histogram of profit and loss for Spread Option 1 and Spread Option 2. 4500 sample paths were used, and each point on each sample path was priced and hedged using 50 000 Monte Carlo simulations. This simulation was distributed over 6 workstations.



4 The Calibration

In general option pricing requires parameters which will apply over the life of the option. However, it is not possible to get forward looking estimates for many of the parameters in the model. The volatility of the spot price and convenience yield can be obtained from forward looking market data, but the remaining parameters must be estimated historically

One of the chief advantages of the Schwartz⁹⁷ model is that it allows for calibration of the parameters to historical term structures. This is not an easy task. Both the factors which drive the model are unobservable and in conjunction with the parameters, form futures curves in a highly non-linear fashion. However, due to the model structure it is possible to calibrate with the use of a Kalman filter. Schwartz [Sch97] states that it is appropriate to implement a Kalman filter because we have an observed process which is a function of Markovian state variables (the two factors of the model). This is because the two factor model can be written as a pair of linear state space equations. Using this state space representation of the model it is possible to apply a Kalman filter to the data and thereby estimate parameters for the model. There are many reviews of the Kalman filter, which was initially published by Kalman [Kal60]. One of the most comprehensive and straight forward introductions is given by Welch and Bishop [WB95].

4.1 The Kalman Filter

4.1.1 State Space Representation of the Schwartz 2 Factor Model

The Kalman filter can be used to essentially minimise the mean square error associated with the model's futures curves and the observed futures curves. In the event that these model residuals are normally distributed then the parameters found will also be maximum likelihood parameter estimates. While the observations can be written as a linear function of the state variables, they are non-linear in the parameters. Consequently, the procedure of calibrating the Schwartz⁹⁷ model implements an extended Kalman filter [Jav03]. Ljung [Lju79] highlights that depending on the implementation it is not guaranteed that the extended Kalman filter will yield unbiased, or indeed even consistent (i.e. convergent in sample size) estimates of the parameters. Ljung [Lju79] proves that divergence can occur if the filtering algorithm used is not sufficiently sensitive to the changes in the parameter estimates, the residuals are large and/or the sample size is small.

Despite these challenges, the Kalman filter is widely used to calibrate the Schwartz⁹⁷ model. Javaheri *et al* [Jav03] discuss in detail the specific issues associated with implementing an extended Kalman filter for the calibration of the Schwartz 2 factor model. The chief challenges are the specification of the initial conditions for the filtering process, the biases introduced when linearising the state space and the sensitivity of the algorithm to the residuals' covariance structure.

The implementation of the Kalman filter for the Schwartz 2 factor model by Erb *et al* [Erb14] is presented here. In order to fully explain the parameter estimation via the Kalman filter we follow the explanation of Zivot [Ziv06]. Erb *et al* [Erb14] build a linear state space representation of the model;

$$\begin{aligned} y_t &= d_t + Z_t \omega_t + \varepsilon_t \\ \omega_t &= c_t + T_t \omega_{t-1} + R_t \eta_t. \end{aligned}$$

where the first equation is termed the measurement equation and the second is termed the transition equation. The measurement equation has $N \times 1$ Gaussian white noise ε_t which has a covariance matrix H_t , the transition equation has a 2×1 vector of serially uncorrelated Gaussian innovations η_t . The observation, y_t , in this context is an $N \times 1$ vector of the log futures prices for maturities T_1, T_2, \dots, T_N , and the state vector, ω_t , is a 2×1 vector of the log spot $X_t = \ln(S_t)$ and convenience yield δ_t . So we can write

$$y_t = \begin{bmatrix} \ln(F(1)) \\ \vdots \\ \ln(F(N)) \end{bmatrix}, \omega_t = \begin{bmatrix} X_t \\ \delta_t \end{bmatrix}.$$

The remaining matrices d_t, Z_t, c_t, T_t and R_t are matrices which specify the parameterisation of the system. The state space parameterisation for the Schwartz97 model is given by Erb *et al* [Erb14] as follows;

$$\begin{aligned} d_t &= \begin{bmatrix} A(T_1 - t) \\ \vdots \\ A(T_N - t) \end{bmatrix}, Z_t = \begin{bmatrix} 1 & B(T_1 - t) \\ \vdots & \vdots \\ 1 & B(T_N - t) \end{bmatrix}, \\ c_t &= \begin{bmatrix} (\mu - \frac{1}{2}\sigma_S^2 - \alpha)\Delta t + \frac{\omega}{\kappa}(1 - e^{-\kappa\delta t}) \\ \alpha(1 - e^{-\kappa\delta t}) \end{bmatrix}, \\ T_t &= \begin{bmatrix} 1 & \frac{1}{\kappa}(e^{-\kappa\Delta t} - 1) \\ 0 & e^{-\kappa\Delta t} \end{bmatrix}, R_t R_t' = \begin{bmatrix} \sigma_X^2(\Delta t) & \sigma_{X\delta}(\Delta t) \\ \sigma_{X\delta}(\Delta t) & \sigma_\delta^2(\Delta t) \end{bmatrix}. \end{aligned}$$

Note that that the covariance structure applicable to X_t and δ_t is enforced by multiplying the independent components of ε_t by R_t , this is conceptually different to the measurement noise which η_t represents. The Kalman filter specifies that the measurement error should be serially uncorrelated across time and within contemporaneous errors. Thus we can write the measurement error covariance matrix H_t as

$$H_t = \begin{bmatrix} h_{11}^2 & 0 & \dots & 0 \\ 0 & h_{22}^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & h_{NN}^2 \end{bmatrix}.$$

4.1.2 Parameter Estimation via the Kalman Filter

For a linear state space with a given set of parameters the Kalman filter supplies estimates of the underlying state processes ω_t . The estimate of ω_{t-1} can be used for an *a priori* prediction of the observation vector y_t . The errors associated with these predictions from the Kalman filter have a Gaussian distribution. So we can find the parameters which maximise the log-likelihood of the prediction errors. This corresponds to minimising the prediction errors associated with the Kalman filter. Thus the calibration process is complex as there are two levels of iteration: the Kalman filter is an iterative process which is wrapped inside an iterative log-likelihood maximisation.

The detailed calculation of the log-likelihood which Zivot [Ziv06] outlines is presented here. Firstly, it is necessary to understand that the Kalman filter is a recursive process which moves forward over time. At each time point two sub calculations take place; a prediction step and an updating step. The prediction step rolls the state process forward, and gives a prediction of the observation process $y_{t|t-1}$. The correction step adjusts the previous state space estimate for the error between the predicted observation and the actual observation $v_t = y_t - y_{t|t-1}$. We distinguish between the *a priori* estimate of the state vector, $w_{t|t-1}$, and the *a posteriori* estimate (or the filtered estimate) of the state vector w_t and their corresponding mean square errors, $P_{t|t-1}$ and P_t . This process is then repeated over the sample.

The equations associated with this process are given in Zivot [Ziv06] as follows. The prediction step equations are given by

$$\begin{aligned} w_{t|t-1} &= E[\omega_t | \mathcal{F}_{t-1}] = T_t w_{t-1} + c_t \\ P_{t|t-1} &= E[(\omega_t - w_{t-1})(\omega_t - w_{t-1})' | \mathcal{F}_{t-1}] \\ &= T_t P_{t-1} T_t' + R_t R_t' \\ y_{t|t-1} &= Z_t w_{t|t-1} + d_t. \end{aligned}$$

The prediction error v_t and its corresponding mean square error F_t are given by

$$\begin{aligned} v_t &= y_t - y_{t|t-1} \\ F_t &= E[v_t v_t'] \\ &= Z_t P_{t|t-1} Z_t' + H_t. \end{aligned}$$

These two components are used to correct the *a priori* state space estimate and its mean square error. The correction step equations are given by

$$\begin{aligned} w_t &= w_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t \\ P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}. \end{aligned}$$

If the assumption of normality in the measurement errors and state space innovations hold, then the prediction errors are also normally distributed. Consequently,

the log-likelihood of the prediction errors as a function of the parameter vector $\theta = (\kappa, \omega, \lambda, \sigma_S, \sigma_\epsilon, \rho)'$ is

$$\ln(\mathcal{L}(\theta|y_t)) = -\frac{Nn}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln(\det(F(\theta))) - \frac{1}{2} \sum_{t=1}^n v_t(\theta) F_t^{-1}(\theta) v_t(\theta).$$

This function is maximised over θ to calibrate the model.

4.1.3 Difficulties Associated with Calibration

Erb *et al* [Erb14] state “Estimation of the Schwartz97 model parameters is statistically fragile and computationally demanding”. This idea is echoed in Javaheri *et al* [Jav03], who also emphasise the difficulty of implementing a Kalman filter to estimate parameters for a Schwartz 2 factor model.

One of the most problematic areas is the choice of starting values for the Kalman Filter, that is the choice of a_0 and P_0 . In the event that ω_t and y_t are covariance stationary then there are analytical expressions for the starting values, as given by Zivot [Ziv06]. However, the Schwartz97 model explicitly has covariance non-stationarity. Thus, it is matter of judgment to choose the starting values. One solution proposed by Javaheri *et al* [Jav03] is to estimate the initial states using the first few observations in the sample. The initial convenience yield can be taken as

$$\delta_0 = r - \frac{\ln(G(t, T_1)) - \ln(G(t, T_2))}{T_1 - T_2}$$

and the initial spot, S_0 can be taken as the price of the shortest-dated futures contract. The estimate of P_0 can be a diagonal matrix, with elements being sample variances of the corresponding points on the futures curve in the first 30 observations.

An additional issue pointed out by Javaheri *et al* [Jav03] is the measurement error covariance matrix H_t . If the volatility of the measurement errors become too unstable then it can lead to numerical difficulties, particularly relating to the invertability of F_t . The use of the linearised state space also means that care must be taken to avoid bias in the parameter estimates.

There are also issues related the log-likelihood and parameters. Erb *et al* [Erb14] highlight that local maxima often exist in log-likelihood function. Consequently it is necessary to allow for many iterations in an optimisation procedure as well start the procedure from different initial parameter estimates. Finally many of the parameters have constraints, which adds another layer of complexity to the estimation.

4.1.4 Calibration Process

Due to the time constraints in place it was necessary to budget the amount of time set aside to implement the Kalman filter parameter estimation. It was necessary

to implement the filter in a short amount of time. Unfortunately, it proved too difficult to develop an estimation procedure within the time constraints. Consequently, an existing package designed specifically for estimating parameters of the Schwartz97 model was used. The package was designed for the R statistical computing language by Erb *et al* [Erb14]. The package implements the procedure described above with additional steps and methodologies to control for the numerical instability inherent in the method.

The package supplies model output which can be used to assess both the robustness of the estimation procedure and the parameter estimates. The process can be started at multiple initial parameters to ensure that the estimates found do not correspond to local maxima. It is also important to note that there is scope to choose which points, on the futures curve, can be pushed into the calibration procedure. The granularity and length of the futures curve used can be altered. Using too granular a futures curve is likely to incur too much residual error which may skew the estimates obtained. Using too sparse a curve may lead to insufficient calibration to the existing term structure. Ideally a set of points which allow for a sufficient representation of the futures term structure without inducing too much residual error should be used. This is an empirical problem and so the estimation stability and final use of the parameters should also be taken into account when deciding what data to use.

A final issue is the choice of a risk-free rate over the sample period. It was decided to use a risk-free rate of 3 per cent over the sample period.

4.2 Applying the Kalman Filter

4.2.1 The Data and Model Results

The model was calibrated to data from the Copper futures curve. 48 Month Copper futures curves observed daily from 1 January 2002 until 30 June 2014 on a daily basis were obtained from Bloomberg. Figures 11 and 12 demonstrate the Samuelson effect and backwardation and contango respectively in the sample. No seasonality in the market for metals was expected. The mean reversion is difficult to assess before parameters are estimated.

It was decided to calibrate the Schwartz97 model using 3 month points on the futures curve for a futures curve of length 1.5 years. This data yielded the most stable parameter estimates in the calibration, in addition it corresponds to the length of the futures curve required to price the spread option in question.

Table 1 shows the parameters found using the Schwartz97 package to implement a Kalman filter log-likelihood maximisation. The parameters seem consistent with market observable phenomena. The historical spot volatility of 0.36 per cent is in line with what is expected from a metal commodity, especially since the calibration period includes the 2008 crisis. The strongly positive correlation parameter implies that there will be mean reversion in the spot price which is also a desir-

Figure 11: Sample volatility structure for Copper from 2002 to 2014.

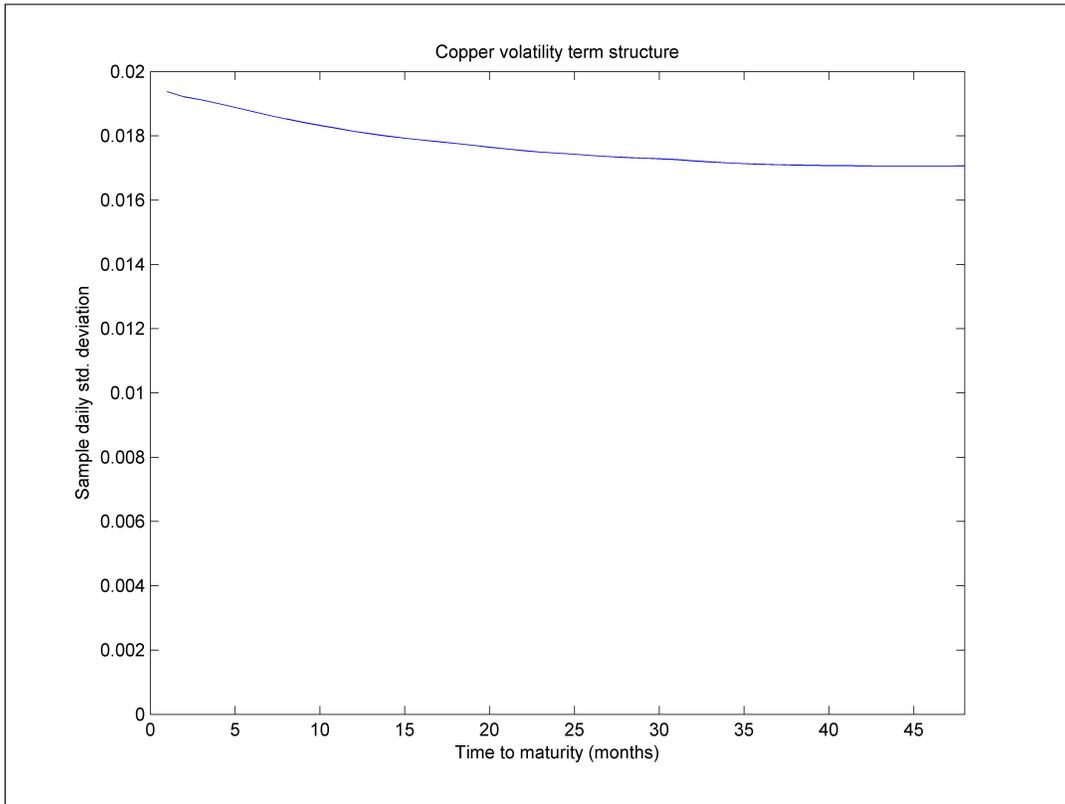
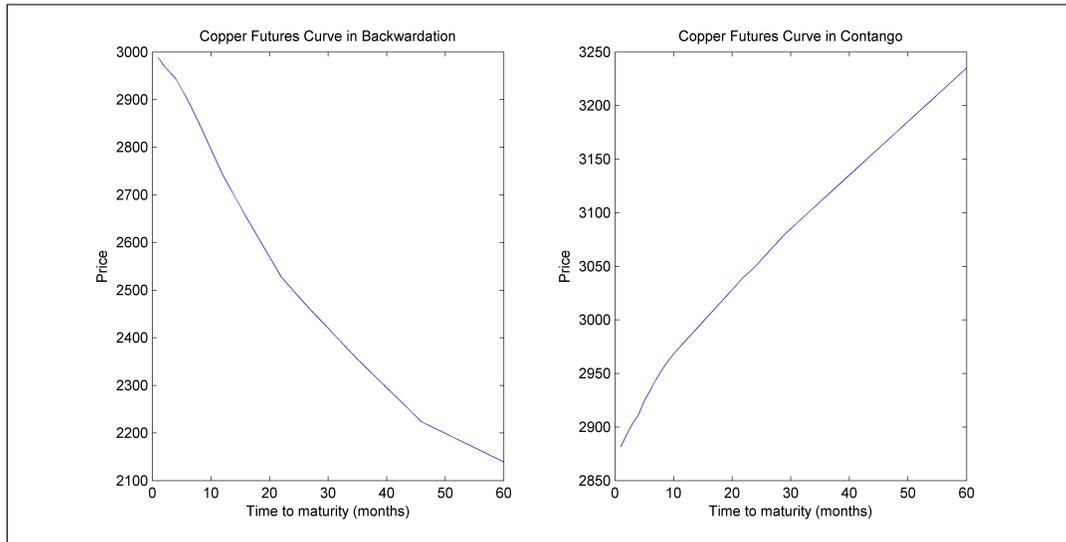


Table 1: Parameter Estimates for the Schwartz 2 factor model

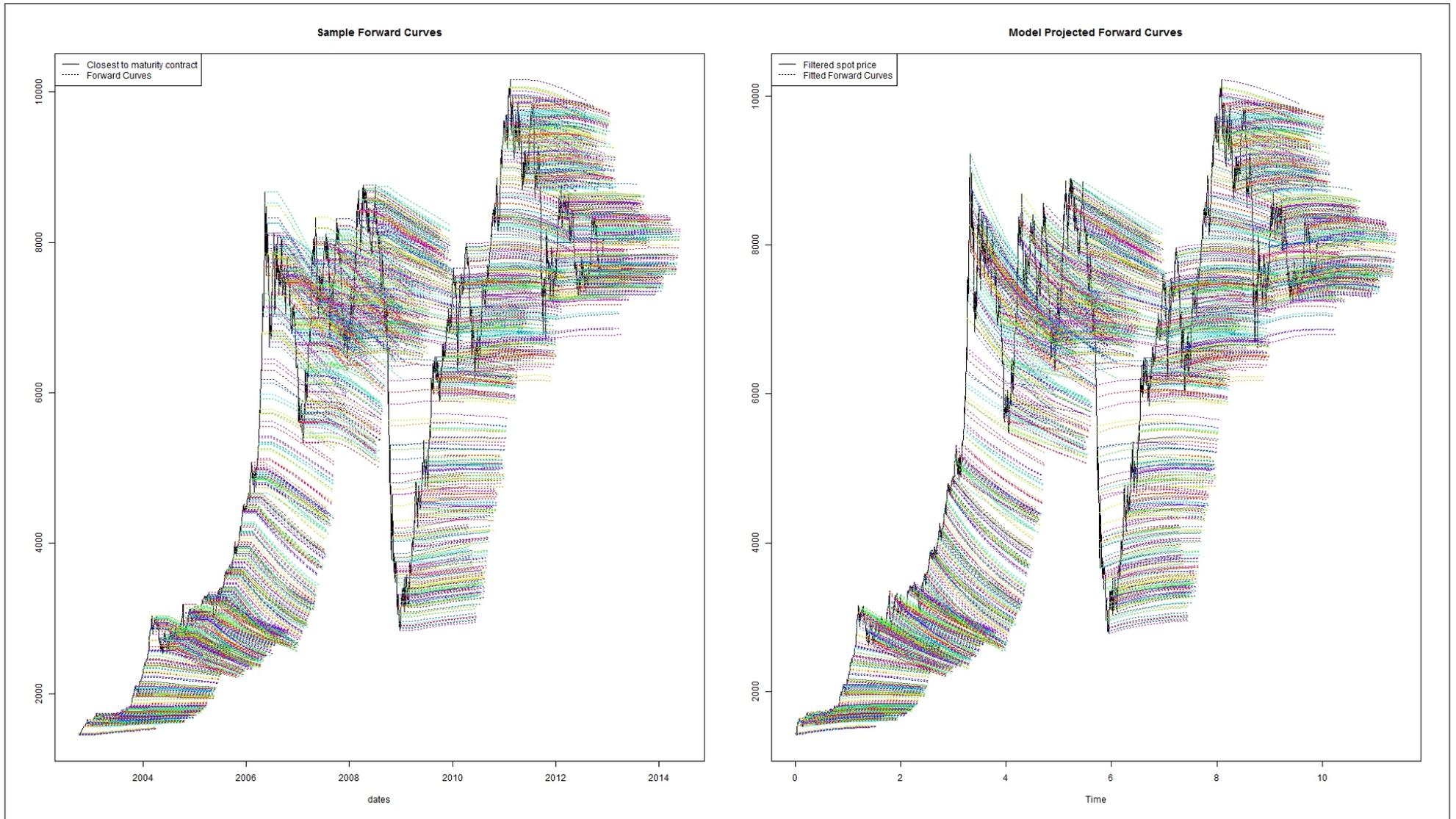
Parameter	Estimate
μ	0.9991
σ_S	0.3626
κ	0.5212
ω	0.3513
σ_ϵ	0.1023
ρ	0.9207
λ	0.1698
α	0.3513

Figure 12: Selected Copper futures curves from 1 Jan 2003 and 1 Jan 2012 exhibiting backwardation and contango respectively.



able factor. In addition, the volatility of the convenience yield sits below the bound discussed in section 2.2, so the Samuelson effect is recovered. Figure 13 shows the actual forward curves and the forward curves projected using the parameters found as well as the filtered estimates of the convenience yield and the spot price. There is a favourable correspondence between the two plots. Much of the behavior of the curve in the sample is captured in the model.

Figure 13: Plots of the actual and model projected forward curves over the sample period



4.2.2 Model Assessment

One of the shortfalls of the Schwartz97 package is that standard errors for the parameter estimates are not readily available. However other forms of model checking are available. The primary method is to assess the distribution of the residuals associated with the Kalman filter. Figure 14 shows the distributions of the standardised residuals for each point on the futures curve. These distributions seem close to normal. There is evidence of skewness in the residuals indicating some structure that the models do not capture. Overall, given the time period over which estimation took place, this is to be expected given the flat interest rate assumption which was made.

The estimation procedures runs successfully, with the parameter estimates converging to their final values. Figure 15 shows this for selected parameters.

5 Bonus: Seasonality

The effect and impact of seasonality in commodity markets has been well documented in the literature, and there exists a small subset which attempts to model the seasonality in futures curves. [Gab91] uses a combination of sine and cosine functions in his models of the spot price and the long-term price, similar to the work done by [RS02]. They implement the Schwartz97 model, however they expand the model by including a stochastic volatility in the spot price which allows for seasonality. [Mil03] takes this approach one step further by attempting to model as closely as possible the term structure of futures and futures curves observed in the market as well as “the current term structure of future and futures volatilities, and the inter-temporal pattern of the volatility of the future and futures prices” as observed in the market. Miltersen models the underlying spot price with geometric Brownian motion and the convenience yield with a mean reverting process, and obtains greater flexibility in the model by “allowing the volatility of the spot commodity price, the speed of mean-reversion parameter, the mean-reversion parameter, and the diffusion parameter of the spot convenience yield all to be time-varying deterministic functions”. Finally, [BG06] suggested a more unique approach to modelling the seasonality in futures curves by modelling the average long-term price, which exhibits no seasonality, as well as a factor similar to the convenience yield which she terms the “seasonal factor”. This factor is responsible for the random changes in the futures curve shape as well as the the seasonal premia that result in the seasonal shape of the curve.

The problem under consideration is the pricing and hedging of the European spread option (as stated above) on a commodity with a seasonal futures curve. The previous methodologies for modelling futures curves all necessitate estimating parameters, mostly by using some form of quasi-likelihood maximisation.

Time constraints prevented the implementation of a new parameter estimation

Figure 14: Plots of the actual and model projected forward curves over the sample period

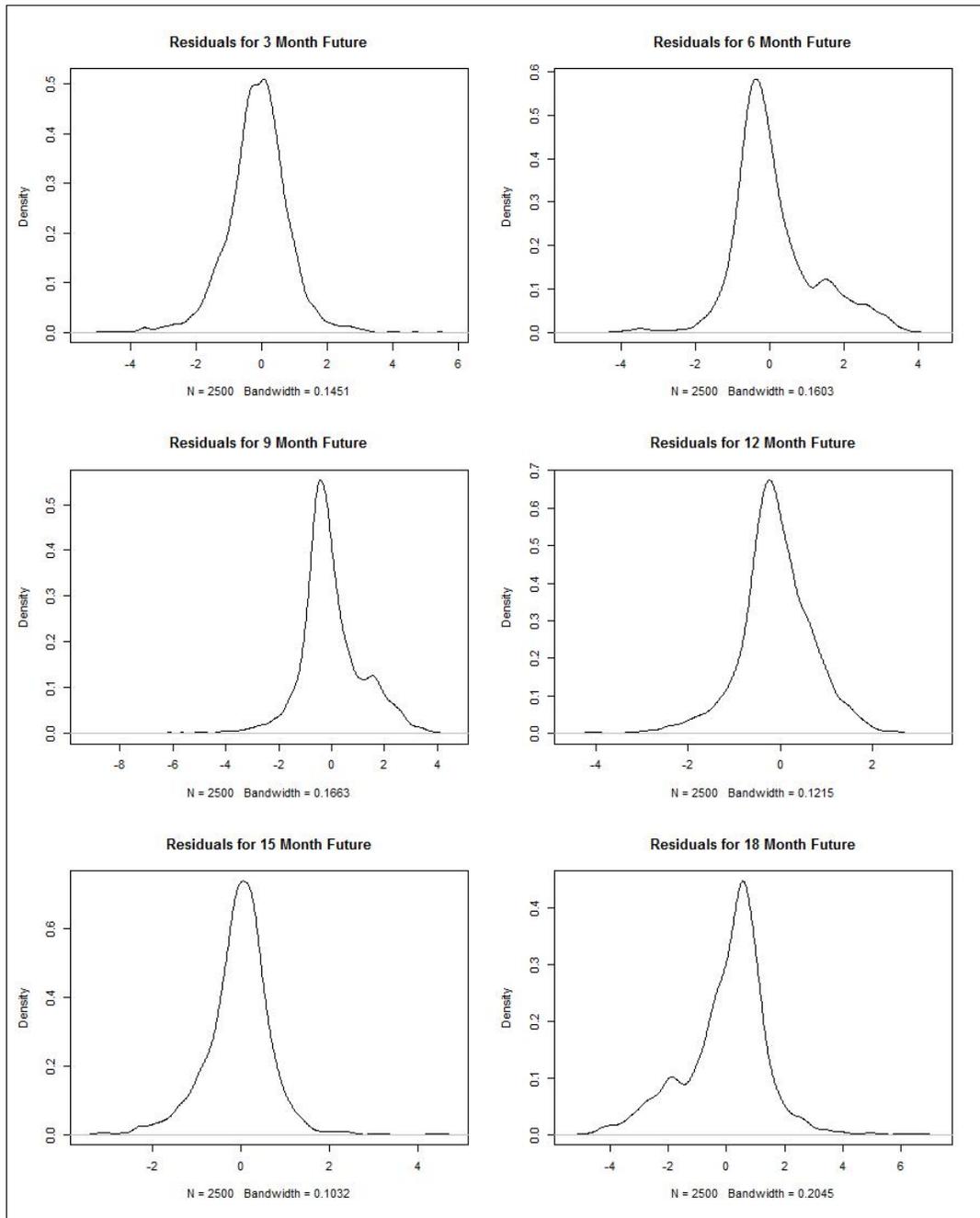


Figure 15: Plots of the parameters as over the likelihood maximisation

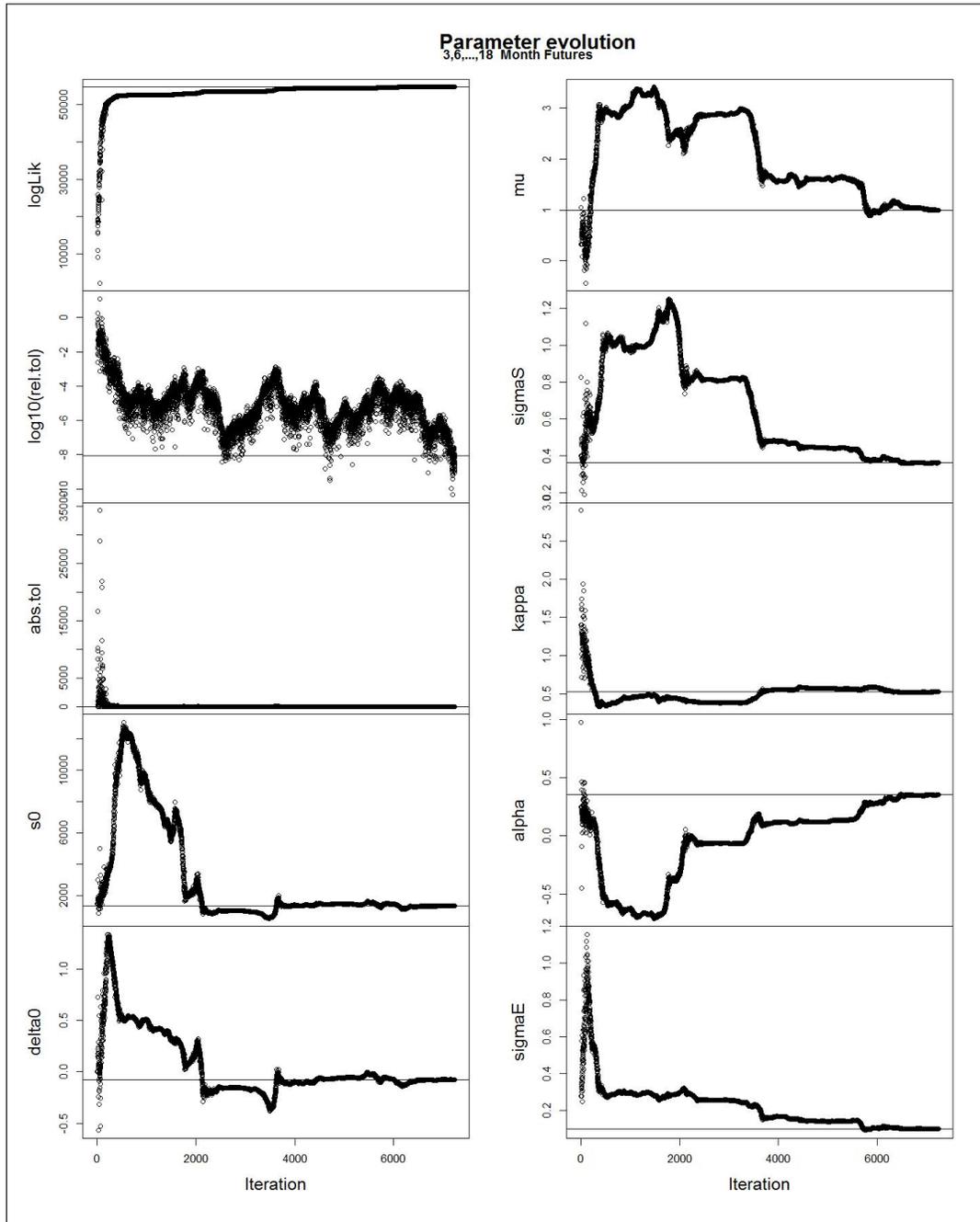
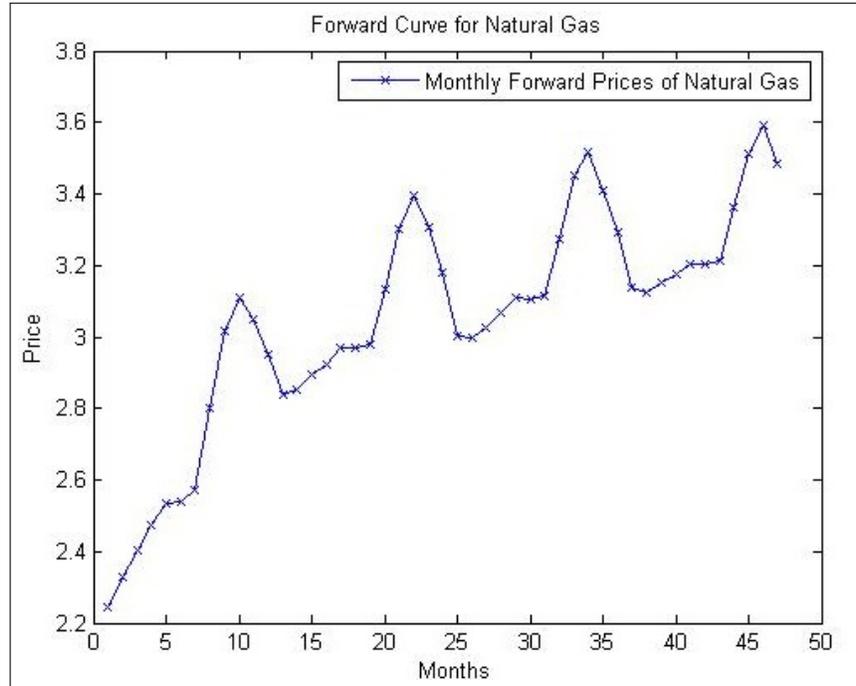


Figure 16: Forward curve for natural gas on 14 February 2002. Monthly forward prices from one month out to 48 months.



procedure (see Section 4) and thus the investigation was limited to models for seasonality whose parameters could be estimated under the current implementation of the Kalman filter.

This constraint was the largest factor preventing the implementation and assessment of the approaches outlined in the literature.

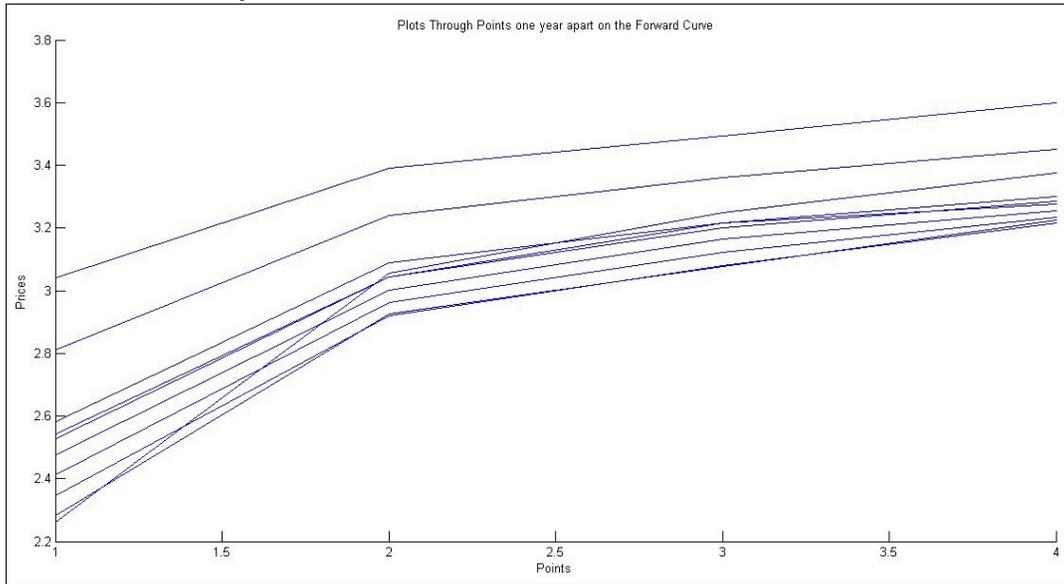
5.1 Modelling Seasonality

Figure 16 is the natural gas curve at 14 February 2002 with monthly points from one month out to 48 months.

The seasonal pattern is clearly visible in Figure 16 above. Empirically, one can see the effect of the increased demand for natural gas in the winter months, hence the higher premium for futures expiring during those months, and during the summer months when demand is slightly lower, the premium is lower. Another important observation is that this seasonal effect appears constant year on year, i.e this yearly seasonal pattern does not appear to be changing over time.

Developing this idea further, if one only connected the future prices one year apart (i.e connecting the one month and 13 month, the 13 month and 25 month etc) one would have a futures curve with, predictably, no seasonality, but more impor-

Figure 17: Illustration of the shape of the forward achieved by connecting points on the forward curve exactly one year apart. This is demonstrated for the forward curve on 7 January 2002.



tantly, the curve would maintain the general shape of backwardation or contango experienced over the period. This is displayed in Figure 17.

Thus each day, if one only looked at the futures curve implied by plotting the yearly future prices from the one month future price today, one would have a futures curve maintaining the general contango or backwardation shape, and to obtain the monthly prices in between these points, one can apply a set of monthly seasonal adjustments.

The approach will be to obtain the monthly adjustments mentioned above, henceforth called 'seasonal premia', and use these to remove the seasonality from the observed market futures curves for natural gas. These futures curves will then be used in the calibration procedure in the same way the copper futures curves were used. The futures curves estimated from using the calibration data will exhibit no seasonality as the model is set up to ignore seasonality and the seasonal premia can be added back to introduce seasonality to the resultant futures curves.

It is assumed that the yearly seasonality is constant over time, which appears to be a justifiable assumption(see Figure 16). Secondly the assumption is made that the seasonality can be expressed as a multiplicative monthly factor or premium. In other words, each day a futures curve exists with monthly future prices from one month onwards. There are 13 seasonal premia that determine the seasonal shape of the curve between the one month and the 13 months future prices, and these same premia dictate the seasonal shape of the curve between the 13 month and 25

month future prices.

5.1.1 The Model

If we let $X^S(t_0, T)$ be the seasonal future price at time t_0 for delivery at time T , and $X^{DS}(t_0, T)$ the corresponding 'deseasonalised' future price, then the model proposed is as follows:

$$X^S(t_0, T) = SP(t_0, m(T))X^{DS}(t_0, T) \quad (9)$$

for $t = 1, 2, \dots, 252, T = 1, 2, \dots$ and

$$m(i) = \begin{cases} \text{Rem}(\frac{i}{12}) & \text{when } \text{Rem}(\frac{i}{12}) > 0, \\ 12 & \text{when } \text{Rem}(\frac{i}{12}) = 0, \end{cases}$$

where $\text{Rem}(\frac{x}{y})$ is the integer remainder of $\frac{x}{y}$.

Hence the seasonal premia will take the form of a 252 by 13 matrix. The function $m(i)$ ensures that no matter how large T is, it is always related to the correct monthly premium. Also note that as a result of the model design and assumptions: $SP(t, 1) = SP(t, 13) = 1$ for all t .

5.1.2 The Data

Daily futures curve data from 2 January 2002 up until 30 June 2014. Each daily futures curve consists of futures prices from one month out to 37 months. In the computation, a trading year of 252 days is assumed, thus 21 days each month. During the period, any years with more than 252 trading days were edited (i.e the 253rd day was removed).

Also it must be noted that the convention used in the natural gas market is that the future price today expiring at some time T is quoted in terms of the price of the nearest to expiry future of term T . In other words the price quoted today and tomorrow for a one month future will both be linked to the nearest to expiry one month future. And once this future expires, the pricing jumps to being in terms of the next nearest one month future.

5.2 Applying the Model

5.2.1 Estimating the Seasonal Premia

The seasonal premia are estimated by first obtaining unique daily seasonal premia and then calculating overall average seasonal premia. This is outlined below.

1. Unique daily seasonal premia.

In this first step, the seasonal premia are estimated using each day's unique futures curve out to 37 months. To estimate the seasonal premia, straight lines are fitted between the one month and the 13 month future prices, the 13 month and the 25 month future prices and the 25 month and the 37 month prices. The prices on the actual futures curve are then divided by the points along the straight line. Hence a per month factor is computed showing how much higher(or lower) that month's price is compared to the straight line. This is then averaged with the corresponding month factors obtained in the 13 month to 25 month segment, and the 25 month to 37 month segment.

A set of 13 seasonal premia for each day from 2 January 2002 to 30 June 2014 were obtained during this step.

2. Overall average seasonal premia.

In this step, the seasonal premia are averaged over the corresponding day in each year, so as a result a set of 252 seasonal premia is obtained, corresponding to the future prices monthly from one month to 13 months.

Figure 18 shows the final seasonal premia for each month. There are 252 lines, each showing the corresponding seasonal premia. The apparent clustering effect is a result of the nature of the natural gas data (see Figure 16). The seasonal effect is clearly visible.

Figure 19 shows a cross-section of Figure 18. Each subplot shows the monthly premium that would be applied at each given day over the course of the 252 day year. It shows early in the year that the premiums on the 9, 10 and 11 month natural gas futures are high, as is to be expected. Also, it is reassuring to see this price premium move in a predictable way as the winter season is approached. Also the red lines shown on the subplots emphasise the stepped nature of the curve, again which is to be expected given the nature of the natural gas data. Essentially, every 21 day sequence is in reference to the same futures contract, hence exactly 12 of these red lines can be superimposed and they conform almost exactly to the stepped nature of the curves. Consider the one month future prices for each of the first 21 trading days of the year. They all give very similar seasonal premia as they reference the same contract. This then shifts as that contract expires and the next 21 observations reference the next one month contract, hence the stepped nature of the curve.

Figure 18: The daily set of seasonal premia. Each line represents one of 252 days, and it indicates the seasonal premia observed in the monthly forward prices, from one month to 13 months.

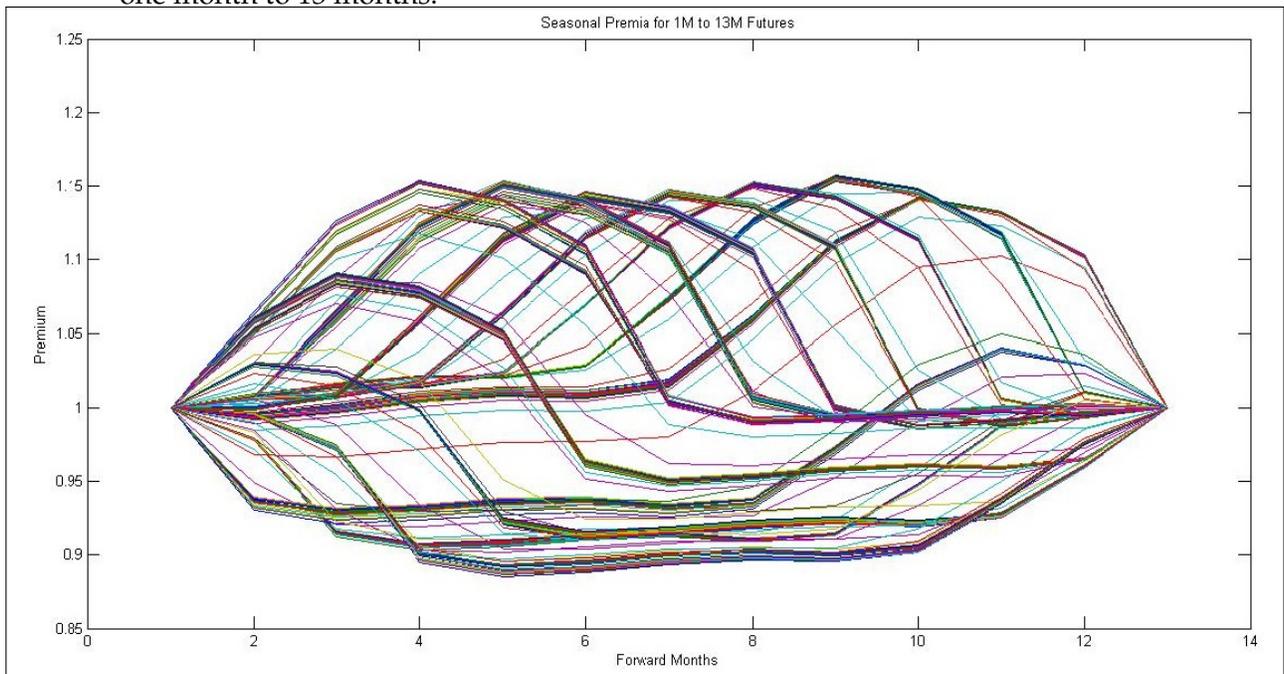


Figure 19: The 13 monthly premia that would be applied to each day. The x-axis shows the day number from one to 252.

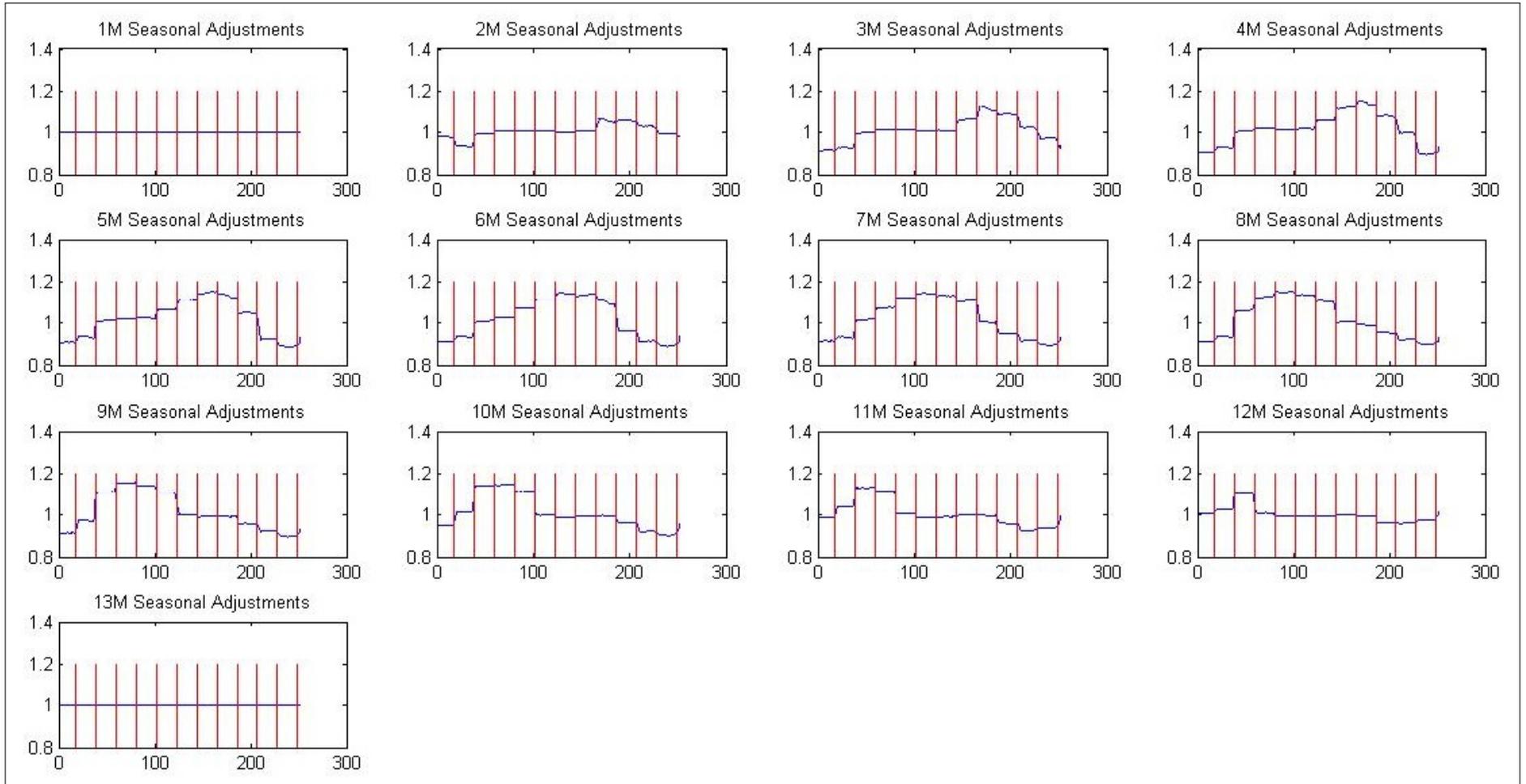
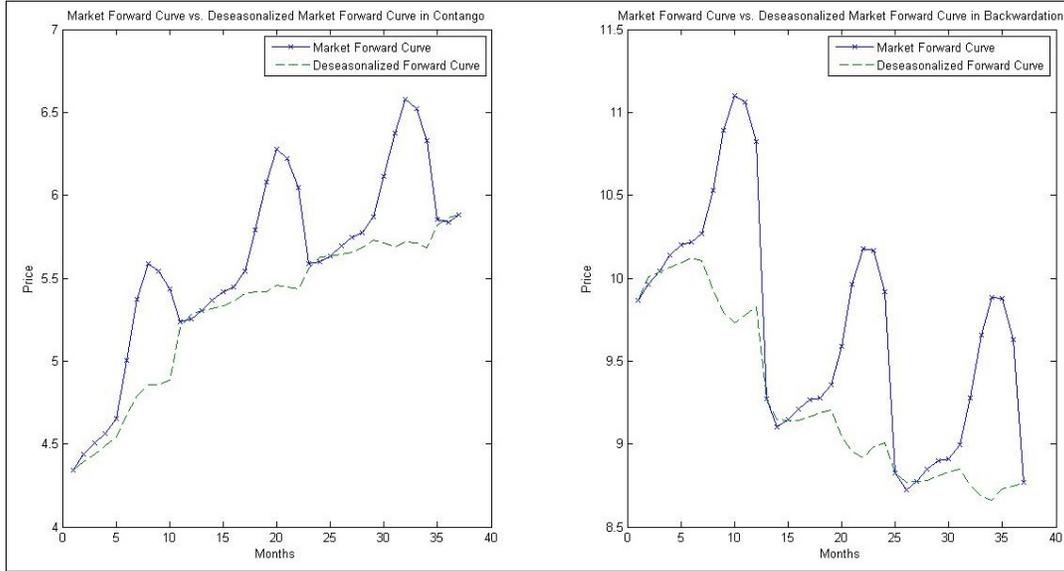


Figure 20: The market forward curves exhibiting seasonality, alongside the same curves with seasonality removed.



5.2.2 Application of Seasonal Premia

First, seasonality is removed from the market futures curves via the inverse of Equation (9),

$$X^{DS}(t_0, T) = (SP(t_0, m(T)))^{-1} X^S(t_0, T).$$

Figure 20 shows two examples (one in contango and one in backwardation) of a futures curve as seen in the market and the same futures curve stripped of seasonality through the application of the proposed method. It is observed that the overall contango and backwardation shape is maintained as well as a large portion of the seasonality has been removed.

Small traces of seasonal patterns still remain to a reasonable extent. These futures curves with seasonality removed were then run through the Kalman filter for calibration. Owing to time constraints, the process was not completed as rigorously as the parameter estimation for the Schwartz97 model. Moreover, the results were not investigated as thoroughly as desired, however, the provisional results of the run through the Kalman filter were favourable. Figure 21 indicates the deseasonalised natural gas futures curves passed through the Kalman filter, whereas Figure 22 shows the curves constructed from the resulting estimated parameters.

There is clear evidence that some seasonality has persisted, particularly in the 2008 period where prices were extremely volatile. The positive result is that the Kalman filter parameters appear to be able to capture the general shape of the nat-

Figure 21: The full range of deseasonalised forward curves that were used for the implementation of the Kalman Filter.

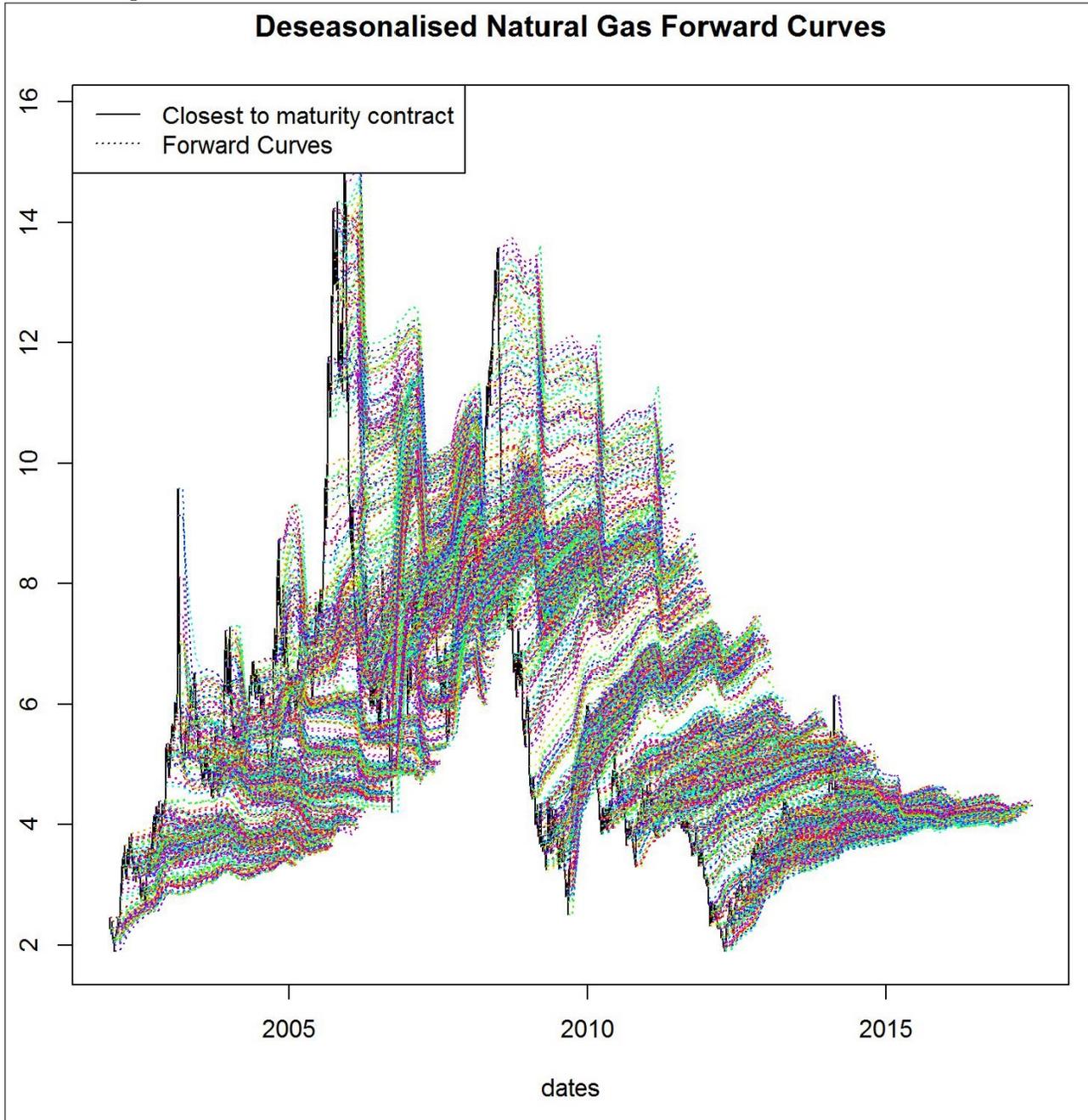


Figure 22: The corresponding forward curves that were computed using the parameters estimated from the Kalman Filter.

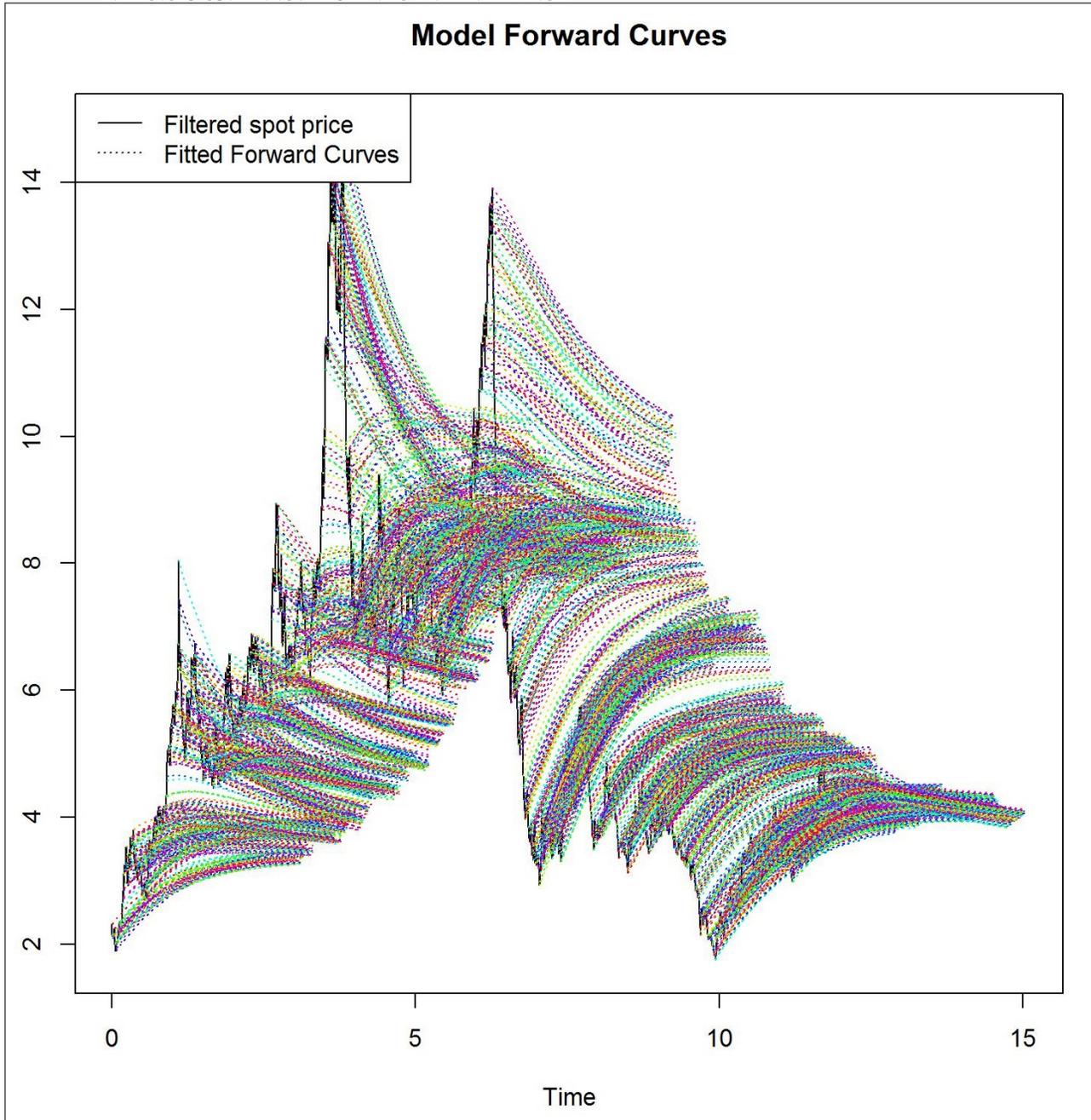
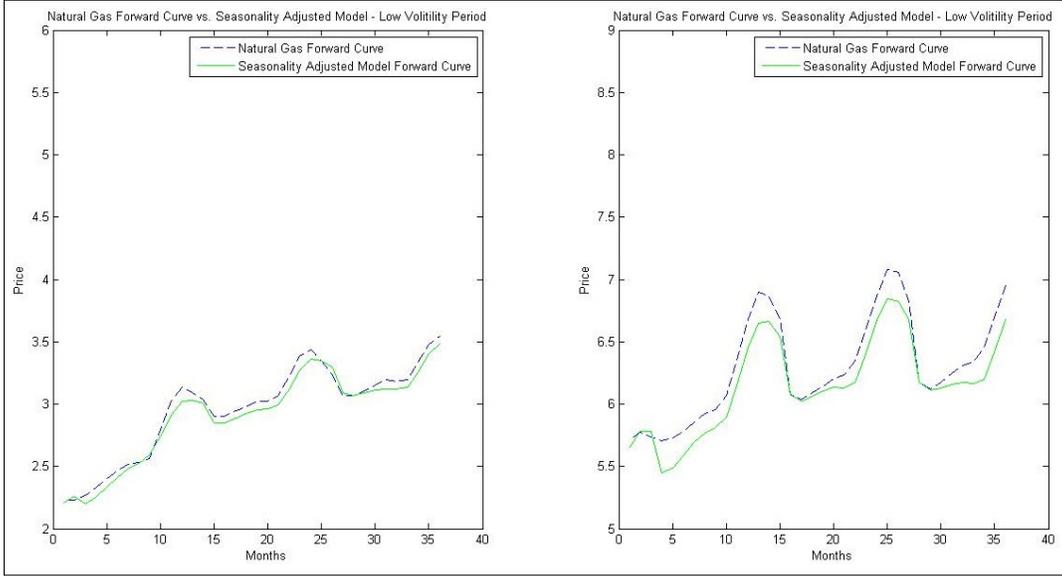


Figure 23: Two plots of the forward curves modelled by the Kalman Filter adjusted for seasonality, plotted against the market-observed forward curves at the same point during a relatively low volatility period.



atural gas futures curves, particularly in relatively stable periods, for example the past hundred trading days.

The futures curves plotted from the Kalman filters, as expected, do not have any seasonal component. This can now be added back, to finally replicate fully the market observable futures curves using the extracted seasonal premia,

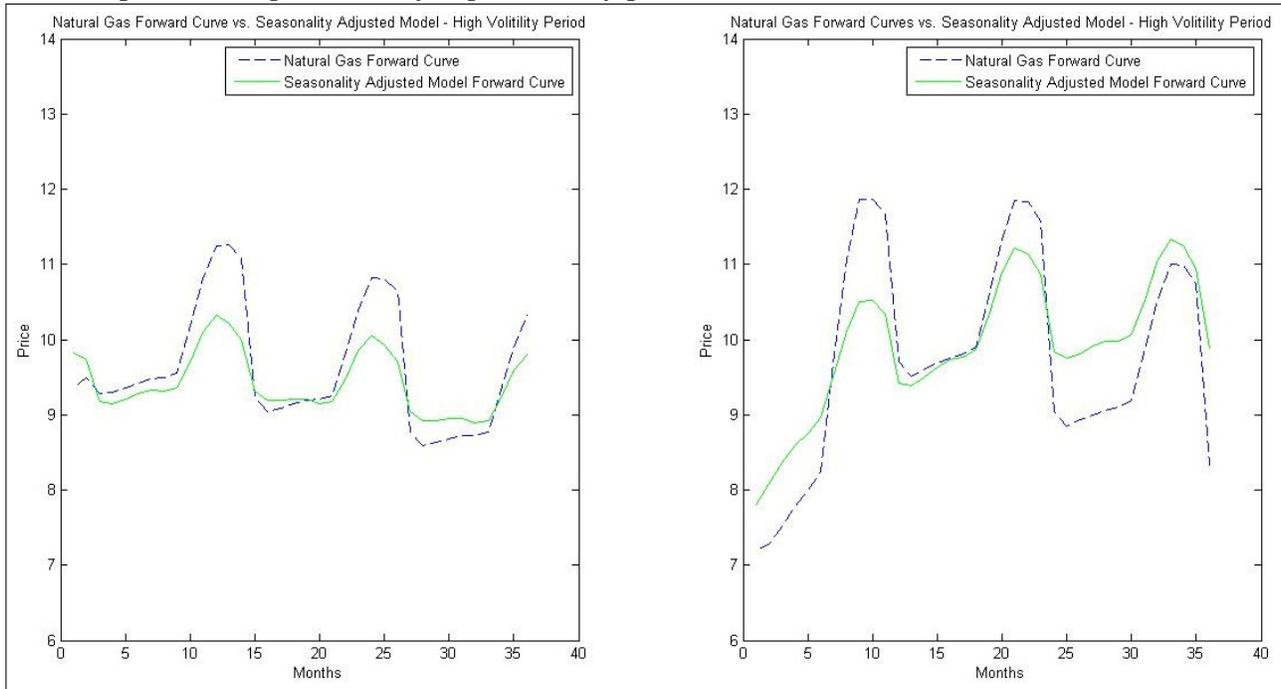
$$\hat{X}^S(t_0, T) = SP(t_0, m(T))\hat{X}^{DS}(t_0, T).$$

Figure 23 shows two cases of the futures curves generated by adding back seasonality against the corresponding original market observable futures curve during a period of relatively low volatility period.

Figure 24 is a similar plot, however taken over a period of relatively high volatility.

The result is an apparent ability to model the general seasonality. However it must be noted that the ability of the futures curves to accurately model the market-observed futures curves during extremely volatile periods (mainly during the market crisis of 2008) was particularly poor. However, for the remainder of the period from 1 January 2002 to 30 June 2014 the model performed adequately.

Figure 24: Two plots of the forward curves modelled by the Kalman Filter adjusted for seasonality, plotted against the market-observed forward curves at the same point during a relatively high volatility period.



6 Hedging

The final test of the effectiveness of the model is its ability to hedge the spread option under the real-world evolution of the futures curve.

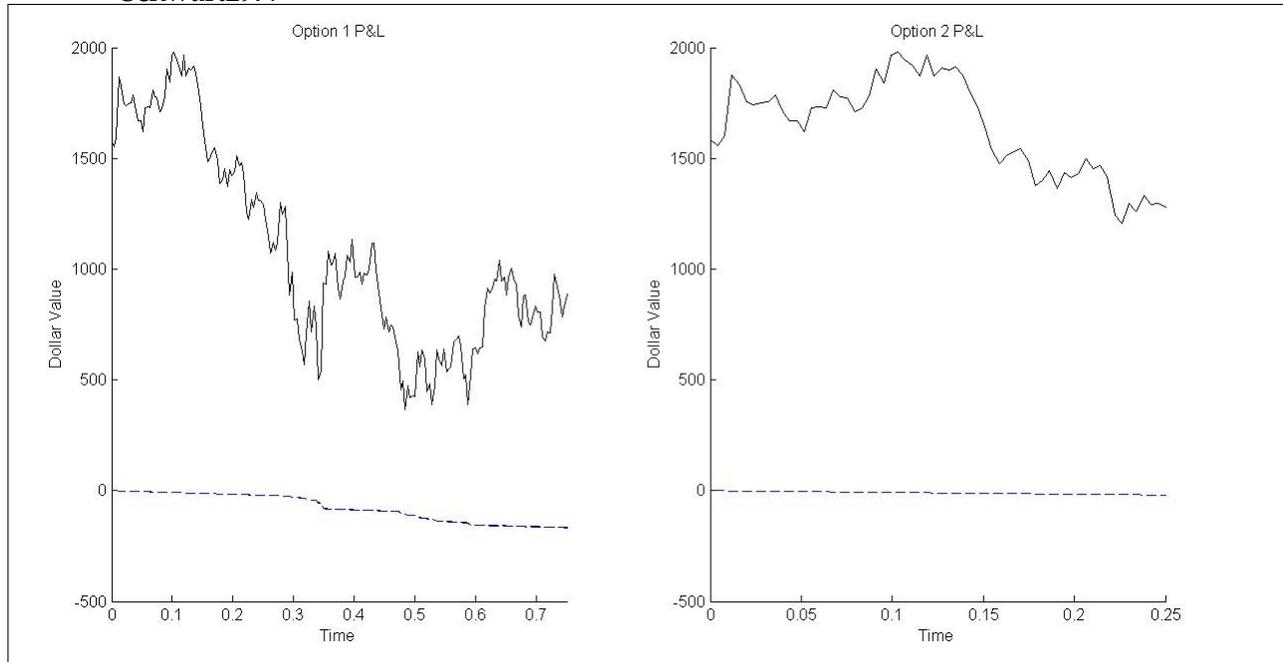
The full time period considered is 9-months of 21 trading days each starting on 11 December 2012 and finishing on 22 September 2013. The data was LME Copper Futures with one month maturities out to 48 months. A historical profit and loss simulation was implemented, pricing and hedging the relevant options during this period, using the model parameters estimated in section 4.

6.1 Historical Profit and Loss

6.1.1 The European Call

Two call options are considered. Option 1 is a 9-month European call on the 12-month future (3-month future at expiry). Option 2 is a 3-month European call on the 9-month future (6-month future at expiry). The option is struck at the money: the value of the LME 12-month and 9-month Copper futures, as on 11 December 2012, respectively.

Figure 25: Historical profit and loss for a call option on a commodity future under Schwartz97.



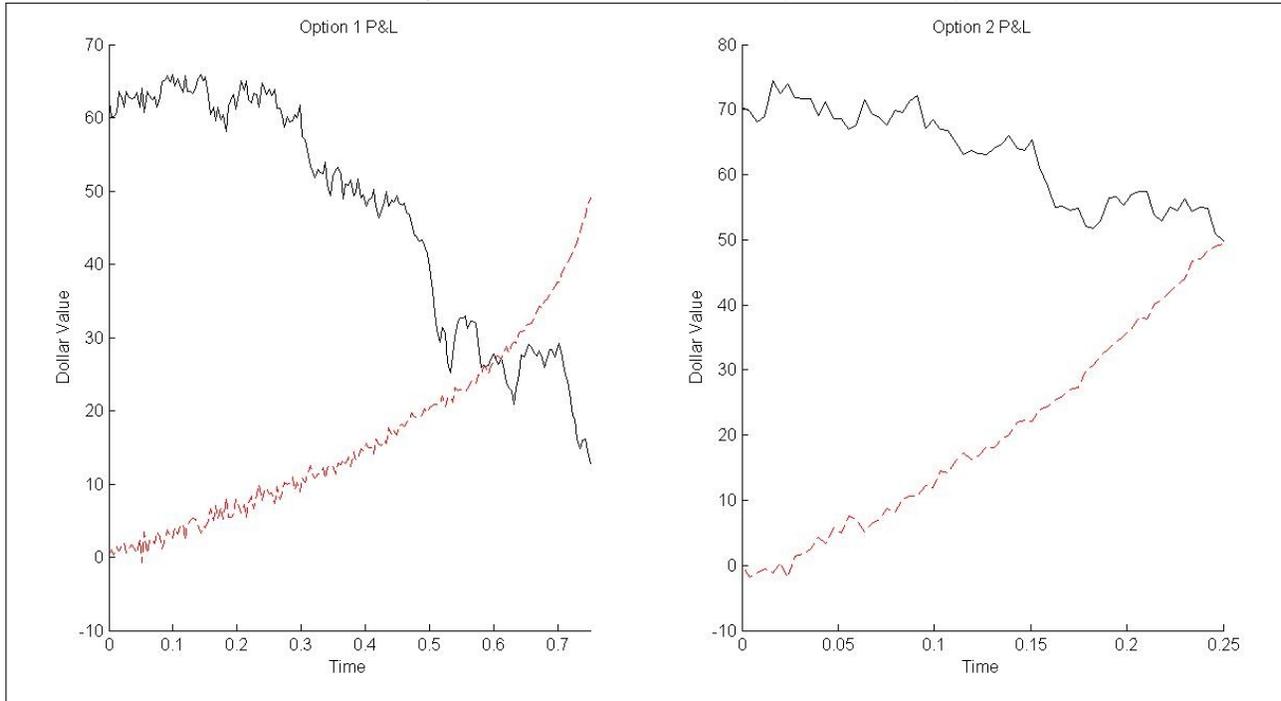
The value of the convenience yield is estimated each day using the Kalman filter. The value of the spot is then reverse-engineered from the current value of the underlying future using the analytical equation for the future (Equation (1)). This ensures that each day's value of S_{t_0} and δ_{t_0} is consistent with the current market price of the underlying future. The results for each option are displayed in Figure 25.

For Option 1, the hedge produces a final loss of capital of approximately 10 per cent of the initial option price. For Option 2, the hedge produces a final loss of approximately 1 per cent of the initial option premium. The results for Option 2 are encouraging, as this is well within expected bounds for daily hedge-slippage. This result would indicate that the model is capturing most of the features of the underlying future's dynamics. The larger loss for Option 1 is primarily owing to several large drops in the option price over short time periods, that occur after the expiry of Option 2. These drops could be considered to be jumps, which the Schwartz97 model fails to capture and thus fails to fully hedge.

6.1.2 The Spread Option

Two spread options are considered. Spread Option 1 is a 9-month European call on the difference between the 16-month and 12-month futures (7-months and 3-

Figure 26: Historical profit and loss for a spread option on commodity futures under Schwartz97, using the historical convenience yield volatility.



months respectively at maturity) and Spread Option 2 is a 3-month European call on the difference between the 18-month and 9-month futures (15-months and 6-months respectively at maturity.)

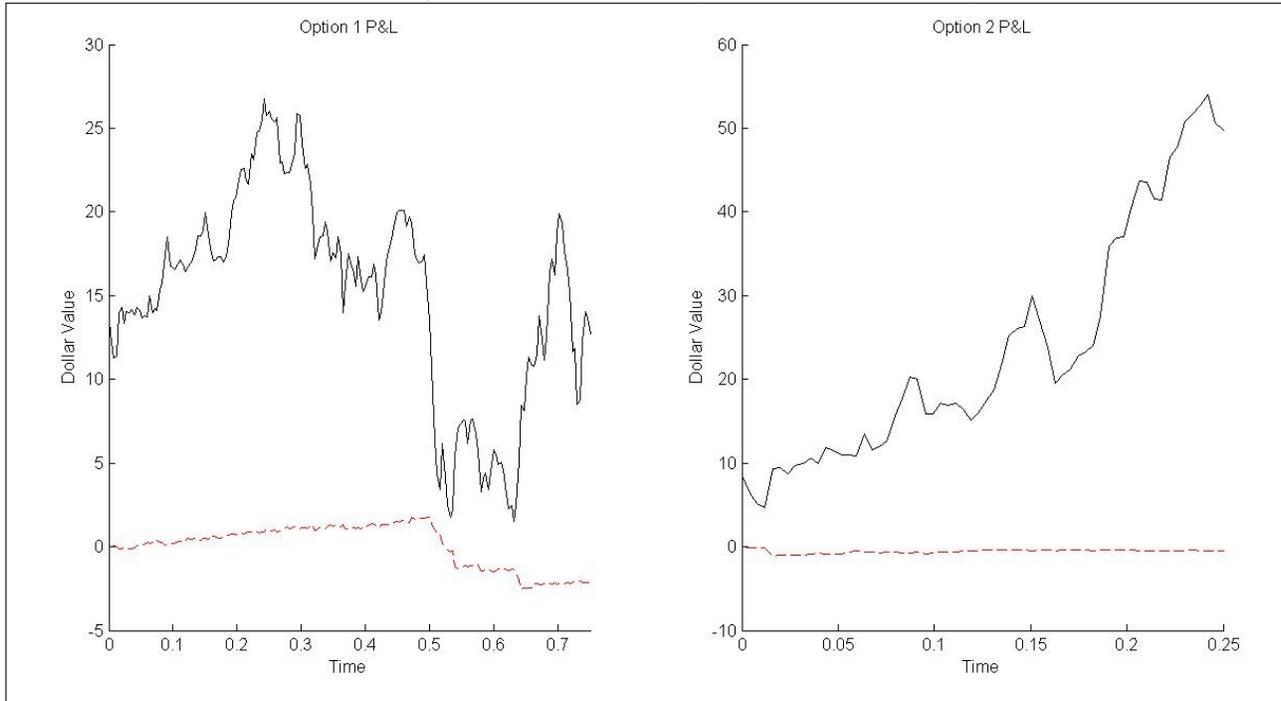
The value of δ_{t_0} and S_{t_0} are determined each day by solving Equation (1) to simultaneously match the market price of each underlying future.

Figure 26 displays the results under the historical parameter estimation. For Spread Option 1 the resultant profit is approximately 78 per cent of the initial option price. For Spread Option 2, the resultant profit is approximately 70 per cent of the initial option price.

The clear conclusion is that both options are being overpriced. After investigation, it was determined that the historical volatility of the convenience yield is approximately five times higher than the realised volatility over the 9-month period of Spread Option 1 and 10 times higher than the realised volatility over the lifetime of Spread Option 2. Whenever an option is sold (and hedged) with a much higher volatility than what is realised, the hedging portfolio will show a large profit.

As a check, the historical profit and loss simulation was re-run using the realised volatility of the convenience yield. The result is displayed in Figure 27. In this case, for Spread Option 1 the resultant portfolio lost 15 per cent of the initial option price and for Spread Option 2 the resultant portfolio lost 5 per cent of the

Figure 27: Historical profit and loss for a spread option on commodity futures under Schwartz97, using the realised volatility of the convenience yield.



initial premium. This is well within reason considering the previously generated profit and loss distribution (see Figure 10).

6.2 Conclusion

Although the call options appear to be hedged appropriately using historical parameter estimates for the volatility of the spot and convenience yield, it is clear from Figures 26 and 27 that a forward-looking volatility is needed to price and hedge the spread options.

This is not surprising, as option pricing should always occur with a view towards capturing the true volatility experienced by the option over its lifetime. Various possible solutions exist to obtaining a forward-looking volatility for the spot and convenience yield, but unfortunately owing to time constraints these solutions could not be thoroughly explored.

In conclusion, the implemented framework provides suitable parameter estimates for those for which forward-looking estimates would be insensible (or impossible). Barring the need for obtaining forward-looking volatilities, it is possible under the implemented framework to price and hedge commodity spread options under the real-world evolution of the futures curve.

Bibliography

- [AV07] Carol Alexander and Aanand Venkatramanan, *Analytic approximations for spread options*, ICMA Centre Discussion Papers in Finance DP2007-11, University of Reading, UK (2007).
- [BG06] Svetlana Borovkova and Helyette Geman, *Seasonal and stochastic effects in commodity forward curves*, Review of Derivatives Research **9** (2006), no. 2, 167–186.
- [BLO97] John Barkoulas, Walter C Labys, and Joseph Onochie, *Fractional dynamics in international commodity prices*, Journal of Futures Markets **17** (1997), no. 2, 161–189.
- [BP13] Janis Back and Marcel Prokopczuk, *Commodity price dynamics and derivative valuation: A review*, International Journal of Theoretical and Applied Finance **16** (2013), no. 06.
- [Bre91] Michael J Brennan, *The price of convenience and the valuation of commodity contingent claims*, Stochastic models and option values **200** (1991), 33–71.
- [BS85] Michael J Brennan and Eduardo S Schwartz, *Evaluating natural resource investments*, Journal of business (1985), 135–157.
- [Erb14] Otziger Erb, Luthi, *Schwartz 1997 two-factor model technical document*, Tech. report, R Foundation for Statistical Computing, 2014.
- [Gab91] Jacques Gabillon, *The term structures of oil futures prices*, Oxford Institute for Energy Studies. Working paper (1991).
- [Gem05] Helyette Geman, *Commodities and commodity derivatives*, Wiley Finance, 2005.
- [GS90] Rajna Gibson and Eduardo S Schwartz, *Stochastic convenience yield and the pricing of oil contingent claims*, The Journal of Finance **45** (1990), no. 3, 959–976.

- [HR98] Jimmy E Hilliard and Jorge Reis, *Valuation of commodity futures and options under stochastic convenience yields, interest rates, and jump diffusions in the spot*, *Journal of Financial and Quantitative Analysis* **33** (1998), no. 01, 61–86.
- [Jav03] Galli Javaheri, Lautier, *Filtering in finance*, *Wilmott Magazine* **5** (2003), 67–83.
- [Kal60] RE Kalman, *A new approach to linear filtering and prediction problems*, *Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.
- [Key30] J.M. Keynes, *A treatise on money vol 2: The applied theory of money*, Macmillan, 1930.
- [Lau03] Delphine Lautier, *Term structure models of commodity prices*, *Cahier de recherche du cereg* (2003), no. 2003-9.
- [Lau05] Lautier, *Term structure models of commodity prices: a review*, *Journal of Alternative Investments* **8** (2005), 42–64.
- [Lju79] Ljung, *Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems*, *IEEE Transactions on Automatic Control* **24** (1979), 35–60.
- [Mil03] Kristian R Miltersen, *Commodity price modelling that matches current observables: A new approach*.
- [MS98] Kristian R Miltersen and Eduardo S Schwartz, *Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates*, *Journal of Financial and Quantitative Analysis* **33** (1998), no. 01, 33–59.
- [NS04] Martin J Nielsen and Eduardo S Schwartz, *Theory of storage and the pricing of commodity claims*, *Review of Derivatives Research* **7** (2004), no. 1, 5–24.
- [Pin01] Robert S Pindyck, *The dynamics of commodity spot and futures markets: a primer*, *The Energy Journal* (2001), 1–29.
- [Pir11] Craig Pirrong, *Commodity price dynamics: A structural approach*, Cambridge University Press, 2011.
- [RS02] Martin Richter and Carsten Sørensen, *Stochastic volatility and seasonality in commodity futures and options: The case of soybeans*, *Journal of Futures Markets* (2002).
- [RSS00] Bryan R Routledge, Duane J Seppi, and Chester S Spatt, *Equilibrium forward curves for commodities*, *The Journal of Finance* **55** (2000), no. 3, 1297–1338.

- [Sam65] Paul A Samuelson, *Proof that properly anticipated prices fluctuate randomly*, *Industrial management review* **6** (1965), no. 2, 41–49.
- [Sch97] E Schwartz, *The stochastic behaviour of commodity prices: implications for valuation and hedging*, *Journal of Finance* **52** (1997), 923–973.
- [WB95] Greg Welch and Gary Bishop, *An Introduction to the Kalman Filter*, ACM, 1995.
- [Ziv06] Zivot, *State space models and the kalman filter*, Tech. report, University of Washington, 2006.

BA Multi-Curve Models with Counterparty Risk

Team 2

Y. JIANG, University College London

J. T. J. MBONGO NKOUNGA, AIMS - Cape Town

T. M. NGUYEN, Université d'Evry

Supervisor:

S. CRÉPEY, Université d'Evry

Contents

1	Introduction	3
2	From Pricing Kernels to bA Models	4
2.1	Bottom-up Risk-Neutral Construction of bA Models	4
3	Swaptions Pricing and Calibration	6
3.1	Pricing	6
3.2	Lognormal Calibration	7
3.3	exp-NIG Calibration	8
4	Counterparty Risk	9
4.1	Exposures	11
4.2	Credit Model	13
4.3	CVA	15
4.4	LVA	16

1 Introduction

The credit crisis and the ongoing European sovereign debt crisis have highlighted the native form of credit risk, namely counterparty risk, valued as the so-called Credit Valuation Adjustment (CVA). Moreover, as banks themselves have become risky, counterparty risk must be understood in a bilateral perspective (CVA and Debt Valuation Adjustment DVA), where the counterparty risk of the two parties are jointly accounted for in the modelling. In this context the classical assumption of a locally risk-free asset, which is used for financing purposes (lending and borrowing as needed), is no longer sustainable. This shortcoming raises a related issue: proper accounting of the funding costs of a position (Funding Valuation Adjustment FVA). See Brigo, Morini, and Pallavicini (2013) and Crépey, Bielecki and Brigo (2014), respectively for a more financial and mathematical perspective, and the latter for recent counterparty risk references in book form. In addition, since August 2007, the emergence of systemic counterparty risk has also been recorded. This type of risk refers to various significant spreads between quantities that were very similar before, e.g. between OIS rates and LIBOR swap rates of different tenors. For a variety of so called multi-curve models, see, among others, Kijima, Tanaka, and Wong (2009), Kenyon (2010), Henrard (2007, 2010), Bianchetti (2010), Mercurio (2010b, 2010a, 2010c), Fujii, Shimada, and Takahashi (2011, 2010), Moreni and Pallavicini (2014), Bianchetti and Morini (2013) and Crépey, Grbac, Ngor and Skovmand (2013). Through its relation with the concept of discounting, this systemic component of counterparty risk has impacted all derivatives markets.

Given its dual regulatory and accounting implications, counterparty risk must be both “measured” under the statistical probability \mathbb{P} and “priced” under a pricing measure \mathbb{Q} . Moreover, due to netting clauses (for the good sake of counterparty credit risk mitigation at the portfolio level), counterparty risk requires consistent modelling across different asset classes. Because of the optional feature of counterparty risk (that only bears on the in-the-money side of the position), the modelling also needs to be dynamic, for which a parsimonious Markov structure is important. In view of these elements, an approach that looks particularly appropriate is based on the modelling of the pricing kernel (see e.g. Rogers (1997), Cochrane (2005), Flesaker and Hughston (1996), Akahori, Hishida, Teichmann and Tsuchiya (2014), Akahori and Macrina (2012), Macrina (2014), Brody, Hughston and Macrina (2008)). Accordingly, this project explores the possible use of such an approach under the form of the so called bA models for dealing with interest rate derivatives in post-crisis markets, including multi-curve and counterparty risk features. We note that other recent uses of related models include Cuchiero, Keller-Ressel and Teichmann (2012), for moment computations in financial applications, and Cheng

and Tehranchi (2014), motivated by stochastic volatility modelling.

Our report is organised as follows. Section 2 situates the bA models in the pricing kernels universe. Section 3 deals with the pricing and calibration of swaptions. Section 4 discusses counterparty risk.

NB: The mathematical parts of this report rely mainly on Crépey et al. (2014) that had been provided to our team as pre-challenge package. The *main contribution* of this work are the numerical sections 3 and 4. All the computations were done in Matlab and we were given access to the code of the related paper Crépey, Grbac, Ngor and Skovmand (2013). Otherwise, such an implementation and experimentation work, by us in six days, would not have been possible.

2 From Pricing Kernels to bA Models

We model a financial market by a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{0 \leq t \leq U})$ where U is a finite time horizon, \mathbb{P} denotes the real probability measure, and $\{\mathcal{F}_t\}_{0 \leq t \leq U}$ is the market filtration. By no arbitrage principle, the price process $\{S_{tT}\}_{0 \leq t \leq T < U}$ of an asset paying dividend stream $\{D_t\}_{0 \leq t \leq T < U}$ up until a maturity T satisfies

$$S_{tT} = \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}} \left[\pi_T S_{TT} + \int_t^T \pi_u D_u du \mid \mathcal{F}_t \right],$$

where $\{\pi_t\}_{0 \leq t \leq U}$ is the pricing kernel that needs to be modelled. In Macrina (2014), the following heat pricing kernel is introduced

$$\pi_t = f_0(t) + f_1(t) \int_0^{U-t} \mathbb{E}[F(t+u, X_{t+u}) \mid X_t] w(t, u) du,$$

where $(X_t)_{0 \leq t \leq U}$ is a Markov process generating the market filtration $\{\mathcal{F}_t\}_{0 \leq t \leq U}$. Once the pricing kernel is specified, the OIS discount bond price $\{P_{tT}\}_{0 \leq t \leq T \leq U}$ is determined by the formula

$$P_{tT} = \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}} [\pi_T \mid \mathcal{F}_t].$$

2.1 Bottom-up Risk-Neutral Construction of bA Models

Under suitable specification of the Markov process (X_t) and the functions $F(t, x)$ and $w(t, u)$, the price processes of the OIS discount bond and of a (non-dividend paying) asset S are shown to be of the following “bA”-form:

$$S_{tT} = \frac{S_{0T} + b_2(T)A_t^{(2)}}{P_{0t} + b_1(t)A_t^{(1)}}, \quad P_{tT} = \frac{P_{0T} + b_1(T)A_t^{(1)}}{P_{0t} + b_1(t)A_t^{(1)}},$$

where the $b_i(t)$ are nonnegative nonincreasing functions and where the $A_t^{(i)} = A_i(t, X_t^{(i)})$ are martingales, not necessarily under \mathbb{P} , but rather under an equivalent auxiliary measure \mathbb{M}_t , where $X_t = (X_t^{(1)}, X_t^{(2)})$ is the market factor. In this section we present a self-contained bottom up, risk-neutral construction of bA models. In this perspective, we model the short risk-free rate r_t and the corresponding risk-neutral discount factor D_t as

$$r_t = -\frac{\dot{P}_{0t} + \dot{b}_1(t)A_t^{(1)}}{P_{0t} + b_1(t)A_t^{(1)}}, \quad D_t = e^{-\int_0^t r_s ds}, \quad (2.1)$$

where $A^{(1)}$ is an \mathcal{F}_t -martingale starting from 0 under a probability measure \mathbb{M} , related to a risk-neutral probability \mathbb{Q} via

$$\frac{d\mathbb{Q}}{d\mathbb{M}}|_{\mathcal{F}_t} = \nu_t^{(1)}, \quad 0 \leq t \leq T, \quad \text{where } \nu_t^{(1)} = \mathcal{E} \left(\int_0^t \frac{b_1(s) dA_s^{(1)}}{c_1(s) + b_1(s)A_s^{(1)}} \right).$$

As established in Crépey et al. (2014, Lemma 2.2), the \mathbb{M} supermartingale $h = D\nu^{(1)}$ satisfies

$$h_t = P_{0t} + b_1(t)A_t^{(1)}, \quad t \geq 0. \quad (2.2)$$

The OIS bond price process satisfies

$$P_{tT} = \frac{1}{D_t} \mathbb{E}^{\mathbb{Q}}[D_T | \mathcal{F}_t] = \frac{1}{D_t \nu_t^{(1)}} \mathbb{E}^{\mathbb{M}}[D_T \nu_T^{(1)} | \mathcal{F}_t] = \frac{1}{h_t} \mathbb{E}^{\mathbb{M}}[h_T | \mathcal{F}_t] = \frac{P_{0T} + b_1(T)A_t^{(1)}}{P_{0t} + b_1(t)A_t^{(1)}}. \quad (2.3)$$

Hence, the process h plays the role of the pricing kernel of the OIS market under the measure \mathbb{M} . Similarly, the Libor \mathbb{M} kernel $L_{S,T}$ can be modelled in terms of another \mathbb{M} -martingale $A^{(2)}$ as

$$L_{S,T} = \frac{L(0; S, T) + b_2(S, T)A_S^{(2)}}{P_{0S} + b_1(S)A_S^{(1)}}, \quad (2.4)$$

which results in a (forward) Libor process

$$L(t; T_{i-1}, T_i) := \frac{1}{h_t} \mathbb{E}_t^{\mathbb{M}}[h_{T_i} L_{T_{i-1}, T_i}]$$

equal, for $t \leq T_{i-1}$, to

$$L(t; T_{i-1}, T_i) = \mathbb{E}_t^{\mathbb{M}} \left[\frac{\mathbb{E}_{T_{i-1}}^{\mathbb{M}}[h_{T_i}] L_{T_{i-1}, T_i}}{h_t} \right] = \frac{L(0; T_{i-1}, T_i) + b_2(T_{i-1}, T_i)A_t^{(2)}}{P_{0t} + b_1(t)A_t^{(2)}}. \quad (2.5)$$

The above results in a HJM setup where, in the spirit of Heath, Jarrow and Morton Heath, Jarrow and Morton (1992), the initial term structures P_{0T_i} and $L(0; T_{i-1}, T_i)$ are fitted by construction.

3 Swaptions Pricing and Calibration

The next questions are the pricing of and calibration to Libor derivatives, especially swaptions that are quite liquidly traded on the market.

3.1 Pricing

First, an interest rate swap (see e.g. Brigo and Mercurio (2006)) is an agreement between two counterparties, where one stream of future interest payments is exchanged for another based on a specified nominal amount N . A popular case of interest rate swaps is the exchange of a fixed rate (contractual swap spread) against the LIBOR rate at the end of successive time intervals (T_{i-1}, T_i) of length δ . Such a swap can also be viewed as a collection of n forward rate agreements. The swap price Sw_t at time $t \leq T_0$ is given by the following model-free formula:

$$Sw_t = \sum_{i=1}^n N\delta[KP_{tT_i} - L(t; T_{i-1}, T_i)].$$

Next, a swaption is an option between two parties to enter the above swap at the expiry T_k (date of maturity of the option). Its price at time $t \leq T_k$ is given by the following \mathbb{M} formula:

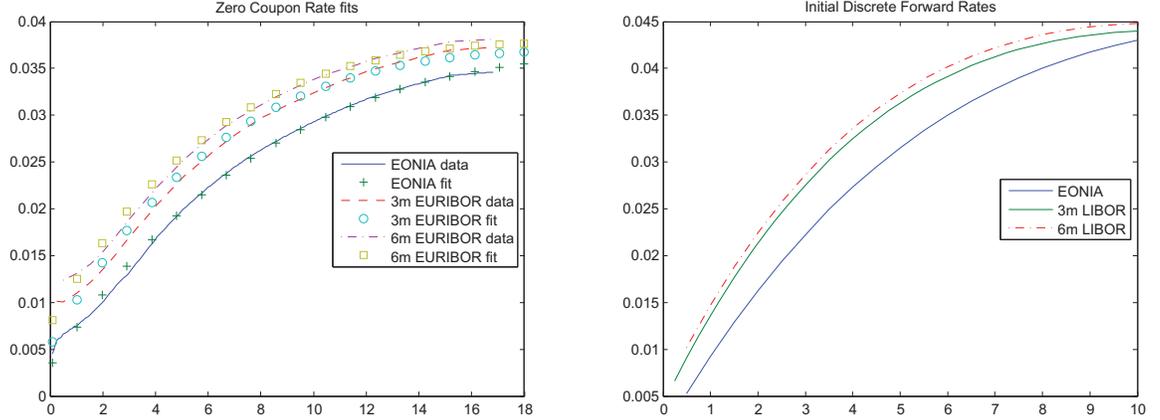
$$\begin{aligned} Sw_{n_t T_k} &= \frac{1}{h_t} \mathbb{E}^{\mathbb{M}}[h_{T_k}(Sw_{T_k})^+ | \mathcal{F}_t] \\ &= \frac{1}{h_t} \mathbb{E}^{\mathbb{M}} \left[h_{T_k} \left(\sum_{i=K+1}^n N\delta[KP_{T_k T_i} - L(T_k; T_{i-1}, T_i)] \right)^+ \middle| \mathcal{F}_t \right], \end{aligned}$$

where $P_{T_k T_i}$ and $L(T_k; T_{i-1}, T_i)$ are determined by (2.3) and (2.5). Based on this formula, swaption prices at time $t = 0$ can be efficiently priced by various numerical schemes, including (depending on the specification of the driving martingales): Fourier schemes based on Eberlein and Raible (1999, Theorem 2.2), Hurd and Zhou (2010, Theorem 1) or Caldana and Fusai (2013, Proposition 1) (possibly also involving the linear boundary approximation of Singleton and Umantsev (2002)) or Monte Carlo schemes, as well as the Jamshidian schemes (see Jamshidian (1996)) studied by Team 3 in their project. In our computations we used Fourier and Monte Carlo schemes.

Calibration

As done by Crépey, Grbac, Ngor and Skovmand (2013), we used the following EUR market Bloomberg data of January 4, 2011 to calibrate our model: Eonia, 3m Euribor and 6m Euribor initial term structures on the one hand (fitted by construction

Figure 1: Initial term structures. *Left*: Zero coupon rates. *Right*: Discrete forward rates.



in our HJM setup, see Fig. 1), 3m and 6m tenor Libor swaptions on the other hand. Since there are no liquid OIS derivative data available in the market today, we simply used deterministic OIS rates r_t , i.e. $A^{(1)} = 0$ (hence, b_1 plays no role either). The calibration of the residual model ingredients b_2 and $A^{(2)}$ is divided into two phases.

1. In the first phase, we calibrate the non-maturity/tenor dependent parameters (parameters of the driving martingales $A^{(1)}, A^{(2)}$) to the smile of the $9y \times 1y$ swaption with (most liquid) tenor $\delta = 3m$. The market smile corresponds to a vector of strikes $[-200, -100, -50, -25, 0, 25, 50, 100, 200]$ bps around the underlying swap spread. This phase also gives us the values of $b_2(9, 9.25)$, $b_2(9.25, 9.5)$, $b_2(9.5, 9.75)$ and $b_2(9.75, 10)$, which we assume to be equal for the calibration efficiency.
2. In the second phase, in order to calibrate the remaining values of b_2 , we use at-the-money swaptions data with tenor $\delta = 3m$ and $6m$, termination $T_n = 10$ years and expiries T_k ranging from 1 to 9 years.

3.2 Lognormal Calibration

As a first example of the market factor, we assume that $A^{(2)}$ has the form

$$A_t^{(2)} = \exp\left(a_2 X_t^{(2)} - \frac{1}{2} a_2^2 t\right), i = 1, 2,$$

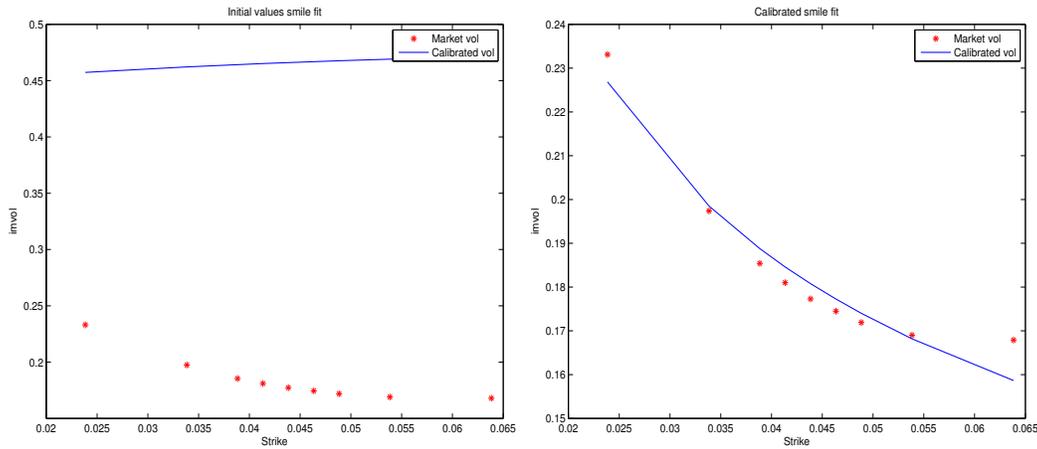
where $(X_t^{(2)})$ is a standard Brownian motion. Hence, in the first phase, we calibrate the parameter a_2 and $b = b_2(9, 9.25) = b_2(9.25, 9.5) = b_2(9.5, 9.75) = b_2(9.75, 10)$. Starting from the initial values

$$a_2 = 0.5, b_2 = 0.03,$$

a nonlinear least squares optimisation¹ of the distance between the swaption market and model implied smile resulted in the following parameters (see Fig. 2-4):

$$a_2 = 0.0544, b_2 = 0.1083.$$

Figure 2: Lognormal calibration. Initial versus calibrated fit to market data



3.3 exp-NIG Calibration

With the above Brownian driven market factor $A^{(2)}$, we were able to fit the level of the volatility smile but the model implied skew is quite different from the market skew. To overcome this, we now use a richer family of Lévy processes, namely normal inverse Gaussian (NIG) processes (see e.g. Cont and Tankov (2003)). The parameters that need to be calibrated at the first phase are ν, θ, σ and $b = b_2(9, 9.25) = b_2(9.25, 9.5) = b_2(9.5, 9.75) = b_2(9.75, 10)$. The initial values are

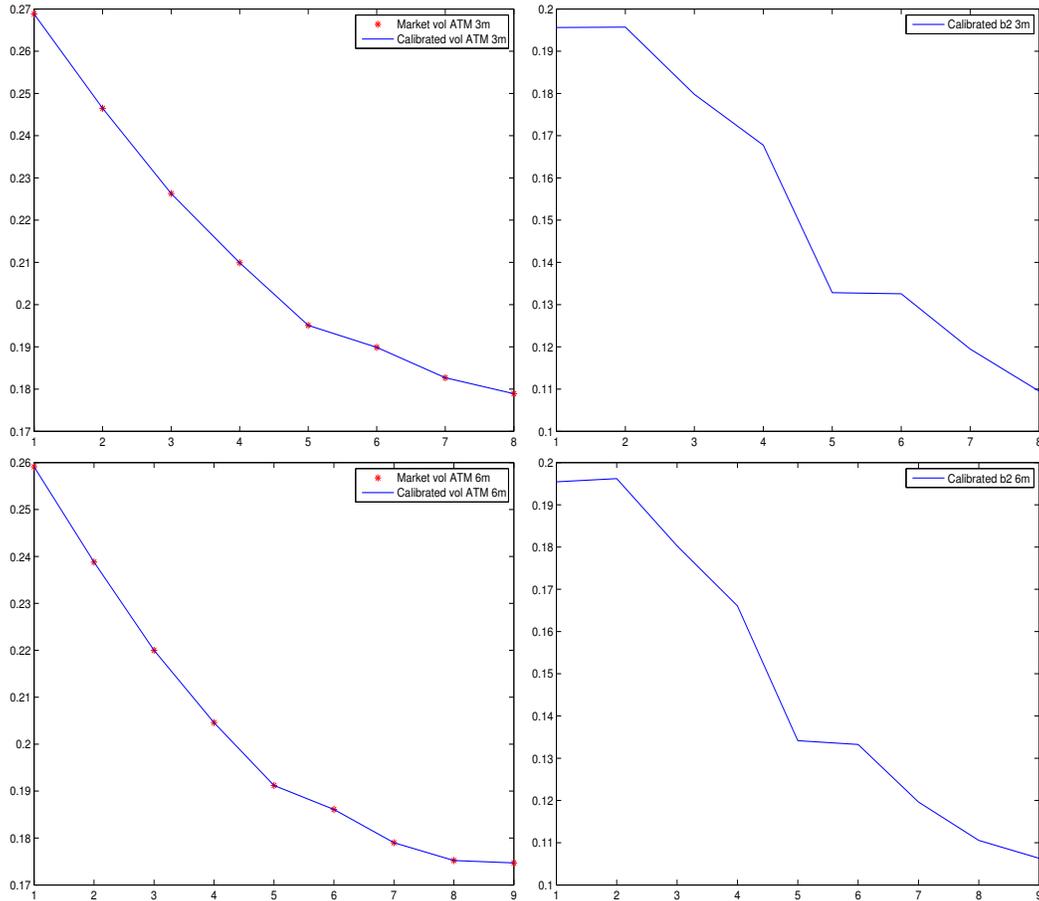
$$b_2 = 0.03, \nu = -0.05, \theta = -1.1, \sigma = 0.8.$$

After the optimisation, we obtain (see Fig. 5-7):

$$b_2 = 0.0426, \nu = 0.2491, \theta = -0.0244, \sigma = 0.1603.$$

¹Using the Matlab function "lsqnonlin".

Figure 3: Lognormal calibration. (Left) Fit to ATM swaption implied volatility term structures. (Right) Calibrated values of the b_2 parameters. (Top) $\delta = 3m$. (Bottom) $\delta = 6m$.



4 Counterparty Risk

So far we focused on so called clean computations, ignoring counterparty risk and funding costs in excess over the risk-free rate. In reality, one or both counterparties in the contract may default, so different adjustments need to be computed to account for counterparty risk and the related excess funding costs. In this section this is illustrated in the case of a basis swap, a typical multicurve product consisting in the exchange of two streams of floating payment based on a nominal N , or more generally, one counterparty pays a floating leg against another floating leg plus a fixed leg. Note that in the classical one-curve setup the time- t value of such a swap is zero. Since the crisis, markets quote positive basis swap spreads that have to be

Figure 4: Lognormal calibration. Histograms based on 10^5 scenarios of the calibrated model implied $L_{9y}(9y, 9.25y)$ (yellow), $L_{9y}(9.25y, 9.5y)$ (blue), $L_{9y}(9.5y, 9.75y)$ (red) and $L_{9y}(9.75y, 10y)$ (turquoise).

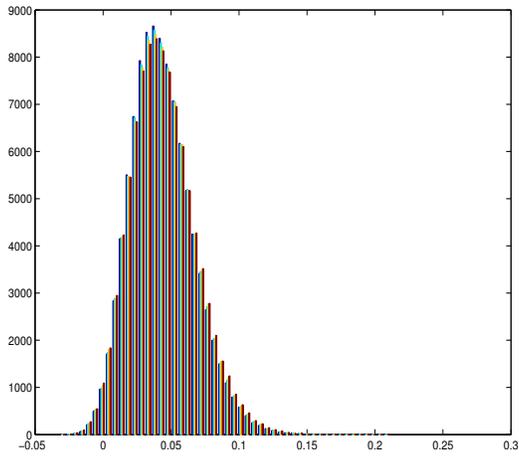


Figure 5: exp-NIG calibration. Initial versus calibrated fit to market data

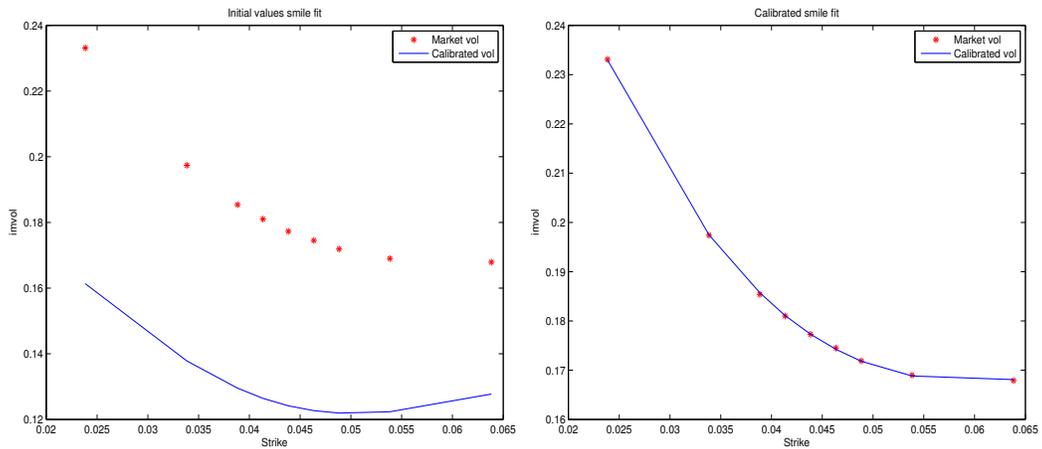
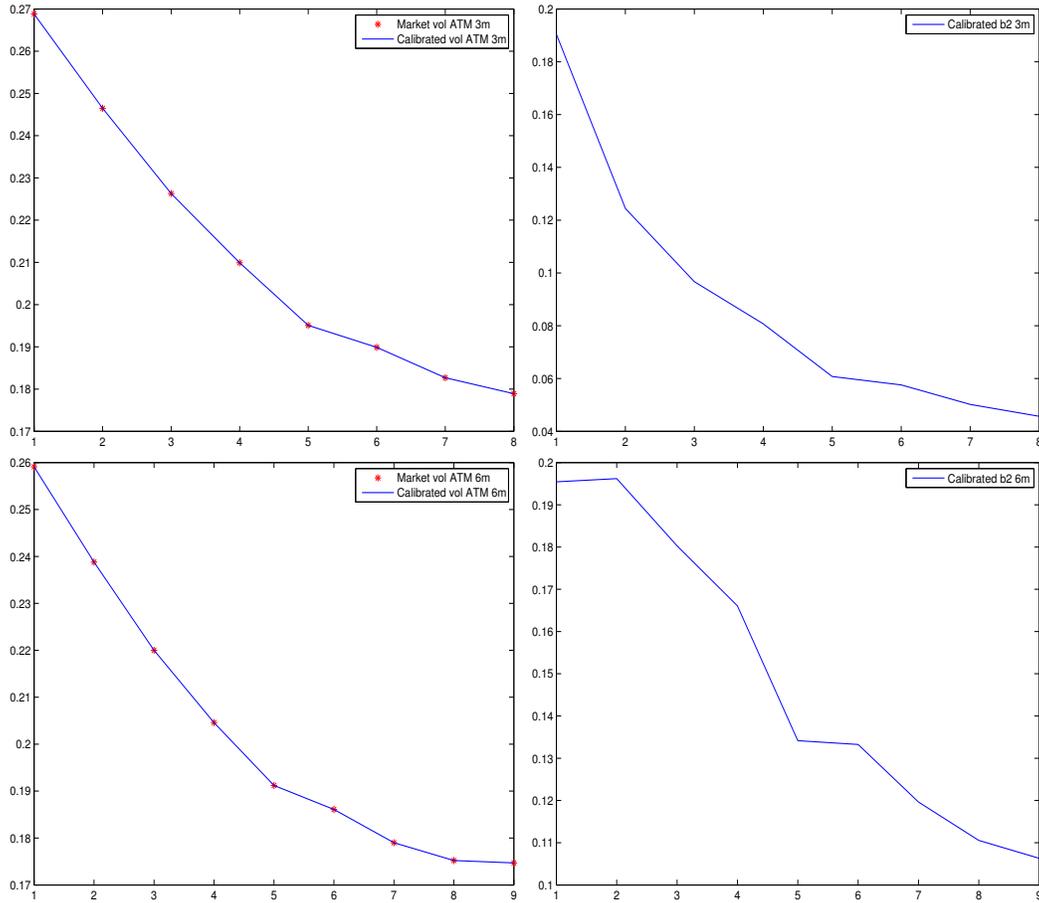


Figure 6: exp-NIG calibration. (Left) Fit to ATM swaption implied volatility term structures. (Right) Calibrated values of the b_2 parameters. (Top) $\delta = 3m$. (Bottom) $\delta = 6m$.

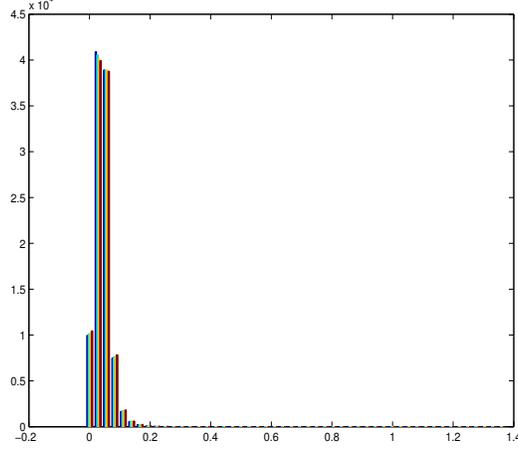


added to the smaller tenor leg.

4.1 Exposures

First, we use the exp-NIG calibrated parameters to study the counterparty risk exposure (price process with mean and quantiles) of a basis swap. We consider a 10-year length basis swap exchanging LIBOR $6m$ tenor payments against LIBOR $3m$ tenor payments plus a fixed spread. The two streams start and end at the same points $T_0 = T_0^1 = T_0^2$, $T = T_{n_1}^1 = T_{n_2}^2$. The time- t value of the basis swap with

Figure 7: exp-NIG calibration. Histograms based on 10^5 scenarios of the calibrated model implied $L_{9y}(9y, 9.25y)$ (yellow), $L_{9y}(9.25y, 9.5y)$ (blue), $L_{9y}(9.5y, 9.75y)$ (red) and $L_{9y}(9.75y, 10y)$ (turquoise).



spread K is given, for $t \leq T_0$, by

$$BS_t = N \left(\sum_{i=1}^{n_1} \delta_i^{6m} L(t; T_{i-1}^1, T_i^1) - \sum_{j=1}^{n_2} \delta_j^{3m} (L(t; T_{j-1}^2, T_j^2) + KB_t(T_j^2)) \right).$$

The time- t value after the initiation, i.e. for $T_0 \leq t < T$, is given by

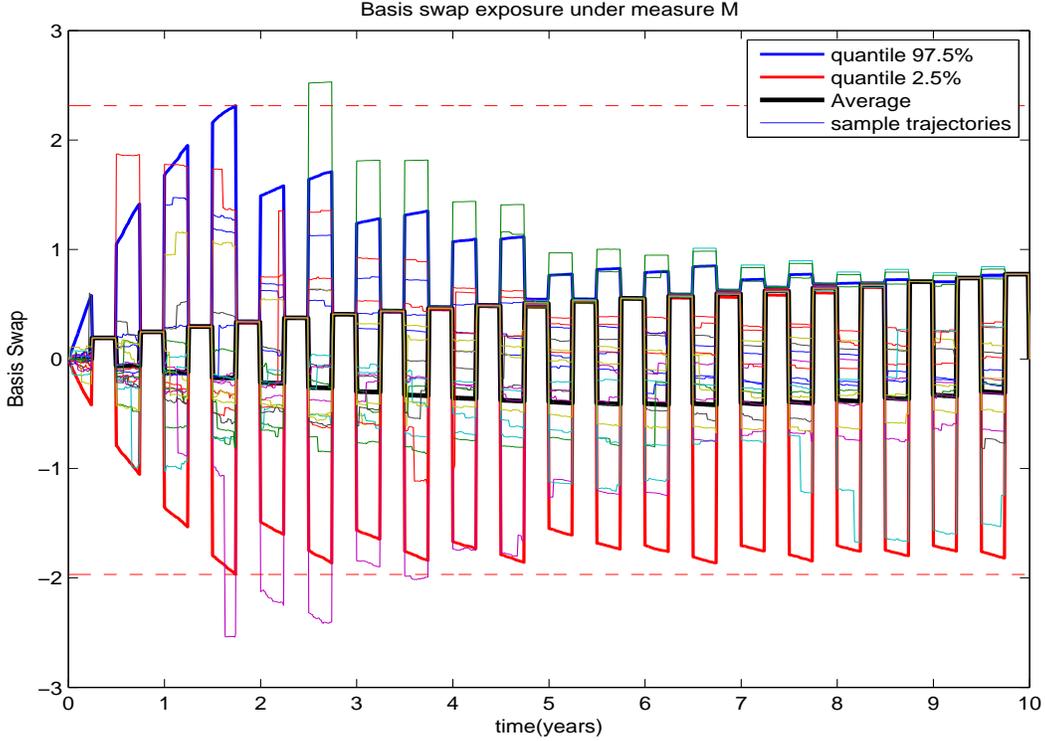
$$BS_t = N \left(\delta_{i_t}^{6m} L(T_{i_t-1}^1; T_{i_t-1}^1, T_{i_t}^1) + \sum_{i=i_t+1}^{n_1} \delta_i^{6m} L(t; T_{i-1}^1, T_i^1) \right. \\ \left. - \delta_{j_t}^{3m} (L(T_{j_t-1}^2; T_{j_t-1}^2, T_{j_t}^2) + P_t(T_{j_t}^2)) - \sum_{j=j_t+1}^{n_2} \delta_j^{3m} (L(t; T_{j-1}^2, T_j^2) + KB_t(T_j^2)) \right),$$

where $T_{i_t}^1$, respectively $T_{j_t}^2$ denotes the smallest T_i^1 , respectively T_j^2 , strictly greater than t . The spread K is chosen to be the fair basis swap spread at T_0 so that the basis swap has nil value at inception, i.e.

$$K = \frac{\sum_{i=1}^{n_1} \delta_i^{6m} L(T_0; T_{i-1}^1, T_i^1) - \sum_{j=1}^{n_2} \delta_j^{3m} L(T_0; T_{j-1}^2, T_j^2)}{\sum_{j=1}^{n_2} \delta_j^{3m} B_{T_0}(T_j^2)}.$$

We illustrate numerically our methodology on a basis swap with notional $N = 100$ and maturity $\bar{T} = 10y$, in the calibrated model of Sect. 3.3. The above formula yields the time-0 basis swap spread $K = 15\text{bps}$, which is added to the 3m leg so that the basis swap is incepted at par. The time 0 value of both legs is then equal to

Figure 8: Exposure of a basis swap (price process with mean and quantiles) in the exp-NIG calibrated model.



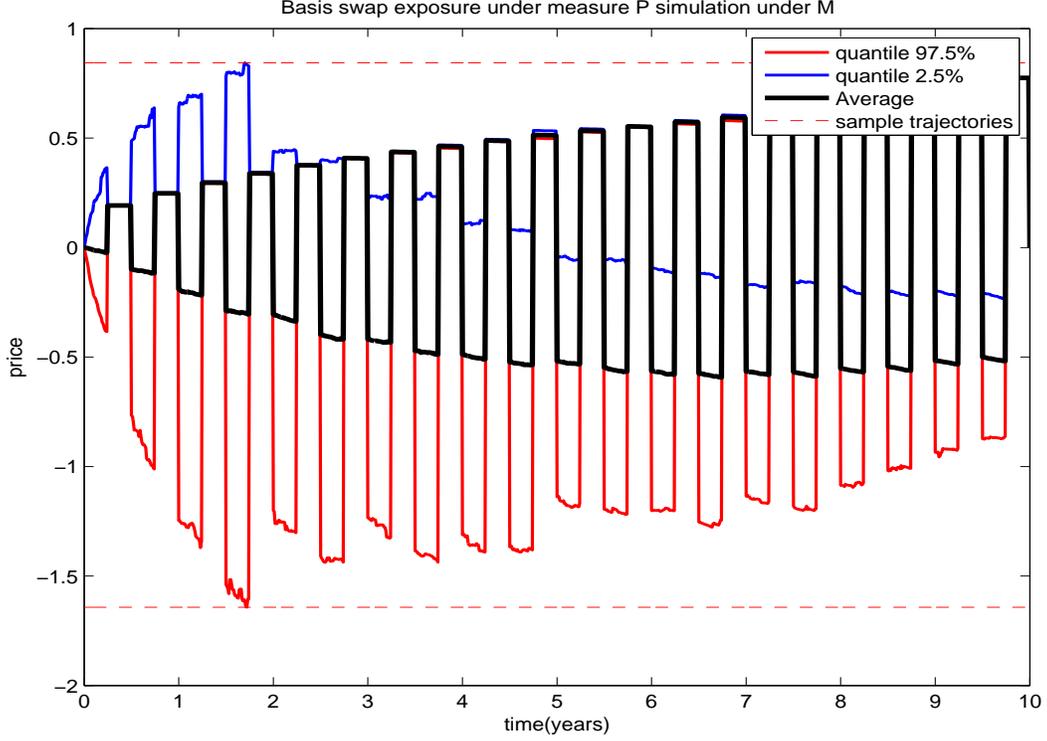
28€89. The resulting exposure is displayed in Fig. 8. Assuming that under the real world probability \mathbb{P} , the market factor $A^{(2)}$ (recall $A^{(1)} = 0$) has the law of a Lévy random bridge generated by the NIG process, the \mathbb{M} -exposure can be converted into \mathbb{P} -exposure by using the change of measure weights defined in Macrina (2014, Proposition 5.1)(see Fig. 9 and Hull, Sokol and White (2014)).

4.2 Credit Model

As we shall see below, the above exposures can be used to effectively compute CVA (credit valuation adjustment), DVA (debit valuation adjustment) and LVA (liquidity funding valuation adjustment). For such computations we shall consider the basic reduced-form approach of Crépey (2012) in which the first default time of a party $\tau = \tau_b \wedge \tau_c$ (where b respectively refer to the bank and its counterparty) is modeled as a Cox time

$$\tau = \inf \left\{ t > 0 \mid \exp \left(- \int_0^t \gamma_s ds \right) \geq \mathcal{E} \right\}, \quad (4.6)$$

Figure 9: Exposure of a basis swap (price process with mean and quantiles) in the exp-NIG-RB calibrated model.



where $\{\gamma_t\}$ is an $\{\mathbb{F}_t\}$ -adapted intensity process and \mathcal{E} is an independent exponentially distributed random variable. For this purpose, we can extend with credit the bottom-up approach of Sect. 2.1 as explained in Crépey et al. (2014, Sect.4). The reference (or market) filtration is defined as $\mathcal{F}_t = \mathcal{X}_t^{(0)} \vee \mathcal{X}_t^{(1,2)}$, $t \geq 0$, where $\{\mathcal{X}_t^{(0)}\}$ and $\{\mathcal{X}_t^{(1,2)}\}$ are the usual-made filtrations of independent \mathbb{M} Markov processes. In addition to the risk-free rate r_t modeled by (2.1) and the LIBOR rates modeled by (2.4), the first-to-default intensity of the two counterparties is modeled as

$$\gamma_t = -\frac{\dot{c}_0(t) + b_0(t)A_t^{(0)}}{c_0(t) + b_0(t)A_t^{(0)}},$$

where the b_0 and c_0 are nonnegative nonincreasing functions of time with $c_0(0) = 1$ and where $A^{(0)}$ (resp. $A^{(1)}, A^{(2)}$) is an $\{\mathcal{X}_t^{(0)}\}$ (resp. $\{\mathcal{X}_t^{(1,2)}\}$) adapted nonnegative $\{\mathcal{F}_t\}$ martingale, hence an $\{\mathcal{X}_t^{(0)}\}$ (resp. $\{\mathcal{X}_t^{(1)}\}$ and $\{\mathcal{X}_t^{(2)}\}$) martingale, starting from 0. In addition, we assume that the pricing measure \mathbb{Q} satisfies

$$\frac{d\mathbb{Q}}{d\mathbb{M}} \Big|_{\mathcal{F}_t} = \nu_t^{(0)} \nu_t^{(1)} \quad 0 \leq t \leq T,$$

where

$$\nu^{(0)} = \mathcal{E} \left(\int_0^\cdot \frac{b_0(t) dA_t^{(0)}}{c_0(t) + b_0(t) A_{t-}^{(0)}} \right), \nu^{(1)} = \mathcal{E} \left(\int_0^\cdot \frac{b_1(t) dA_t^{(1)}}{c_1(t) + b_1(t) A_{t-}^{(0)}} \right).$$

Denoting by $Z_t = e^{-\int_0^t \gamma_s ds}$ the Azéma supermartingale of τ , the credit analog of (2.2) yields the \mathbb{M} supermartingale $k = Z\nu^{(0)}$ (counterparty risk \mathbb{M} kernel) explicitly given, for $T \geq 0$, by

$$k_T = c_0(T) + b_0(T) A_T^{(0)}.$$

In the end, the primitives of the multi-curve Libor model with counterparty risk (recalling also Sect. 2.1) are the counterparty risk kernel k_T , the OIS kernel h_T and the Libor kernel $L_{S,T}$ (kernels under \mathbb{M} , the related \mathbb{P} pricing kernels being given as M times the latter, e.g. $\pi = Mh$).

4.3 CVA

In the simplest case of unilateral counterparty risk without collateralisation, the time-0 CVA on interest rate derivatives with mark-to-market P_t is (assuming zero recovery rate of the counterparty)

$$CVA_0 = \mathbb{E}^{\mathbb{Q}} \int_0^T P_s^+ D_s Z_s \gamma_s ds = \int_0^T \mathbb{E}^{\mathbb{Q}}(D_s P_s^+) \mathbb{E}^{\mathbb{Q}}(Z_s \gamma_s) ds,$$

where

$$\mathbb{E}^{\mathbb{Q}}(Z_s \gamma_s) = -\partial_s \mathbb{E}^{\mathbb{Q}}(Z_s) = -\partial_s \mathbb{E}^{\mathbb{M}}(\nu_s^{(0)} Z_s) = -\partial_s \mathbb{E}^{\mathbb{M}}(k_s) = -\dot{c}_0(s),$$

and

$$\mathbb{E}^{\mathbb{Q}}(D_s P_s^+) = \mathbb{E}^{\mathbb{M}}(h_s P_s^+).$$

Hence,

$$CVA_0 = - \int_0^T \mathbb{E}^{\mathbb{M}}(h_s P_s^+) \dot{c}_0(s) ds,$$

In the case of an interest-rate swap, $\mathbb{E}^{\mathbb{M}}(h_s P_s^+)$ corresponds to the mark-to-market of the swaption with expiry s on the swap, which can be recovered analytically if available in the model specification at hand (see section 3). In general, it can be retrieved numerically by simulation (see Sect. 4.1). For instance, in case of a counterparty with a constant $\gamma = 100\text{bps}$, we obtain for the above basis swap $CVA_0 = 0.1626$.

4.4 LVA

To simplify the LVA computation, people in the industry often neglect the involved nonlinearity (see Crépey, Bielecki and Brigo (2014)), working with a LVA coefficient exogenously proxied by

$$lva_t = \bar{b}_t \Gamma_t^+ - b_t \Gamma_t^- + \tilde{\lambda}_t (P_t - \Gamma_t)^+ - \lambda_t (P_t - \Gamma_t)^-, \quad (4.7)$$

where

- $\Gamma_t = \Gamma_t^+ - \Gamma_t^-$, where Γ_t^+ (respectively Γ_t^-) represents the value of the collateral posted by the counterparty to the bank (respectively by the bank to the counterparty), e.g. $\Gamma_t = 0$ (used henceforth unless otherwise stated) or $\Gamma_t = P_t$,
- \bar{b}_t and b_t are the spreads over the OIS (risk-free) short rate r_t for the remuneration of the collateral Γ_t^+ and Γ_t^- posted by the counterparty and the bank to each other,
- λ_t (respectively $\tilde{\lambda}_t$) is the liquidity funding spread over the OIS short rate r_t corresponding to the remuneration of the external funding loan (respectively debt) of the bank. By liquidity funding spreads we mean that these are free from credit risk.

The data Γ_t, b_t and \bar{b}_t are specified in a credit support annex (CSA) contracted between the two parties. The time 0 linearised LVA resulting from (4.7) is given as

$$LVA_0 = \mathbb{E}^{\mathbb{Q}} \int_0^T D_s Z_s lva_s ds = \int_0^T \mathbb{E}^{\mathbb{Q}}(D_s lva_s) \mathbb{E}^{\mathbb{Q}}(Z_s) ds$$

where

$$\mathbb{E}^{\mathbb{Q}}(D_s lva_s) = \mathbb{E}^{\mathbb{M}}(h_s lva_s), \quad \mathbb{E}^{\mathbb{Q}}(Z_s) = \mathbb{E}^{\mathbb{M}}(k_s) = c_0(s).$$

Hence,

$$LVA_0 = \int_0^T \mathbb{E}^{\mathbb{M}}(h_s lva_s(0)) c_0(s) ds.$$

In case of no collateralisation, $\Gamma_s = 0$, we have

$$lva_s(0) = \tilde{\lambda}_s P_s^+ - \lambda_s P_s^-.$$

In case of continuous collateralation, $\Gamma_s = P_s$, the formula becomes

$$lva_s(0) = \bar{b}_s P_s^+ - b_s P_s^-.$$

Hence, in each of these cases, as for the CVA, the LVA exposure is controlled by the clean process P , but for “scaling parameters” depending on the case under consideration. For instance, for $\tilde{\lambda} = 200\text{bps}$ and $\lambda = 0$, and still with a constant $\gamma = 100\text{bps}$ so that $c_0(t) = e^{-\gamma t}$, the LVA on basis swap (if non-collateralised) is 0.0325.

Acknowledgements The team expresses its gratitude to the University of Cape Town for the workshop, particularly professors David Taylor and Andrea Macrina for the organisation and support.

Bibliography

- Akahori, J., Y. Hishida, J. Teichmann and T. Tsuchiya, A Heat Kernel Approach to Interest Rate Models, *Japan Journal of Industrial and Applied Mathematics*, DOI 10.1007/s13160-014-0147-3 (2014).
- Akahori, J. and A. Macrina, Heat Kernel Interest Rate Models with Time-Inhomogeneous Markov Processes, *International Journal of Theoretical and Applied Finance* Vol. 15, Special Issue on Financial Derivatives and Risk Management (2012).
- Bianchetti, M.. Two curves, one price. *Risk Magazine*, August 74–80 (2010).
- Bianchetti, M. and M. Morini (Eds.) *Interest Rate Modelling After the Financial Crisis* (2013). Risk Books.
- Brigo, D. and F. Mercurio, *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer Verlag (2006).
- Brigo, D., M. Morini, and A. Pallavicini. *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes* (2013). Wiley.
- Brody, D. C., L. P. Hughston, A. Macrina, Information-Based Asset Pricing, *International Journal of Theoretical and Applied Finance*, 11, 107-142 (2008).
- Caldana, R. and G. Fusai, A general closed-form spread option pricing formula, *Journal of Banking and Finance*, 37 (12), 4893-4906 (2013).
- Cheng, S. and M. R. Tehranchi, Polynomial Models for interest rates and stochastic volatility, *arXiv* 1404.6190 (2014).
- Cochrane, J. H., *Asset Pricing*. Princeton University Press (2005).
- Cont, R. and P. Tankov. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC Press (2003).
- Crépey, S., Bilateral counterparty risk under funding constraints—Part II: CVA. *Mathematical Finance*, DOI 10.1111/mafi.12005 (2012, forthcoming).

- Crépey, S., T. R. Bielecki, and D. Brigo. *Counterparty Risk and Funding—A Tale of Two Puzzles*. Chapman and Hall/CRC Financial Mathematics Series (2014).
- Crépey, S., Z. Grbac, N. Ngor, and D. Skovmand (2013). A Lévy HJM multiple-curve model with application to CVA computation. *Quantitative Finance*, forthcoming.
- Crépey, S., Z. Grbac, and H. N. Nguyen. A multiple-curve HJM model of interbank risk. *Mathematics and Financial Economics* 6 (3), 155–190 (2012).
- Crépey, S., A. Macrina, T. M. Nguyen and D. Skovmand. Rational Multi-Curve Models with Counterparty-risk Valuation Adjustments. *Université d'Evry, University College London, Copenhagen Business School working paper* (2014).
- Cuchiero, C., M. Keller-Ressel, J. Teichmann (2012). Polynomial processes and their applications to mathematical Finance. *arXiv* 0812.4740v2.
- Eberlein, E. and S. Raible, Term structure models driven by general Lévy processes. *Mathematical Finance* 9 31-53 (1999).
- Filipović, D. and A. B. Trolle. The term structure of interbank risk. *Journal of Financial Economics* 109(3), 707–733 (2013).
- Flesaker, B. and L. P. Hughston, Positive Interest, *Risk Magazine*, 9, 46-49 (1996).
- Fujii, M., Y. Shimada, and A. Takahashi. A note on construction of multiple swap curves with and without collateral. *FSA Research Review* 6, 139–157 (2010).
- Fujii, M., Y. Shimada, and A. Takahashi. A market model of interest rates with dynamic basis spreads in the presence of collateral and multiple currencies. *Wilmott Magazine* 54, 61–73 (2011).
- Fujii, M. and A. Takahashi. Choice of collateral currency. *Risk Magazine*, January 120–125 (2011).
- Heath, D., R. Jarrow, A. Morton, Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, *Econometrica*, 60, 77-105 (1992).
- Henrard, M.. The irony in the derivatives discounting. *Wilmott Magazine* 30, 92–98(2007).
- Henrard, M.. The irony in the derivatives discounting part II: the crisis. *Wilmott Magazine* 2, 301–316 (2010).
- Hoyle, E., L. P. Hughston, A. Macrina, Lévy Random Bridges and the Modelling of Financial Information, *Stochastic Processes and their Applications*, 121, 856-884 (2011).

- Hughston, L. P. and A. Macrina, Pricing Fixed-Income Securities in an Information-Based Framework, *Applied Mathematical Finance*, 19, 361-379 (2012).
- Hull, J. C., A. Sokol and A. White, Modeling the short rate: the real and risk-neutral worlds, *Rotman School of Management Working Paper* no. 2403067. Available at SSRN: <http://ssrn.com/abstract=2403067> or <http://dx.doi.org/10.2139/ssrn.2403067> (2014).
- Hurd, T. R. and Z. Zhou. A Fourier transform method for spread option pricing. *SIAM Journal on Financial Mathematics* 1, 142–157 (2010).
- Jamshidian, F., Bond, Futures and Option Evaluation in the Quadratic Interest Rate Model, *Applied Mathematical Finance*, 3, 93-115 (1996).
- Kenyon, C.. Short-rate pricing after the liquidity and credit shocks: including the basis. *Risk Magazine*, November 83–87 (2010).
- Kijima, M., K. Tanaka, and T. Wong. A multi-quality model of interest rates. *Quantitative Finance* 9(2), 133–145 (2009).
- Macrina, A., Heat Kernel Models for Asset Pricing, *International Journal of Theoretical and Applied Finance*, 17 (7) (2014), DOI:10.1142/S0219024914500484.
- Mercurio, F. A LIBOR market model with stochastic basis. *Risk Magazine*, December 84–89 (2010).
- Mercurio, F. Interest rates and the credit crunch: new formulas and market models. Technical report, Bloomberg Portfolio Research Paper No. 2010-01-FRONTIERS (2010).
- Mercurio, F. Modern Libor Market Models: Using Different Curves for Projecting Rates and for Discounting, *International Journal of Theoretical and Applied Finance*, 13, 113-137 (2010).
- Moreni, N. and A. Pallavicini. Parsimonious multi-curve HJM modelling with stochastic volatility. In M. Bianchetti and M. Morini (Eds.), *Interest Rate Modelling After the Financial Crisis*. Risk Books (2013).
- Moreni, N. and A. Pallavicini. Parsimonious HJM modelling for multiple yield-curve dynamics. *Quantitative Finance* 14(2), 199–210 (2014).
- Pallavicini, A., D. Perini, and D. Brigo. Funding, collateral and hedging: uncovering the mechanics and the subtleties of funding valuation adjustments. <http://dx.doi.org/10.2139/ssrn.2161528> (2012).
- Rogers, L. C. G. The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates, *Mathematical Finance* 7, 157-176 (1997).

Singleton, K. and L. Umantsev. Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance* 12, 427–446 (2002).

Pricing Kernels, Multi-Curve Models and Swaption Pricing

Team 3

A. BACKWELL, University of Cape Town
M. KATEREGGA, University of Cape Town
F. PASINI, University College London
T. VON BOETTICHER, University of Johannesburg

Supervisors:

A. MACRINA, University College London
C. LABUSCHAGNE, University of Johannesburg

African Collaboration for Quantitative Finance and Risk Research

Contents

1	Introduction	3
2	Background	3
	2.1 Multi-curve models	3
	2.2 Pricing kernels	4
3	Weighted heat kernel model	6
4	Auxiliary measure \mathbb{M}	9
5	Application to swaption pricing	11
6	Numerical implementation of expectation	14
7	Calibration	16
8	Discussion	22

1 Introduction

This paper investigates the use of a pricing kernel model to price a multi-curve derivative, specifically a swaption written on LIBOR. We use the weighted heat kernel framework outlined by Macrina[13] and Crepey et al. [7]. Direct modelling of the pricing kernel is one approach to interest rate modelling, and can be extended to more general asset pricing. In particular, it can be used to tackle a multi-curve problem, where there is a risky rate in addition to the risk-free rate. This does not involve accounting for counterparty risk in the Credit Value Adjustment (CVA) sense, but only for the spread over the risk-free rate.

In section 2, background regarding multi-curve modelling and the pricing kernel approach is developed. Section 3 introduces the weighted heat kernel model from [13], both for interest rate modelling and general asset pricing. This model uses an auxiliary probability measure which is introduced and expounded in section 4 – we add a small note to [13]. Section 5 specialises the model to a LIBOR swaption, following [7] and applying the findings of section 4. Section 6 tackles the evaluation of the derived swaption expression, using numerical integration. Section 7 looks at the significant challenge of calibrating the swaption model. Section 8 concludes the paper.

2 Background

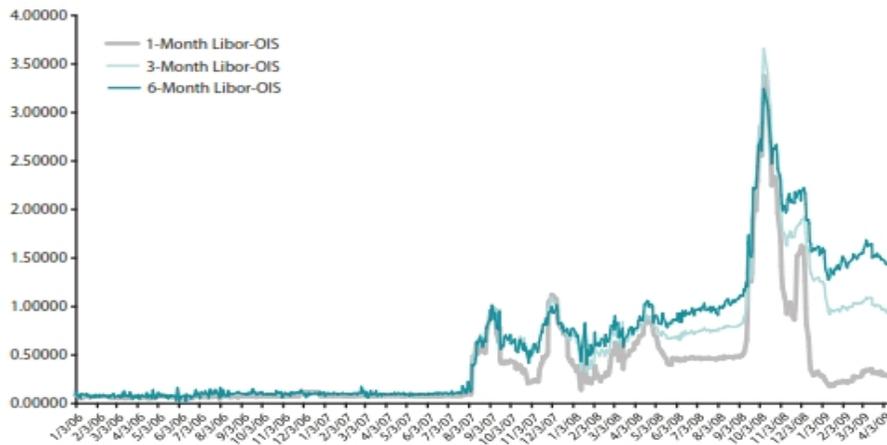
2.1 Multi-curve models

The classical theory of derivative pricing laid down by Black-Scholes [3] is relatively straightforward – a hedging argument allows us to discount the expectation of future cash flows under the risk-neutral measure using the assumed-constant risk-free rate. This theory was extended to interest rate derivatives in the Black [4] model, where the risk-free rate is also assumed to be constant and deterministic. The Black model for the pricing of interest rate derivatives can be extended to relax the assumption of a constant risk-free rate, and instead assume that both the underlying and the risk-free interest rate are random processes.

‘Multi-curve’ models attempt to model both the interest rate underlying a derivative, as well as the discounting interest rate. The motivation for these models is the idea that a widely applicable risk-free rate is no longer a tenable assumption - the 2008 financial crisis has shown that large corporations, banks and even governments, which were thought of as being ‘too big to fail’, are prone to default, which means that a truly risk-free interest rate does not exist. It is common for practitioners and academics to use the overnight indexed swap (OIS) rate as a suitable proxy for the theoretical risk-free rate, because the event of an ‘overnight’ default of a institution is relatively unlikely. The model contained in this paper holds onto this notion, and our pricing kernel (a notion explained below), is related to the OIS

interest rate. We will proceed to tackle derivatives written on LIBOR, the ‘London Interbank Offered Rate’. This rate is the average interest rate estimated by leading banks in London that they would be charged if borrowing from other banks. Before the financial crisis, the difference between the OIS rate and LIBOR was negligible, however, since 2007 a significant spread has been observed between OIS and LIBOR (see figure 1). The spread between OIS and LIBOR indicates that banks are

Figure 1: LIBOR-OIS spread for different tenors. Source: Federal Reserve Bank of St. Louis.



not risk-free, which requires the pricing of derivatives to take into account the risk of both parties. In the model introduced in this paper, we will postulate a model under the statistical probability measure \mathbb{P} , and will proceed, in effect, to price the derivative under a pricing measure \mathbb{Q} . We will also postulate an auxiliary measure \mathbb{M} under which the OIS rate, LIBOR and counterparty risk become tractable and independent.

2.2 Pricing kernels

There are many approaches to interest rate modelling, and a treatise such as Brigo & Mercurio [5] can be consulted to see the landscape of possible approaches. While specifying a diffusion for the short rate of interest (such as in the pre-eminent Hull & White [10]) is popular, here the lesser known approach of directly specifying a model for the *pricing kernel* is briefly introduced. While a rigorous axiomatisation is quite involved (see, for example, [9]), a pricing kernel is a fairly intuitive mathematical object: the pricing kernel is a positive process $\{\pi_t\}$ that when multiplied by a tradeable non-dividend-paying asset $\{S_t\}$ results in a \mathbb{P} -martingale. Writing $\{\mathcal{F}_t\}$ for the relevant market filtration, for an established pricing kernel we therefore

have

$$\pi_t S_t = \mathbb{E}^{\mathbb{P}}[\pi_T S_T | \mathcal{F}_t]. \quad (1)$$

The pricing kernel is also referred to as a stochastic discount factor or (misleadingly in the authors' opinion) the state price density. Intuitively, the pricing kernel models market time-value-of-money discounting and, simultaneously, market-aggregate risk preferences. Accordingly, a model for $\{\pi_t\}$ immediately gives rise to an interest-rate model; we can specialise the above equation to the case of a discount bond, for which, at time t with maturity T , we write P_{tT} . We have

$$\pi_t P_{tT} = \mathbb{E}^{\mathbb{P}}[\pi_T P_{TT} | \mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\pi_T | \mathcal{F}_t]$$

as the discount bond is pulled to par. It is advantageous to specify a model for $\{\pi_t\}$ of the form $\pi_t = \pi(t, X_t)$ where $\{X_t\}$ is a $(\{\mathcal{F}_t\}, \mathbb{P})$ -Markov process that generates the filtration. We can then write

$$\begin{aligned} P_{tT} &= \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}}[\pi_T | X_t] \\ &=: P(t, T, X_t). \end{aligned} \quad (2)$$

An examination of equation (2) reveals the key advantage of a pricing kernel approach to interest rate modelling: $0 \leq P_{tT} \leq 1$ for all t and T if and only if $\{\pi_t\}$ is a \mathbb{P} -supermartingale (modulo integrability concerns). This is a fairly straightforward and transparent mechanism to exclude negative rates (and therefore to exclude 'under-the-mattress arbitrage', which is borrowing at a negative rate and 'lending' at the greater rate of zero) – an issue that plagues diffusion models for the short rate. Also, the use of a Markov model is advantageous, as only the Markov processes need to be tracked in order to track the bond system (this puts the bond model into the Hunt, Kennedy & Pelsser [11] form – see their paper for a thorough discussion of this issue).

If we suppose that $\{X_t\}$ is an Itô process (so that it is possible to find the stochastic differential equations of the bond system, in principle at least – it could be difficult to find $P(\cdot, \cdot, \cdot)$) and assume relevant differentiability (i.e. assume the standard definitions for the instantaneous forward rates f_{tT} and short rate r_t exist), we can see how the model for π_t has given rise to a *complete* interest rate model:

$$\begin{aligned} f_{tT} &= -\partial_T \ln(P_{tT}) \\ r_t &= f_{tt} \\ dP_{tT}/P_{tT} &= \mu_{tT} dt + \sigma_{tT} dW_t^{\mathbb{P}} \\ dP_{tT}/P_{tT} &= (r_t + \lambda_t \sigma_{tT}) dt + \sigma_{tT} dW_t^{\mathbb{P}}, \end{aligned}$$

as the \mathbb{P} -dynamics have forced the market-price of $W^{\mathbb{P}}$ -risk $\lambda_t = \frac{\mu_{tT} - r_t}{\sigma_{tT}}$, where we use the fact that λ_t does not depend on T – a classic result.

A specification of a pricing kernel model is essentially the specification of a (suitably flexible) supermartingale, giving rise to a hopefully-tractable bond and interest model. Below we direct the reader to a few important references in the pricing kernel model literature. In addition to exploiting the supermartingale-positivity link, these pricing kernel models can give rise to tractable and useful models. It will be shown in the next section how the approach can be extended to general asset pricing in a natural way, and this paper will study the use of a pricing kernel approach to a multi-curve problem – the pricing kernel will model the risk-free rate (for which the OIS rate is a proxy) and the extended model will take care of the LIBOR.

The first pricing kernel model in the literature can be seen in [6], and building on this work Rogers [14] specified a canonical pricing kernel model of the form

$$\pi_t := \mathbb{E}^{\mathbb{P}} \left[\int_t^{\infty} e^{-\alpha s} g(X_s) ds \middle| X_t \right]$$

for a time-homogeneous Markov process $\{X_t\}$, a positive and integrable function g and a positive real number α . This specification can easily be shown to be a supermartingale and can recover a remarkable variety of non-negative interest rate models, especially when one allows $\{X_t\}$ to be multi-dimensional. Notice how positive rates are guaranteed while we still have flexibility over the driving Markov process $\{X_t\}$. In a short rate diffusion, you can guarantee positivity only by restricting the driver (for example using a log-normal diffusion).

Another noteworthy pricing kernel model by [8], expounded in [15] and [12], which was initially a direct model of the bond price family, and was shown to come from a pricing kernel supermartingale in a natural way. In the next section we will present the pricing kernel specification of Macrina [13], which will be used in the remainder of this work. This specification built upon [2] and [1].

3 Weighted heat kernel model

In this section we present the [13] pricing kernel model that uses weighted heat kernels. This is the basis for our approach to pricing LIBOR swaptions.

As a special case of the general $\pi_t = \pi(t, X_t)$ paradigm, the following pricing kernel form is postulated:

$$\pi_t = g_1(t) + f_1(t) \int_0^{U-t} \mathbb{E}^{\mathbb{P}}[F_1(t+u, X_{t+u}) | X_t] w_1(t, u) du,$$

where $\{X_t\}$ is an $(\{\mathcal{F}_t\}, \mathbb{P})$ -Markov process that generates the filtration $\mathcal{F}_t = \sigma(\{X_s\}_{0 \leq s \leq t})$. The model requires that $g_1(t)$ and $f_1(t)$ be deterministic, non-increasing, positive functions for $t \leq U$, where U is some point on the strictly

positive real line. The function $F_1(t, x)$ is chosen to be positive, such that the product of the conditional expectation of $F_1(t, x)$ and the weight given by $w_1(t, u)$ is integrable. It is required that $w_1(t, u - s) \leq w_1(t - s, u)$ for $s \leq \min(t, u)$.

As a result of the structure of the π_t specification and the requirements on $g_1(t)$, $f_1(t)$ and $F_1(t, x)$, it is easy to show that $\{\pi_t\}$ is a \mathbb{P} -supermartingale (the proof in [2] can be easily adapted). We therefore have a positive interest rate model.

Because no risky rates have been introduced (not yet - notice the subscripts of 1), this pricing kernel, used as the stochastic discounting term, can be thought of as the OIS-related model.

To see the model at the level of bonds, we put π_t into the form of equation (2). First it is useful to define

$$Y_{tT}^{(1)} := \int_{T-t}^{U-t} \mathbb{E}^{\mathbb{P}}[F_1(t+u, X_{t+u})|X_t]w_1(T, u - T + t)du.$$

We then have

$$\pi_t = g_1(t) + f_1(t)Y_{tt}^{(1)}.$$

Calibration to the initial term structure is done with g_1 . To see this, consider

$$\begin{aligned} P_{0t} &= \frac{1}{\pi_0} \mathbb{E}^{\mathbb{P}}[\pi_t|X_0] \\ &= \frac{\mathbb{E}^{\mathbb{P}}[g_1(t) + f_1(t)Y_{tt}^{(1)}|X_0]}{\pi_0} \\ &= \frac{g_1(t) + f_1(t)Y_{0t}^{(1)}}{\pi_0}, \end{aligned}$$

where the last step uses the Tower property and a change of variables in the integral $Y^{(1)}$. We can solve for $g_1(t)$ and substitute back into π_t , which is now automatically calibrated to the initial term structure,

$$\pi_t = \pi_0(P_{0t} + \frac{f_1(T)}{\pi_0}(Y_{tt}^{(1)} - Y_{0t}^{(1)})).$$

Calibration freedom for derivative data resides in f_1 . We can now see a general bond expression given by

$$\begin{aligned} P_{tT} &= \frac{\mathbb{E}^{\mathbb{P}}[\pi_T|X_t]}{\pi_t} \\ &= \frac{\mathbb{E}^{\mathbb{P}}[P_{0T} + \frac{f_1(T)}{\pi_0}(Y_{TT}^{(1)} - Y_{0T}^{(1)})|X_t]}{P_{0t} + \frac{f_1(T)}{\pi_0}(Y_{tt}^{(1)} - Y_{0t}^{(1)})} \\ &= \frac{P_{0T} + \frac{f_1(T)}{\pi_0}(Y_{tT}^{(1)} - Y_{0T}^{(1)})}{P_{0t} + \frac{f_1(t)}{\pi_0}(Y_{tt}^{(1)} - Y_{0t}^{(1)})}, \end{aligned}$$

where again we have used the Tower property and a change of variables in $Y^{(1)}$.

We are now in a position to write expressions for interest-rate derivatives with equation (1), as we have an expression for the bond family and the kernel. In [13], one can see how elegant simplifications occur here.

Certain choices for F_1 and w_1 allow $Y^{(1)}$ to be calculated in closed-form and the above to become

$$P_{tT} = \frac{P_{0T} + b_1(T)A_t^{(1)}}{P_{0t} + b_1(t)A_t^{(1)}},$$

where $A_t^{(1)} = A^{(1)}(t, X_t)$ is a \mathbb{P} -martingale. To show this, one would need to manipulate the integral $Y^{(1)}$ and use the change of variables alluded to above – we refer the reader to section 3 of [13], in which one can also see different choices of functions that can be integrated analytically. This is a very useful form as expectations in derivative expressions can be taken with respect to the martingale $A_t^{(1)}$ in a remarkably neat way. The class of models that can be written in this form will be referred to as ‘bA’ models.

Next we extend this framework to price general assets. These generalised assets will be specialised to LIBOR in section 5. We will use assets S_{tT} with a finite time horizon $T \leq U$. The natural extension of the above framework is to model the martingale $m_{tT} = \pi_t S_{tT}$ by adjusting our supermartingale machinery

$$m_{tT} = g_2(T) + f_2(T)Y_{tT}^{(2)},$$

where $g_2(T)$ and $f_2(T)$ are deterministic, and $Y_{tT}^{(2)}$ is defined in an analogous way to $Y_{tT}^{(1)}$, with new deterministic functions $F_2(t, x)$ and $w_2(t, u)$. It is required for $F_2(t, x)$ to be measurable, and that the product of the conditional expectation of $F_2(t, x)$ and $w_2(t, u)$ be integrable. We also require $w_2(t, u - s) = w_2(t - s, u)$ for $s \leq \min(t, u)$. It can then be shown that m_{tT} is a \mathbb{P} -martingale (see [13], proposition 6.1). Repeating the steps above, we have an asset price model given by

$$S_{tT} = \frac{S_{0T} + \frac{f_2(T)}{\pi_0} \left(Y_{tT}^{(2)} - Y_{0T}^{(2)} \right)}{P_{0t} + \frac{f_1(t)}{\pi_0} \left(Y_{tt}^{(1)} - Y_{0t}^{(1)} \right)}, \quad (3)$$

which is calibrated to the initial term structure and asset price. We will consider cases where equation (3) can be expressed in bA form,

$$S_{tT} = \frac{S_{0T} + b_2(T)A_t^{(2)}}{P_{0t} + b_1(t)A_t^{(1)}}, \quad (4)$$

where both $A_t^{(1)}$ and $A_t^{(2)}$ are \mathbb{P} -martingales. In section 5, this framework will be specialised to a LIBOR swaption.

4 Auxiliary measure \mathbb{M}

This section can be considered a note to [13], which makes extensive use of an auxiliary measure \mathbb{M} to make the model more tractable in light of the fact that the conditional expectation in $Y^{(1)}$ and $Y^{(2)}$ might be difficult to take under \mathbb{P} . Following [13], we define

$$Y_{tT}^{\mathbb{M}(1)} := \int_{T-t}^{U-t} \mathbb{E}^{\mathbb{M}}[F_1(t+u, X_{t+u}) | X_t] w_1(T, u - T + t) du, \quad (5)$$

and $Y_{tT}^{\mathbb{M}(2)}$ in the same way. $\{X_t\}$ could be specified as a Lévy random bridge under \mathbb{P} , and then we could identify the measure \mathbb{M} under which $\{X_t\}$ behaves like a Brownian bridge or Brownian motion, making the conditional expectation in Y much easier to compute. Focusing on the bA form, the pricing kernel is then modelled as

$$\begin{aligned} \pi_t &= \frac{\pi_0}{M_0} (P_{0t} + \frac{f_1(T)}{\pi_0} (Y_{tt}^{\mathbb{M}(1)} - Y_{0t}^{\mathbb{M}(1)})) M_t \\ &= \frac{\pi_0}{M_0} [P_{0t} + b_1(t) A_t^{(1)}] M_t, \end{aligned} \quad (6)$$

where M_t is the change-of-measure density $\frac{d\mathbb{M}}{d\mathbb{P}}|_{\mathcal{F}_t}$. The machinery of the previous section is used to produce an \mathbb{M} - (rather than \mathbb{P} -) supermartingale, and the change-of-measure density ensures that π_t is a \mathbb{P} -supermartingale, as we require. $A_t^{(1)} = A^{(1)}(t, X_t)$ is an \mathbb{M} -martingale, and M_t will allow all derivative expectations to be taken under \mathbb{M} , and over the \mathbb{M} -martingale $A_t^{(1)}$. Simple steps and a cancellation allows us to recover the form

$$P_{tT} = \frac{P_{0T} + b_1(T) A_t^{(1)}}{P_{0t} + b_1(t) A_t^{(1)}}. \quad (7)$$

This auxiliary measure extension is also applicable to the general asset framework. We set

$$\begin{aligned} m_{tT} &= \frac{M_t}{M_0} (g_2(T) + f_2(T) Y_{tT}^{\mathbb{M}(2)}) \\ &= \frac{\pi_0}{M_0} (S_{0T} + b_2(T) A_t^{(2)}) M_t, \end{aligned} \quad (8)$$

where M_t ensures that we get a \mathbb{P} -martingale from the \mathbb{M} -martingale inside the brackets. We can recover the form

$$S_{tT} = \frac{S_{0T} + b_2(T) A_t^{(2)}}{P_{0t} + b_1(t) A_t^{(1)}}. \quad (9)$$

Up to here we have followed [13], but now we clarify the effect of using the auxiliary measure. It is crucial to have a clear understanding of the kernel specification in equation (6). This is clearly a generalisation of the case that does not appeal to an auxiliary measure – we can see this by setting $\mathbb{P} = \mathbb{M}$ and $M_t = 1$ for all t and recovering the model outlined in section 3. However, it is *not* a generalisation in the sense that the modeller can choose any convenient auxiliary measure without affecting the model. We claim that the choice of the measure \mathbb{M} is in fact a modelling ingredient, just like the driver $\{X_t\}$ and the functions F_1 and F_2 . In other words, once the model has been fully specified under \mathbb{P} , for example $\{X_t\}$ being a particular random bridge, two modellers that use different auxiliary measures (one under which the random bridge behaves like a Brownian bridge, one under which it behaves like a Brownian motion, say), will get different models.

To see this, suppose that all the modelling ingredients are specified except the auxiliary measure, which, below, we view as generic. The kernel is then given by

$$\pi_t = \frac{M_t}{M_0} [g_1(t) + f_1(t) \int_0^{U-t} \mathbb{E}^{\mathbb{M}}[F_1(t+u, X_{t+u})|X_t] w_1(t, u) du].$$

One can try to get this back to a form that does not refer to \mathbb{M} – this would demonstrate that the choice of the measure is not integral to the model. However, we get

$$\begin{aligned} \pi_t &= \frac{M_t}{M_0} [g_1(t) + f_1(t) \int_0^{U-t} \mathbb{E}^{\mathbb{M}}[F_1(t+u, X_{t+u})|X_t] w_1(t, u) du] \\ &= \frac{M_t}{M_0} \mathbb{E}^{\mathbb{M}} [g_1(t) + f_1(t) \int_0^{U-t} F_1(t+u, X_{t+u}) w_1(t, u) du | X_t] \\ &= \frac{1}{M_0} \mathbb{E}^{\mathbb{P}} [M_U (g_1(t) + f_1(t) \int_0^{U-t} F_1(t+u, X_{t+u}) w_1(t, u) du | X_t)] \\ &= \frac{1}{M_0} [M_t g_1(t) + f_1(t) \int_0^{U-t} \mathbb{E}^{\mathbb{P}} [M_U F_1(t+u, X_{t+u}) | X_t] w_1(t, u) du] \\ &=: \frac{1}{M_0} [M_t g_1(t) + f_1(t) \int_0^{U-t} \mathbb{E}^{\mathbb{P}} [\tilde{F}_1^{\mathbb{M}}(t+u, X_{t+u}) | X_t] w_1(t, u) du]. \end{aligned}$$

The M_U term appears above because the terms in the expectation are \mathcal{F}_U -measurable. These steps show that one is unable to get back to the specification that does not involve the auxiliary measure. Moreover, we can see exactly how the choice of \mathbb{M} changes the model – *it modifies the function F* . The effect on the first term is not important, as it will disappear under \mathbb{P} -expectation.

This has a few implications. Firstly, when reading the specification equation (6) from [13], it is unclear whether this is a definition or a result from the basic model. We can now answer definitively: it is a definition – specifically, it is a definition

of a more general model that preserves the essential property of π_t (namely, \mathbb{P} -supermartingality). If it were a result rather than a definition, we would be able to recover the specification under \mathbb{P} for any choice of \mathbb{M} . This is not the case – instead the function F is modified by the choice of the measure.

This means that one can vary the choice of the measure and experiment with the model, just as one can vary the choice of the function F . Below, we will do exactly that – the Brownian bridge and Brownian motion measure will be compared.

One last comment here can be made regarding equations (7) and (9). The M_t terms have cancelled in the rational form (it is possible to use different measures for the OIS and the asset model and not get this cancellation, but we have not explored this). It is interesting to note that if we make a change to $\{X_t\}$ under \mathbb{P} , for example change the terminal distribution of the random bridge, it will not show up in equations (7) and (9) (and it will not affect derivative expressions, as we will see below). It will only affect the \mathbb{P} -dynamics of P_{tT} and S_{tT} , as they are functions of $\{X_t\}$.

5 Application to swaption pricing

This section will apply the developed framework to a multi-curve context, and we will specialise to the pricing of swaptions – the original problem of this work. This follows [7]. The price process underlying the swaption will be the LIBOR, and the pricing kernel will relate to the OIS rate. $L(T_i; T_{i-1}, T_i)$ is used to denote LIBOR for the period from T_{i-1} to T_i , which is revealed at T_{i-1} and is therefore $\mathcal{F}_{T_{i-1}}$ -measurable, and gives rise to the time T_i cashflow in a forward rate agreement (FRA) with fixed rate K ,

$$H_{T_i} = \delta_i(K - L(T_i; T_{i-1}, T_i)),$$

where $\delta_i = T_i - T_{i-1}$ and the notional is set to one. The FRA process is then given by

$$\begin{aligned} H_{tT_i} &= \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}}[\pi_{T_i} H_{T_i} | X_t] \\ &= \delta_i(K P_{tT_i} - L(t; T_{i-1}, T_i)), \end{aligned}$$

where the LIBOR process is defined as

$$L(t; T_{i-1}, T_i) = \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}}[\pi_{T_i} L(T_i; T_{i-1}, T_i) | X_t].$$

Even though the LIBOR is not tradeable, this definition is appropriate because LIBOR is synthetically tradeable through the FRA and speaks to the time t forward

LIBOR rate for the period T_{i-1} to T_i . [7] uses the asset framework developed above to model $L(T_i; T_{i-1}, T_i)$:

$$L(T_i; T_{i-1}, T_i) = \frac{L(0; T_{i-1}, T_i) + b_2(T_{i-1}, T_i)A_{T_{i-1}}^{(2)}}{P_{0T_i} + b_1(T_i)A_{T_{i-1}}^{(1)}}.$$

The indexing is slightly delicate – heuristically, the martingale subscripts arise from the $\mathcal{F}_{T_{i-1}}$ -measurability of the rate; the T_i subscript of the OIS process arise from the fact that the cashflow is made at T_i and the double argument of b_2 is necessary to keep track of the beginning and end of the FRA.

Note that $A^{(1)}$ and $A^{(2)}$ are \mathbb{M} -martingales. To recapitulate the message of section 4: even though we cannot see the measure \mathbb{M} in the $L(T_i; T_{i-1}, T_i)$ model, it will affect the functional form of $A_t^{(1)} = A^{(1)}(\cdot, \cdot)$ and $A_t^{(2)} = A^{(2)}(\cdot, \cdot)$, and we need consistency between these forms and the measure in the expectation.

Applying the definition of $L(t; T_{i-1}, T_i)$, or by analogy to the general asset pricing model outlined in section 3, we have

$$L(t; T_{i-1}, T_i) = \frac{L(0; T_{i-1}, T_i) + b_2(T_{i-1}, T_i)A_t^{(2)}}{P_{0t} + b_1(t)A_t^{(1)}}.$$

A swap at time t is the sum of n forward rate agreements (FRAs) over various T_i ,

$$Sw_t = \sum_{i=1}^n H_{tT_i} = \sum_{i=1}^n \delta_i \left[KP_{tT_i} - L(t; T_{i-1}, T_i) \right].$$

Using the developed framework, the price of a receiver swaption (as our FRA receives the fixed leg) at time t with maturity coinciding with the first reset date, T_0 ,

of the underlying swap is given by

$$\begin{aligned}
Sw_{n_t T_0} &= \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}} \left[\pi_{T_0} \left(Sw_{T_0} \right)^+ \middle| X_t \right] \\
&= \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}} \left[\pi_{T_0} \left(\sum_{i=1}^n \delta_i \left[K P_{T_0 T_i} - L(T_0; T_{i-1}, T_i) \right] \right)^+ \middle| X_t \right] \\
&= \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}} \left[\left(\sum_{i=1}^n \delta_i \left[K \mathbb{E}^{\mathbb{P}}[\pi_{T_i} | X_{T_0}] - \mathbb{E}^{\mathbb{P}}[\pi_{T_i} L(T_i; T_{i-1}, T_i) | X_{T_0}] \right] \right)^+ \middle| X_t \right] \\
&= \frac{1}{P_{0t} + b_1(t) A_t^{(1)}} \mathbb{E}^{\mathbb{M}} \left[\left(\sum_{i=1}^n \delta_i \left[K \mathbb{E}^{\mathbb{M}}[P_{0T_i} + b_1(T_i) A_{T_i}^{(1)} | X_{T_0}] \right. \right. \right. \\
&\quad \left. \left. \left. - \mathbb{E}^{\mathbb{M}}[L(0; T_{i-1}, T_i) + b_2(T_{i-1}, T_i) A_{T_{i-1}}^{(2)} | X_{T_0}] \right] \right)^+ \middle| X_t \right] \\
&= \frac{1}{P_{0t} + b_1(t) A_t^{(1)}} \mathbb{E}^{\mathbb{M}} \left[\left(\sum_{i=1}^n \delta_i \left[K (P_{0T_i} + b_1(T_i) A_{T_0}^{(1)}) \right. \right. \right. \\
&\quad \left. \left. \left. - L(0; T_{i-1}, T_i) - b_2(T_{i-1}, T_i) A_{T_0}^{(2)} \right] \right)^+ \middle| X_t \right]. \quad (10)
\end{aligned}$$

Many steps have been taken above: the relationships between the variables have been exploited, the positive quantity π_{T_0} has been taken into the maximum function, the measure has been changed in the way that the specifications allow, and the martingale property of the A s under \mathbb{M} is used.

We will now show the necessary steps that need to be taken in order to evaluate this integral:

- 1) We first choose a multivariate $(\{\mathcal{F}_t\}, \mathbb{P})$ -Markov process $\{X_t\}$, which generates the filtration $\mathcal{F}_t = \sigma(\{X_s\}_{0 \leq s \leq t})$. Only the behaviour of the processes $\{X_t^{(i)}\}$ under \mathbb{M} is relevant for us. We have in mind a Lévy random bridge for the \mathbb{P} specification, but we would only need precise specification under \mathbb{P} to simulate paths. Our choice for $\{X_t\}$ will be standard Brownian motion, i.e. $X_{t+u} | \mathcal{F}_t \sim \mathcal{N}_{\mathbb{M}}(X_t, u)$.
- 2) The pricing kernel is given by equation (6) (even though we do not need to see M_t explicitly for our purposes). Our choices for the F and w functions of the model will be given by $F_i(t, x) = x^2$, and $w_i(t, u) = (U - t - u)$, for $i = 1, 2$. These choices satisfy the relevant conditions and result in \mathbb{M} -martingales $A^{(i)}$ as we will see.
- 3) We evaluate the conditional expectations of $F_i(t + u, X_{t+u})$ under our auxil-

inary measure \mathbb{M} :

$$\begin{aligned}\mathbb{E}^{\mathbb{M}}[F_i(t+u, X_{t+u})|\mathcal{F}_t] &= \mathbb{E}^{\mathbb{M}}[X_{t+u}^2|X_t] = X_t^2 + u, \quad \text{and} \\ \mathbb{E}^{\mathbb{M}}[F_i(u, X_u)|\mathcal{F}_0] &= \mathbb{E}^{\mathbb{M}}[X_u^2|X_0] = u.\end{aligned}$$

- 4) Next, we need to find the expressions for $b_i(t)$ and $A_t^{(i)}$. We use the definition of the price process given in equation (8) in order to compute the $b_i(t)$ and $A_t^{(i)}$ for the price process of $L(0; T_{i-1}, T_i)$. Suppressing a great deal of computation, from equation (8) and (9) we can deduce that

$$\begin{aligned}b_1(T)A_t^{(2)} &= \frac{f_2(T)}{\pi_0} \left(Y_{tT}^{(2)} - Y_{0T}^{(2)} \right) \\ &= \frac{f_2(T)}{1 + f_1(0)\frac{U^3}{6}} \left[\frac{1}{2} (U - T)^2 (X_t^2 - t) \right],\end{aligned}$$

such that

$$\begin{aligned}b_1(T) &= \frac{f_2(T)(U - T)^2}{2[1 + f_1(0)\frac{U^3}{6}]}, \quad \text{and} \\ A_t^{(2)} &= X_t^2 - t.\end{aligned}$$

Analogously we have that

$$\begin{aligned}b_1(t) &= \frac{f_1(t)(U - t)^2}{2[1 + f_1(0)\frac{U^3}{6}]}, \quad \text{and} \\ A_t^{(1)} &= X_t^2 - t.\end{aligned}$$

These all correspond to Brownian motion. In [13], a measure related to a Brownian bridge is used and expressions analogous to the above are calculated. Having derived our own above, we are in a position to use the Brownian bridge for comparison purposes.

- 6) We substitute the expressions for $b_i(t)$ and $A_t^{(i)}$ and evaluate the price of the swaption by solving the expectation. The expression for the swaption price is automatically calibrated to the initial term structure and forward LIBOR rates.

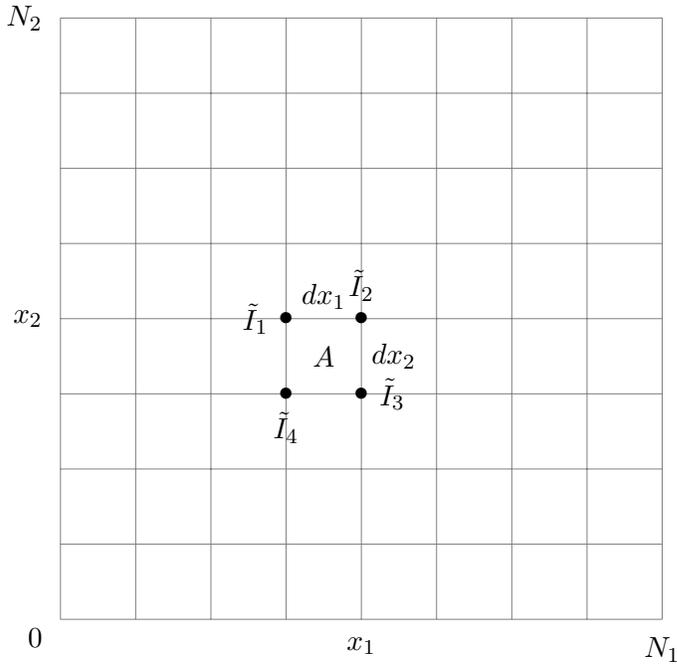
6 Numerical implementation of expectation

This section will provide a method of calculating the expectation in the expression of the price for a swaption. We suppose that $\{X_t^{(i)}\}$ follows a normal distribution given by $\mathcal{N}(\mu_i, \sigma_i)$, where μ_i and σ_i is the mean and standard deviation of $\{X_t^{(i)}\}$.

Taking an expectation is equivalent to calculating an integral of the function inside the expectation with respect to the probability measure. In this particular case this means that we multiply the function inside the expectation with the probability density function of the random variable and integrate the product over all real numbers under the Lebesgue measure. Since we cannot integrate over all real numbers, we define a 99.97% confidence interval for $\{X_t^{(i)}\}$ given by x_i , where

$$x_i = [-3\sigma_i + \mu_i, 3\sigma_i + \mu_i].$$

Next, we choose arbitrary large numbers (N_1+1) and (N_2+1) and divide x_1 and x_2 into N_1 and N_2 sub-intervals, respectively. We will denote the grid point distance of x_i by dx_i . This allows us to form a mesh grid as follows.



For each of the N_1 points in the interval x_1 , and each of the N_2 points in the interval x_2 , we evaluate the expression in the expectation of equation (10) over all the legs of the swaption, i.e.

$$I(k, j) = \sum_{i=1}^n \delta_i \left[K \left(P_{0T_i} + b_1(T_i) A_1(T_0, x_1(k)) \right) - L(0; T_{i-1}, T_i) - b_2(T_{i-1}, T_i) A_2(T_0, x_2(j)) \right],$$

where we index x_1 over k , x_2 over j . Having established $I(k, j)$ for all $k \in \{0, \dots, N_1\}$, and all $j \in \{0, \dots, N_2\}$, we set all negative results to zero and multiply the result by its relevant probability density at the point of x_1 and x_2 ,

$$\tilde{I}(k, j) = \max(I(k, j), 0) \mathcal{N}_1(x_1(k), \mu_1, \sigma_1) \mathcal{N}_2(x_2(j), \mu_2, \sigma_2).$$

Having applied the max function to all points on the grid, we now estimate the integral across each field on the grid. For example, consider the grid field A in the grid depicted above. We assign to A the averaged value of the points given by \tilde{I}_i , $i = 1, 2, 3, 4$, i.e.

$$A = \frac{1}{4} \sum_{i=1}^4 \tilde{I}_i.$$

We multiply A by the grid size $dx_1 dx_2$. The calculations applied to the grid point A are now applied throughout the entire grid, and each solution is summed. This will give us a value for the expectation given in equation (10). To further clarify the calculation, refer to the code extract below.

```
N1=100; % number of steps for x1
N2=100; % number of steps for x2

x1=linspace(-3*sigma+mu1, 3*sigma+mu1, N1); % x1 interval
x2=linspace(-3*sigma+mu2, 3*sigma+mu2, N2); % x2 interval
dx1=x1(2)-x1(1); % Change in x1
dx2=x2(2)-x2(1); % Change in x2

swaption=0;

for i=0:N1 % Summation over x1
    for j=0:N2 % Summation over x2
        integrand(i, j)=sum(delta*(K*(POT(T(1:n)) + b1(T(1:n))*A1(T0, x1(i))
            -L(0, T(0:n-1), T(1:n))-b2(T(0:1-1), T(1:n))*A2(T0, x2(j)))]
        integrand(i, j)=max(integrand(i, j), 0)*dist1(x1(i), mu1, sigma1)
            *dist2(x2(j), mu2, sigma2)

        mean=[integrand(i, j)+integrand(i-1, j)
            +integrand(i, j-1)+integrand(i-1, j-1)]/4
        swaption=swaption+mean*dx1*dx2
    end
end
```

7 Calibration

In this section we implement and calibrate the model obtained for the price of a swaption given by equation (10), which is automatically calibrated to the initial risk-free and forward LIBOR term structure (implicitly using g_1 and g_2).

First we calibrate to primary market data (initial term structure) and investigate the model at this level. Then we tackle calibration to derivative (swaption) data.

Calibration to primary market prices

In the first calibration procedure we make use of the initial term structure only, by fixing both f_1 and f_2 arbitrarily. The purpose of this is to investigate the different influences on the model, such as the significance of the parameter U , or the choice of the Markov process $\{X_t\}$. To find the swaption prices, we first calculate $L(0; T_{i-1}, T_i)$ using the market bond prices, i.e.

$$L(0; T_{i-1}, T_i) = \frac{1}{\delta_i} \left(\frac{P_{libor}(0, T_{i-1})}{P_{libor}(0, T_i)} - 1 \right) P_{ois}(0, T_i), \quad (11)$$

where the 'ois' and 'libor' subscripts indicate the bonds on the OIS interest rate and the LIBOR respectively. Using the quadratic F function, the expressions for $A_t^{(i)}$ and $b_i(t)$ are found by following the steps in the previous sections. A particular swaption price can now be calculated by evaluating the expression given in equation (10), estimated numerically as described in the previous section. We compare two different Markov processes under the measure \mathbb{M} , namely a standard Brownian bridge, where we assume that U is the terminal point of the Brownian bridge, and a standard Brownian motion. The results appear below.

Figure 2: The swaption prices versus strike prices for swaptions with maturity of 3 months, where the underlying swap has 3 month legs over a tenor of 3 years. We choose $U = 25$, $F(t, x) = x^2$, and different Markov processes, namely a standard Brownian bridge process and a standard Brownian motion.

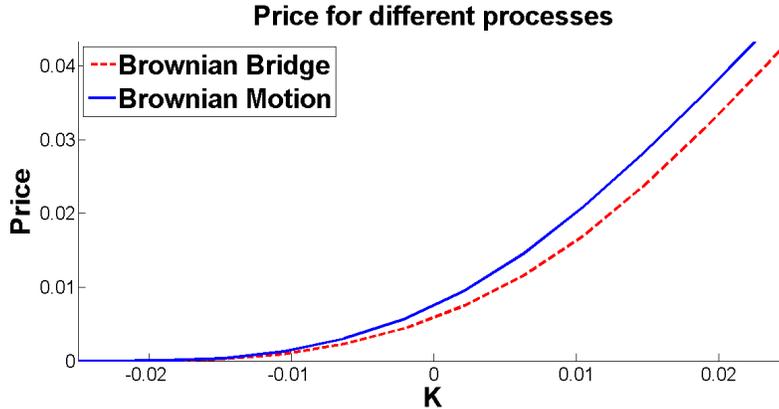
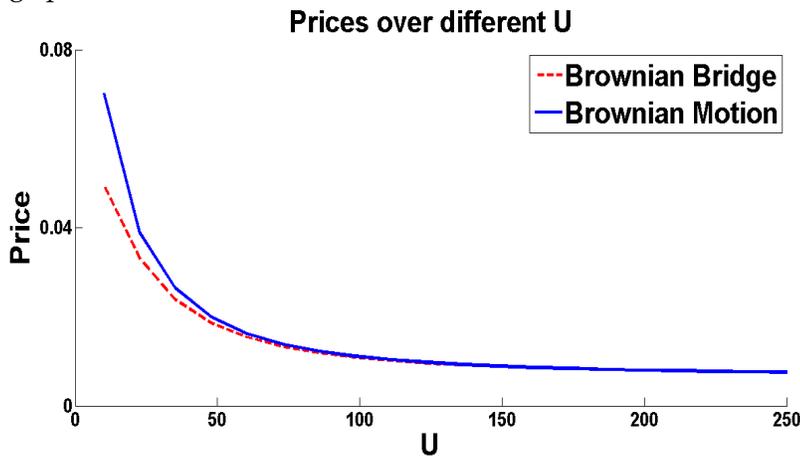


Figure 2 shows receiver swaption prices as a function of strike for different choices of $\{X_t\}$. The standard Brownian motion generally has a higher price than

the standard Brownian bridge. The graph shows that the price of a receiver swaption is positive even when the strike K is less than zero. This may seem counter-intuitive, but is caused by the fact that the model allows negative LIBOR. Only the risk-free rate is guaranteed to be positive; this might be a possible drawback with the model presented in this paper.

Figure 3: The swaption prices versus the choice of U for swaptions with maturity of 3 months, where the underlying swap has 3 month legs over a tenor of 3 years. We choose $F(t, x) = x^2$, and different Markov processes, namely a standard Brownian bridge process and a standard Brownian motion.

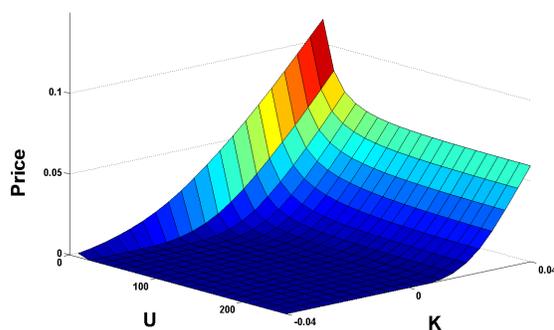


In the figure above we can see the influence of the choice of U on the model. For identical choices of $F(t, x)$ and $w(t, u)$, the two different processes produce different swaption prices. As U increases, the prices produced by the two processes converge. The convergence can be attributed to the fact that the Markov process of a standard Brownian bridge, as the terminal point U becomes large, converges to a standard Brownian motion in distribution.

Figure 4 displays a pricing surface of the swaption defined in the previous figures in terms of the strike price K and the choice in U , for the standard Brownian motion Markov process. It becomes clear the swaption becomes more expensive for a smaller choice of U . The price of a swaption versus the strike price produces a characteristic curve for a large choice of U .

The choice of the Markov process and U influence the price of the swaption, as well as the choice of the function of F . This is intuitively clear, since the A_i terms in the bA model depend on the conditional expectation of F .

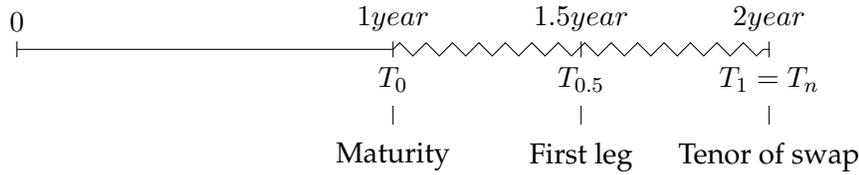
Figure 4: The swaption price surface as a function of the choice of U and strike price K , for swaptions with maturity of 3 months, where the underlying swap has 3 month legs over 3 years. We choose $F(t, x) = x^2$, and define the Markov processes to be a standard Brownian motion.



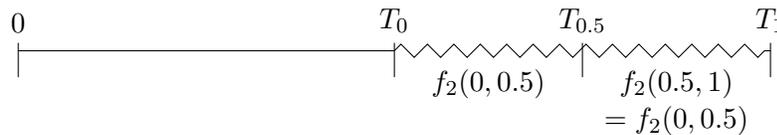
Calibration to swaption market prices

The model will now be calibrated to swaption market prices. The model has two free parameters for calibration, namely $f_1 = f_1(T_{i-1}, T_i)$ and $f_2 = f_2(T_{i-1}, T_i)$, which are contained in the functions b_1 and b_2 . Recall that f_1 related to the OIS rate, and since derivatives on the OIS rate do not exist, and we want to calibrate the model to LIBOR swaptions only, we set it equal to a constant for now, i.e. $f_1(T_{i-1}, T_i) = 1$.

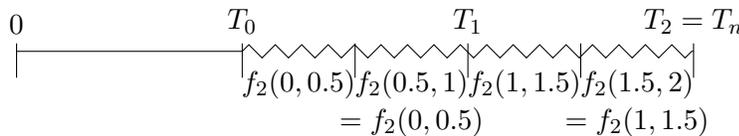
The model can now only be calibrated to the swaption market prices through the parameter f_2 . We define by $Swn(T_0, \tau, T_n, K)$ a swaption with maturity T_0 , where the underlying swap has a tenor of T_n , and the legs of the swap have length τ . The strike price of the swap underlying the swaption is given by K . We will illustrate the calibration with an example. Suppose that the duration of an at-the-money swap underlying the swaption with maturity $T_0 = 1$ is 1 year, so that $T_n = 2$ years from today, where the legs of the swap are $\tau = 0.5$ years (6 months) in length, as shown on the time-line below.



The swaption is now fully specified. We make use of the OIS and LIBOR bond prices in order to evaluate $L(0; T_{i-1}, T_i)$ as in equation (11). Since we have two legs in our underlying swap, we will have to calibrate both $f_2(0, 0.5)$ and $f_2(0.5, 1)$. This means that we have two unknowns in one known. The f_2 function only appears in the b_2 term in the expression for the price of a swaption, given in equation (10). Examining this equation, one can see that the expectation term is taken to be a sum over the b_2 terms, which are all functions of f_2 . Therefore we can, without a loss of degrees of freedom or generality, set $f_2(0, 0.5) = f_2(0.5, 1)$. We now have to solve only one unknown, namely $f_2(0, 0.5) = f_2(0.5, 1)$, to match the swaption price given by the market. This can be done numerically through several different root finding algorithms such as the Newton-Raphson method.

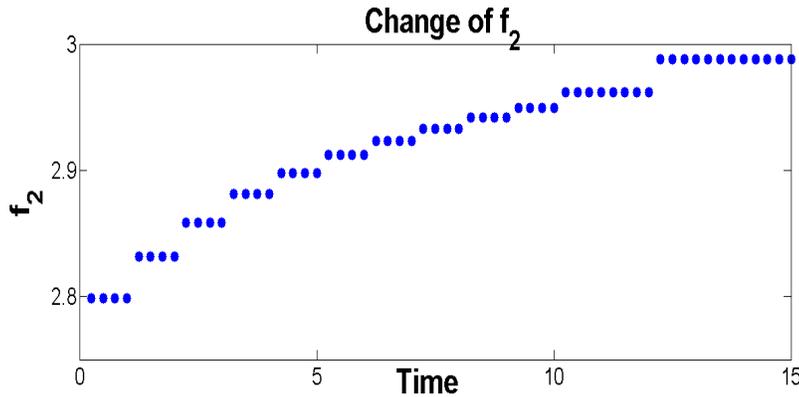


A swaption with the same maturity T_0 , but a greater tenor will then 're-use' the values of f_2 which were obtained for the previous swaption. For example, consider an at-the-money swaption with the same maturity and leg length as the above swaption, but with a swap duration of 2 years, so that $T_n = 3$ years. Suppose that we have found the values of $f_2(0, 0.5) = f_2(0.5, 1)$. To keep the f_2 term structure for the shorter tenor swaptions consistent, we keep the f_2 terms which we have already calibrated. We now want to calibrate the remaining f_2 values given by $f_2(1, 1.5)$ and $f_2(1.5, 2)$. We set the unknown f_2 variables equal, i.e. $f_2(1, 1.5) = f_2(1.5, 2)$. Since the previous f_2 parameters are already known, we need to solve one unknown parameter, namely $f_2(1, 1.5) = f_2(1.5, 2)$, to match the swaption market price.



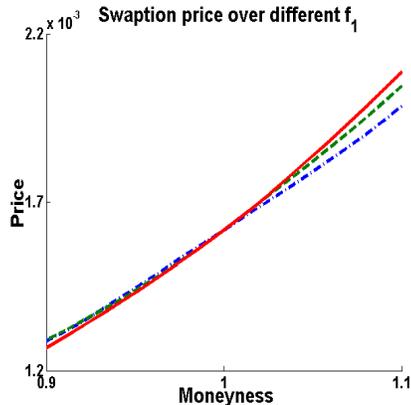
In this way, we can calibrate to all tenors for any given maturity, and figure 5 shows how for a fixed, arbitrary maturity, we have determined f_2 so that we are calibrated to at-the-money swaptions for all traded tenors.

Figure 5: The calibrated f_2 values for a swaption with a fixed maturity given by $T_0 = 0.5$ years, and different tenors. The underlying swaps have 3 month legs. We choose $U = 25$, $F(t, x) = x^2$, and a standard Brownian bridge process for the choice of $\{X_t\}$.



All our calibration is for a particular maturity, but the remaining dimension to explore is across strike. The model returns prices for swaptions in- and out-the-money, around our calibrated at-the-money price, and we would like to have some degree of freedom to calibrate here, but it appears we are out of viable calibration parameters. We noted that f_1 should be calibrated to OIS-related derivatives, but since these do not exist we can consider using this to affect our prices in- and out-the-money. An examination of equation (10) reveals that this is possible to an extent, but only the sum of the relevant f_1 s affects the price. Because the sum of f_1 s, residing in b_1 , is multiplied to the strike K , we can see intuitively that increasing this sum should steepen the price function (anchored at-the-money to the market price). Therefore calibration across the smile seems possible, albeit the somewhat weak form of being able to rotate the function. Figure 6 illustrates this. However, our ability to rotate the smile appears bounded and time limitations preclude further investigation here.

Figure 6: The swaption prices versus strike prices K , for different f_1 values. The swaption has a fixed maturity given by $T_0 = 0.5$ years, and the underlying swap has a duration of 1 year. The underlying swaps have 3 month legs. We choose $U = 25$, $F(t, x) = x^2$, and a standard Brownian motion process for the choice of $\{X_t\}$.



Our calibration logic appears sound: the relevant sum from f_1 is free for rotation as described, and then one could explore using various measures and functions F to find the most appropriate shape, which can be rotated to match the market as closely as possible. Alternatively one could look at inserting a free parameter in the function F or the Markov process itself – this might allow us to control both the rotation and curvature. Further study could interrogate this more detailed calibration and the boundedness of the rotation.

8 Discussion

Here we briefly summarise and discuss the paper, which started with an outline of pricing kernel theory, for both interest rate and general asset modelling. A contribution is clarification of the auxiliary measure used in [13]. We go on to specialise the framework to a LIBOR swaption, following [7]. As a general comment, the pricing kernel framework gives rise to tractable model – the benefits of the bA class are enjoyed when applied to swaptions, as well as the Markov nature and positive rates-supermartingale link. The derived expectation is then numerically integrated in a neat and efficient way. We find that calibration is possible to an extent, despite the dimensionality challenge of the swaption market. Further research could pursue our attempt at matching the market across strike (perhaps undertaking a more detailed investigation of different measures and functions F) and also investigating whether our calibration is consistent across maturities.

Bibliography

- [1] J. AKAHORI, Y. HISHIDA, J. TEICHMANN, T. TSUCHIYA (2014), A heat kernel approach to interest rate models, *Japan Journal of Industrial and Applied Mathematics*, 1-21.
- [2] J. AKAHORI, A. MACRINA (2012), Heat kernel interest rate models with time-inhomogeneous Markov processes, *International Journal of Theoretical and Applied Finance*, 15 (1).
- [3] F. BLACK, M. SCHOLES (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, 81 (3) , 637-659.
- [4] F. BLACK (1976), The pricing of commodity contracts, *Journal of Financial Economics*, 3, 167-179.
- [5] D. BRIGO, F. MERCURIO (2007), Interest rate models - theory and practice: with smile, inflation and credit, *Springer*.
- [6] G. CONSTANTINIDES (1992), A theory of the nominal term structure of interest rates, *Review of Financial Studies*, 5 (4), 531-552.
- [7] S. CREPEY, A. MACRINA, T. M. NGUYEN, D. SKOVMAND (2014), Rational Multi-Curve Models with Counterparty-risk Valuation Adjustments, *Université d'Evry, University College London, Copenhagen Business School working paper*
- [8] B. FLESAKER, L. HUGHSTON (1996), Positive interest, *Risk*, 9 (1), 46-49.
- [9] P. HUGHSTON, A. RAFAILIDIS (2005), A chaotic approach to interest rate modeling, *Finance and Stochastics*, 9 (1), 43-65.
- [10] J. HULL, A. WHITE (1993), One-factor interest-rate models and the valuation of interest-rate derivative securities, *Journal of Financial and Quantitative Analysis*, 28 (2), 235-254.
- [11] P. HUNT, J. KENNEDY, A. PELSSER (2000), Markov-functional interest rate models, *Finance and Stochastics*, 4 (4), 391-408.

- [12] Y. JIN, P. GLASSERMAN (2001), Equilibrium positive interest rates: a unified view, *Review of Financial Studies*, 14 (1), 187-214.
- [13] A. MACRINA (2014), Heat Kernel Models for Asset Pricing, *International Journal of Theoretical and Applied Finance*, 17 (7), DOI:10.1142/S0219024914500484.
- [14] L. ROGERS (1996), The potential approach to the term structure of interest rates and foreign exchange, *Mathematical Finance*, 7, 157-176.
- [15] M. RUTKOWSKI (1997), A note on the Flesaker-Hughston model of the term structure of interest rates, *Applied Mathematical Finance*, 4 (3), 151-163.

Portfolio Diversification using Higher moment Measures - Entropy and the Diversification Delta

Team 4

O. MAHOMED, University of Cape Town
W. USHAN, University of Cape Town
M. KHUMALO, University of Johannesburg

Supervisors:

D. R. TAYLOR, University of Cape Town
G. VAN DER LINDE, Prescient Securities
M. PILLAY, Prescient Securities

Contents

1	Problem	3
2	Introduction	4
3	Naive and Low Moment Measures	4
3.1	Weight-Based Measures	4
3.1.1	Weight Entropy	4
3.1.2	Herfindahl Index	4
3.2	Correlation and Variance-Based Measures	5
3.2.1	Intra-Portfolio Correlation	5
3.2.2	Portfolio Diversification Index	6
3.3	Effective Number of Bets	6
4	Higher moment Measures	7
4.1	Entropy	7
4.2	Diversification Delta	8
4.3	Entropy, Variance and Higher moments	9
4.4	Motivation for Higher moment Measures	11
5	Estimation of Higher-Moment Measures	13
5.1	Entropy and the Diversification Delta	13
5.1.1	Estimation of entropy using a histogram density estimator	14
5.1.2	Estimation of entropy using a kernel density estimator	16
5.1.3	Estimation of entropy using k-d partitioning	19
5.2	Application of diversification delta to South African market data	20
6	Application and Critique	25
6.1	Comparison of different measures	25
6.2	Efficient Frontiers	31
6.3	Optimised Portfolios	35
7	Consequences and Conclusions	38

1 Problem

Portfolio diversification measures may be segregated into three broad categories:

- **Naive Measures** - these are based on the premise that an optimally diversified portfolio is one that is equally weighted across all assets in the investment universe;
- **Lower moment Measures** - these are measures based almost entirely on the second moment of the asset distribution, most commonly in the guise of variances and covariances, with the premise being that lower metrics for these measures imply lower return/value dispersion and therefore better diversification; and,
- **Higher moment Measures** - these are measures based on higher moments of the return/value distribution, which incorporate more information about the structure and consistency of the distribution as opposed to just information about the first two moments and second order, statistical relationships.

In response to Samuelson (1967), who observed that the measurement of diversification using the first two moments of a distribution may be too restrictive and crude, numerous enhancements to lower moment measures, and new, higher moment measures of portfolio diversification have since been proposed. Two measures that have been advocated recently are the Effective Number of Bets (ENB) (Meucci, 2009) and the Diversification Delta measure (DD) (Vermorken et al., 2012). The primary research problem is **to compare the efficacy of these two measures, either theoretically or practically, via application to the South African financial market data.**

This report is structured in the following manner:

- **Literature Review and Survey of Measures** - a review of the literature relating to portfolio diversification, along with proposed measures for quantification;
- **Motivation for Higher moment Measures** - mathematical and simulation-based evidence justifying the use of higher moment measures to quantify portfolio diversification;
- **Estimation of Entropy** - estimation of the information entropy measure for discrete and continuous random variables; and,
- **Application and Critique** - an application of the reviewed and defined measures to the South African financial market data and the implications thereof.

2 Introduction

Modern Portfolio Theory centers on minimising the risk given a certain return. The minimisation of risk in a portfolio is achieved primarily through diversification. Diversification is a technique that reduces unrewarded risk in a portfolio by investing in a variety of assets and asset classes. A diversified portfolio will, on average, yield higher returns and pose a lower risk than the individual constituents of the portfolio. The benefits of portfolio diversification are evident from the fact that when different assets are combined, changes in the value of the portfolio are not as adverse as the changes in the value of the individual constituents.

Despite the intuitive need for diversification, there is no unique measure that adequately quantifies diversification and allocates assets optimally in a portfolio. In the following sections, some of the existing diversification measures that are commonly used will be surveyed.

3 Naive and Low Moment Measures

In terms of which risky assets to consider, some investors follow the naive rule of portfolio diversification, which allocates an equal amount of wealth to every asset available in the investment universe, at each rebalancing date. The naive rule is easy to implement, as it does not rely on moments of asset returns or optimisation, and is widely considered to be a benchmark.

3.1 Weight-Based Measures

3.1.1 Weight Entropy

Weight entropy is a measure that quantifies the risk in a portfolio using the weights of the constituents to compute its entropy. It is defined by:

$$WE = \exp \left(- \sum_{n=1}^N w_n \ln w_n \right),$$

where w_i is the weight of the i -th asset in the portfolio. The portfolio weights can be seen as the probability of being invested in a certain asset. One could then argue that the entropy difference between these probabilities and the uniform distribution is a measure of information content and diversification.

3.1.2 Herfindahl Index

Herfindahl Index (HI) is a weight-based measure of concentration. It is defined by:

$$HI = \sum_{i=1}^N w_i^2$$

where N is the number of assets, and w_i is the weight the i -th stock in the portfolio.

Qualitatively, there is little difference between these weight-based approaches of measuring risk as they lead to the same decision regarding the selection of assets.

These weight-based measures are heavily dependent on the weights and do not account for the differing risk-return characteristics of the portfolio. This can lead to the selection of a portfolio with risky assets (Kirchner and Zunckel, 2011).

3.2 Correlation and Variance-Based Measures

In his groundbreaking work on Portfolio Theory, Markowitz (1952) quantified the risk of a portfolio using variance, which quickly became standard practice in the financial industry. Variance measures the variability of returns from their mean return. Consider a portfolio Π with N securities where R_i is the return on each security, and w_i is the weight of the i -th security. The variance of the portfolio is defined by:

$$\sigma^2 = \sum_{i=1}^N w_i (R_i - R_{\Pi})^2,$$

where R_{Π} is the return on the portfolio. An investor aims to achieve a portfolio with low variance as this translates to low risk. Correlation demonstrates the benefits of diversification by measuring the movement of securities in relation to each other. The lower the correlation of the securities in a portfolio the greater the diversification.

3.2.1 Intra-Portfolio Correlation

Intra-Portfolio Correlation (IPC) is another commonly used weight-based measure. Unlike the other weight-based measures the IPC takes into account the correlation of assets and can be viewed as a weighted, standardised correlation coefficient. There are many competing definitions in the literature for the IPC, but it is most commonly defined by:

$$IPC = \sum_i \sum_j w_i w_j \rho_{ij} \quad i \neq j,$$

where w_i and w_j are the weights of assets i and j that are invested in the portfolio, and ρ_{ij} is the correlation between asset i and j . The IPC ranges from -1 to 1 , with -1 being most diversified and 1 the least (Livingston, 2013). Although the IPC focuses on correlation (i.e., the standardised measure of co-movement), it is clearly flawed because it ignores the standard deviations, which are major contributors to the risk of a portfolio.

3.2.2 Portfolio Diversification Index

The Portfolio Diversification Index (PDI) measures diversification using principal component analysis, which calculates the number of independent components in a portfolio. Given N securities the portfolio diversification index is defined by:

$$PDI = 2 \sum_{k=1}^N kw_k - 1,$$

where w_k are the ordered and normalised covariance eigenvalues of the correlation matrix of the securities. PDI indicates the following:

- A fully diversified portfolio ($w_k = \frac{1}{N}$ for all k) has a PDI = N ;
- A completely concentrated portfolio ($w_1 = 1, w_k = 0$ for all $k > 1$) has a PDI = 1.

The PDI summarises the diversification of large number of securities using a single statistic, and can compare the diversification across different portfolios or time periods. Although the PDI conveniently summarises diversification, it is difficult to find an optimised PDI portfolio, as the measure is independent of weights and is completely dependent on the covariance.

3.3 Effective Number of Bets

Consider a market of N securities. Let the returns of these securities be given by the N -dimensional vector R . A portfolio Π in this market has weights given by the vector w , and R_Π is the return on the portfolio Π , and is defined by $R_\Pi = w'R$. The covariance matrix of returns Σ is decomposed into uncorrelated risk sources using principal composition analysis:

$$E'\Sigma E = \Lambda$$

In the above equation the diagonal matrix $\Lambda \equiv \text{diag}(\lambda_1 \dots \lambda_n)$, contains the eigenvalues of Σ , sorted in decreasing order. The columns of the matrix $E \equiv (e_1 \dots e_N)$ are the respective eigenvectors. The eigenvectors can be seen as N uncorrelated principal portfolios whose returns $\tilde{R} \equiv E^{-1}R$ are decreasingly responsible for the randomness in the market. The portfolio with weights w can also be a combination of the uncorrelated principal portfolios with weights $\tilde{w} \equiv E^{-1}w$. Since the principal portfolios are uncorrelated, the variance of the n -th principal component can be defined by:

$$v_n = \tilde{w}_n^2 \lambda_n,$$

and the total portfolio variance by:

$$\text{Var}(R_w) = \sum_{n=1}^N v_n.$$

After normalising the variance, the diversification distribution is given by:

$$p_n = \frac{v_n}{\text{Var}(R_w)}$$

where p_n equals the R^2 from a regression of the total portfolio return on the n -th principal component. The diversification distribution can be interpreted as a set of probability masses associated with the uncorrelated principal portfolios.

From the diversification distribution it is evident that if the probability masses p_n are equal (i.e., uniformly distributed) then the portfolio is fully diversified. If the portfolio is concentrated on a single principal component, then it is not fully diversified, as all the risk will be due to that single component. Therefore portfolio diversification can be represented by the dispersion of the diversification distribution, as defined by the exponential of its entropy:

$$N_{Ent} = \exp\left(-\sum_{n=1}^N p_n \ln p_n\right).$$

N_{Ent} represents the true number of uncorrelated bets in a portfolio in a general market;

- a fully diversified portfolio has a maximum value of $N_{Ent} = \ln N$ when the risk is homogeneously spread over all principal components (i.e., $p_j = p_i = \frac{1}{N}$).
- A completely concentrated portfolio has a value of $N_{Ent} = 0$ and indicates that risk is due to a single principal component.

Meucci (2009) gives a more detailed analysis on the dynamics of finding the optimal weights which will maximise diversification in the presence of weight constraints.

4 Higher moment Measures

4.1 Entropy

Entropy is a measure of the uncertainty associated with a random variable. This measure was initially defined by Shannon (1948) for discrete probability distributions and is defined by:

$$H(X) = -\sum_i p(x_i) \log p(x_i)$$

for a discrete random variable X .

For a continuous random variable X , the entropy is defined as an expectation by:

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx.$$

where $f(x)$ is the probability density function and $H(X)$ is called differential or continuous entropy.

The greater the uncertainty in the outcome of a random variable, the higher its associated entropy. A random variable that has only one possible outcome has no uncertainty, and therefore will have an entropy of zero. In the discrete case, entropy depends on the probabilities associated with the different outcomes. Therefore, the uniform distribution, defined over some interval, has the highest uncertainty amongst all discrete distributions defined over that same interval, as each of the outcomes under the uniform distribution are equally likely.

In the continuous case, for a given mean and variance, the normal distribution has the greatest entropy across all distributions defined on the real line.

Furthermore, since probabilities lie between 0 and 1, the entropy for a discrete random variable will always be positive, while in the case of a continuous random variable the differential entropy may be allowed to be negative.

4.2 Diversification Delta

The Diversification Delta (DD) was introduced by Vermorken et al. (2012) as a measure of the diversification on the portfolio. The DD is defined by:

$$DD(P) = \frac{\exp\left(\sum_{i=1}^N w_i H(X_i)\right) - \exp\left(H\left(\sum_{i=1}^N w_i X_i\right)\right)}{\exp\left(\sum_{i=1}^N w_i H(X_i)\right)},$$

where w_i is the weight of the i -th asset in the portfolio. It is the ratio of the weighted average entropy of the individual assets minus the entropy of the portfolio, divided by the weighted average entropy of the assets (Vermorken et al., 2012).

The DD therefore measures the difference in the entropy between the case where the assets are held individually and the case where they are combined, in their appropriate weightings, to create a portfolio. The DD is a ratio that varies between zero and one. A value of one indicates that all non-systematic risk has been diversified away by forming the portfolio, while a value of zero indicates that there has been no improvement in diversification as a result of forming the portfolio.

Vermorken et al. (2012) state that exponentiation of entropy in estimation of the diversification delta retains the characteristics of the measure that are required, thus not causing any distortions. Their motivation for the exponentiation is to avoid the non-singular case where the entropy of the portfolio reaches zero.

4.3 Entropy, Variance and Higher moments

A natural question to consider is the relationship between entropy and variance. This relationship may be understood if one considers an approximation of the probability density function, $f_X(x)$, using a Fourier-Legendre series expansion, first proposed by Ebrahimi et al. (1999). The Legendre polynomials $\{P_n(x)\}$ are n -th degree polynomials, which form an orthogonal system over the interval $[-1, 1]$, and may be defined by:

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad (1)$$

where the orthogonal property yields the following:

$$\int_{-1}^1 P_m(x) P_n(x) dx = \frac{2}{2n+1} \delta_{mn}, \quad (2)$$

where δ_{mn} denotes the Kronecker delta function, and $P_n(1) = 1$. Each $P_n(x)$ satisfies the Legendre differential equation:

$$\frac{d}{dx} \left[(1-x^2) \frac{d}{dx} P_n(x) \right] + n(n+1) P_n(x) = 0. \quad (3)$$

Assume the density function, $f_X(x)$, has support $[-1, 1]$, then a Fourier-Legendre series expansion yields the following approximation for the density function:

$$f_X(x) = \sum_{n=0}^{\infty} a_n P_n(x). \quad (4)$$

Using (1) or (2), the first three Legendre polynomials are:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1). \end{aligned}$$

$P_1(x) = x$ and $\frac{1}{3}[2P_2(x) + P_0(x)] = x^2$. Equation (4) yields:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-1}^1 x f_X(x) dx \\
&= \int_{-1}^1 P_1(x) \left(\sum_{n=0}^{\infty} a_n P_n(x) \right) dx \\
&= \sum_{n=0}^{\infty} \int_{-1}^1 a_n P_1(x) P_n(x) dx \\
&= \int_{-1}^1 a_1 P_1(x) P_1(x) dx \\
&= \frac{2}{3} a_1,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{-1}^1 x^2 f_X(x) dx \\
&= \int_{-1}^1 \frac{1}{3} [2P_2(x) + P_0(x)] \left(\sum_{n=0}^{\infty} a_n P_n(x) \right) dx \\
&= \frac{1}{3} \sum_{n=0}^{\infty} \int_{-1}^1 a_n [2P_2(x) + P_0(x)] P_n(x) dx \\
&= \frac{1}{3} \left[\int_{-1}^1 2a_2 P_2(x) P_2(x) dx + \int_{-1}^1 a_0 P_0(x) P_0(x) dx \right] \\
&= \frac{1}{3} \left[\frac{4}{5} a_2 + 2a_0 \right].
\end{aligned}$$

The resulting variance is given by:

$$\mathbb{V}[X] = \frac{1}{3} \left[\frac{4}{5} a_2 + 2a_0 \right] - \frac{4}{9} a_1^2.$$

From the above equations, it is obvious that a_1 will be calibrated to the mean, which implies that the variance is determined by a_2 and a_0 , with all other a_n being irrelevant. Using the Fourier-Legendre expansion, the information entropy of the random variable X is defined by:

$$\mathbb{H}[X] = - \sum_{n=0}^{\infty} \int_{-1}^1 a_n P_n(x) \ln \left(\sum_{n=0}^{\infty} a_n P_n(x) \right) dx.$$

It can be shown that the partial derivative with respect to a_n yields:

$$\frac{\partial}{\partial a_n} \mathbb{H}[X] = - \int_{-1}^1 P_n(x) \ln \left(\sum_{n=0}^{\infty} a_n P_n(x) \right) dx. \quad (5)$$

From equation (5), it can be seen that the information entropy depends on all of the coefficients of the Fourier-Legendre, which corroborates that it is a higher moment measure, therefore incorporating and providing more information with regard to the structural uncertainty or dispersion of the distribution.

4.4 Motivation for Higher moment Measures

A rather simple justification as to why one would consider a higher moment measure, such as entropy, as a measure of diversification is presented below.

Consider a financial market where an investor has the choice between two investments, denoted by X and Y , with both expiring at a fixed, finite future time. The terminal return on investment X has a Laplace distribution, i.e., $X \sim \mathcal{L}(\alpha, \beta)$, with probability density function and parameters:

$$f_X(x) = \frac{1}{2\beta} \exp\left(-\frac{|x - \alpha|}{\beta}\right),$$

with parameters α and $\beta = \sqrt{\frac{\mu^2 + \sigma^2}{2}}$. The terminal return on investment Y is a mixture of normal distribution, i.e., $Y \sim \mathcal{MN}(n, \{\mu_i\}, \{\sigma_i\}, \{p_i\})$, with probability density function:

$$f_Y(y) = \sum_{i=1}^n \frac{p_i}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu_i}{\sigma_i}\right)^2\right). \quad (6)$$

The parameters are $n = 2$, $\mu_1 = -\mu$, $\mu_2 = \mu$, $\sigma_1 = \sigma$, $\sigma_2 = \sigma$ and $p_1 = p_2 = \frac{1}{2}$. The mean and variance of these two options are:

$$\begin{aligned} \mathbb{E}[X] &= \alpha, \\ \mathbb{V}[X] &= 2\beta^2 = \mu^2 + \sigma^2, \\ \mathbb{E}[Y] &= -\frac{1}{2}\mu + \frac{1}{2}\mu = 0, \\ \mathbb{V}[Y] &= \frac{1}{2}((-\mu)^2 + \sigma^2) + \frac{1}{2}(\mu^2 + \sigma^2) = \mu^2 + \sigma^2, \end{aligned}$$

which reveals that both returns have the same variance, but investment Y has an expected return of α and X an expected return of 0. In the absence of any further information, assuming that the investor is rational and makes investment decisions purely on the trade-off between risk and return, he would be indifferent between the two investments if $\alpha = 0$.

Now consider the information entropy measure of the return distribution for each

of these investments. For investment X , the entropy measure is:

$$\begin{aligned}
\mathbb{H}[X] &= - \int_{-\infty}^{\infty} \frac{1}{2\beta} \exp\left(-\frac{|x-\alpha|}{\beta}\right) \ln\left[\frac{1}{2\beta} \exp\left(-\frac{|x-\alpha|}{\beta}\right)\right] dx \\
&= - \int_{-\infty}^{\alpha} \frac{1}{2\beta} \exp\left(\frac{x-\alpha}{\beta}\right) \ln\left[\frac{1}{2\beta} \exp\left(\frac{x-\alpha}{\beta}\right)\right] dx \\
&\quad - \int_{\alpha}^{\infty} \frac{1}{2\beta} \exp\left(\frac{-x+\alpha}{\beta}\right) \ln\left[\frac{1}{2\beta} \exp\left(\frac{-x+\alpha}{\beta}\right)\right] dx \\
&= 1 + \ln(2\beta).
\end{aligned}$$

For investment Y , it is first useful to note that the probability density function (6), with parameters as specified above, may be rewritten as:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x^2 + \mu^2)}{2\sigma^2}\right) \cosh\left(\frac{\mu x}{\sigma^2}\right).$$

Then:

$$\begin{aligned}
\mathbb{H}[X] &= - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x^2 + \mu^2)}{2\sigma^2}\right) \cosh\left(\frac{\mu x}{\sigma^2}\right)\right] dx \\
&\quad - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \left[\exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right)\right] \ln\left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x^2 + \mu^2)}{2\sigma^2}\right) \cosh\left(\frac{\mu x}{\sigma^2}\right)\right] dx \\
&= \frac{1}{2} \ln(2\pi e\sigma^2) + (\lambda^2 - I),
\end{aligned}$$

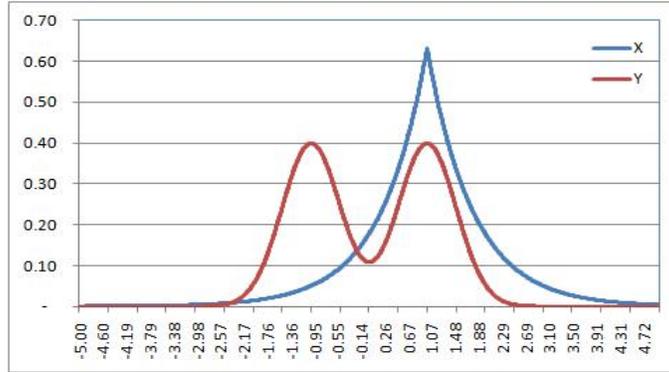
where $\lambda = \mu/\sigma$. I is an integral with no analytical solution defined by:

$$I = \frac{2}{\lambda\sqrt{2\pi}} \exp\left(-\frac{\lambda^2}{2}\right) \int_0^{\infty} \exp\left(-\frac{y^2}{2\lambda^2}\right) \cosh(y) \ln(\cosh(y)) dy.$$

The first part of the above expression, for the entropy of the mixture distribution, is just the entropy of a normal distribution with variance σ^2 . In fact, when $\mu = 0$ the mixture reduces to a normal distribution with mean 0 and variance σ^2 , and hence the above expression reduces to just the first term. Since the integrand in I is always positive, we know that $I \geq 0$. Further, Michalowicz et al. (2008) show that I is bounded with analytic upper and lower bounds, that $(\lambda^2 - I)$ monotonically increases from 0 to $\ln(2)$, and provide numerically computed values for I for different values of λ . These numerical estimates will prove to be critical in the motivation for higher moment measures of diversification.

Assume now that $\alpha = 1$, $\mu = 1$ and $\sigma = \frac{1}{2}$, then the terminal distributions of the two investment options are shown in the figure below.

Figure 1: The terminal distribution of investments X and Y .



One can now show, using the results derived in this section that both these investments have the same variance, which is equal to 1.25. Therefore, an investor who considers risk and diversification purely from the perspective of the variance measure will be indifferent between these two investment options. Visually, it is clear that the risk profile of these two investment options are significantly different. Measuring the entropy of these two investment alternatives, yields a measure of 1.45815 for X , and 1.64151 for the alternative investment Y (using the numerical estimate for $I = 0.111$ as provided by Michalowicz et al. (2008)). Based on this measure, a rational investor will surely prefer investment X to investment Y . Note that the mean of investment X has played no role in the discussion with regard to risk, which is a common feature with symmetric continuous distributions.

5 Estimation of Higher-Moment Measures

5.1 Entropy and the Diversification Delta

The entropy of a random variable was defined in Section 4.1.

The entropy of a random sample from a distribution can be naively estimated by using the empirical probabilities for each outcome of the random variable. This and other procedures for estimating entropy of discrete random variables is not considered any further as this report concentrates on entropy estimation for continuous random variables.

Bierlant et al. (1997) provide an overview of several nonparametric methods for the estimation of differential entropy for a continuous random variable based on a sample drawn from the distribution. They also discuss the convergence properties of each of these entropy estimates.

A common method is to “plug-in” estimates, where the density function $f(x)$ is replaced by an estimate $f_n(x)$ based on a sample of size n . In the simplest case, the estimate is a histogram density estimator. An alternative is the kernel density es-

timator, which constructs a smoother estimate of the density and generally reflects the underlying theoretical probability distribution more accurately.

The integral estimate of entropy is one of the “plug-in” estimates discussed and is defined by:

$$H_n(f) = - \int_{A_n} f_n(x) \log f_n(x) dx,$$

where the set A_n usually does not include tail values of the kernel estimator f_n . Using the kernel density estimator, the integral estimate exhibits strong consistency in that, as the sample size approaches infinity, the estimate approaches the differential entropy, almost surely. Evaluating the integral however, is difficult and requires numerical integration. The estimation is even more difficult when dimensionality is greater than two. In the case where the histogram density estimator is used, the estimate is more easily evaluated, and Bierlant et al. (1997) further mention that strong consistency has been proved.

Other “plug-in” estimates discussed by Bierlant et al. (1997) are the resubstitution estimate, splitting data estimate and cross-validation estimate.

In addition to “plug-in” estimates, Bierlant et al. (1997) also discuss estimates of entropy-based on sample spacings. They define these estimates only for the one-dimensional case. For the sample X_1, \dots, X_n , order statistics are given by $X_{n,1} < X_{n,2} < \dots < X_{n,n}$, and a spacing of order m is given by $X_{n,i+m} - X_{n,i}$. The following density estimate can be constructed using the m -order spacings:

$$f_n(x) = \frac{m}{n} \frac{1}{X_{n,im} - X_{n,(i-1)m}},$$

where $x \in [X_{n,(i-1)m}, X_{n,im})$. Using this density estimate, one can obtain an entropy estimate using the “plug-in” estimate approach. Alternatively, Bierlant et al. (1997) specify two m -spacing estimates of entropy that do not require explicit estimation of the density function.

Miller (2003) extends these m -spacing estimates to new classes of entropy estimators for the case where the probability density is multi-dimensional. For the purposes of this report, these estimators are not of particular relevance as the data considered is of one dimension.

A further non-parametric method for the estimation of entropy due to Stowell and Plumbley (2009) is called k - d partitioning and is used by Vermorken et al. (2012) in their estimation of entropy and diversification delta.

5.1.1 Estimation of entropy using a histogram density estimator

This approach uses the idea that the sample can be used to produce a histogram, which can then be used to find the discrete entropy of the sample. In order to do this, an appropriate number or spacing of bins needs to be specified. Using daily stock returns from January 2002 to June 2014 on each stock in the FTSE/JSE Top40

Index, the entropy was computed for each individual stock as well as for a portfolio based on the stocks. Figure 2 shows the behaviour of the entropy of a single stock and the portfolio, as the number of bins is varied.

Figure 2: Variation of entropy with number of bins used in histogram density estimator

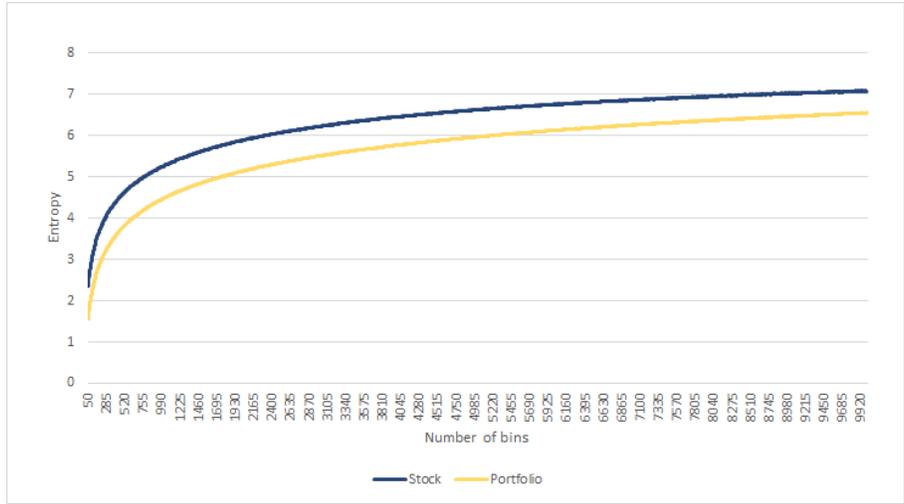


Figure 3: Variation of diversification delta with number of bins used in histogram density estimator

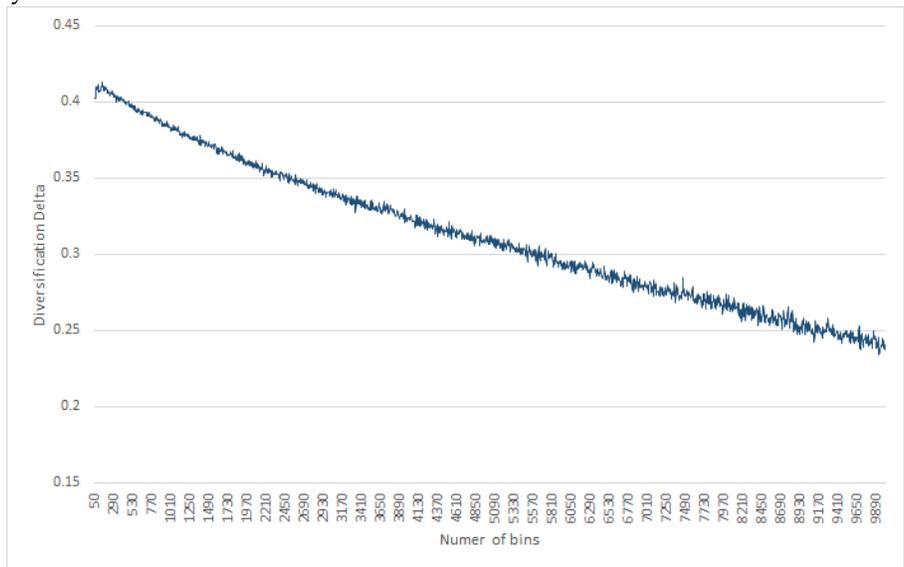


Figure 2 shows that the entropy of the portfolio is consistently lower than that

of the single stock. This reflects diversification and a lower degree of uncertainty associated with the return on the portfolio. Furthermore, the entropy of both the stock and portfolio returns increases as the number of bins are increased, converging to $\log(N)$ where N is the number of returns for each stock. This is because as the number of bins increases, the distribution tends to that of the uniform distribution which has maximum entropy for discrete distributions. The diversification delta has been plotted for the portfolio in Figure 3. It decreases with increasing entropy, implying poorer diversification.

Choice of the number of bins is non-trivial, and there is no unique measure that defines the optimal number of bins. As can be seen from Figure 2, using a large number of bins will split the data very finely leading to a representation of the uniform distribution, while using too small a number of bins will give insufficient precision to the density estimation. There are several rules-of-thumb to determine the “optimal” number of bins to use, but these are generally dependent on sample size. For the purposes of comparing entropy of returns across different estimation periods (and hence different sample sizes) choosing a number of bins in accordance with these rules is inappropriate because each estimation window will use a different number of bins. This results in different ranges for the histograms, causing the entropy estimates for the different estimation windows to be inconsistent with each other and hence incomparable. To counteract this, the number of bins and intervals are fixed, and the same histogram structure is applied for all estimations of entropy to ensure that the estimates are consistent with one another.

5.1.2 Estimation of entropy using a kernel density estimator

The kernel density estimator for a probability density function for a sample of size n drawn from an unknown distribution is defined as:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

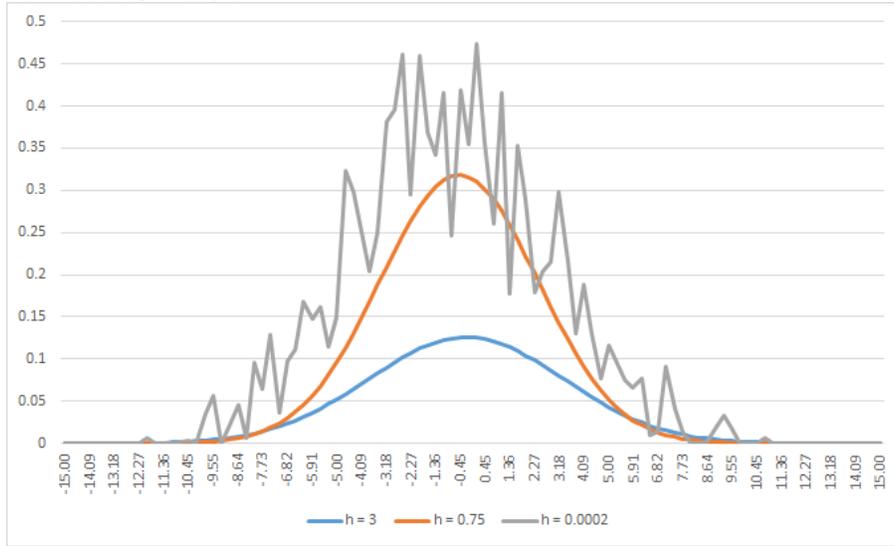
where $K(\cdot)$ is the kernel and h is the smoothing parameter. The kernel can be any symmetric, positive function that integrates to 1, i.e., a probability density function. Silverman (1986) provides an overview of this approach to estimating density, as well as several others, including the histogram density estimate.

A popular kernel is the standard normal distribution, and will be the kernel used in this investigation.

Akin to choosing the number of bins for the histogram in the histogram density estimate, a smoothing parameter h (called the bandwidth) must be chosen for kernel density estimation. Ideally, the bandwidth should be small, but not so small that it represents the data with spurious accuracy. Too large a value of the bandwidth will result in over-smoothing and will distort the structure of the underlying

distribution. Figure 4 shows kernel density estimators for a sample of normal variates and the progression from under-smoothing to over-smoothing as the bandwidth is varied.

Figure 4: Structure of the kernel density estimate of the underlying distribution as the bandwidth is varied.



Like Figure 2, Figure 5 shows how the entropies for a single stock on the FTSE/JSE Top40 Index and for a portfolio vary with the smoothing parameter. As the smoothing parameter increases, the bias in the density estimate is more profound and hence the entropy increases. Since normal kernels were used for the density estimation, the density estimate approaches normality as h increases, and the entropy therefore approaches that of a normal distribution where the standard deviation is a function of the bandwidth, h . This occurs for both the single stock and the portfolio, since at the higher values of h , the density estimate has been distorted and appears to be the same for all stocks and the portfolio. This is reflected in the diversification delta (Figure 6) as it approaches 0, showing no diversification.

Visually, it appears that only for values of the bandwidth, h , ranging from zero to 0.5 is there any difference in the entropy and diversification delta, and hence uncertainty, between the single stock and the portfolio. It may thus be that the “optimal” value for the bandwidth lies within this range for the purposes of this investigation.

Figure 5: Variation in the entropy estimated using a kernel density estimator with changes in the bandwidth.

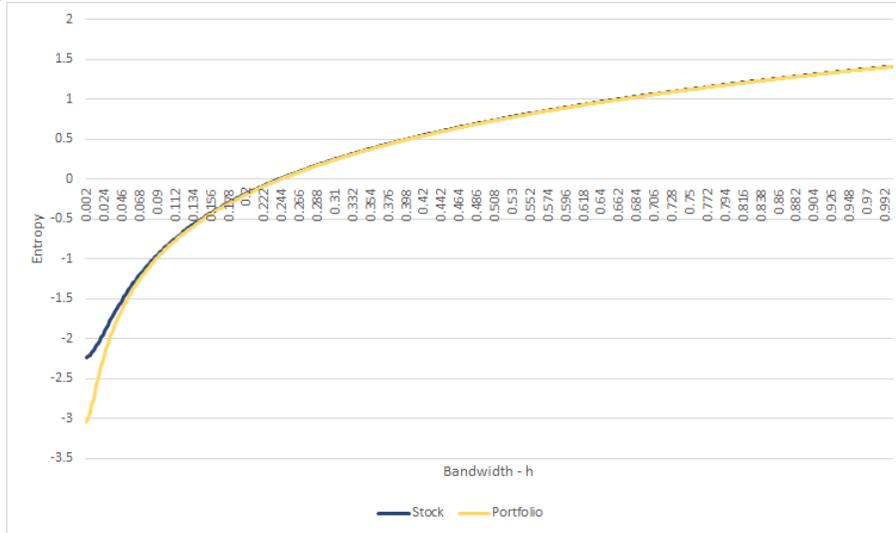
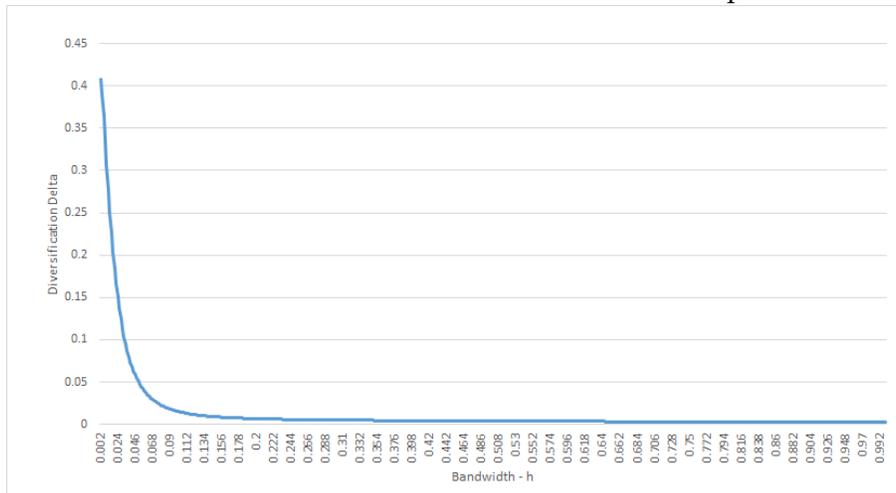


Figure 6: Variation in the diversification deltas the bandwidth parameter is varied.



Choosing a value for the bandwidth, like choosing the number of bins, is non-trivial. For the case where a normal kernel is used and it is believed that the underlying distribution is normal, there is an analytical value for the optimal bandwidth and is given by:

$$h_{opt} = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{\frac{1}{5}}$$

Silverman's (1986) rule of thumb is that for distributions that appear to be unimodal and symmetric, the optimal bandwidth generated by assuming that the underlying distribution is normal will be not too far off from the actual optimal bandwidth. In such a case, it may be justified to use the optimal bandwidth assuming normally distributed variables.

In the case where it cannot be justified that the sample points come from a distribution close to that of the normal distribution, there are several other approaches to estimating an optimal bandwidth. Raykar and Duraiswani (2006) provide an overview of several of these methods. The most common criterion involves minimising the AMISE (Asymptotic Mean Integrated Square Error) such that the optimal bandwidth is given by:

$$h_{AMISE} = \left(\frac{R(K)}{Nm_2(K)^2 R(f'')} \right)^{\frac{1}{5}}$$

where $R(g) = \int g(x)^2 dx$ and $m_2(K) = \int x^2 K(x) dx$ and f'' is the second derivative of the underlying density function f .

The unknown quantity in the above expression is $R(f'')$ requires an estimate of the second derivative of the underlying density. This is estimated using a new bandwidth, although estimation of this second derivative in turn depends on higher derivatives of the underlying density. Hence, at some level, a smoothing parameter is chosen with respect to a reference distribution, usually normal, and the derivatives are computed recursively backwards to eventually obtain the optimal bandwidth for the density estimate. Generally, as the number of levels in the estimation procedure increases, the variance in the estimate of the bandwidth increases. According to Raykar and Duraiswani (2006) the most common choice is to use only two levels.

For the purposes of this report, the data appear to be symmetrically distributed around zero and are bell-shaped. Therefore a bandwidth is chosen using Silverman's rule of thumb.

5.1.3 Estimation of entropy using k-d partitioning

Stowell and Plumbley (2009) developed a new nonparametric method which requires the range of the sample to be partitioned into cells that have equal probability mass, and then estimate the density function for each interval. They use a method called k - d partitioning which involves recursively splitting the data using the median. The algorithm stops when each interval passes the uniformity test and there are at least \sqrt{n} points in each partition.

For each interval A_j , the density function is estimated as:

$$f_A(x) = \frac{p_j}{\mu(A_j)}$$

for $x \in A_j$ and where $\mu(A_j)$ is the length of the j -th interval.

The strength of the k - d partitioning method for estimating the density and entropy is that there is no need to make any choices of parameters as in the previous cases where the bandwidth parameter and bin size or number of bins had to be chosen. However, it should be noted that requiring a minimum of \sqrt{n} data points in each partition does indeed create an implicit assumption on the number and range of partitions used.

5.2 Application of diversification delta to South African market data

In this section, the measures of entropy and diversification delta are applied to South African financial market data. In particular, two South African stock market indices are considered: the FTSE/JSE Top40 and FTSE/JSE Shareholder Weighted Top40 Indices.

The FTSE/JSE Top40 and FTSE/JSE Shareholder Weighted Top40 stock market indices constitute the 40 largest “blue-chip” companies listed on the Johannesburg Stock Exchange by market capitalisation. These indices have the same constituents but differ in terms of weighting, with the Top40 index adjusting total market capitalisation for the percentage of “free floated” shares, i.e. shares that are available for trading, while the shareholder weighted version further adjusting for shareholding within South Africa. Given the characteristics of these indices, they serve as key benchmarks within the South African market serving as benchmarks for most equity-managed funds and being a tradable index underlying a sizeable proportion of the outstanding equity derivatives exposure at any point in time.

These indices are rebalanced on a quarterly basis, according to the FTSE/JSE ground rules, and as such are subject to constant fluctuation in terms of constituency, shares in issues and free float and shareholder adjust factors - all of which contribute to the construction of the index. The data set considered here is a time-series of daily data of the constituents of these indices as at 30 June 2014. Due to time constraints, it was not possible to accurately reconstruct these indices retrospectively - as such, proxy indices were constructed, using current shares in issue, free float and shareholding factors. Furthermore, 27 of the 42 constituents were selected to construct the proxy versions of these indices, as these shares had the longest uninterrupted histories, going back to the inception of the FTSE/JSE South African index series (January 2002).

The aim of this section is to compute, compare and interpret diversification delta estimates produced using each of the three different methods of estimating entropy discussed above. This is repeated for the FTSE/JSE Top40 and FTSE/JSE Shareholder Weighted Top40 Indices.

Figures 7 to 9 show the estimates of the diversification delta on the FTSE/JSE

Top40 Index, using estimation windows of 1 month, 3 months, 6 months, 9 months and 12 months.

Figure 7: Diversification deltas on the FTSE/JSE Top40 Index estimated using the histogram density estimate.

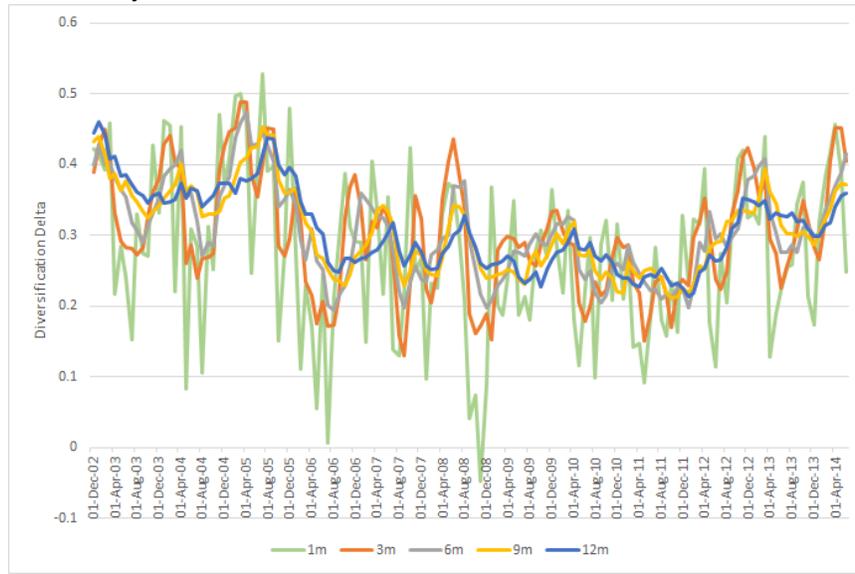


Figure 8: Diversification deltas on the FTSE/JSE Top40 Index estimated using the kernel density estimate.

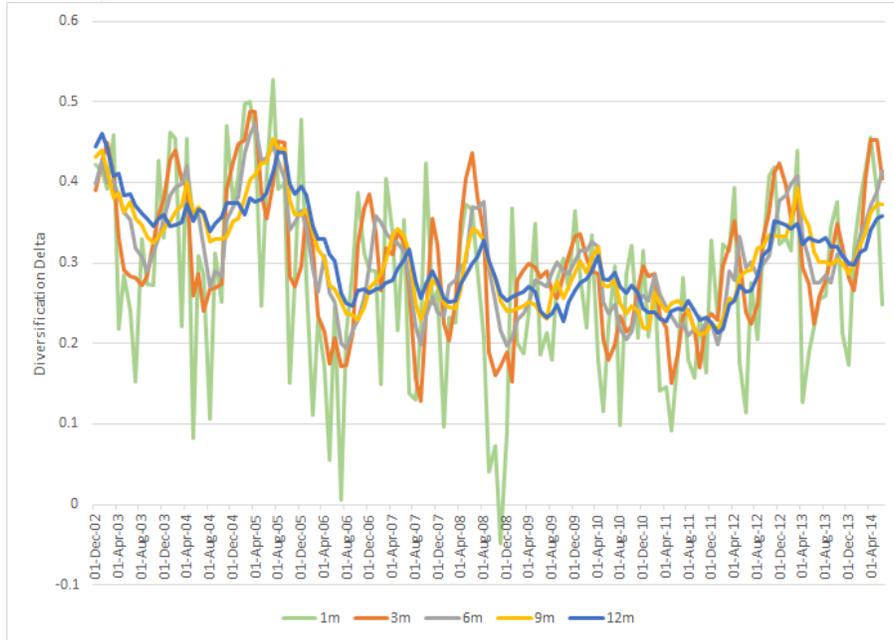
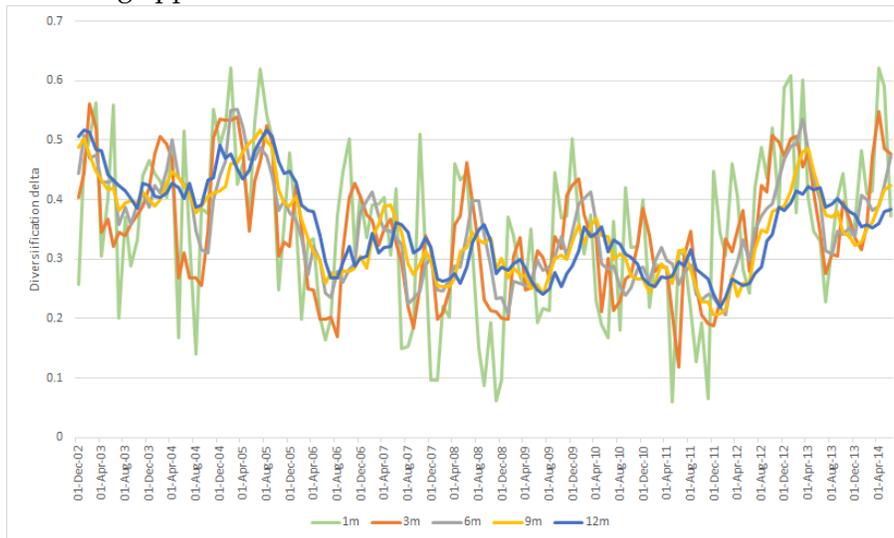


Figure 9: Diversification deltas on the FTSE/JSE Top40 Index estimated using the $k-d$ partitioning approach.



Comparing Figures 7 to 9 it is interesting to note that the estimates of the diversification delta using entropy estimates from the histogram density, kernel density

and $k-d$ partitioning approach give similar results. Therefore, the method used to estimate the entropy appears to have little effect on the final estimate of the diversification delta produced. This is important as it reduces the subjectivity associated with choosing a particular method to estimating entropy.

The estimates for the diversification delta appear to lie between 0.2 and 0.5 for the histogram and kernel density approaches, while the $k-d$ partitioning approach produces estimates that reach up to 0.6.

The estimates of the diversification delta are most stable when using an estimation window of 12 months, and least stable in the case of one month, which is as expected. In the case where estimation windows of one month have been used, the diversification delta appears to dip slightly below zero when using the histogram density and kernel density approaches to estimating entropy. However, this happens only for the one month estimation window and therefore can be attributed to an issue of few data points.

The diversification delta also appears to vary as expected, for example in 2008, there is a reasonable drop in the measure reflecting the poorer diversification in the market at that time.

Diversification deltas are expected to lie between zero and one, and in this case the estimates reach only up to 0.6. This may be a result of the data itself, and it is possible that repeating the investigation using a different index could produce diversification delta estimates that are higher or lower, reflecting greater and poorer diversification than the FTSE/JSE Top40 Index, respectively.

Figures 10 to 12 below show the estimates of the diversification delta repeated, as previously, but now on the FTSE/JSE Shareholder Weighted Top40 Index. The results obtained are similar to the FTSE/JSE Top40 Index in that the estimates of the diversification delta are consistent with each other, irrespective of the measure used to estimate the entropy. Also, the $k-d$ partitioning approach again appears to give a wider range of diversification deltas.

Diversification delta estimates based on the histogram density and kernel density approach lie between 0.1 and 0.6, while when using the $k-d$ partitioning method, the estimates approach 0.7.

Based on the diversification delta estimates, it appears that the FTSE/JSE Top40 and FTSE/JSE Shareholder Weighted Top40 Indices have a similar levels of diversification in the South African market, with the FTSE/JSE Shareholder Weighted Top40 Index performing slightly better than the FTSE/JSE Top40 in terms of diversification.

Figure 10: Diversification deltas on the FTSE/JSE Shareholder Weighted Top40 Index estimated using the histogram density estimate.

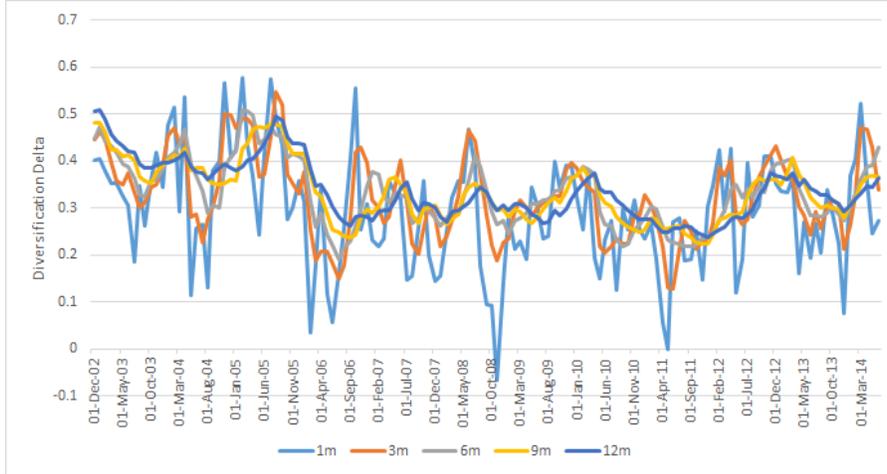


Figure 11: Diversification deltas on the FTSE/JSE Shareholder Weighted Top40 Index estimated using the kernel density estimate.

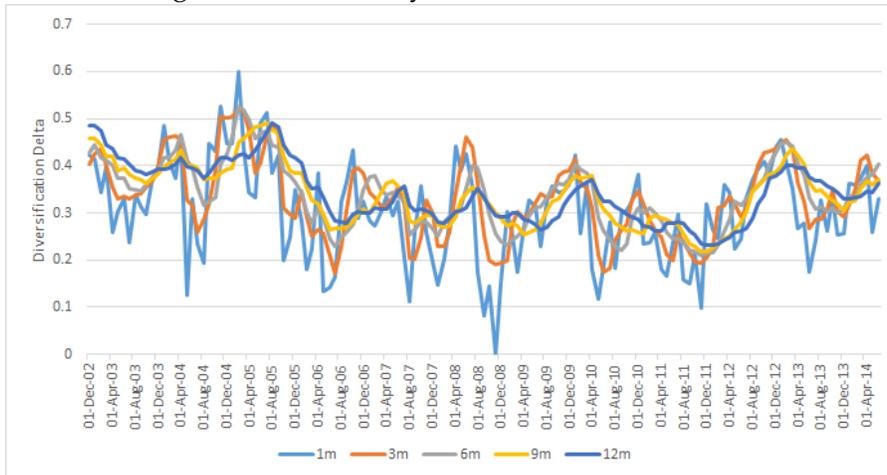
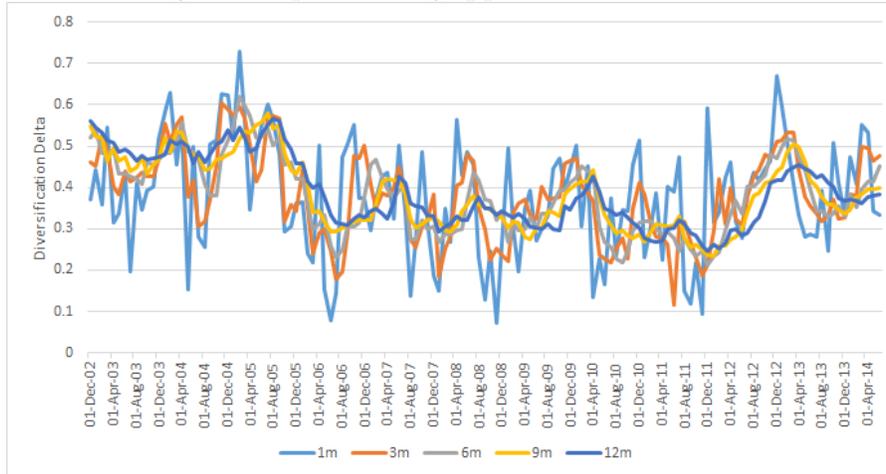


Figure 12: Diversification deltas on the FTSE/JSE Shareholder Weighted Top40 Index estimated using the k - d partitioning approach.



6 Application and Critique

In this section, a practical analysis is undertaken whereby the various measures of diversification are applied to South African financial market data. The purpose of this section is two-fold:

- (i) the diversification measures defined in the previous sections are applied to two South African stock market indices, in order to assess the coherence of the informational content of each of the measures; and,
- (ii) efficient frontiers and optimised portfolios are generated for a selected set of the measures in order to assess the relative behaviour of each of these respective measures from a risk and performance perspective.

As before, two South African stock market indices are considered: the FTSE/JSE Top40 and FTSE/JSE Shareholder Weighted Top40 Indices.

6.1 Comparison of different measures

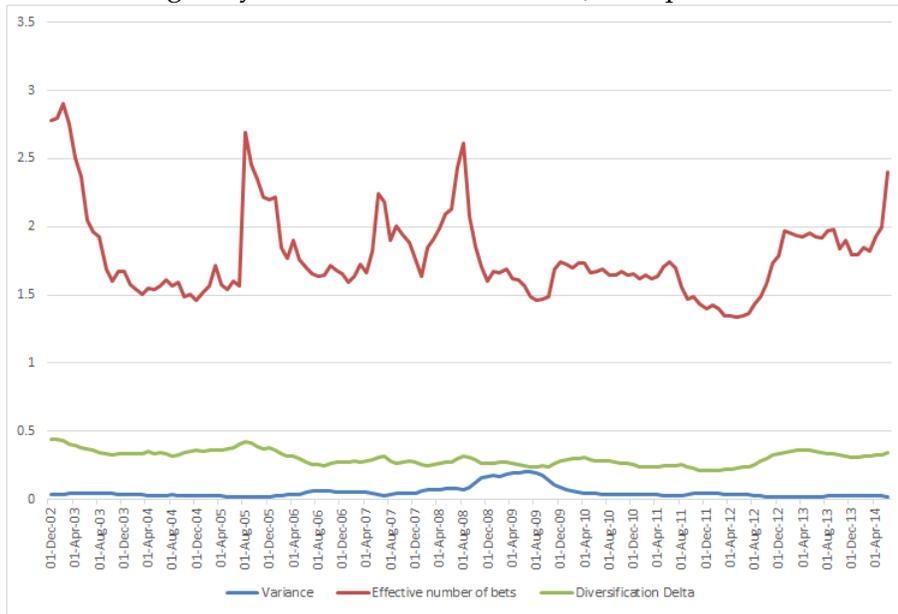
An analysis of several of the diversification measures mentioned in this report was conducted using South African financial market data as mentioned above.

The focus of this section will be on the following measures: variance, effective number of bets and the diversification delta.

Figure 13 below shows the estimates of the variance, effective number of bets and the diversification delta computed using the stock returns and an estimation period of 12 months. However, the three measures of diversification are incomparable for two main reasons. Firstly there is a scaling factor associated with the measures. It can be seen that the effective number of bets lies on a range that is

much higher than both variance and diversification delta. Secondly, the measures are inconsistent in their interpretation of diversification. While an increase in variance implies greater uncertainty and poorer diversification, an increase in either the diversification delta or the effective number of bets actually implies an increase in the diversification of the index.

Figure 13: Crude estimates of diversification: the variance, effective number of bets and variance, using daily return data on the FTSE/JSE Top40 Index.



To counteract these issues, a re-based “diversification index” was created for each measure first. The returns on each of these diversification indices was then computed over time, using an estimation window of 1 month, 6 months and 12 months. Furthermore, the returns on the variance “diversification index” were multiplied by -1 to ensure that for all three measures, movements in a particular direction has consistent implications on diversification.

Figures 14 to 16 below show the returns on each of these diversification indices for estimation periods of 1, 3 and 6 months respectively. At the lower estimation periods, due to the smaller number of data points, the return on the diversification indices is more erratic than that for the longer estimation periods.

Figure 14: Returns on rebased “diversification index” for each diversification measure, using daily return data on the FTSE/JSE Top40 Index, and estimation window of one month.

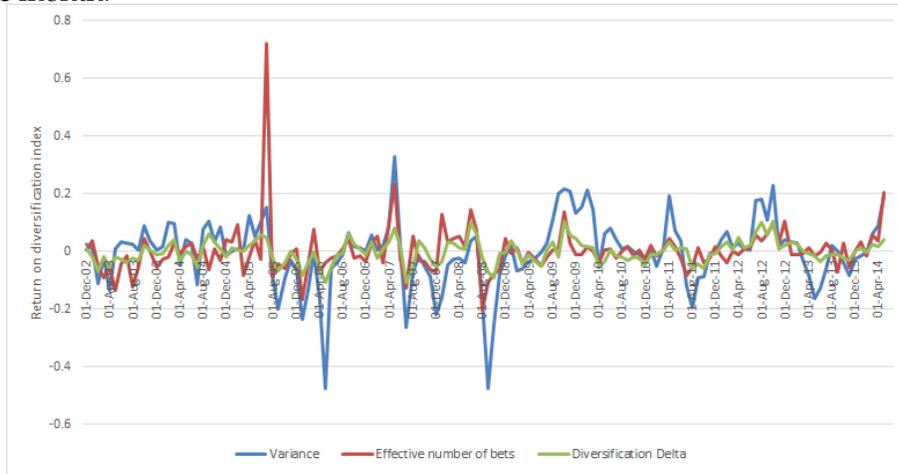


Figure 15: Returns on rebased “diversification index” for each diversification measure, using daily return data on the FTSE/JSE /Top40 Index, and estimation window of six months.

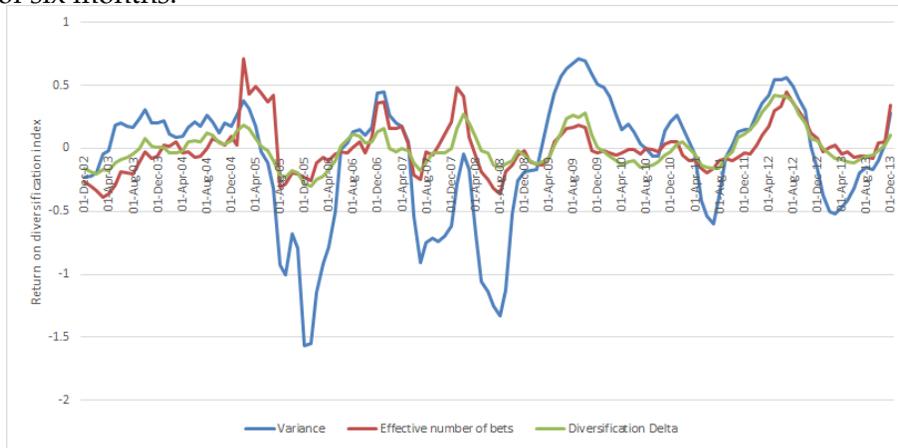
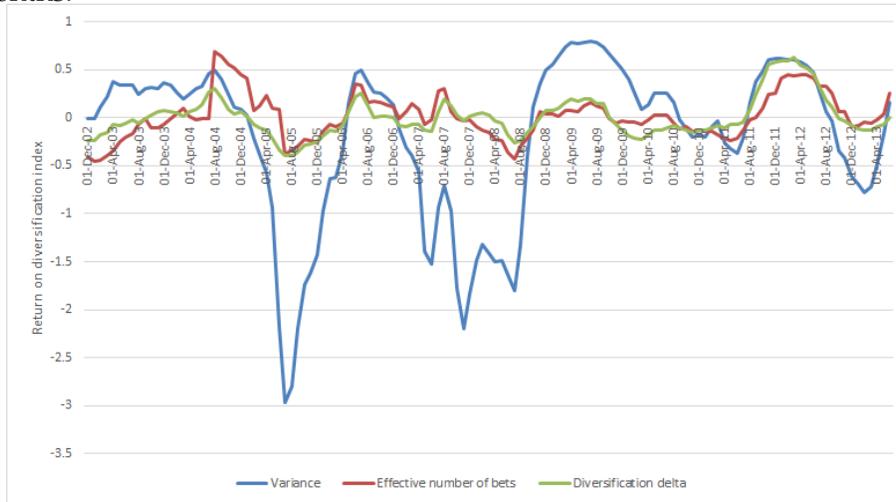


Figure 16: Returns on rebased “diversification index” for each diversification measure. Using return data on the FTSE/JSE Top40 Index, and estimation window of 12 months.



From Figures 14 to 16 it can be seen that all three diversification measures - variance, effective number of bets and diversification delta - rebased as mentioned above, give consistent results. All three measures move in the same direction, such that when there is indeed a change in the diversification, it is portrayed by all three measures. As an example, in the year 2008 it can be seen that all three measures drop, reflecting the poor diversification that existed in the market at the time.

Furthermore, the variance appears to be more erratic than both the effective number of bets and the diversification delta, with the diversification delta being the most stable of the three. This could be as a result of the fact that the variance is much more susceptible to outliers, since it is only a second moment measure. The diversification delta and effective number of bets take into account more information than the variance when computing the variability of the returns since they are higher moment measures.

An identical analysis was conducted for the FTSE/JSE Shareholder Weighted Top40 Index and the corresponding results are displayed in below Figures 17 to 20.

Figure 17: Crude estimates of diversification: the variance, effective number of bets and variance, using daily return data on the FTSE/JSE Shareholder Weighted Top40 Index.

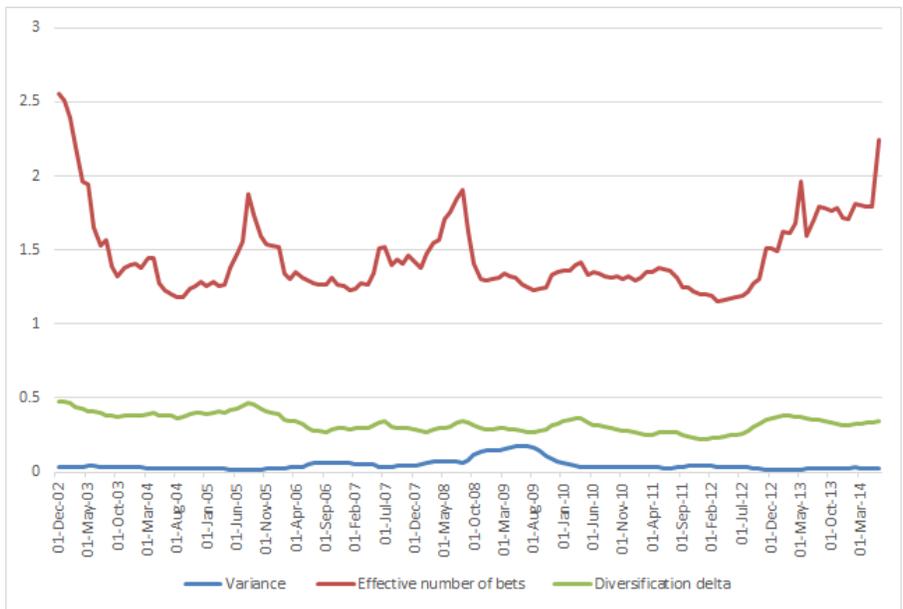


Figure 18: Returns on rebased "diversification index" for each diversification measure. Using return data on the FTSE/JSE Shareholder Weighted Top40 Index, and estimation window of one month.

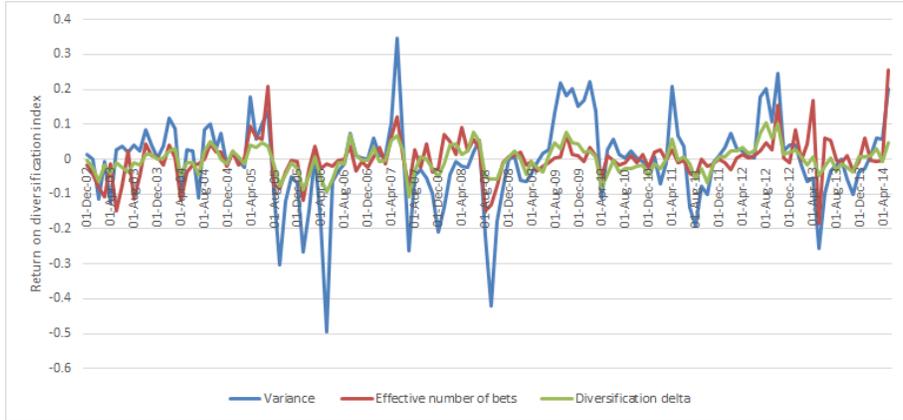


Figure 19: Returns on rebased “diversification index” for each diversification measure. Using return data on the FTSE/JSE Shareholder Weighted Top40 Index, and estimation window of 6 months.

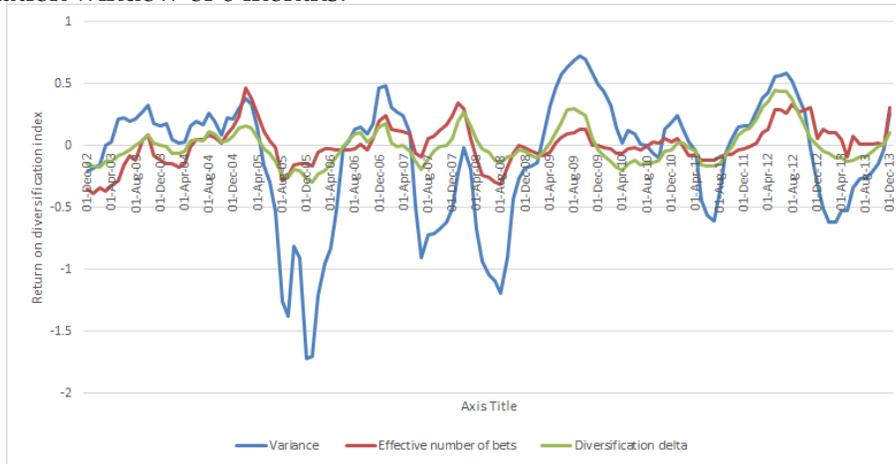
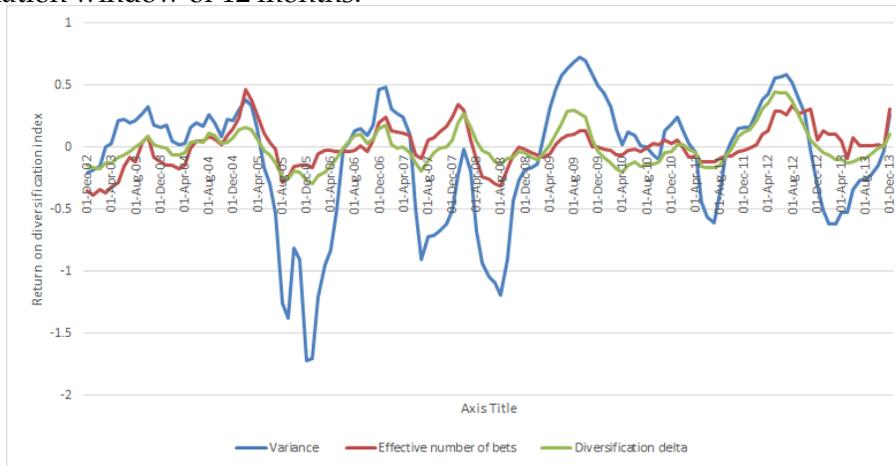


Figure 20: Returns on rebased “diversification index” for each diversification measure. Using return data on the FTSE/JSE Shareholder Weighted Top40 Index, and estimation window of 12 months.



Similar results were obtained from these data, in particular that all three measures are consistent with each other and give similar results for the diversification of the indices. Also, the variance again appears to behave more erratically, and the possible reasons are as were discussed previously.

6.2 Efficient Frontiers

In order to obtain a better understanding of the relationships and subtleties underpinning each of the key diversification measures under consideration, viz. variance, ENB and diversification delta, mean-diversification efficient frontiers are presented in this section. All efficient frontiers presented here are based on the full 27 stock universe, using daily data from January 2002 to June 2014 in order to estimate expected returns, variances, covariances and all other required statistical data. In addition, it should be noted that no scaling for time-horizon was effected, and therefore all statistics presented graphically and otherwise are for a one-day time horizon. The **long-only constraint** along with the usual **budget consideration** were the only optimisation and investment restrictions that were effected in the generation of all frontiers.

Having created mean-variance, mean-ENB and mean-DD efficient frontiers, the next six Figures depict these optimised portfolios across all four of these statistical measures, viz. the mean, variance, ENB and diversification delta. Each Figure also depicts the optimally diversified portfolio-based on each of the respective measures.

Figure 21: Mean-diversification efficient frontiers in mean-variance co-ordinates.

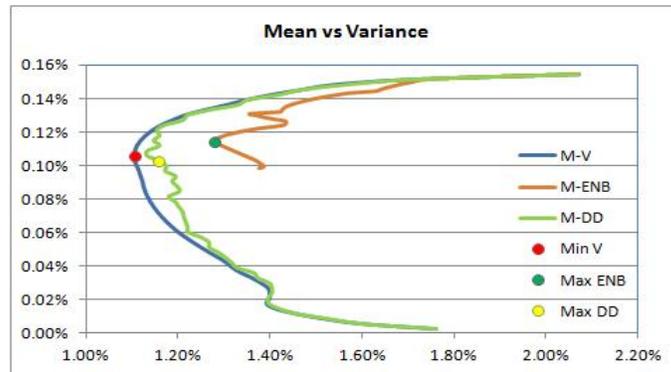


Figure 22: Mean-diversification efficient frontiers in mean-ENB co-ordinates.

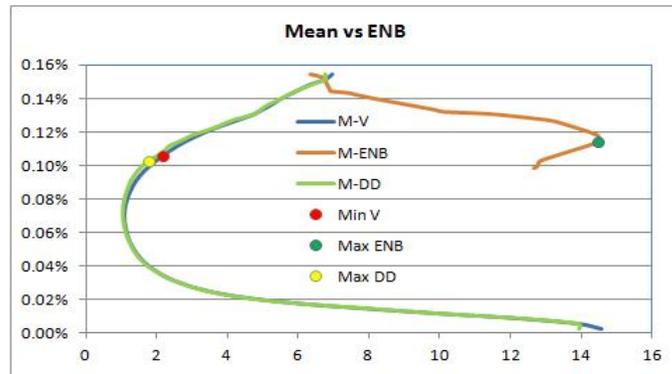


Figure 23: Mean-diversification efficient frontiers in mean-DD co-ordinates.

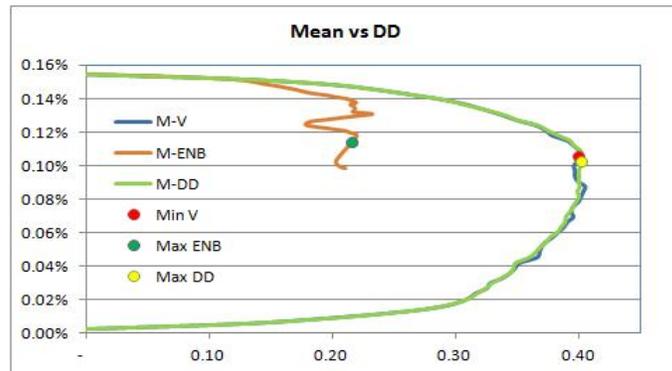


Figure 24: Mean-diversification efficient frontiers in mean-ENB co-ordinates.

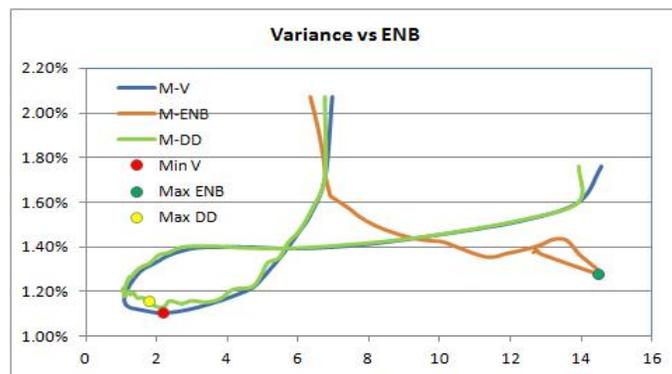


Figure 25: Mean-diversification efficient frontiers in variance-ENB co-ordinates.

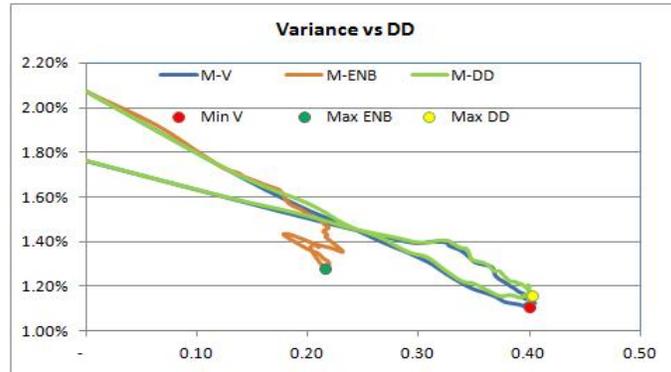
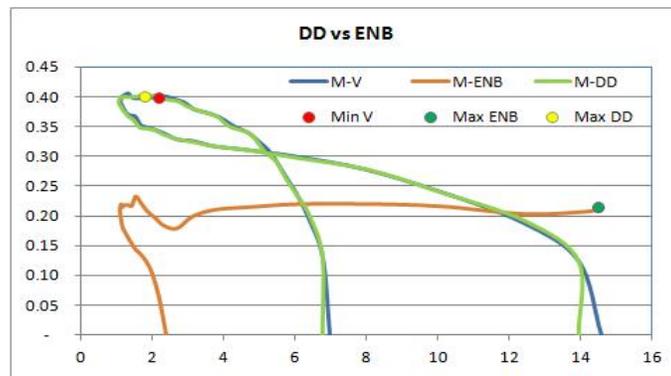


Figure 26: Mean-diversification efficient frontiers in diversification delta-ENB co-ordinates.



Evidence from Figure 21 alone already illuminates the difference between the ENB measure and the others, along with the surprising similarity between the variance and diversification delta measures. Intuitively, based on the notion that ENB is effectively a lower moment measure, one would have expected that the structure and behaviour of the optimal ENB portfolio would closely resemble the minimum variance portfolio. Nonetheless, the eigenvalue decomposition of the asset return space into uncorrelated portfolios appears to incorporate distributional information that is understated by the other two measures. Of course this assertion is not well-founded, being based solely on visual evidence of a single snap shot of an emerging equity market - it could be the case that incorporating the entire informational content of the return sample via principal component analysis actually imparts a large degree of noise into the analysis, hence imparting spuriousness into the inferences that one can make from the ENB measure. Nonetheless, these results cannot be completely overlooked either - the structure of the mean-ENB index is consistent with that presented by Meucci (2009), and while the mean-DD frontier

has not appeared in the literature, neither theoretically not practically, it is difficult to believe that the structure would deviate significantly from that presented here, in general.

In the next section, an interrogation of the longitudinal performance of these measures is effected, through the construction of maximum diversification portfolios. This section is drawn to conclusion with the presentation of the portfolio weight composition and diversification distribution of the respective optimal portfolios.

Figure 27: The portfolio weight distribution of the optimally diversified portfolios across the 27 stocks.

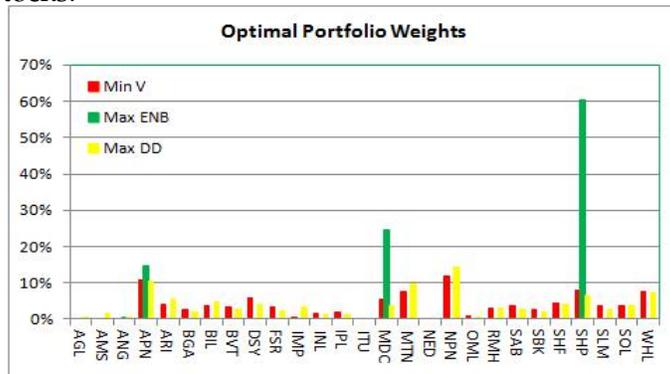
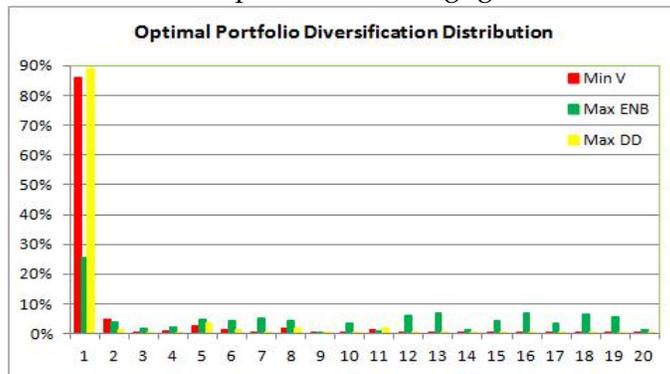


Figure 28: The diversification distribution of the optimally diversified portfolios across the 27 uncorrelated portfolios - only 20 are depicted here, since the contribution to the last 7 uncorrelated portfolios are negligible.

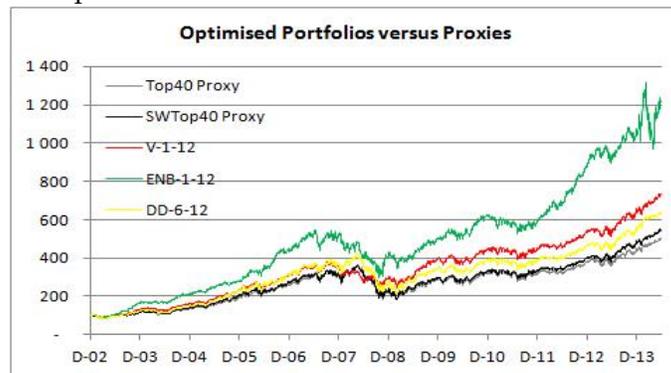


It is interesting to note that the diversification optimal ENB portfolio has a concentrated weight structure, with the portfolio only comprising of 3 stocks, viz. SHP (60.45%), MDC (24.58%) and APN (14.64%), yet it has the best diversification distribution.

6.3 Optimised Portfolios

In the previous section, the cross-sectional problem of generating mean-diversification efficient frontiers at a specific point in time was considered. Here the longitudinal problem of generating optimally diversified portfolios through time is considered. To this end, we use the same data set as before, i.e. a 27 stock universe consisting of daily data ranging from January 2002 to June 2014. Again we focus on just three key measures of diversification, viz. variance, ENB and diversification delta. Five different estimation tenors, 1-month, 3-month, 6-month, 9-month and 12-month, were considered and diversification optimal portfolios were computed-based on each the above-mentioned measures on a rolling monthly basis from January 2003 up until June 2014. Five different rebalancing frequencies, 1-month, 3-month, 6-month, 9-month and 12-month, were then also considered in the out-of-sample testing of the optimised portfolios. This created 75 portfolios in total, 25 based on each diversification measure. For the sake of brevity, only 3 portfolios are presented here, one for each diversification measure, with these reflecting the best risk-adjusted performance characteristics over the entire out-of-sample period. The graph below reflects the performance of each of these portfolios along with the proxied versions of the FTSE/JSE Top40 and Shareholder Weighted Top40 indices mentioned before.

Figure 29: Out-of-sample performance of the best risk-adjusted diversification optimised portfolios for each measure versus the proxied FTSE/JSE Top40 and Shareholder Weighted Top40 indices.



The best risk-adjusted variance- and ENB-based portfolios are optimised using a 1-month estimation period, while the DD-based portfolio is estimated using a 6-month period, all of the portfolios are rebalanced annually. The short estimation period for the variance- and ENB-based portfolios suggests that there may be a significant amount of volatility in the estimated weight distributions over time, due to substantial short-term volatility effects in the South African equity market. This is indeed true, as is reflected in the graphs below.

Figure 30: Weight distribution through time of the variance-based optimal portfolio.

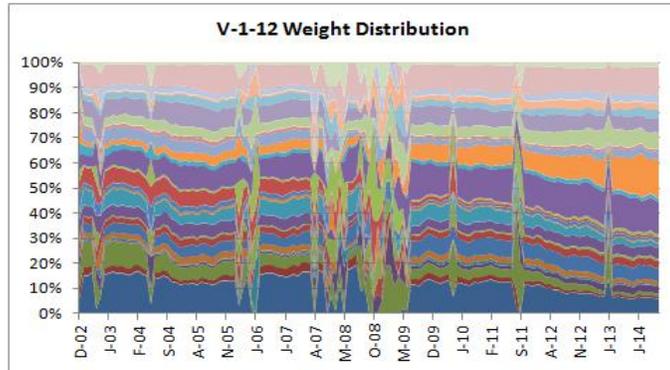


Figure 31: Weight distribution through time of the ENB-based optimal portfolio.

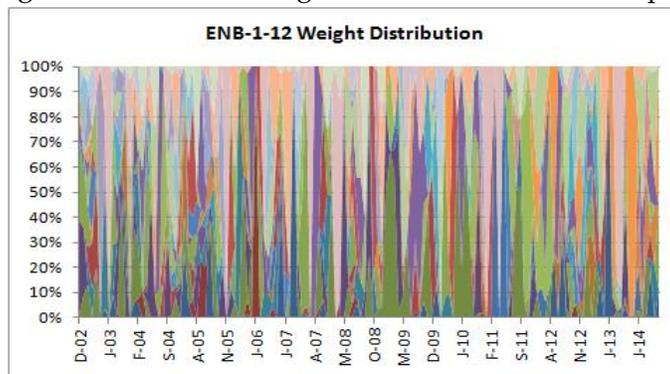
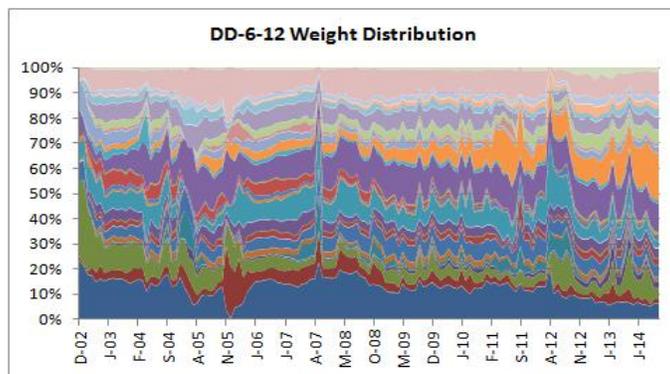
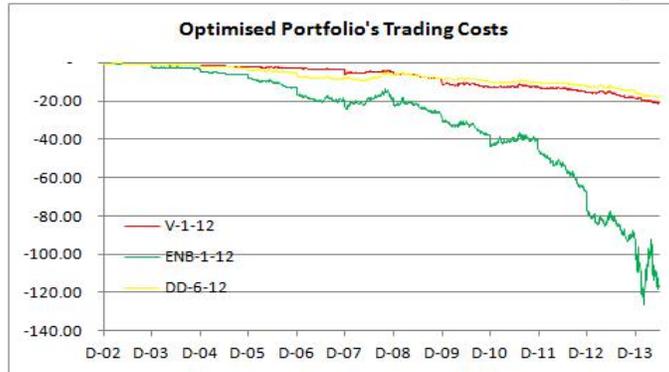


Figure 32: Weight distribution through time of the diversification delta-based optimal portfolio.



As the performance graph clearly reflects; the ENB-optimised portfolio outperforms the other two portfolios significantly, however, the volatility in the weight distribution of the ENB-optimised portfolio is substantial. This seems to be a feature of the ENB-based methodology, as volatility does not reduce significantly as one increases the length of the estimation period. In order to assess the effect of such volatile swings in the weight distribution, the graph below reflects an estimation of the trading costs involved - these costs are incurred annually in the simulation, upon rebalancing of each of the portfolios. As one would expect, the trading costs of the ENB-based portfolio dwarfs that of the other two, with the reason for this being two-fold: (i) the volatile weight structure; and (ii) the significant capital growth in the portfolio over time. Both of these factors contribute to higher trading costs at points of rebalancing. In order to generate the graph below, an assumption that each purchase and sale of any quantity of a share would attract a cost of 50 basis points - in reality, this is akin to bid-ask spread costs and brokerage, amongst other fees.

Figure 33: Estimated traded cost on each of the diversification optimised portfolios.



It is fairly evident, that even after an adjustment for trading costs, the ENB-based portfolio still outperforms the other two. It is intriguing that the ENB-based portfolio has exhibited the best risk-adjusted performance, despite the volatile optimal weight distribution structure. Moreover, one should also note that the DD-based portfolio underperforms the variance-based portfolio. Accordingly, these longitudinal results corroborate well with the results of the cross-sectional report in the previous section, which seems to suggest that the ENB is a better measure of diversification than the diversification delta. Of course this is circumstantial evidence and may just be particular to the given data set - nonetheless, there appears to be sufficient indications that one should, at the least, not discount measures such as the PDI or the ENB when assessing the risk of a portfolio of financial assets.

7 Consequences and Conclusions

This paper has investigated the landscape of existing risk and diversification measures for portfolio selection, beginning from naive weight-based measures, and then moving on to higher order moment measures. Qualitatively, there is little difference between the weight-based approaches of measuring risk as they lead to the same decision regarding the selection of stocks. These weight-based measures are heavily dependent on the weights and do not account for the differing risk-return characteristics of the portfolio. Covariance-based measures give a broader understanding of diversification as they link diversification with the risk-return characteristics of the stocks. However, although the covariance-based measures give an intuitive feel of the benefits of diversification, they are highly volatile measures when correlations between returns are high. To mitigate the volatility of covariance-based measures, entropy-based measures were considered which took into account the higher order moments of measurement.

There are several different approaches in the literature with respect to the measurement of entropy. In this investigation, three techniques were considered: the histogram density estimate approach, the kernel density estimate approach and the k - d partitioning method. Issues pertaining to the subjectivity associated with these estimates of entropy were discussed, in particular, the problems of selecting an optimal bin size and/or bin interval in the case of the histogram density estimate approach, and selecting an optimal smoothing parameter h in the case of the kernel density estimate approach were discussed. For the purposes of the analysis, these estimates were selected using rules-of-thumb.

Each of the entropy estimate measures were used to estimate the diversification delta using South African financial market data. The interesting result is that although the entropy estimates measured using the three different approaches are incomparable, the estimates of the diversification delta are consistent and comparable across all three approaches. This is important as it somewhat eliminates the subjectivity associated with estimating entropy.

A thorough analysis of all the diversification measures mentioned in this paper was conducted, and the following were chosen to be compared: variance, effective number of bets and diversification delta. Since the measures are directly incomparable, a "diversification index" was created for each of the measures and the returns on the indices were measured. The three measures gave consistent results, although the variance appeared to have the most erratic behaviour, while the diversification delta appeared to be the most stable. The reason suggested for this result is that the diversification delta takes higher moment measures into account, while the variance does not, thus allowing for more information, and is also less easily affected by outliers.

However, given the circumstantial evidence provided by the cross-sectional and longitudinal analysis in Sections 6.2 and 6.3, it appears as though the diver-

sification delta measure does not provide a significant amount of additional information regarding the distributional structure of the market, in order to improve one's assessment of diversification for **linear assets**. Nonetheless, a mathematically rigorous proof of this assertion is required to solidify this notion. The Effective Number of Bets measure interprets the concept of diversification in an almost polar opposite manner, in comparison to the traditional and common measures of dispersion. Moreover, the decomposition of the empirical sample return distribution into a set of orthogonal or *uncorrelated portfolios*, with the aim of structuring a portfolio in such a way to firstly meet the prespecified budget, mandate or other market-related constraints and secondly to optimise the distribution of your portfolio exposure across the aforementioned uncorrelated portfolios, seems counter-intuitive. One could actually be optimising toward idiosyncratic exposures that are a temporary manifestation of anomalous market behaviour. Therefore, it is difficult to confidently advocate that the Effective Number of Bets measure is indeed a better measure of diversification than the diversification delta. The only assertion that can be made with any semblance of confidence, is the fact that these two measures provides one with tools to quantify and analyse diversification from almost diametrically opposing viewpoints - thereby enhancing and illuminating the overall risk management process.

Bibliography

- Bierlant, J., Dudewicz, E. J., Györfi, L., van der Meulen, E. C., 1997. Nonparametric entropy estimation: an overview. *International Journal of Mathematical and Statistical Science* 6 (1), 17–39.
- Ebrahimi, N., Maasoumi, E., Soofi, E., 1999. Ordering univariate distributions by entropy and variance. *Journal of Econometrics* 90 (2), 317–336.
- Kirchner, U., Zunckel, C., 2011. Measuring portfolio diversification. arXiv preprint arXiv:1102.4722.
- Livingston, L. S., 2013. Intraportfolio correlation: an application for investments students. *Business Education and Accreditation* 5 (1), 91–105.
- Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7, 77–91.
- Meucci, A., 2009. Measuring portfolio diversification. *Risk* May 2009, 74–79.
- Michalowicz, J., Nichols, J., Bucholtz, F., 2008. Calculation of differential entropy for a mixed gaussian distribution. *Entropy* 10 (3), 200–206.
- Miller, E. G., 2003. A new class of entropy estimators for multi-dimensional densities. In: *International Conference on Acoustics, Speech and Signal Processing*. Vol. 3. pp. 297–300.
- Raykar, V., Duraiswami, R., 2006. Fast optimal bandwidth selection for kernel density estimation. *SDM*, 524–528.
- Samuelson, P., 1967. General proof that diversification pays. *Journal of Financial and Quantitative Analysis* 2 (1), 1–13.
- Shannon, C. E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 379–423, 623–656.
- Silverman, B. W., 1986. *Density estimation for statistics and data analysis*. CRC Press.

- Stowell, D., Plumbley, M. D., 2009. Fast multidimensional entropy estimation by k-d partitioning. *IEEE Signal Processing Letters* 16(6), 537–540.
- Vermorken, M., Medda, F., Schröder, T., 2012. The diversification delta: A higher moment measure for portfolio diversification. *The Journal of Portfolio Management* 39(1), 67–74.