ABSTRACT

Title of Dissertation: TRADING OPTION MODEL PARAMETERS AND CLIQUET PRICING USING OPTIMAL TRANSPORT

Shahnawaz Khalid
Doctor of Philosophy, 2022

Dissertation Directed by: Professor Dilip B. Madan
Department of Finance

This dissertation consists of two independent topics. Chapter 1 titled, "Trading Option Model Parameters" describes two methods of constructing a portfolio of vanilla options that is sensitive to only one parameter for any kind of option pricing model. These special portfolios can be constructed for any parameter and move in the same direction as that specific parameter, while being resistant to changes in all others. We use the Variance Gamma model and Bilateral Gamma model as examples and show both methods yield portfolios with similar payoff structure at maturity. In addition we show that the value of these portfolios remains unchanged when all but one parameter is perturbed. We conclude by assessing the viability of using these methods as a trading or hedging strategy. Chapter 2 titled "Pricing Cliquets using Martingale Optimal Transport" applies the theory of Optimal Transport to pricing forward starts and cliquets. We develop models based on relative entropy minimization that provide close fits to market data using information based on just the marginal distributions. We prove a duality result that provides an explicit form of the optimal distribution. Furthermore we provide an algorithm and a convergence

result for iteratively computing the dual. Chapter 3 titled “Martingale Optimal Transport under Acceptability” addresses the issue of narrowing the no arbitrage price bounds for a cliquet by introducing the concept of acceptable risk. We prove a duality result based on acceptability and show how to numerically compute acceptable bounds.

TRADING OPTION MODEL PARAMETERS AND CLIQUET PRICING USING OPTIMAL TRANSPORT

by

Shahnawaz Khalid

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Dilip B. Madan, Chair/Advisor
Professor Steve Heston
Professor Mark Loewenstein
Professor Alexander Barg
Professor Charles D. Levermore

Dedication

To Ami and Abu. For their uncountably infinite love and support.

Acknowledgments

A great debt of gratitude is owed to my PhD advisor, Professor Dilip Madan, for being exceedingly generous with his time and help throughout my PhD. He was always willing to meet me and answer my (sometimes trivial!) questions whenever I asked him. I have always felt inspired by the depth of his knowledge, sharp insights and infectious passion for research. I've learnt a tremendous amount from him and am immensely grateful to him for his help outside of university and towards my career as well.

I wish to express my deepest thanks to Professors Steve Heston, Mark Loewenstein, Dave Levermore and Alexander Barg for sparing their time and agreeing to serve on my committee. Their valuable suggestions have helped improve the quality of this dissertation. I am also very grateful to my friends and fellow Math Finance RIT members; Sahil, Zhang Zhang, Yiran, Yoshihiro and Ran. They made the weekly RITs a great learning experience for me.

I would like to acknowledge Professors Asad Abidi, Nabil Mustafa, Arnulf Jenstzen and Stephen Wu who mentored and encourage me at different stages of my life and ultimately lead me to pursue a PhD in the first place.

In addition I want to thank Shareef, Ami, Abu, Dado, Nani for their unwavering support during this six journey. Their love and prayers have helped me keep moving forward. I am also indebted to my wife Anam for the support and motivation she has provided me. She is an inspiration to me by always leading by example.

I also wish to thank my friends, Zeyad, Kyle, Yuchen, Anqi, Sarah, Eric, Tessa, Jordan and other students because of whom my time at College Park was an enjoyable one. I also want to thank my students, especially from Math 140 Fall 2016 and Math 246 Fall 2021, for making me realize how much I love teaching and greatly enhancing my PhD experience.

Lastly, I wish to thank all the family and friends who have kept me in their prayers during these six years and I am very grateful to God for helping me at every step.

Table of Contents

List of Tables

List of Figures

Chapter 1: Trading Option Model Parameters

1.1 Introduction

Over a span of few decades, there has been an explosive increase in option trading and other derivatives. We now find up to 3000 options being traded each day for an index option. This has led to a vast literature on option pricing models. To adequately describe the variation in option prices in strike and maturity, models can require numerous parameters that need to be calibrated to the market data. Examples of popular models are stochastic volatility ([1], [2]), local volatility ([3]), jump-diffusion ([4], [5]) and pure jump models ([6], [7]). The day to day evolution of model parameters is an important consideration in knowing how the options surface behaves.

In this chapter, we are interested in constructing portfolios that are sensitive to only one specific parameter for any given option pricing model. We want the value of such a portfolio to change in the same direction as a change in the selected parameter while being resistant to changes in the rest. This method can be used to develop a trading strategy assuming we can accurately predict the moves of the model parameters. In addition it can also provide an important tool to hedge against large deviations in parameters.

We study two ways of constructing these portfolios: a regression based approach and one based on using the score functions of the risk neutral density. We restrict ourselves to two models:

Variance Gamma [6] and Bilateral Gamma model [7]. Both methods use Lévy processes as a model for the underlier and are examples of a pure jump model. Pricing under both models can be efficiently using the fast Fourier method [8].

1.2 Option pricing models

1.2.1 Lévy Processes

We use Lévy processes to model the dynamics of the underlying asset. Lévy processes are stochastic processes with independent and stationary increments that generalize the notion of Brownian motion. Every Lévy process has a càdlàg modification and we always work with such a version of the process.

Definition 1.2.1. *A càdlàg stochastic process $\{X_t, t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $X_0 = 0$ is called a Lévy process if it satisfies:*

1. *For any $n \in \mathbb{N}$ and $0 < t_1 < t_2 < \ldots < t_n$, the increments $X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{(n-1)}}$ are independent.*

2. *For any $s, t \in (0, \infty)$ the distribution of $X_{t+s} - X_s$ does not depend on t.*

3. *For every $t \in (0, \infty)$ and $\epsilon > 0$ it holds that $\lim_{s \to t} \mathbb{P}(|X_t - X_s| > \epsilon)$ = 0.*

Lévy processes have the important property that for any $t \in (0, \infty)$, X_t has an *infinitely divisible* distribution.

Definition 1.2.2. *A random variable X has an infinitely divisible distribution if for all $n \in \mathbb{N}$,* *there exist i.i.d random variables* $X_1^{(1/n)}, \ldots, X_n^{(1/n)}$ *such that*

$$X =^{law} X_1^{(1/n)} + \ldots + X_n^{(1/n)}$$

Equivalently, let $\phi_X(u) := \mathbb{E}[e^{iuX}]$ *be the characteristic function of* X. *Then* X *has an infinitely divisible distribution if for all* $n \in \mathbb{N}$ *there exists a random variable* $X^{(1/n)}$ *such that*

$$\phi_X(u) = \left(\phi_{X^{(1/n)}}(u)\right)^n.$$

For a Lévy process $\{X_t, t \geq 0\}$, infinite divisibility implies that for any $t \in (0, \infty)$, it holds that

$$\phi_{X_t}(u) = (\phi_{X_1}(u))^t.$$

Another important property of Lévy processes is their representation via the *Lévy-Khintchine* formulation.

Theorem 1. *Let* $\{X_t, t \geq 0\}$ *be a Lévy process. Then there exists a triplet* (b, c^2, ν)*, with* $b \in \mathbb{R}$*,* $c^2 \in \mathbb{R}_{\geq 0}$ *and a measure* ν *with the property that* $\nu(\{0\}) = 0$ *and* $\int_{\mathbb{R}}(1 \wedge |x|^2)\nu(dx)$ *such that it holds that*

$$\mathbb{E}[e^{iuX_t}] = \mathbb{E}[e^{tiuX_1}] = \exp\left(t\left(ibu - \frac{u^2c^2}{2} + \int_{\mathbb{R}}(e^{iux} - 1 - iux1_{|x|<1})\nu(dx)\right)\right).$$

The Lévy-Khintchine formula allows us to decompose a Lévy process into three independent parts: a linear deterministic drift with rate b, a Brownian motion with volatility c and a pure jump process with Lévy measure $\nu(dx)$. The Lévy measure dictates how jumps occur, i.e., jumps of size in the set A occur according to a Poisson process with intensity $\int_A \nu(dx)$.

In this study, we will model the underlier with processes without a diffusion component. [9] suggest that while price processes must have a jump component, the diffusion component is not necessary. A Lévy process with infinite activity, i.e., $\int_{\mathbb{R}} \nu(dx) = \infty$ is able to account for the frequent small moves that are typically modelled by a diffusion.

1.2.2 Using Lévy processes to model underlier

We can model the process of the underlying asset using a Lévy process. In the same spirit as the Black-Scholes model, we assume that the log returns of an underlying $\{S_t, t \geq 0\}$ are Lévy:

$$S_t = S_0 e^{X_t}.$$

In order to price derivatives, we must use the representation of S_t under the risk-neutral measure. In contrast to Black-Scholes, the market is incomplete under Lévy models (with the exception of Poisson process). This means that there is no unique risk-neutral measure.

One convenient way of obtaining a risk-neutral measure is by mean correcting the exponential of Lévy process so that the expectation of the forward at any time is simply the spot price. Hence it holds that the risk-neutral dynamics of $\{S_t, t \geq 0\}$, with spot S_0, interest rate r and dividend yield q, under a Lévy process $\{X_t, t \geq 0\}$ is given by

$$S_t = S_0 e^{(r-q)t-\omega+X_t}$$

where $\omega := \ln \mathbb{E}[e^{iuX_t}] = \ln \phi_{X_t}(-i)$.

1.2.3 Variance Gamma model

The variance gamma process [6] is a three parameter pure jump process with infinite activity. It is defined by time changing a Brownian motion with drift θ and volatility σ by an independent gamma process with unit mean rate, and variance rate ν. More precisely, let $W(t)$ be a standard Brownian motion and let $G(t;\nu)$ be an independent gamma process with unit mean rate and ν variance rate. Then the variance gamma process $X_{VG}(t;\sigma,\theta,\nu)$ is defined as

$$X_{VG}(t;\sigma,\theta,\nu) = \theta G(t;\nu) + \sigma W(G(t;\nu)).$$

The characteristic function for $X_{VG}(t;\sigma,\theta,\nu)$ is given by

$$\phi_{VG}(u;\sigma,\theta,\nu,t) = \left(1 - iu\theta\nu + \frac{\sigma^2 u^2}{2}\nu\right)^{-\frac{t}{\nu}}.$$

Let $\Gamma(z)$ denote the Gamma function and let $K_\nu(z)$ be the modified Bessel function of the second kind defined as

$$K_\nu(z) = \frac{\Gamma(\nu+\frac{1}{2})(2z)^\nu}{\sqrt{\pi}} \int_0^\infty \frac{\cos(t)}{(t^2+z^2)^{\nu+\frac{1}{2}}} dt.$$

Then, the density function for $X_{VG}(t;\sigma,\theta,\nu)$, defined for $x \in (-\infty,\infty)$ is given by

$$f_{VG}(x;\sigma,\theta,\nu,t) = \frac{2\, e^{\frac{\theta x}{\sigma^2}}}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})\sigma\sqrt{2\pi}} \left(\frac{x^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{(\frac{t}{\nu}-\frac{1}{2})}\left(\frac{1}{\sigma^2}\sqrt{x^2(\frac{2\sigma^2}{\nu}+\sigma^2)}\right).$$

The variance gamma model is flexible enough to capture the skewness and excess kurtosis found in market data and is hence able to provide much better fits compared to Black-Scholes.

1.2.4 Bilateral Gamma model

The Bilateral Gamma process introduced in [7] is another pure jump Levy process, characterized by four parameters, that is the difference of two independent Gamma subordinators. [10] show that the upwards and downwards motion of the stock are significantly different and recommend using the bilateral Gamma process as it gives control over the speed of the up and down moves.

The characteristic function for a variable with bilateral Gamma distribution at time t, $\phi_{BG}(u; b_p, c_p, b_n, c_n, t)$ is given by

$$\phi_{BG}(u; b_p, c_p, b_n, c_n, t) = (1 + ib_n u)^{-c_n t}(1 - ib_p u)^{-c_p t}.$$

Let $W_{\lambda,\mu}(z)$ denote the Whittaker function given by

$$W_{\lambda,\mu}(z) = \frac{z^{\lambda} e^{-\frac{z}{2}}}{\Gamma(\mu - \lambda + \frac{1}{2})} \int_0^{\infty} t^{\mu-\lambda-\frac{1}{2}} e^{-t\left(1+\frac{t}{2}\right)^{\mu+\lambda-\frac{1}{2}}} dt$$

where $\mu - \lambda > \frac{1}{2}$. The density $f_{BG}(x; b_p, c_p, b_n, c_n, t)$ for a bilateral Gamma random variable at time t is defined on $x \in (0, \infty)$ as

$$\begin{aligned} f_{BG}(x; b_p, c_p, b_n, c_n, t) =& \frac{b_p^{-c_p t}\, b_n^{-c_n t}}{(\frac{1}{b_p} + \frac{1}{b_n})^{\frac{1}{2}(c_p t + c_n t)}\, \Gamma(c_p t)}\, x^{\frac{1}{2}(c_p t + c_n t) - 1}\, e^{-\frac{x}{2}(\frac{1}{b_p} - \frac{1}{b_n})} \\ & \times W_{\frac{1}{2}(c_p t - c_n t), \frac{1}{2}(c_p t + c_n t - 1)}\left(x\left(\frac{1}{b_n} + \frac{1}{b_p}\right)\right). \end{aligned}$$

$f_{BG}(x; b_p, c_p, b_n, c_n, t)$ can be extended to $(-\infty, 0)$ using the symmetry relation

$$f(x; b_p, c_p, b_n, c_n, t) = f(-x; b_n, c_n, b_p, c_p, t).$$

1.2.5 Pricing using Fast Fourier transform

Pricing options under the variance gamma and bilateral gamma model using their density is computationally expensive and becomes infeasible when calibrating option surfaces. However, they both posses simple characteristic functions which makes option pricing possible by Fourier inversion. We use the *Carr-Madan* [8] formula to write the price of a call option in terms of the characteristic function. The Fast-Fourier Transform [11] can then be used to compute the integrals which prices options on multiple strikes in one run, making it especially useful during model calibration.

Let K be the strike and let $k = \ln(K)$. Let $\phi(u; T)$ and $q(x; T)$ be the characteristic function and risk neutral density of $\ln(S_T)$ respectively. The risk neutral price $C_T(k)$ of a call option at strike k is given by

$$C_T(k) = e^{-rT} \int_{-\infty}^{\infty} (e^x - e^k) q(x; T) dx.$$

The function $C_T(k)$ is not square integrable, which is necessary for Fourier inversion. Therefore, for $\alpha \in (0, \infty)$ we introduce a dampening factor $e^{\alpha k}$ and define the modified call price $c(k; T)$ to be

$$c(k; T) = e^{\alpha k}\ C_T(k; T).$$

The Fourier transform $\mathcal{F}_c(u)$ of $c(k;T)$ can hence be computed as

$$\mathcal{F}_c(u) = \int_{-\infty}^{\infty} e^{iuk} c(k;T)\, dk$$
$$= e^{-rT} \frac{\phi(u - i(\alpha+1);T)}{\alpha^2 + \alpha - u^2 + iu(2\alpha+1)}$$

Therefore, we can express $C(k;T)$ as

$$C(k;T) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} e^{iuk} \mathcal{F}_c(u) du.$$

Let $\{0, \eta, 2\eta, \ldots, (N-1)\eta\}$ be an N points grid that we use to approximate the integral using simpson's rule. The integral can be approximated by

$$C(k,T) \approx \frac{e^{-\alpha k}}{\pi} \sum_{j=1}^{N} e^{(-iu_j k)} \mathcal{F}(u_j) \frac{\eta}{3} \left(3 + (-1)^j - \delta_{j-1}\right), \quad u_j := \eta\,(j-1).$$

To calculate option prices on multiple strikes, we let the log strike k range from $-b$ to b on N grid points given by

$$k_n = -b + \lambda(n-1), \quad n = 1, \ldots, N$$

where $\lambda = 2b/N$. By choosing λ and η to be $\lambda\eta = 2\pi/N$ and N to be a power of 2 we can express as (1.2.5) as

$$C(k_n,T) \approx \frac{e^{-\alpha k_n}}{\pi} \sum_{j=1}^{N} e^{-\frac{2\pi i}{N}(j-1)(n-1)}\, e^{ibu_j} \mathcal{F}(u_j) \frac{\eta}{3} \left(3 + (-1)^j - \delta_{j-1}\right).$$

which can readily be solved using the Fast Fourier Transform. In our experiments we take $N = 2^{12}$, $\alpha = 1.5$ and $\eta = 0.25$.

1.3 Building parameter sensitive portfolios

We now consider the problem of building portfolios that are sensitive to moves in only one model parameter. This means we need the partial derivative of the portfolio value to be positive for a chosen parameter and zero for all other parameters. We explore two such methods and apply them to the variance gamma and bilateral gamma models.

1.3.1 Regression based approach

Let $\mathcal{M}(\theta)$ be an option pricing model characterized by m model parameters $\theta \in \mathbb{R}^m$. Let $w(\theta) \in \mathbb{R}^n$ be a collection of n vanilla options (calls and puts) that are priced under $\mathcal{M}(\theta)$. Let $\alpha \in \mathbb{R}^n$ denote the weight of each option in a portfolio. Then the value of the portfolio under $\mathcal{M}(\theta)$ is given by $\alpha^T \cdot w(\theta)$.

For a specific model parameter $\theta_i \in \mathbb{R}$, we wish to find weights $\alpha_i \in \mathbb{R}^n$ with the property that

$$\frac{\partial}{\partial \theta_j} \left(\alpha_i^T \cdot w(\theta) \right) = \delta_{i,j}, \quad j \in \{1, \ldots, m\}.$$

We can do this by considering $A = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T \in \mathbb{R}^{n \times m}$ to be an $n \times m$ matrix with the property that

$$A \frac{\partial w}{\partial \theta} = \mathbb{I}, \quad \text{with } \frac{\partial w}{\partial \theta} = \begin{bmatrix} \frac{\partial w_1}{\partial \theta_1} & \cdots & \frac{\partial w_1}{\partial \theta_m} \\ \vdots & \ddots & \\ \frac{\partial w_n}{\partial \theta_1} & & \frac{\partial w_n}{\partial \theta_m} \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

This ensures that each α_i in A corresponds to the weights of a portfolio that is only sensitive to

θ_i. A can then be easily computed by

$$A = \left(\frac{\partial w}{\partial \theta}^T \frac{\partial w}{\partial \theta} \right)^{-1} \frac{\partial w}{\partial \theta}^T.$$

1.3.2 Score function based approach

In this method we first seek a terminal payoff c which is sensitive to changes in only one model parameter. By replicating this payoff using hedging instruments we are then able to get the desired portfolio. Our approach is similar to the Gram-Schmidt orthogonalization. We assume that we have an explicit representation of the risk neutral density in the form of $f(x;\theta)$, where $\theta \in \mathbb{R}^m$ are m model parameters. For a given claim with a discounted payoff c at maturity, it's present value $w(\theta)$ is given by

$$w(\theta) = \mathbb{E}[c] = \int c(x) f(x;\theta) dx.$$

Taking the partial derivative with respect to some θ_i it holds that

$$\begin{aligned}
\frac{\partial w}{\partial \theta_i} &= \int c(x) \frac{\partial f(x;\theta)}{\partial \theta_i} dx \\
&= \int c(x) \frac{\partial \ln f(x;\theta)}{\partial \theta_i} f(x;\theta) dx \\
&= \mathbb{E}\left[c(x) \frac{\partial \ln f(x;\theta)}{\partial \theta_i} \right].
\end{aligned}$$

where the interchange of the derivative and integral is justified by the assuming that $f(x;\theta)$ is continuously differentiable locally for each θ_i. Note that for a payoff $c(x) = \frac{\partial \ln f(x;\theta)}{\partial \theta_i}$ (known as the score function) the expectation is always non-negative which implies that the value $w(\theta)$ for such a payoff always moves in the same direction as a change in θ_i.

In order for $w(\theta)$ to be resistant to changes in other parameters we seek $c(x)$ such that

$$\mathbb{E}\left[c(x)\frac{\partial \ln f(x;\theta)}{\partial \theta_i}\right] = 1 \quad \text{and} \quad \mathbb{E}\left[c(x)\frac{\partial \ln f(x;\theta)}{\partial \theta_j}\right] = 0,\ \forall j \neq i. \tag{1.1}$$

Let $c(x)$ be of the form

$$c(x) = \sum_{j=1}^{m} \beta_j \frac{\partial \ln f(x;\theta)}{\partial \theta_j}, \quad \beta_j \in \mathbb{R}.$$

We can re-frame (1.3.1) as a system of equations to find coefficients $\{\beta_j\}_{j=1}^m$ that satisfy

$$\sum_{j=1}^{m} \beta_j \mathbb{E}\left[\frac{\partial \ln f(x;\theta)}{\partial \theta_j} \cdot \frac{\partial \ln f(x;\theta)}{\partial \theta_i}\right] = \delta_{i,j}.$$

The value of the resulting payoff $c(x)$ therefore moves in the same direction as θ_i and does not change with respect to any other parameter. We can formalize the above arguments using the proposition below.

Proposition 1.3.1. *Let $f(x;\theta)$ be the risk neutral density of an m-parameter model $\mathcal{M}(\theta)$. Let $\mathbb{I} \in \mathbb{R}^{m\times m}$ be the identity matrix and let $B = \{\beta_j^{(i)}\} \in \mathbb{R}^{m\times m}$ be a matrix that satisfies*

$$B \cdot \mathbb{E}\left[\frac{\partial \ln f(x;\theta)}{\partial \theta} \otimes \frac{\partial \ln f(x;\theta)}{\partial \theta}\right] = \mathbb{I}. \tag{1.2}$$

Let $c^{(i)}$ be a function that represent the discounted terminal payoff of a contingent claim given by

$$c^{(i)}(x) = \sum_{j=1}^{m} \beta_j^{(i)} \frac{\partial \ln f(x;\theta)}{\partial \theta_j}$$

Then the value $w^{(i)}(\theta)$ *of such a payoff satisfies*

$$\frac{\partial w^{(i)}}{\partial \theta_i} = 1 \quad \textit{and} \quad \frac{\partial w^{(i)}}{\partial \theta_j} = 0 \tag{1.3}$$

Proof. The proof is straightforward by observing that for any i, (1.2) implies that

$$\sum_{j=1}^{m} \beta_j^{(i)} \mathbb{E}\left[\frac{\partial \ln f(x;\theta)}{\partial \theta_i} \cdot \frac{\partial \ln f(x;\theta)}{\partial \theta_j}\right] = \delta_{i,j}.$$

and the expectation of the corresponding payoff $c^{(i)}$ satisfies the conditions (1.3). □

Any portfolio that replicates c will have the same properties. Therefore, since we have assumed c to be continuous we can replicate it via a hedging portfolio consisting of a bond, stock and a series of vanilla out of money options i.e.

$$c(x) = \alpha_0 + \alpha_1 x + \sum_{i=2}^{N} \alpha_i C_i(x)$$

where C_i is the payoff of options and $\{\alpha_i\}_{i=0}^N$ are the weights of the portfolio. By discretizing x on a grid we can compute the weights $\{\alpha_i\}_{i=0}^N$ using regression.

1.4 Results

1.4.1 Payoff profiles

We calibrate the Variance Gamma and Bilateral Gamma model daily on each maturity using options on S&P500. Our calibration procedure involves running the global Differential Evolution algorithm [12] for a few iterations to arrive at a suitable neighborhood of the global minimum. We then use the result of this global search as a starting point for our local optimizer. For each model we plot the payoff profiles at maturity of the portfolios resulting from both methods. For the regression approach we use 20 calls and 20 puts with strikes going up to 30% out of the money. The derivatives are computed using a central difference scheme evaluated at the parameter value for that day and we take the step size to be 1% of the parameter value. All option prices are computed using the Fast Fourier method. For the score function method, we obtain the resulting payoff from the procedure described in Proposition 1.3.1. We observe that the payoff profile for both methods are similar. In addition we did not observe any significant change from varying the maturity.

Moreover, we observe how the value of the sensitive portfolios moves if we change the value of each parameter. In almost all cases we see that the parameter specific portfolio moves linearly only when corresponding parameter changes and is resistant to changes from all the others.

1.4.2 Trading strategy

We develop a trading strategy by assuming that we know how the parameters will change the next day. This admittedly is an unrealistic assumption, however, we are primarily interested in knowing how well such a strategy would work in the best case scenario.

We calibrate the parameters for each day and for each parameter we compute the respective weights for a portfolio of 20 calls and 20 puts. We then go long the portfolio if the parameter increases the next day and short it if it falls. The next day we liquidate our positions.

The performance for such a strategy is plotted along with the daily change in the portfolio

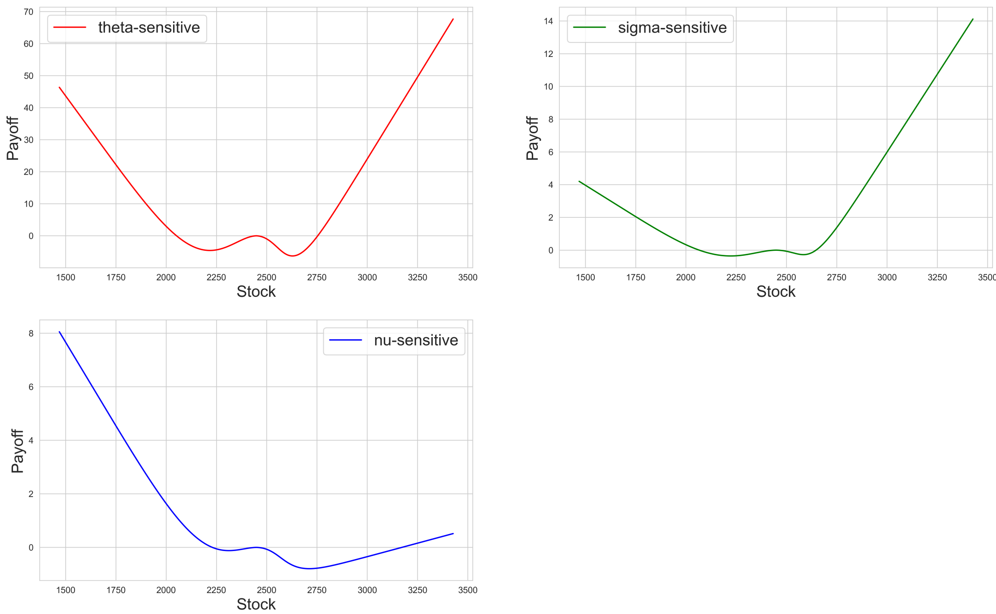


Figure 1.1: Payoff of sensitive portfolios (VG) at maturity from regression

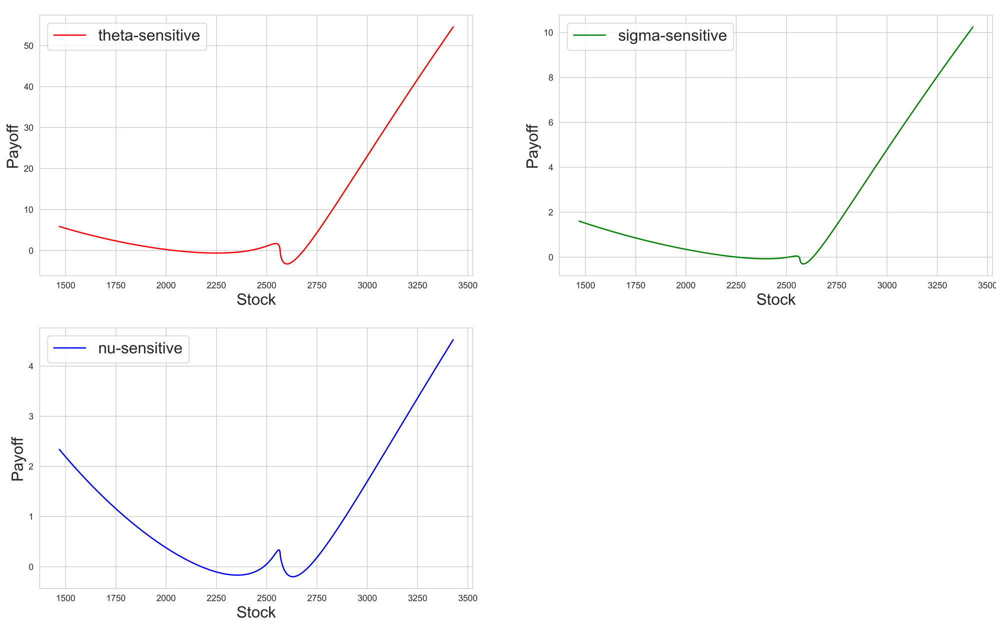


Figure 1.2: Payoff of sensitive portfolios (VG) at maturity from score function

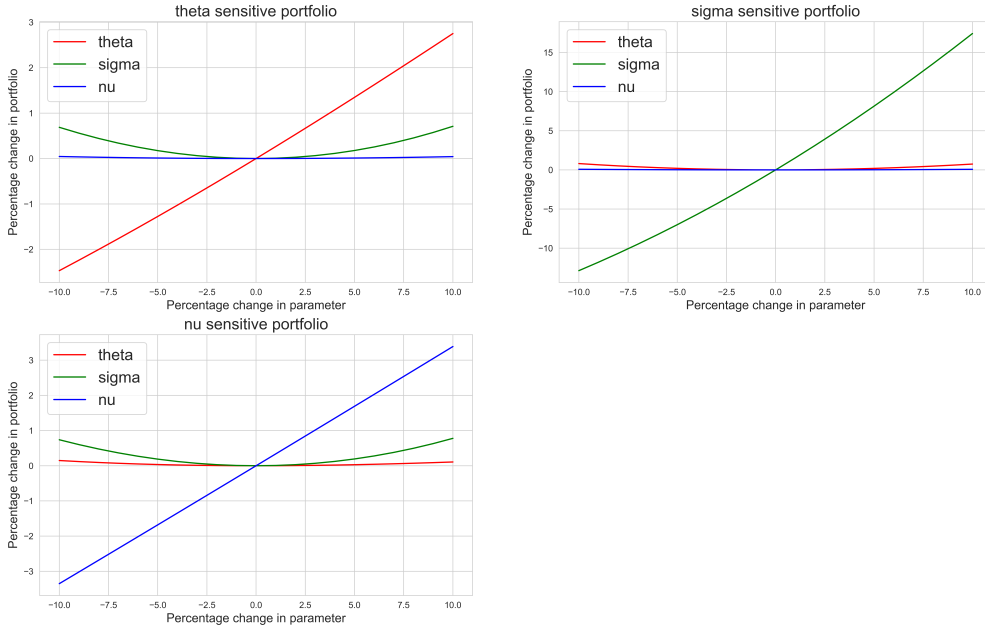


Figure 1.3: Change in portfolio values with different VG parameters (regression)

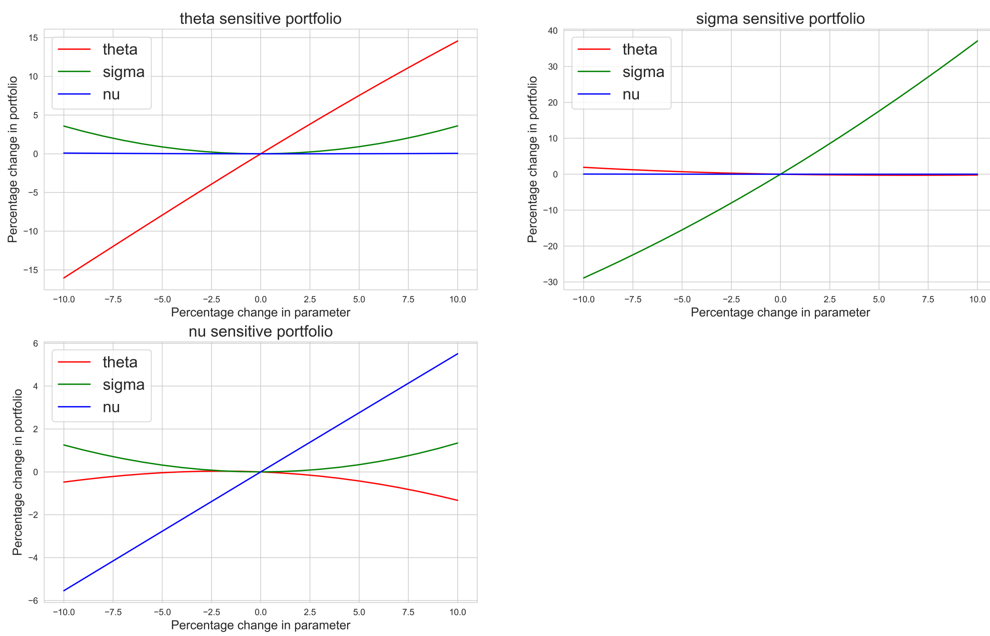


Figure 1.4: Change in portfolio values with different VG parameters (score function)

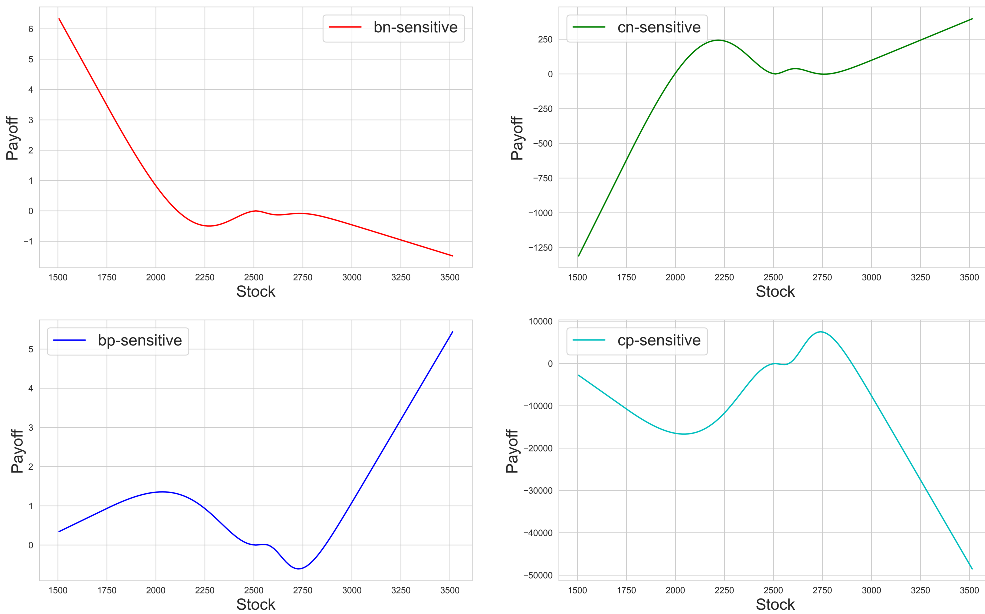


Figure 1.5: Payoff of sensitive portfolios (BG) at maturity from regression

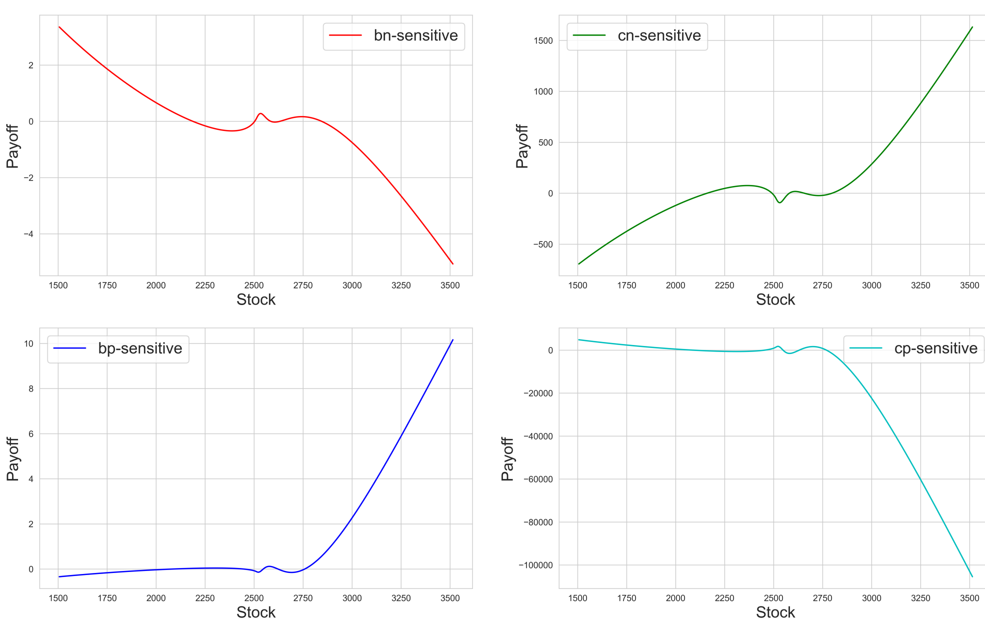


Figure 1.6: Payoff of sensitive portfolios (BG) at maturity from score function

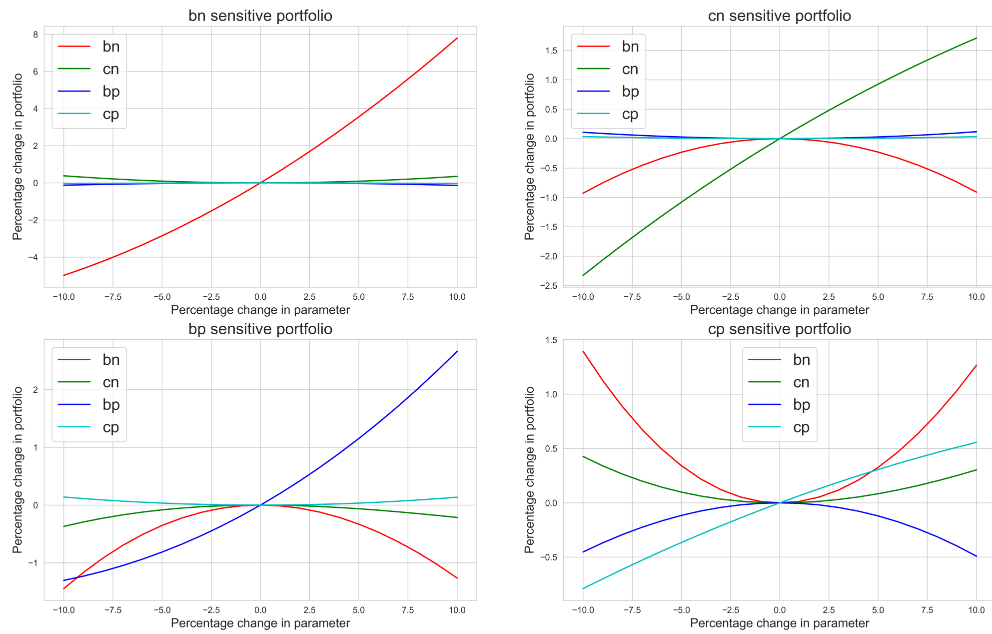


Figure 1.7: Change in portfolio values with different BG parameters (regression)

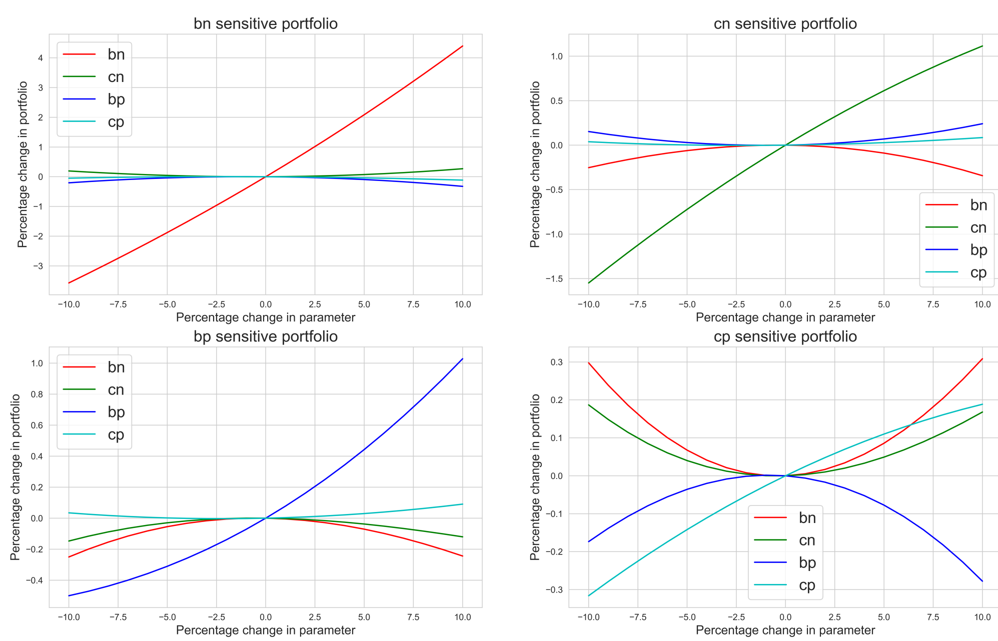


Figure 1.8: Change in portfolio values with different BG parameters (score function)

value. We observe that this strategy is only viable for certain parameters and not all parameters can be traded. Specifically, the parameters with the highest sensitivity; σ for variance gamma and b_n for bilateral gamma are good candidates.

1.5 Conclusions and Future Research

This study leaves us with a number of avenues for future research. First, an important question is whether the two approaches described in this study are equivalent. This would mean that the regression based method will approach the score function payoff as we consider options over a continuum of strikes.

Second, we would like to develop a way to predict the movement of parameters the next day. A possible method is to regress the daily changes in option values given by Δw_i against the matrix of partial derivatives $\frac{\partial w_i}{\partial \theta_j}$ to get the expected change in parameter values for the next day. That is, we wish to find $\Delta\theta$ such that

$$\Delta w = \frac{\partial w}{\partial \theta} \, \Delta\theta.$$

We can compare our results to the actual changes in parameters to assess quality of this predictor.

Lastly, we observe large variability in parameter values during calibration. The reason being that model calibration is an ill-posed inverse problem. One possible solution is to add a regularization term. [13] suggest to add a relative entropy with respect to some prior as a penalty term. The prior can be chosen to be the previous day's parameters. The parameters from this calibration can then be used in our study to see if they result in any improvements. Lastly, we can also extend this method to other kinds of models and examine their performance.

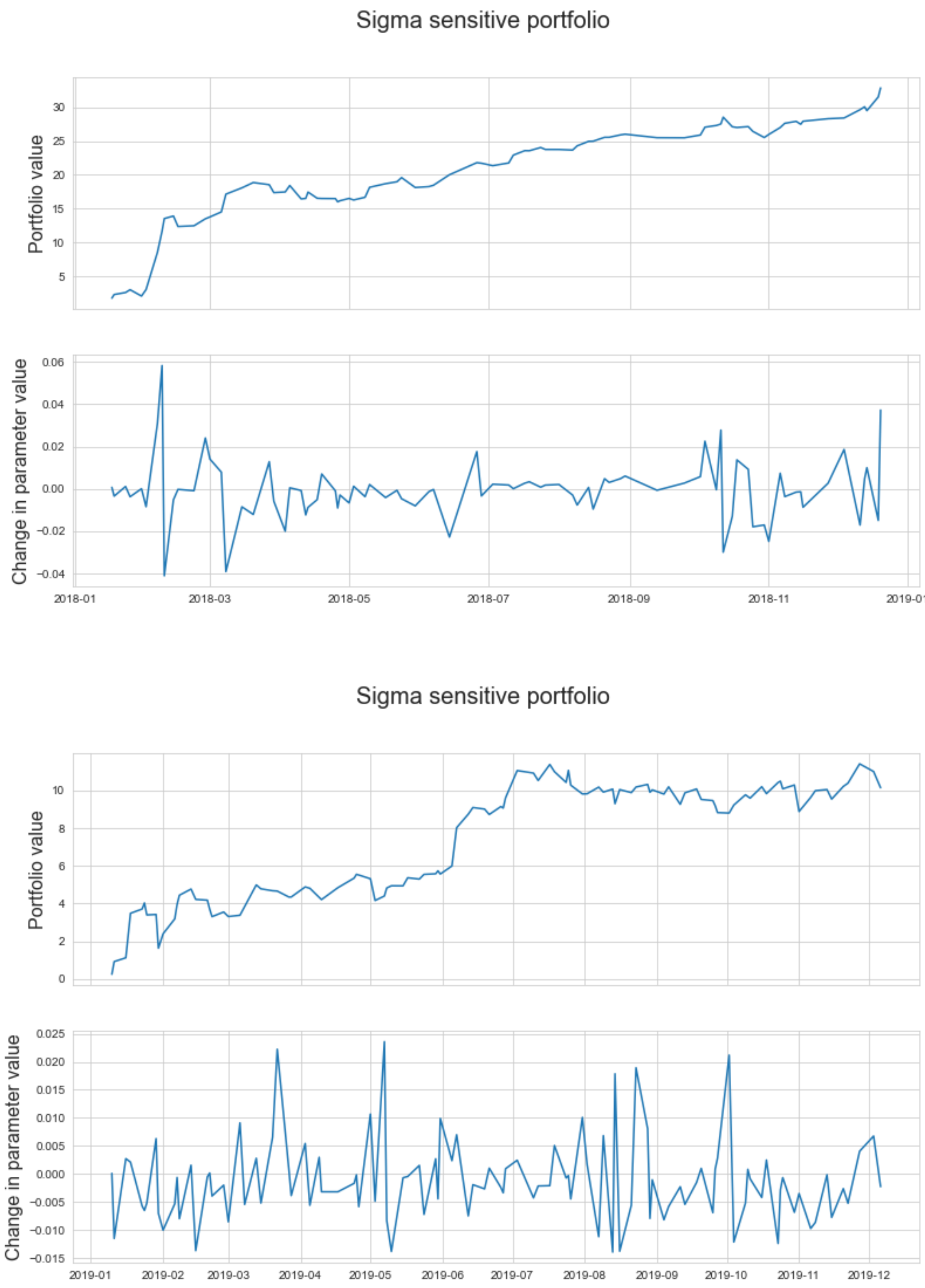
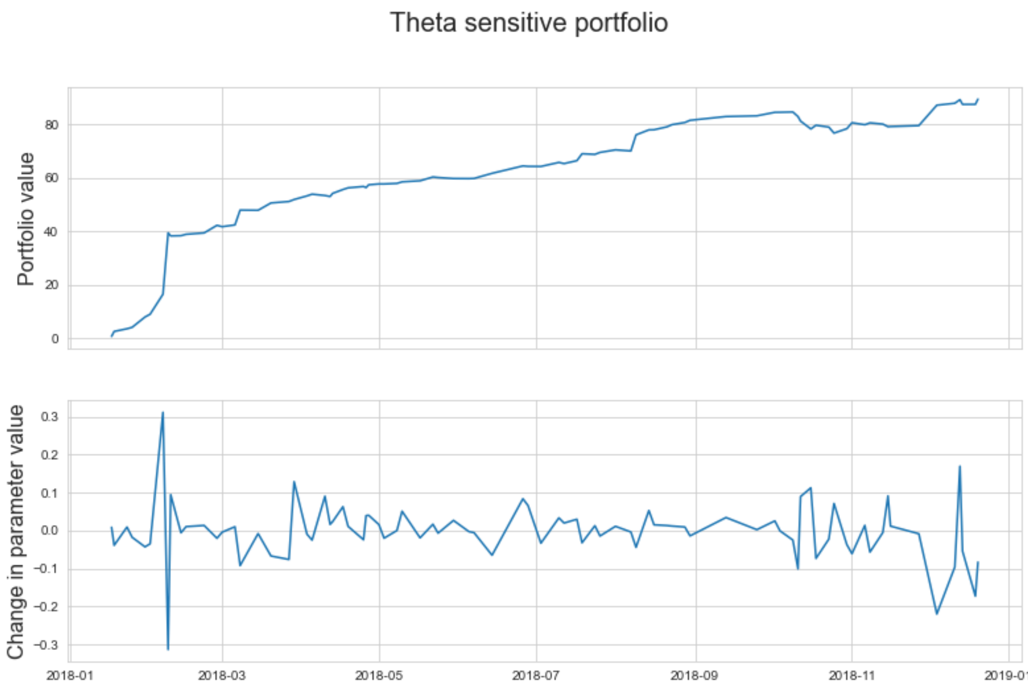


Figure 1.9: Trading sigma sensitive portfolio



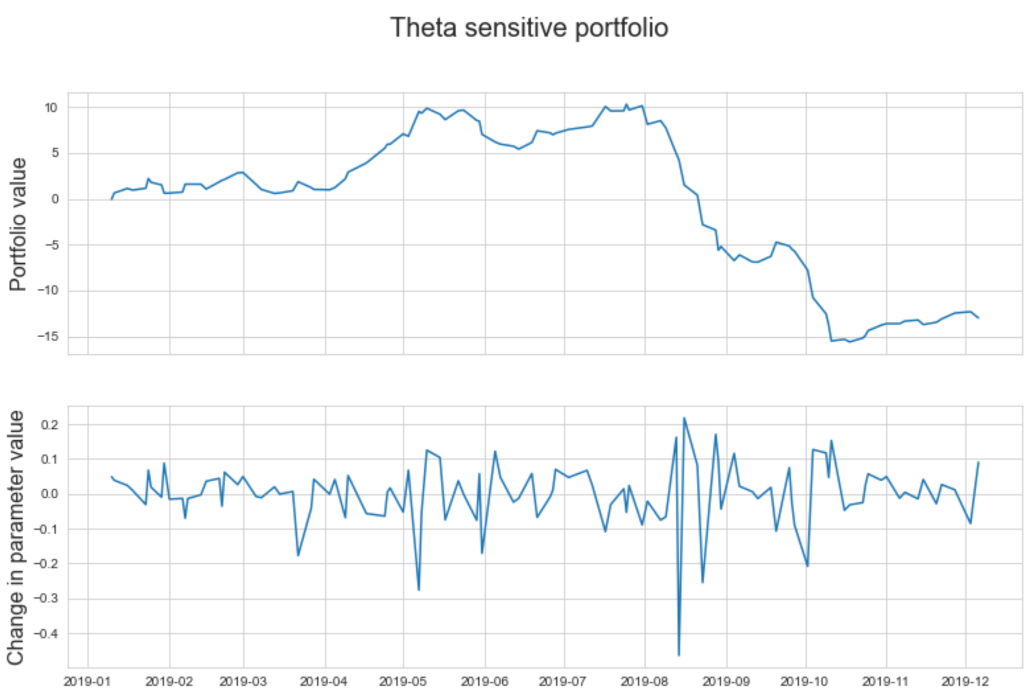


Figure 1.10: Trading theta sensitive portfolio

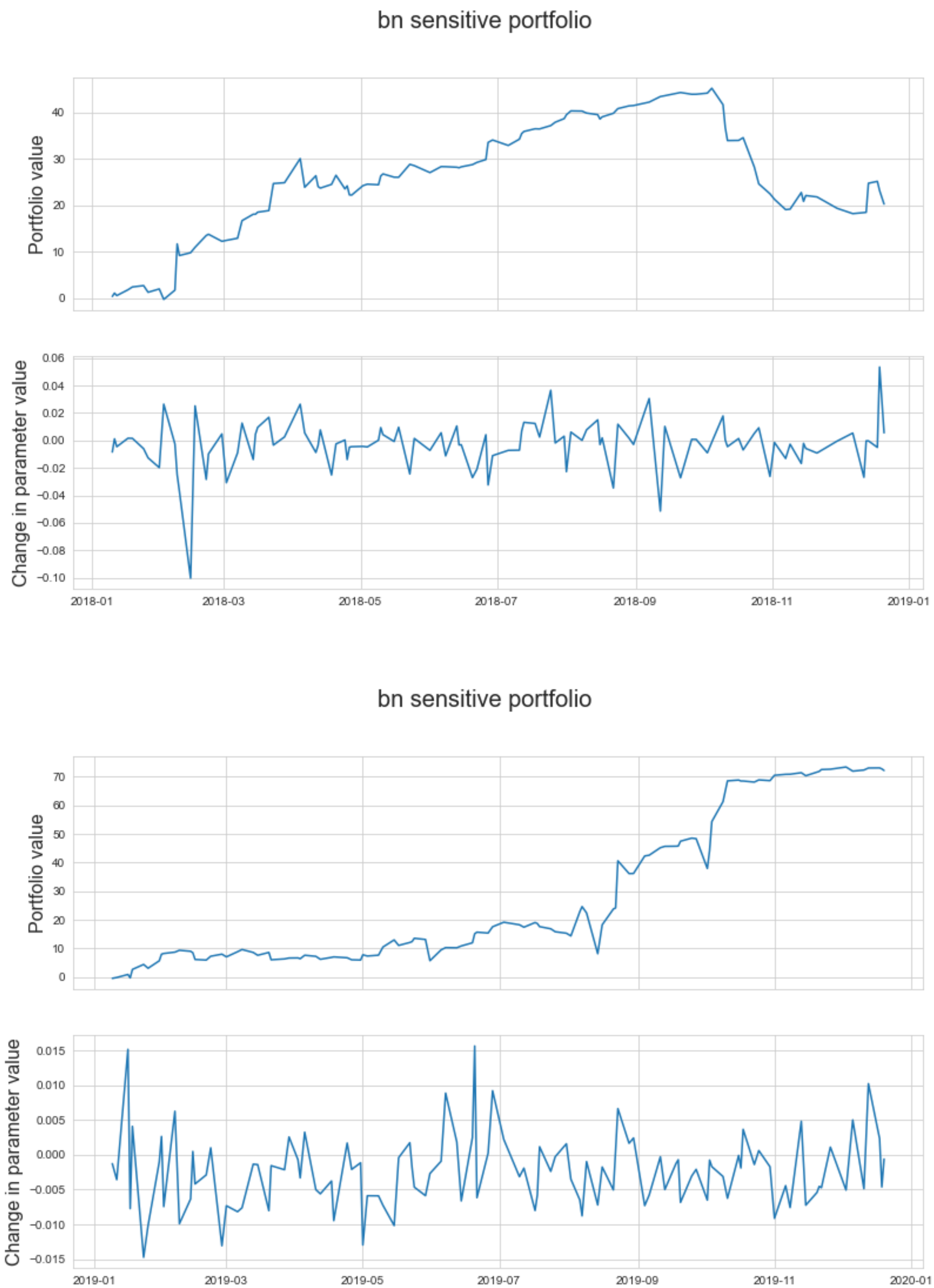


Figure 1.11: Trading bn sensitive portfolio

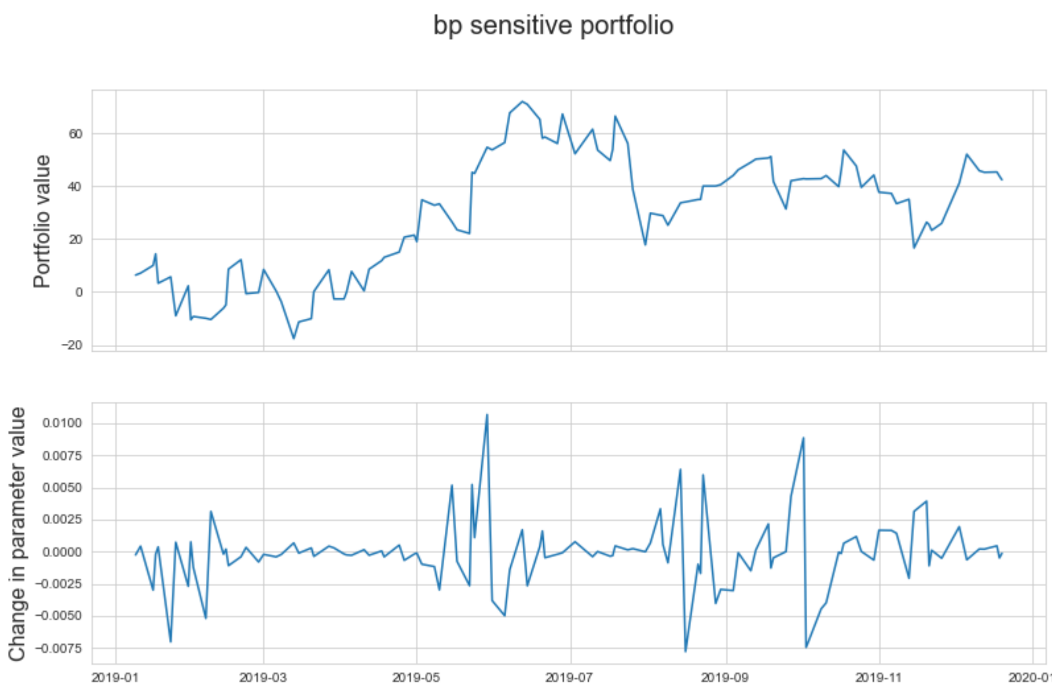


Figure 1.12: Trading bp sensitive portfolio

Chapter 2: Pricing Cliquets using Martingale Optimal Transport

2.1 Introduction

According to the first fundamental theorem of asset pricing, the no arbitrage condition is equivalent to the existence of a probability measure under which the underlying process is a martingale. The choice of such a measure is the subject of a wide range of models that date back to Black and Scholes. In their seminal work, Black and Scholes [14] and Merton [15] model the dynamics of the underlying asset by a geometric Brownian motion and derive an option pricing formula by dynamically replicating a vanilla pay-off. The assumptions made under the Black Scholes model, such as normality of log returns, do not hold empirically and many models have since been proposed that are more realistic and relax some of the restrictive assumptions (see Chapter 1 for a list). Most of these models are parametric that can be calibrated to the market data and are able to fit vanilla options with a high degree of accuracy. As a result, the dynamics of the underlying process implied by these models have similar marginal distributions.

However, these models do not agree on the prices of exotic options. As demonstrated by Schoutens et al. [16] pricing exotics under different models can lead to values that differ substantially even though they provide the same price for vanilla options. The reason being that exotics depend on factors such as forward volatility and forward skew which are modelled differently by different models. This motivates the use of model-independent methods that have gained popularity since the work of Hobson [17]. Hobson provides bounds on the price of exotic derivatives, specifically Lookback options, using call prices. They do not make any assumptions on the dynamics of the underlying asset and hence the bounds are model-free. The appeal of such methods is that they are resistant to model mis-specification.

These results can be extended to a larger class of exotic options by using tools from Optimal Transport. The problem of optimal transport deals with finding a transport plan that moves a mass from one location to another while minimizing a cost function. Put differently, it can be considered as finding an optimal joint distribution that fit pre-specified marginals and where optimality is defined under some objective function. That is, let μ and ν be two probability distributions on $\mathbb{R}$. Let $\Pi(\mu, \nu)$ be the set of all probability measures $\mathbb{P}$ (termed as transport plans) that satisfy

$$\mathbb{P}[A \times \mathbb{R}] = \mu(A) \quad \text{and} \quad \mathbb{P}[\mathbb{R} \times A] = \nu(A), \quad \text{for all } A \in \mathcal{B}(\mathbb{R}).$$

Given a function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, optimal transport seeks to minimize

$$\mathbb{E}^{\mathbb{P}}[c(x, y)]$$

over all $\mathbb{P} \in \Pi(\mu, \nu)$. The optimal transport was first introduced by Monge in 1781 [18] and then given its current formulation by Kantorovich [19] in 1948. Kantorovich showed that this problem admits a dual formulation and he proved the duality result that

$$\inf_{\mathbb{P} \in \Pi(\mu,\nu)} \mathbb{E}^{\mathbb{P}}[c(x, y)] = \sup_{(\phi,\psi) \in \Phi_c} \int_{\mathbb{R}} \phi \, d\mu + \int_{\mathbb{R}} \psi \, d\nu,$$

where $\Phi_c \subset L^1(\mathbb{R}, \mu) \times L^1(\mathbb{R}, \nu)$ is the set of all functions (ϕ, ψ) that satisfy for μ-almost all x and ν-almost all y

$$\phi(x) + \psi(y) \leq c(x, y).$$

Since then optimal transport has found applications in diverse fields such as image processing, machine learning and economics. We refer to [20] by Villani for an excellent monograph on the topic.

Beiglböck et al. [21] introduced a new research area, *Martingale Optimal Transport* where they study transport plans that are also martingales. Specifically, they consider minimizing the expected cost $\mathbb{E}^{\mathbb{P}}[c(S_1, S_2)]$ over

$$\mathcal{M}(\mu, \nu) \coloneqq \{\mathbb{P} \in \Pi(\mu, \nu) \;:\; \mathbb{E}^{\mathbb{P}}[S_2|S_1] = S_1\},$$

where S_1 and S_2 denote the value of the underlier at two points in the future. This allows one to compute model-independent bounds for a large class of exotics using infinite dimensional linear programming methods ((2.1)) and can be easily extended to the multi-period case. They establish a dual formulation of the problem that has a natural financial interpretation in terms of semi-static hedging (Theorem 2). Therefore, one can easily create a portfolio that leads to arbitrage should the price of the exotic exceed these bounds.

In practice however, these bounds are much too wide for practical use. The resulting probability measures for any cost function are sparsely supported and not appropriate for pricing. We therefore turn our attention to finding a suitable measure that can price exotics using as few assumptions as possible. This leads us to replacing the cost function with the Kullback-

Leibler divergence (i.e., relative entropy) with respect to some reference measure and minimizing over all martingales consistent with given marginals. This problem is similar to the one used by Guyon [22] and is termed as the *Martingale Schrödinger problem*. They are interested in jointly calibrating the S&P 500 and the VIX and their problem involves probability measures on $(\mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0})$ with additional constraints. Henry-Labordère [23] consider the martingale optimal transport problem over all measures that are not too far off from a prior measure in the entropic sense. That is, they minimize the expected cost over all $\mathbb{P}$ such that $\mathbb{P} \in \mathcal{M}(\mu, \nu)$ and $H(\mathbb{P}|\mathbb{P}_0) < \lambda$, where $\mathbb{P}_0$ is the prior measure and $\lambda \in [0, \infty]$.

In this chapter we are mainly interested in pricing cliquets (see (2.18) for the definition of cliquet) and comparing them to market data. Our approach is applicable to any exotic whose value depends on the value of the underlying at finitely many time points. Our contribution is two-fold. First, we prove a duality result for the martingale Schrödinger problem (Theorem 3) and provide an explicit form for the minimizing distribution. Then, we show that it can computed iteratively by using Sinkhorns algorithm (2.11) and provide a convergence result (Proposition 2.4.6). Second, we propose different reference measures and compare the pricing performance of the resulting distributions to real data. We price different kinds of cliquets with varying rolling periods. To the best of our knowledge, this is the first study that compares models generated from optimal transport with market data. We demonstrate that we get good fits using certain reference measures. Therefore our approach shows how to obtain exotic prices consistent with observed data, using information from just the marginals.

2.2 Setting

Let $\Omega := \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ be the sample space and let $\mathcal{B}(\Omega)$ be the Borel σ-algebra on Ω. Let $\mathcal{P}(\Omega)$ be the set of all probability measures on $(\Omega, \mathcal{B}(\Omega))$. Let $S : \{1,2\} \times \Omega \to \mathbb{R}_{>0}$ be a coordinate process with the property that for all $(s_1, s_2) \in \Omega$ it holds that

$$S_1(s_1, s_2) = s_1 \quad \text{and} \quad S_2(s_1, s_2) = s_2.$$

We will let S represent the underlying asset with S_1 and S_2 denoting its value at two maturities $0 < t_1 < t_2$ and let S_0 be the current spot price. We assume zero interest rate and dividends for simplicity unless stated otherwise.

Given two probability distributions μ, ν on $(\mathbb{R}_{>0}, \mathbb{B}(\mathbb{R}_{>0}))$ define the set of transport plans $\Pi(\mu, \nu) \subseteq \mathcal{P}(\Omega)$ to be the set of all probability measures with marginals μ and ν. More explicitly, $\pi \in \Pi(\mu, \nu)$ if and only if π is a non-negative measure satisfying

$$\pi(A \times \mathbb{R}_{>0}) = \mu(A) \quad \text{and} \quad \pi(\mathbb{R}_{>0} \times B) = \nu(B)$$

for all measurable $A, B \in \mathcal{B}(\mathbb{R}_{>0})$.

We assume that there are vanilla options with maturities t_1 and t_2 that imply the existence of two risk neutral distributions μ_1 and μ_2 on $(\mathbb{R}_{>0}, \mathcal{B}(\mathbb{R}_{>0}))$ (due to Breeden et al. [24]) with the property that

$$\mathbb{E}^{\mu_1}[S_1] = \mathbb{E}^{\mu_2}[S_2] = S_0.$$

Lastly, we introduce the shorthand notation that for arbitrary functions $u_1, u_2 : \mathbb{R}_{>0} \to \mathbb{R}$

and $h : \mathbb{R}_{>0} \to \mathbb{R}$ we denote $\forall (x, y) \in \mathbb{R}_{>0}$:

$$(u_1 \oplus u_2)(x, y) = u_1(x) + u_2(y) \;\text{ and }\; h^{\otimes}(x, y) = (y - x)h(x)$$

2.3 Martingale Optimal Transport

2.3.1 Primal Problem

Let $c : \Omega \to \mathbb{R}$ be the payoff of a contingent claim that depends on the value of an underlying asset S at two time points, S_1 and S_2. An example of such a claim is a *forward start* option with the payoff, $c(s_1, s_2) = \left(\frac{s_1}{s_2} - 1\right)^+$. In order to determine the fair value under no-arbitrage, we postulate a model $\mathbb{Q} \in \mathcal{P}(\Omega)$ under which S is a martingale adapted to the natural filtration. The fair value for c is then given by

$$\mathbb{E}^{\mathbb{Q}}[c] = \int_{\Omega} c(s_1, s_2) d\mathbb{Q}(s_1, s_2).$$

We assume that the model is calibrated to vanilla options with maturities t_1 and t_2. Hence, the one dimensional marginals of $\mathbb{Q}$ satisfy

$$\mathbb{Q} \circ S_i^{-1} = \mu_i \quad \text{for } \; i \in \{1, 2\}.$$

Moreover we assume absence of calendar arbitrage which is equivalent to μ_1 and μ_2 being in convex order, i.e., for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ it holds that

$$\mathbb{E}^{\mu_1}[\phi] \leq \mathbb{E}^{\mu_2}[\phi].$$

Let $\mathcal{M}(\mu_1, \mu_2) \subseteq \Pi(\mu_1, \mu_1)$ be the set of all transport plans that satisfy the martingale condition such that $\mathbb{E}^{\mu}[S_2|S_1] = S_1$ for all $\mu \in \mathcal{M}(\mu_1, \mu_2)$. By Strassen's theorem [25] $\mathcal{M}(\mu_1, \mu_2)$ is non-empty if and only if μ_1 and μ_2 are in convex order.

The no-arbitrage lower bound price P for c is hence given by *Martingale Optimal Transport* primal problem:

$$P = \inf\{\mathbb{E}^{\mu}[c(s_1, s_2)] : \mu \in \mathcal{M}(\mu_1, \mu_2)\}$$

We can explicitly compute the lower bound by reframing this as a linear program

$$\begin{aligned} &\inf_{\mu} \int c(s_1, s_2)\mu(ds_1, ds_2) \\ \text{subject to } &\int \mu(s_1, ds_2) = \mu_1(s_1)\ \forall s_1 \in \mathbb{R}_{>0} \\ &\int \mu(ds_1, s_2) = \mu_2(s_2)\ \forall s_2 \in \mathbb{R}_{>0} \\ &\int (s_2 - s_1)\mu(s_1, ds_2) = 0\ \forall s_1 \in \mathbb{R}_{>0}. \end{aligned} \tag{2.1}$$

2.3.2 Dual Problem

The dual to martingale optimal transport corresponds to an upper bound on the value of a sub-replicating portfolio for c. Let $\mathcal{U}$ be the collection of portfolios (u_1, u_2, Δ) defined by $\mathcal{U} := \mathbb{L}^1(\mathbb{R}_{>0}, \mu_1) \times \mathbb{L}^1(\mathbb{R}_{>0}, \mu_2) \times \mathcal{C}_b(\mathbb{R}_{>0})$ where $C_b(\mathbb{R}_{>0})$ is the space of all continuous and bounded functions. Let $\mathcal{U}_c \subseteq \mathcal{U}$ be the set of sub-replicating portfolios of c with the property that for all $(s_1, s_2) \in \Omega$ it holds that

$$c(s_1, s_2) \geq (u_1 \oplus u_2)(s_1, s_2) + \Delta^{\otimes}(s_1, s_2). \tag{2.2}$$

The dual problem is given by:

$$D = \sup\{\mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] : \exists\, \Delta \ s.t.\ c(s_1, s_2) \geq (u_1 \oplus u_2)(s_1, s_2) + \Delta^{\otimes}(s_1, s_2)\}.$$

The dual provides an upper bound D by considering the highest value of a sub-replicating portfolio of c that consists of two vanilla options u_1, u_2 with maturities t_1, t_2 and a position in the forward given by Δ.

We can formally derive the dual by introducing Lagrange multipliers and using a minimax argument to switch the supremum and infimum:

$$\begin{aligned}
&\inf_{\mu \in \mathcal{M}(\mu_1, \mu_2)} \mathbb{E}^{\mu}[c(s_1)] \\
&= \inf_{\mu \in \mathcal{P}(\Omega)} \sup_{u \in \mathcal{U}} \Big\{ \mathbb{E}^{\mu}[c(s_1, s_2)] + (\mathbb{E}^{\mu_1}[u_1(s_1)] - \mathbb{E}^{\mu}[u_1(s_1)]) \\
&\qquad\qquad + (\mathbb{E}^{\mu_2}[u_2(s_2)] - \mathbb{E}^{\mu}[u_2(s_2)]) - \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1, s_2)] \Big\} \\
&= \sup_{u \in \mathcal{U}} \Big\{ \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&\qquad\qquad + \inf_{\mu \in \mathcal{P}(\Omega)} \left\{ \mathbb{E}^{\mu}[c(s_1, s_2) - u_1(s_1) - u_2(s_2) - \Delta^{\otimes}(s_1, s_2)] \right\} \Big\} \\
&= \sup_{u \in \mathcal{U}_c} \mathbb{E}^{\mu_1}[u_1(S_1)] + \mathbb{E}^{\mu_2}[u_2(S_2)].
\end{aligned}$$

The last equality is true because the infimum over $\mathcal{P}(\Omega)$ goes to $-\infty$ if the expression inside is negative and (2.2) does not hold. These arguments can be made rigorous to prove the absence of a duality gap, i.e., $P = D$.

2.3.3 Absence of Duality gap

Beiglböck et al. [21] show that under mild assumptions on the payoff c, the primal and dual values are equal. In particular they prove:

Theorem 2. *Assume that $\mathcal{M}(\mu_1, \mu_2)$ is nonempty. Let $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a lower semi-continuous function such that*

$$c(s_1, s_2) \geq -K \cdot (1 + |s_1| + |s_2|)$$

for some constant K. Then, there is no duality gap i.e.,

$$P = \inf_{\mu \in \mathcal{M}(\mu_1, \mu_2)} \mathbb{E}^{\mu}[c(s_1, s_2)] = \sup_{(u_1, u_2, \Delta) \in \mathcal{U}_c} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] = D.$$

Moreover, the primal value P is attained, i.e., there exists a minimizing martingale measure $\mu \in \mathcal{M}(\mu_1, \mu_2)$ such that $P = \mathbb{E}^{\mu}[c]$. The dual supremum in general is not attained.

It is sufficient to consider u_1, u_2 that are linear combinations of finitely many call options and Δ is be taken to be continuous and bounded. A similar result holds by switching the infimum to a supremum in the primal and by assuming that c is upper semi-continuous.

Corollary 2.1. *Assume that $\mathcal{M}(\mu_1, \mu_2)$ is nonempty. Let $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be an upper semi-continuous function such that*

$$c(s_1, s_2) \leq K \cdot (1 + |s_1| + |s_2|)$$

for some constant $K \in \mathbb{R}$. Then there is no duality gap

$$\sup_{\mu \in \mathcal{M}(\mu_1,\mu_2)} \mathbb{E}^{\mu}[c(s_1, s_2)] = \inf_{(u_1,u_2,\Delta)\in\mathcal{U}_c} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)]$$

and the supremum is attained, i.e., there exists a maximizing martingale measure.

The martingale optimal transport allows us to compute model-free lower and upper bounds for any given payoff c given by

$$\sup_{\mu \in \mathcal{M}(\mu_1,\mu_2)} \mathbb{E}^{\mu}[c(_1, S_2)] \quad \text{and} \quad \inf_{\mu \in \mathcal{M}(\mu_1,\mu_2)} \mathbb{E}^{\mu}[c(S_1, S_2)].$$

In practice however, these bounds are often too wide and not appropriate for pricing such claims. Moreover, for any given function c, the optimal joint distribution is payoff dependent and has sparse support which implies that S_2 is largely determined by S_1 which is an unrealistic assumption. Henry-Labordère et al. [26] show that for a payoff satisfying $\partial_{s_1,s_2,s_2} c(s_1, s_2) > 0$ the optimal joint distribution is supported on two graphs. Similarly, based on numerical experiments, De March [27] conjecture that the optimal coupling for any $c(s_1, s_2)$ is supported on at most two graphs. Therefore, any objective function that is linear in probability is unsuitable for pricing.

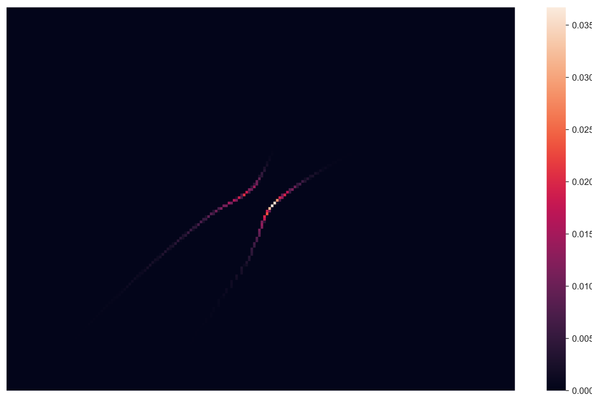


Figure 2.1: Optimal joint distribution from maximizing the payoff $c(s_1, s_2) = (\frac{S_2}{S_1} - 1)^+$.

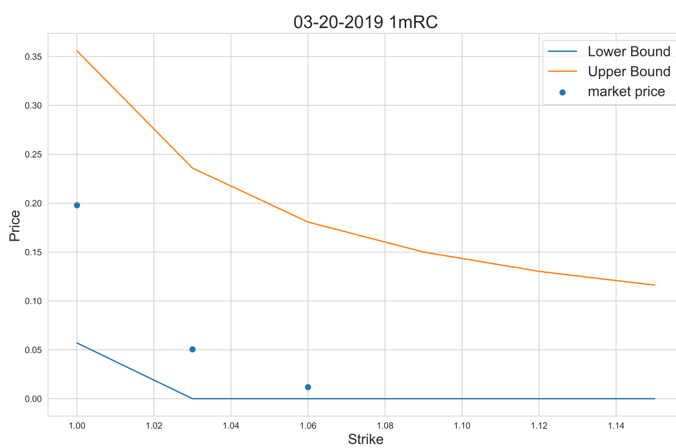


Figure 2.2: Upper and Lower bound for $c(s_1, s_2) = (\frac{S_2}{S_1} - 1)^+$ compared to market prices under Optimal Transport.

2.4 Martingale Schrödinger Problem

Our goal is to find a suitable objective function with which we can obtain a joint distribution that prices claims consistent with market data. As such, we look for a function that is non-linear in probability. Natural candidates include Shannon entropy and the Kullback-Leibler relative entropy. We consider minimizing the relative entropy with respect to some reference measure.

Given two probability measures μ and ν, the relative entropy $H(\mu|\nu)$ of μ with respect to ν is defined as:

$$H(\mu|\nu) \coloneqq \begin{cases} \mathbb{E}^{\mu}\left[\ln \frac{d\mu}{d\nu}\right] & \text{if } \mu \ll \nu \\ \\ +\infty & \text{otherwise} \end{cases}$$

Under this objective function, the problem of finding a pricing measure can be formulated as

$$\inf_{\mu \in \mathcal{M}(\mu_1,\mu_2)} H(\mu|\bar{\mu}),$$

where $\bar{\mu}$ is a given reference measure. Following Guyon [22] we term this as the *Martingale Schrödinger problem*. Unlike the martingale optimal transport, here we are interested in finding an optimal measure as opposed to the lower and upper bounds of a given claim. This problem is strictly convex and assuming that there exists $\mu \in \mathcal{M}(\mu_1,\mu_2)$ such that $H(\mu|\bar{\mu}) < \infty$, it holds that the infimum is uniquely attained. The modern interpretation of the original problem as proposed by Schrödinger ([28], [29]) involves finding a probability measure over the path space that matches given marginal densities and minimizes the relative entropy with respect to the Wiener measure. A detailed survey on the Schrödinger problem is provided by [30] and [31].

Similar to the martingale optimal transport, we can formally derive the dual for the martingale Schrödinger problem by observing that,

$$\begin{aligned}
&\inf_{\mu\in\mathcal{M}(\mu_1,\mu_2)} H(\mu|\bar{\mu}) \\
&= \inf_{\mu\in\mathcal{P}(\Omega)} \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} \Big\{ H(\mu|\bar{\mu}) + (\mathbb{E}^{\mu_1}[u_1(s_1)] - \mathbb{E}^{\mu}[u_1(s_1)]) \\
&\qquad\qquad + (\mathbb{E}^{\mu_2}[u_2(s_2)] - \mathbb{E}^{\mu}[u_2(s_2)]) - \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1,s_2)] \Big\} \\
&= \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} \Big\{ \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&\qquad\qquad + \inf_{\mu\in\mathcal{P}(\Omega)} \big\{ H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[(u_1\oplus u_2)(s_1,s_2) + \Delta^{\otimes}(s_1,s_2)] \big\} \Big\}
\end{aligned}$$

The inner infimum can be computed explicitly (see [23], Proposition 7.3) using the proposition below.

Proposition 2.4.1. *Let μ and $\bar{\mu}$ be two probability measures such that $\mu \ll \bar{\mu}$. Let X be a random variable with the property that $\mathbb{E}^{\bar{\mu}}[e^X] < +\infty$. Then it holds that*

$$\inf_{\mu\in\mathcal{P}(\Omega)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[X] = -\ln \mathbb{E}^{\bar{\mu}}[e^X].$$

The infimum is attained at $\mu(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2)\frac{e^X}{\mathbb{E}^{\bar{\mu}}[e^X]}$.

Proof. Define a probability measure ν_X given by the density $\frac{d\nu_X}{d\bar{\mu}} = \frac{e^X}{\mathbb{E}^{\bar{\mu}}[e^X]}$. Then it holds that

$$\begin{aligned}
H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[X] &= \mathbb{E}^{\mu}\left[\ln\frac{d\mu}{d\nu_X} + \ln\frac{d\nu_X}{d\bar{\mu}}\right] - \mathbb{E}^{\mu}[X] \\
&= H(\mu|\nu_X) + \mathbb{E}^{\mu}\left[\ln\frac{d\nu_X}{d\bar{\mu}} - X\right] \\
&= H(\mu|\nu_X) - \ln\mathbb{E}^{\bar{\mu}}[e^X].
\end{aligned}$$

Since $H(\mu|\bar{\mu}) \geq 0$ from Jensen's inequality, it holds that

$$\inf_{\mu \in \mathcal{P}(\Omega)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[X] = -\ln \mathbb{E}^{\bar{\mu}}[e^X]$$

and the infimum is attained when $\mu = \nu_X$. $\square$

Therefore, the dual can be framed as,

$$\begin{aligned} \inf_{\mu \in \mathcal{M}(\mu_1,\mu_2)} H(\mu|\bar{\mu}) = \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} & \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_1}[u_2(s_2)] \\ & - \ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1\oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}\right] \end{aligned}$$

2.4.1 Duality

We now prove the absence of a duality gap. Guyon [22] proves a similar result under additional VIX constraints.

Theorem 3. *Let $\bar{\mu} \in \mathcal{P}(\Omega)$ be a reference measure that satisfies $\bar{\mu} \sim \mu_1 \otimes \mu_2$ and $H(\mu_1 \otimes \mu_2|\bar{\mu}) < \infty$. Then it holds that*

$$\begin{aligned} \inf_{\mu \in \mathcal{M}(\mu_1,\mu_2)} H(\mu|\bar{\mu}) = \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} & \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_1}[u_2(s_2)] \\ & - \ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1\oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}\right] \end{aligned} \tag{2.3}$$

If the problem is finite, then the infimum is uniquely attained, i.e., there exists a unique $\mu^* \in \mathcal{M}(\mu_1, \mu_2)$ such that

$$H(\mu^*|\bar{\mu}) = \inf_{\mu \in \mathcal{M}(\mu_1,\mu_2)} H(\mu|\bar{\mu}).$$

Furthermore, suppose that the supremum is attained and let $(u_1^*, u_2^*, \Delta^*) \in \mathcal{U}$ be the maximizer. Then the problem is finite and the unique minimal probability measure $\mu^* \in \mathcal{M}(\mu_1, \mu_2)$ for the primal is given by

$$\mu^*(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2) \frac{e^{(u_1^* \oplus u_2^*)(s_1,s_2) + \Delta^{*\otimes}(s_1,s_2)}}{\mathbb{E}^{\bar{\mu}}\big[e^{(u_1^* \oplus u_2^*)(s_1,s_2) + \Delta^{*\otimes}(s_1,s_2)}\big]} \quad \bar{\mu} - a.s. \tag{2.4}$$

The proof of Theorem 3 relies mainly on two results: a duality result for the Schrödinger problem without the martingale constraint and Sion's minimax theorem which allows us to interchange the infimum and supremum.

Theorem 4. *([32]. Theorem 3.2). Assume the same setting as in Theorem 3. Then it holds that*

$$\inf_{\pi \in \Pi(\mu_1, \mu_2)} H(\mu|\bar{\mu}) = \sup_{u_1 \in L^1(\mu_1), u_2 \in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1 \oplus u_2)(s_1,s_2)}].$$

Theorem 5. *(Sion's minimax theorem [33], Corollary 3.3). Let X and Y be convex subsets of a linear topological space. Let X be compact and let $f : X \times Y \to \mathbb{R}$ be a functions such that*

1. *For all $x \in X$, $f(x, \cdot)$ is upper semi-continuous and quasi-concave on Y.*

2. *For all $y \in y$, $f(\cdot, y)$ is lower semi-continuous and quasi-convex on X.*

Then it holds that

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

Theorem 5, can be justified by the following two propositions.

Proposition 2.4.2. *([21], Lemma 2.2). Let $c : \mathbb{R}^2 \to \mathbb{R}$ be continuous and assume there exists a* *constant $K > 0$ such that for all $(x_1, x_2) \in \mathbb{R}^2$ it holds that*

$$|c(x_1, x_2)| \leq K(1 + |x_1| + |x_2|).$$

Then for any $\pi \in \Pi(\mu, \nu)$, the mapping $\pi \mapsto \int_{\mathbb{R}^2} c(x_1, x_2) d\pi$ is continuous on $\Pi(\mu, \nu)$.

Proposition 2.4.3. *([34], Theorem D.13). Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Then the mapping $\mu \mapsto H(\mu|\nu)$ is lower semi-continuous in the weak topology.*

Lastly, the following two auxiliary results are also required for the proof.

Proposition 2.4.4. *([21], Lemma 2.3). Let $\Pi(\mu, \nu)$ be the set of all probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν. Let $\pi \in \Pi(\mu, \nu)$. Then the following are equivalent:*

1. $\pi \in \mathcal{M}(\mu, \nu)$
2. *For every continuous and bounded function $\Delta : \mathbb{R} \to \mathbb{R}$ it holds that*

$$\int_{\mathbb{R}^2} \Delta(x_1)(x_2 - x_1) d\pi(x_1, x_2) = 0 \tag{2.5}$$

Lemma 2.4.1. *Let $\bar{\mu} \in \mathcal{P}(\Omega)$. Let $c : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}$ be a function with the property that there exists a constant $K_c \in \mathbb{R}$ such that it holds for all $(s_1, s_2) \in \Omega$ that $c(s_1, s_2) \leq K_c \cdot (1 + s_1 + s_2)$. Then it holds that*

$$\begin{aligned} &\inf_{\mu \in \Pi(\mu_1, \mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[c(s_1, s_2)] \\ &= \sup_{u_1 \in L^1(\mu_1),\, u_2 \in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1 \oplus u_2)(s_1, s_2) + c(s_1, s_2)}] \end{aligned}$$

Proof. We first prove for the case when $\mathbb{E}^{\bar{\mu}}[e^{c(s_1,s_2)}] < \infty$. Let $\bar{\mu}_c \in \mathcal{P}(\Omega)$ be a density defined through

$$\frac{d\bar{\mu}_c}{d\bar{\mu}} = \frac{e^{c(s_2,s_2)}}{M_c} \quad \bar{\mu} - a.s.,$$

where $M_c := \mathbb{E}^{\bar{\mu}}[e^{c(s_2,s_2)}]$ is the normalizing constant. Then it holds that

$$\inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[c(s_1,s_2)] = \inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}_c) - \ln M_c.$$

Applying Theorem 4 to $H(\mu|\bar{\mu}_c)$ it holds that

$$\begin{aligned}
&\inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}_c) - \ln M_c \\
&= \sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}_c}[e^{(u_1\oplus u_2)(s_1,s_2)}] - \ln M_c \\
&= \sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[M_c^{-1} e^{(u_1\oplus u_2)(s_1,s_2)+c(s_1,s_2)}] - \ln M_c \\
&= \sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+c(s_1,s_2)}].
\end{aligned}$$

Let $\hat{c} : \mathbb{R}_{>0}\times\mathbb{R}_{>0} \to \mathbb{R}$, be a function with the property that $\hat{c}(s_1,s_2) = c(s_1,s_2) - K_c(2+s_1+s_2)$. Clearly, $\hat{c}(s_1,s_2) \leq 0$ for all $(s_1,s_2) \in \Omega$. Since $\mathbb{E}^{\bar{\mu}}[e^{\hat{c}(s_1,s_2)}] < \infty$, we can apply the first part of the lemma to show that

$$\begin{aligned}
&\inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[\hat{c}(s_1,s_2)] \\
&= \sup_{u_1\in L^1(\mu_1),\, u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+\hat{c}(s_1,s_2)}]
\end{aligned}$$

We conclude the proof of the lemma by observing that

$$\begin{aligned}&\inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[\hat{c}(s_1,s_2)]\\ =&\inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[c(s_1,s_2)] + \mathbb{E}^{\mu_1}[K_c(1+s_1)] + \mathbb{E}^{\mu_2}[K_c(1+s_2)].\end{aligned} \tag{2.6}$$

Similarly, it holds that

$$\begin{aligned}&\sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+\hat{c}(s_1,s_2)}]\\ =&\sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+c(s_1,s_2)-K_c(2+s_1+s_2)}].\end{aligned}$$

Note that for $i\in\{1,2\}$ it holds $u_i(s_i)-K_c(1+s_i)\in L^1(\mu_i)$. We can rewrite the above as

$$\begin{aligned}&\sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+c(s_1,s_2)}]\\ &\qquad + \mathbb{E}^{\mu_1}[K_c(1+s_1)] + \mathbb{E}^{\mu_2}[K_c(1+s_2)].\end{aligned} \tag{2.7}$$

The result follows from observing the equality between equations (2.6) and (2.7). $\square$

The proof of Theorem 3 now follows.

Proof. It holds that

$$
\begin{aligned}
\inf_{\mu\in\mathcal{M}(\mu_1,\nu_2)} H(\mu|\bar{\mu}) &= \inf_{\mu\in\Pi(\mu_1,\mu_2)} \sup_{\Delta\in\mathcal{C}_b} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1,s_2)] \\
&= \sup_{\Delta\in\mathcal{C}_b} \inf_{\mu\in\Pi(\mu_1,\mu_2)} H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1,s_2)] \\
&= \sup_{\Delta\in\mathcal{C}_b} \sup_{u_1\in L^1(\mu_1),u_2\in L^1(\mu_2)} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&\quad - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}] \\
&= \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&\quad - \ln \mathbb{E}^{\bar{\mu}}[e^{(u_1\oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}].
\end{aligned}
\tag{2.8}
$$

The first equality follow from Proposition 2.4.4. The second equality uses Sion's Minimax theorem (Theorem 5) where we let $X = \Pi(\mu_1,\mu_2)$, equipped with the weak topology induced by bounded continuous functions on $\mathbb{R}$ and $Y = \mathcal{C}_b$, equipped with the $\sup$ norm. Both are convex sets and $\Pi(\mu_1,\mu_2)$ is a compact subset of finite Borel measures under the weak topology. Let $f_{\bar{\mu}} : \Pi(\mu_1,\mu_2)\times\mathcal{C}_b \to \mathbb{R}$ be a function defined by

$$
f_{\bar{\mu}}(\mu,\Delta) \coloneqq H(\mu|\bar{\mu}) - \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1,s_2)].
$$

For any $\nu \in \mathcal{P}(\Omega)$ it holds that $H(\cdot|\nu)$ is strictly convex and is lower semi-continuous in $\Pi(\mu_1,\mu_2)$ by Proposition 2.4.3. Furthermore, the mapping $\mu \mapsto \mathbb{E}^{\mu}[\Delta^{\otimes}(s_1,s_1)]$ is convex and continuous by Proposition 2.4.2. Therefore for all $\Delta\in\mathcal{C}_b$, $f_{\bar{\mu}}(\cdot,\Delta)$ is lower semi-continuous and convex on $\Pi(\mu_1,\mu_2)$. Likewise, for all $\mu\in\Pi(\mu_1,\mu_2)$, $f_{\bar{\mu}}(\mu,\cdot)$ is linear and continuous and hence also upper semi-continuous and concave on $\mathcal{C}_b$. Therefore we can justify switching the supremum and infimum in the second equality. Lastly, the third equality follows from Lemma 2.4.1. This proves the duality result (2.3).

Suppose that the problem (2.3) is finite. It follows from the compactness of $\mathcal{M}(\mu_1, \mu_2)$ ([21], Proposition 2.4) and the lower semi-continuity of $\mu \mapsto H(\mu|\bar{\mu})$ that the infimum is attained. In addition, it is unique due to the strict convexity of $H(\mu|\bar{\mu})$.

The proof of (2.4) is similar to [22], Theorem 22. Let $D_{\bar{\mu}} : \mathcal{U} \to \mathbb{R}$ be a function with the property that for all $(u_1, u_2, \Delta) \in \mathcal{U}$ it holds that

$$D_{\bar{\mu}}(u_1, u_2, \Delta) = \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_1}[u_2(s_2)] - \ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1 \oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}\right] \tag{2.9}$$

Now suppose that there exist $(u_1^*, u_2^*, \Delta^*) \in \mathcal{U}$ such that it holds

$$D_{\bar{\mu}}(u_1^*, u_2^*, \Delta^*) = \sup_{(u_1,u_2,\Delta)\in\mathcal{U}} D_{\bar{\mu}}(u_1, u_2, \Delta).$$

It is easy to see that $D_{\bar{\mu}}(u_1^*, u_2^*, \Delta^*) \geq 0$. Hence it follows that

$$\mathbb{E}^{\bar{\mu}}\left[e^{(u_1 \oplus u_2)(s_1,s_2)+\Delta^{\otimes}(s_1,s_2)}\right] < \infty$$

which in turn implies that $D_{\bar{\mu}}(u_1^*, u_2^*, \Delta^*) \leq \infty$ and that the problem is finite. Therefore the infimum is attained and there exists a unique $\mu^* \in \mathcal{M}(\mu_1, \mu_2)$ such that

$$H(\mu^*|\bar{\mu}) = \mathbb{E}^{\mu_1}[u_1^*(s_1)] + \mathbb{E}^{\mu_1}[u_2^*(s_2)] - \ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1^* \oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right].$$

It holds that

$$
\begin{aligned}
&\ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right] \\
\geq &\ln \mathbb{E}^{\bar{\mu}}\left[1_{\{\frac{d\mu^*}{d\bar{\mu}}>0\}}e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right] \\
= &\ln \mathbb{E}^{\bar{\mu}}\left[1_{\{\frac{d\mu^*}{d\bar{\mu}}>0\}}\frac{d\mu^*}{d\bar{\mu}}e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}}\right] \\
= &\ln \mathbb{E}^{\mu^*}\left[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}}\right] \\
\geq &\mathbb{E}^{\mu^*}\left[(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}\right] \\
= &\mathbb{E}^{\mu_1}[u_1^*(s_1)]+\mathbb{E}^{\mu_2}[u_2^*(s_2)]-H(\mu^*|\bar{\mu}) \\
= &\ln \mathbb{E}^{\bar{\mu}}\left[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right].
\end{aligned}
$$

Therefore, we get the equality:

$$
\mathbb{E}^{\bar{\mu}}\left[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right]=\mathbb{E}^{\bar{\mu}}\left[1_{\{\frac{d\mu^*}{d\bar{\mu}}>0\}}e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}\right],
$$

which implies that $\frac{d\mu^*}{d\bar{\mu}}>0, \bar{\mu}-a.s.$ and that $\mu^*\sim\bar{\mu}$, i.e., the two measures are equivalent. Moreover, it holds that

$$
\ln \mathbb{E}^{\mu^*}\left[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}}\right]=\mathbb{E}^{\mu^*}\left[\ln\left(e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}}\right)\right]
$$

By Jensen's inequality and strict concavity of the logarithm, it holds that $(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)-\ln\frac{d\mu^*}{d\bar{\mu}}$ must by constant $\mu^*-a.s.$. Hence it follows that

$$
\mu^*(ds_1,ds_2)=\bar{\mu}(ds_1,ds_2)\frac{e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}}{\mathbb{E}^{\bar{\mu}}[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}]}\quad \bar{\mu}-a.s.
$$

$\square$

2.4.2 Solving the Dual using Sinkhorn's Algorithm

Let $D_{\bar{\mu}} : \mathcal{U} \to \mathbb{R}$ be as defined in (2.9). The dual formulation (2.3) effectively transforms the problem of minimizing $H(\mu|\bar{\mu})$ over a constrained set $\mathcal{M}(\mu, \nu)$ to a problem of maximizing a strictly concave function $D_{\bar{\mu}}$ over an unconstrained set. Solving the dual is usually less computationally expensive than finding the primal. Under a discrete setting, the primal involves computing an $N \times N$ matrix whereas the dual is an $N \times 3$ matrix.

Note that for any $x_0 \in \mathbb{R}$ it holds that

$$D_{\bar{\mu}}(u_1, u_2, \Delta) = D_{\bar{\mu}}(u_1 + x_0, u_2, \Delta).$$

Hence, if (u_1^*, u_2^*, Δ^*) maximizes $\sup_{(u_1,u_2,\Delta)\in\mathcal{U}} D_{\bar{\mu}}$ then so does $(u_1^* + c, u_2^*, \Delta^*)$. Therefore by taking $x_0 = -\mathbb{E}^{\bar{\mu}}[e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)}]$ we can always work with a normalized form of the minimizing measure μ^* given by

$$\mu^*(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2)e^{(u_1^*\oplus u_2^*)(s_1,s_2)+\Delta^{*\otimes}(s_1,s_2)} \quad \bar{\mu} - a.s. \tag{2.10}$$

The maximizing dual variables (u_1^*, u_2^*, Δ^*) satisfy

$$\frac{\partial D_{\bar{\mu}}}{\partial u_1(s_1)}\Big|_{u_1=u_1^*} = \frac{\partial D_{\bar{\mu}}}{\partial u_2(s_2)}\Big|_{u_2=u_2^*} = \frac{\partial D_{\bar{\mu}}}{\partial \Delta(s_1)}\Big|_{\Delta=\Delta^*} = 0.$$

This yields a system of equations provided by the following proposition.

Proposition 2.4.5. *Let* (u_1^*, u_2^*, Δ^*) *be the maximizing dual variables to* (2.3)*. Then they satisfy the system of equations*

$$
\begin{aligned}
u_1(s_1) &= \ln \mu_1(s_1) - \ln \int \bar{\mu}(s_1, ds_2) e^{u_2(s_2)+(s_2-s_1)\Delta(s_1)}, \ \ \textit{for all } \mu_1 - a.a.\ s_1, \\
u_2(s_2) &= \ln \mu_2(s_2) - \ln \int \bar{\mu}(ds_1, s_2) e^{u_1(s_1)+(s_2-s_1)\Delta(s_1)}, \ \ \textit{for all } \mu_2 - a.a.\ s_2, \\
0 &= \int \bar{\mu}(s_1, ds_2)(s_2 - s_1) e^{u_2(s_2)+(s_2-s_1)\Delta(s_1)}, \ \ \textit{for all } \mu_1 - a.a.\ s_1.
\end{aligned}
$$

The proof is a straightforward adaptation of [22], Proposition 44 and is hence omitted. The above equations can be solved iteratively using an extension of Sinkhorn's algorithm which is a fixed point method that iterates over the computation of one-dimensional gradients to approximate the optimizer. Sinkhorn's algorithm [35] was popularised by Cuturi [36] in the context of quickly solving optimal transport problems. More recently it's been used by [37] to build arbitrage free smile interpolations and by [22] to jointly calibrate the S&P 500 and VIX smile. Starting from an initial guess $u^{(0)} = \left(u_1^{(0)}, u_2^{(0)}, \Delta^{(0)}\right)$, we can define $u^{(n+1)}$ using $u^{(n)}$ by

$$
\begin{aligned}
u_1^{(n+1)}(s_1) &= \ln \mu_1(s_1) - \ln \int \bar{\mu}(s_1, ds_2) e^{u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)} \\
u_2^{(n+1)}(s_2) &= \ln \mu_2(s_2) - \ln \int \bar{\mu}(ds_1, s_2) e^{u_1^{(n+1)}(s_1)+(s_2-s_1)\Delta^{(n)}(s_1)} \\
0 &= \int \bar{\mu}(s_1, ds_2)(s_2 - s_1) e^{u_2^{(n+1)}(s_2)+(s_2-s_1)\Delta^{(n+1)}(s_1)}
\end{aligned}
\tag{2.11}
$$

till we get convergence. $\Delta^{(n+1)}$ is defined implicitly and can be solved using a one-dimensional root solver. We now prove that the density $\mu^{(n)}$ defined using $\left(u_1^{(n)}, u_2^{(n)}, \Delta^{(n)}\right)$ converges to μ^*.

Lemma 2.4.2. *Given* $\left(u_1^{(0)}, u_2^{(0)}, \Delta^{(0)}\right) \in \mathcal{U}$*, define* $\left(u_1^{(n)}, u_2^{(n)}, \Delta^{(n)}\right)$ *recursively as in* (2.11)*.*

Let $\mu^{(0)} := \bar{\mu} \in \mathcal{P}(\Omega)$ *and let* $\mu^{(3n)}, \mu^{(3n+1)}, \mu^{(3n+2)} \in \mathcal{P}(\Omega)$ *for* $n \in \mathbb{N}$ *be defined as*

$$\mu^{(3n)}(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2)\, e^{u_1^{(n)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)} \tag{2.12}$$

$$\mu^{(3n+1)}(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2)\, e^{u_1^{(n+1)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)} \tag{2.13}$$

$$\mu^{(3n+2)}(ds_1, ds_2) = \bar{\mu}(ds_1, ds_2)\, e^{u_1^{(n+1)}(s_1)+u_2^{(n+1)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)}. \tag{2.14}$$

Then for all $n \in \mathbb{N}$ *it holds that*

$$H\left(\mu^{(3n)}|\mu^{(3n-1)}\right) = 0,$$

$$H\left(\mu^{(3n+1)}|\mu^{(3n)}\right) = \mathbb{E}^{\mu_1}\left[u_1^{(n+1)} - u_1^{(n)}\right],$$

$$H\left(\mu^{(3n+2)}|\mu^{(3n+1)}\right) = \mathbb{E}^{\mu_2}\left[u_2^{(n+1)} - u_2^{(n)}\right].$$

Proof. Observe that (2.12) - (2.14) correspond to densities satisfying exactly one constraint of the martingale Schrödinger problem. That is, the first marginal of $\mu^{(3n+1)}$ is μ_1, the second marginal of $\mu^{(3n+2)}$ is μ_2 and $\mu^{(3n)}$ satisfies the martingale condition.

It holds that

$$\begin{aligned} H\left(\mu^{(3n)}|\mu^{(3n-1)}\right) &= \int \ln\left(\frac{e^{u_1^{(n)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)}}{e^{u_1^{(n)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n-1)}(s_1)}}\right) d\mu^{(3n)} \\ &= \int (s_2 - s_1) d\mu^{(3n)} \\ &= 0. \end{aligned}$$

Similarly,

$$
\begin{aligned}
H\left(\mu^{(3n+1)}|\mu^{(3n)}\right) &= \int \ln\left(\frac{e^{u_1^{(n+1)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)}}{e^{u_1^{(n)}(s_1)+u_2^{(n)}(s_2)+(s_2-s_1)\Delta^{(n)}(s_1)}}\right) d\mu^{(3n+1)} \\
&= \int u_1^{(n+1)} - u_1^{(n)} d\mu^{(3n+1)} \\
&= \mathbb{E}^{\mu_1}\left[u_1^{(n+1)} - u_1^{(n)}\right]
\end{aligned}
$$

Using a similar argument we conclude that

$$
H\left(\mu^{(3n+2)}|\mu^{(3n+1)}\right) = \mathbb{E}^{\mu_2}\left[u_2^{(n+1)} - u_2^{(n)}\right].
$$

□

Proposition 2.4.6. *Let $\mu^* \in \mathcal{M}(\mu_1, \mu_2)$ be the solution to the martingale Schrödinger problem under a reference measure $\bar{\mu} \in \mathcal{P}(\Omega)$. It holds for all $n \in \mathbb{N}$ that*

$$
H(\mu^*|\mu^{(n)}) = H(\mu^*|\bar{\mu}) - \sum_{i=0}^{n-1} H(\mu^{(i)}|\mu^{(i-1)}) - \mathbb{E}^{\mu_1}\left[u_1^{(0)}\right] - \mathbb{E}^{\mu_2}\left[u_2^{(0)}\right]. \tag{2.15}
$$

In particular, $H(\mu^*|\mu^{(n)})$ is decreasing in n.

Proof. From Lemma (2.4.2) it holds that

$$
\sum_{i=0}^{3n} H(\mu^{(i)}|\mu^{(i-1)}) = \mathbb{E}^{\mu_1}\left[u_1^{(3n)} - u_1^{(0)}\right] + \mathbb{E}^{\mu_2}\left[u_2^{(3n)} - u_2^{(0)}\right]. \tag{2.16}
$$

Observe that,

$$\begin{aligned}
H(\mu^*|\bar{\mu}) - H(\mu^*|\mu^{(3n)}) &= \mathbb{E}^{\mu^*}\left[\ln\left(\frac{d\mu^{(3n)}}{d\bar{\mu}}\right)\right] \\
&= \mathbb{E}^{\mu^*}\left[u_1^{(3n)} + u_2^{(3n)} + (s_2 - s_1)\Delta^{(3n)}\right] \\
&= \mathbb{E}^{\mu_1}[u_1^{(3n)}] + \mathbb{E}^{\mu_2}[u_2^{(3n)}]
\end{aligned}$$

Substituting (2.16), it holds that

$$H(\mu^*|\mu^{(3n)}) = H(\mu^*|\bar{\mu}) - \sum_{i=0}^{3n} H(\mu^{(i)}|\mu^{(i-1)}) - \mathbb{E}^{\mu_1}\left[u_1^{(0)}\right] - \mathbb{E}^{\mu_2}\left[u_2^{(0)}\right].$$

We obtain a similar result by replacing $3n$ with $3n+1$ and $3n+2$. Hence (2.15) follows. Finally, observe that $H(\cdot|\cdot) \geq 0$ and so $H(\mu^*|\mu^{(n)})$ is decreasing in n.

□

2.5 Bilateral Gamma Sato marginals

Pricing cliquets requires the marginal distributions of the underlier at specific maturities. One practical issue is that there may not be any option expiring at the required maturities. To circumvent this, we fit a Bilateral Gamma Sato process to the nearest two maturities for which we have data and use it to obtain the distribution for the specific maturity.

2.5.1 Bilateral Gamma Process

The Bilateral Gamma process has been described in Chapter 1.

2.5.2 Sato Process

Sato processes [38] allow for the modelling of option prices across strikes and maturities. They are related to self-decomposable laws which are a sub-class of infinitely divisible distributions. The distribution of a random variable X is self-decomposable ([39], page 90, Definition 15.1) if for any constant $c \in (0,1)$, there exists an independent random variable $X^{(c)}$ such that

$$X =^{\text{law}} cX + X^{(c)},$$

i.e., it has the same distribution as the sum of a scaled down version of itself (cX) and an independent residual random variable ($X^{(c)}$). Self-decomposable laws are suitable candidates for the return distribution because they are unimodal, infinitely divisible and can be characterized as limit laws.

A self-similar process $\{Y_s, s \geq 0\}$ is a stochastic process with the property that for any $\alpha > 0$ there exists $\beta(\alpha)$ such that for all $s > 0$ it holds that

$$Y_{\alpha s} =^{\text{law}} \beta(\alpha) Y_s. \tag{2.17}$$

It can be proven that $\beta(\alpha) = \alpha^\gamma$ for some $\gamma \geq 0$. Such a process is called self-similar with index γ.

An additive process is a process with inhomogeneous and independent increments. The Levy process is a special case of an additive process with homogeneous increments. Sato ([40]) showed that a distribution is self-decomposable if and only if it is the distribution of a self-similar additive process at unit time. In other words, if X is a self-decomposable law then there exists a self-similar additive process $Y_s, s \geq 0$ such that $Y_1 =^{\text{law}} X$. Such a process is called a Sato process. Taking $\alpha = t$ and $s = 1$, it follows that the law of a Sato process at time t is given by

$$Y_t =^{\text{law}} t^\gamma X.$$

2.5.3 Bilateral Gamma Sato Process

Let X be a random variable with bilateral Gamma distribution with the characteristic function $\phi_X(u)$ and density $f_X(x)$. It can be shown that X is self-decomposable and hence we can construct a bilateral Gamma Sato process $Y = \{Y_t, t \geq 0\}$. It holds that the characteristic function $\phi_{Y_t}(u)$ is given by

$$\begin{aligned}
\phi_{Y_t}(u) &= \mathbb{E}[\exp(iuY_t)] \\
&= \mathbb{E}[\exp(iut^\gamma X)] \\
&= \phi_X(ut^\gamma) \\
&= (1 + ib_n ut^\gamma)^{-c_n}(1 - ib_p ut^\gamma)^{-c_p},
\end{aligned}$$

and the density $f_{Y_t}(x)$ is

$$f_{Y_t}(x) = f_X\left(\frac{x}{t^\gamma}\right)\frac{1}{t^\gamma}.$$

2.6 Discretizing schemes for marginals

In practice, the marginal distributions need to be decomposed over a finite dimensional basis. It is important for the quantization scheme to preserve the convex ordering of the marginals in order for the problem to remain feasible. Alfonsi et al ([41], Corollary 2.2) provide a useful result that gives necessary and sufficient conditions for two probability measures with finite support to be in the convex order.

Proposition 2.6.1. *Let* $\mu = \sum_{i=1}^{I} \mu_i \delta_{x_i}$ *and* $\nu = \sum_{j=1}^{J} \nu_j \delta_{y_j}$ *be two probability measures on* $\mathbb{R}$*. Let* $F_\mu(x) = \mu((-\infty, x])$ *denote the cumulative distribution and let* $\phi_\mu(t) = \int_{-\infty}^{t} F_\mu(x)dx$*. Without loss of generality, assume that* $x_1 < \ldots < x_I$*,* $y_1 < \ldots < y_J$ *and* $\mu_1\mu_I\nu_1\nu_J > 0$*. Then it holds that* $\mu \leq_{cx} \nu$ *if and only if*

1. $y_1 \leq x_1$ *and* $x_I \leq y_J$
2. *for all* $j \in \{1, \ldots, I\}$ *such that* $x_1 \leq y_j \leq x_I$ *it holds that* $\phi_\mu(y_j) \leq \phi_\nu(y_j)$
3. $\sum_{i=1}^{I} \mu_i x_i = \sum_{j=1}^{J} \nu_j y_j$

We apply the *Dual Quantization* scheme as introduced by Pagès et al. ([42]) to discretize the marginals distributions. This method assumes that the probability has bounded support which is easily achieved by truncating the tails. Other discretization schemes that preserve convex ordering include ([41], [43], [44]).

Let $\{x_i\}_{i=1}^{I} \in \mathbb{R}$ be a set of grid points with $x_1 < \ldots < x_I$ and suppose μ and ν are two probability distributions in increasing convex order ($\mu \leq_{cx} \nu$) with bounded support in $[x_1, x_I]$. The dual quantization of μ is defined by

$$\hat{\mu} := \mu(\{x_1\})\delta_{x_1} + \sum_{i=1}^{I-1} \int_{(x_i, x_{i+1}]} \frac{x - x_i}{x_{i+1} - x_i} \mu(dx)\delta_{x_{i+1}} + \int_{(x_i, x_{i+1}]} \frac{x_{i+1} - x}{x_{i+1} - x_i} \mu(dx)\delta_{x_i}.$$

This is equivalent to the law

$$\hat{X} := X1_{X=x_1} + \sum_{i=1}^{I-1} 1_{X\in(x_i,x_{i+1}]}\left(x_{i+1}1_{U\le\frac{X-x_i}{x_{i+1}-x_i}} + x_i 1_{U>\frac{X-x_i}{x_{i+1}-x_i}}\right),$$

where U is uniformly distributed on $[0,1]$ independent of X. Following [41] it can be shown that dual quantization preserves convex ordering. Let $\phi:\mathbb{R}\to\mathbb{R}$ be a convex function and define $\hat{\phi}:[x_1,x_I]\to\mathbb{R}$ by $\hat{\phi}(x)=\frac{x-x_i}{x_{i+1}-x_i}\phi(x_{i+1})+\frac{x_{i+1}-x}{x_{i+1}-x_i}\phi(x_i)$ for $x\in[x_i,x_{i+1}], i=1,\dots,I-1$. This is a convex function and it holds that

$$\mathbb{E}^{\hat{\mu}}[\phi(x)] = \mathbb{E}^{\mu}[\hat{\phi}(x)] \le \mathbb{E}^{\nu}[\hat{\phi}(x)] = \mathbb{E}^{\hat{\nu}}[\phi(x)].$$

2.7 Implementation Details

2.7.1 Forward starting options and Cliquets

Forward starting call and put options with strike K are options with the payoff

$$\left(\frac{S_2}{S_1}-K\right)^+ \quad \text{and} \quad \left(K-\frac{S_2}{S_1}\right)^+ \text{(resp.)} \tag{2.18}$$

Valuing such a payoff requires the risk neutral joint distribution of (S_1,S_2) which can be computed using the martingale Schrödinger problem if the marginal distributions of S_1 and S_2 are known. However, forward starts are not traded and we only have market data on cliquets which are essentially a series of forward starting options.

We consider two types of cliquets. Rolling cliquets consist of a series of forward starts with increasing maturities and a given strike K, where the return from the earlier option rolls onto the next one. Given a set of maturities $0 < T_1 < T_2 < \ldots < T_n$, let $S_1, S_2, \ldots S_n$ be the value of the underlier at each maturity. The payoff for a rolling cliquet (call) is given by

$$\sum_{i=1}^{n-1} \left(\frac{S_{i+1}}{S_i} - K \right)^+ .$$

We can price a cliquet by obtaining $n-1$ joint distributions $\mathbb{Q}_i(S_{i+1}, S_i)$ and thereby computing

$$\sum_{i=1}^{n-1} \mathbb{E}^{\mathbb{Q}_i(S_{i+1},S_i)} \left[\left(\frac{S_{i+1}}{S_i} - K \right)^+ \right] .$$

We also consider at the money cliquets with a local cap (LC) and a global floor (GF) that are defined by a payoff

$$\max \left(\sum_{i=1}^{n-1} \min \left(\frac{S_{i+1} - S_i}{S_i}, LC \right), GF \right) .$$

These can be priced using Monte Carlo methods once we have computed the joint distribution.

2.7.1.1 Description of market data

Cliquet prices are obtained using the Totem service by IHS Markit. Totem provides consensus based prices for a variety of over the counter derivatives. In this study we look at cliquets on the S&P500 (SPX) with a 1 year maturity. We consider market prices of cliquets on six days in 2019: *20-03-2019*, *18-04-2019*, *19-06-2019*, *22-07-2019*, *19-09-2019*, *17-12-2019*. In total the data consisted of 180 options.

The rolling cliquets were rolled after 1-month, 3-month and 6-month gaps and the strikes ranged from 0.86 to 1 for puts and 1 to 1.14 for calls. On each day there are 3 rolling puts and 3 rolling calls. The cliquets with local cap/global floor consisted of 9 options with a global floor of 0 and local cap ranging from 1% to 3% for a 1-month rolling period and 2% to 6% for 3-months.

2.7.2 Computing discrete marginal densities

The marginal densities are obtained by fitting the bilateral Gamma Sato model for each of the 12 maturities $t_i \in \{\frac{1}{12}, \frac{2}{12}, \ldots, 1\}$ on S&P500 (SPX) out of the money vanilla options. In case there are no options for a specific maturity t_i, we calibrate the model to the nearest available maturities above and below t_i.

Assuming non-zero interest rate and dividends, let S_i and F_i represent the underlier and forward at maturity t_i. Let $X_i := \frac{S_i}{F_i}$ and let $f_{X_i}(x)$ denote the bilateral Gamma Sato risk neutral-density. For all $i \in \{1, \ldots, 12\}$ it holds that

$$\mathbb{E}^{f_{X_i}}[X_i] = 1.$$

For any X_i, X_j, $i < j$ a joint density $\mu_{(X_i,X_j)}(x_i, x_j)$ is a martingale if and only if

$$\mathbb{E}^{\mu_{(X_i,X_j)}}[X_j|X_i] = X_i$$

or alternatively, for all $x_i \in \mathbb{R}_{>0}$ it holds

$$\int (x_j - x_i)\mu_{(X_j,X_j)}(x_i, x_j)dx_j = 0.$$

A non-uniform grid is used to discretize the densities with $N = 100$ points. Compared to a uniform grid, the non-uniform grid is desirable in modelling as it gives a higher resolution at a given location, which is near the spot S_0 in our case. We follow Mijatović and Pistorius [45] to construct the grid. Let a be the lower bound and b the upper bound of the state space and let g be a density parameter that controls how closely packed the grid points are around S_0. Let

$$c_1 = \operatorname{arcsinh}\left(\frac{a - S_0}{g}\right)$$
$$c_2 = \operatorname{arcsinh}\left(\frac{b - S_0}{g}\right)$$

Let $k \in \{1, \dots, N/2\}$. The lower half of the grid is then defined as

$$s_k = S_0 + g \sinh\left(c_1\left(1 - \frac{k-1}{N/2 - 1}\right)\right)$$

and the upper half as

$$s_{k+N/2} = S_0 + g \sinh\left(c_2 \frac{k}{N/2}\right).$$

The densities $f_{X_i}(x)$ are defined over the entire interval $\mathbb{R}$ and must be truncated before they can be discretized using Dual Quantization. For a given f_{X_i} and an interval $[a_i, b_i]$, define the truncated density $\hat{f}_{X_i}$ to be

$$\hat{f}_{X_i}(x) = \frac{1}{N} f_{X_i}\left(x\frac{M}{N}\right) 1_{x \in \left[a_i \frac{N}{M}, b_i \frac{N}{M}\right]} \frac{M}{N},$$

where $N := \int_a^b f(x)dx$ and $M := \int_a^b x f(x)$. It is easy to see that $\hat{f}_{X_i}(x)$ is a density with mean 1 over a compact support. We take $a_1 = 0.6$ and $b_1 = 1.4$ for the support of X_1 and keep expanding the support for each successive X_i so that for all $i < j$, $\operatorname{supp}(X_i) \subset \operatorname{supp}(X_j)$.

2.7.3 Computing discrete joint densities

The discrete marginals allow us to run the optimization problem and obtain the joint density required for pricing cliquets. For the cliquets with 1-month gaps we compute 11 separate joint densities corresponding to the 11 marginal density pairs (X_i, X_{i+1}), $i \in \{1, \ldots 11\}$. Likewise, there are 4 joint densities for the 3-month cliquets using the pairs $\{(X_1, X_3), (X_3, X_6), (X_6, X_9), (X_9, X_{12})\}$ and 2 joint densities for the 6-month cliquets i.e., $\{(X_1, X_6), (X_6, X_{12})\}$.

By linearity, the price of rolling cliquets can be decomposed into a sum of forward contracts whose value can be computed individually. In contrast, the cliquets with a global floor cannot be separated as individual contracts and must therefore be computed using Monte Carlo methods. We use $N = 2^{14}$ realizations in our experiments.

2.7.4 Choice of reference measure

Given (X_i, X_j) for $i < j$, with marginals $\mu_i(x_i), \mu_j(x_j)$, the reference measure typically used in entropic optimal transport is the product measure $(\mu_i \otimes \mu_j)(x_i, x_j)$. This is equivalent to maximizing the joint Shannon entropy. However, $(\mu_i \otimes \mu_j)(x_i, x_j)$ implies independent marginals which is an unrealistic assumption in finance. Using a similar approach to [22] we seek a reference measure $\bar{\mu}(x_i, x_j)$ that is a martingale and matches μ_i. We vary $\bar{\mu}(x_i, x_j)$ for each (X_i, X_j) and hence each optimization has its own reference measure. Therefore, let $\bar{\mu}(x_i, x_j) = \mu_i(x_i)T(x_j|x_i)$, where $T(x_j|x_i)$ is a transitional kernel satisfying for all x_i, $\int x_j T(x_j|x_i)dx_j = x_i$. For a fixed x_i, let $T(x_j|x_i)$ be the distribution of x_iR, where R is a random variable with unit mean. Four different reference measures are constructed by choosing the distribution of R to be:

1. X_i, the BG-Sato marginal at maturity t_i (denoted by BGS - t_1).

2. X_j, the BG-Sato marginal at maturity t_j (denoted by BGS - t_2).

3. $X_{\frac{1}{12}}$ the BG-Sato marginal at maturity $t = \frac{1}{12}$ (denoted by BGS - t_0).

4. Lognormal, as $\exp\left(\sigma\sqrt{T}N - \frac{1}{2}\sigma^2 T\right)$ with $N \sim \mathcal{N}(0,1)$, $T = \frac{1}{12}$ and $\sigma = 20$ (denoted by Lognorm - 20).

We consider two additional reference measures: $\mu_i \otimes \mu_j$ (denoted by Max - Ent) and following Madan [46], computing two joint densities that maximize and minimize the expected squared log return, i.e., $\mathbb{E}[(\log \frac{S_2}{S_1})^2]$ and using their average as a reference measure denoted by Log - Sq.

2.8 Results

We compare the cliquet market prices with the ones generated from the six reference measures in the following figures. The average pricing error (APE) given by

$$\frac{1}{\text{mean option price}} \sum_{i}^{N} \frac{|model_i - market_i|}{N}$$

is reported for each day.

Date	Max Ent.	BGS T2	BGS T1	GBM 20
03-20	0.13	0.056	0.054	0.146
06-19	0.131	0.061	0.05	0.12
09-19	0.116	0.058	0.038	0.124
12-17	0.109	0.078	0.084	0.161

Table 2.1: APE for 1 month rolling cliquets

Date	Max Ent.	BGS T2	BGS T1	GBM 20
03-20	0.143	0.027	0.043	0.139
06-19	0.152	0.039	0.025	0.131
09-19	0.096	0.024	0.075	0.169
12-17	0.069	0.099	0.159	0.216

Table 2.2: APE for 3 month rolling cliquets

Date	Max Ent.	BGS T2	BGS T1	GBM 20
03-20	0.082	0.029	0.122	0.148
06-19	0.085	0.019	0.111	0.151
09-19	0.07	0.022	0.111	0.169
12-17	0.083	0.115	0.177	0.215

Table 2.3: APE for 6 month rolling cliquets

Date	Max Ent.	BGS T2	BGS T1	BGS T0	GBM 20
03-20	0.556	0.416	0.386	0.102	0.542
04-18	0.552	0.383	0.345	0.076	0.54
06-19	0.591	0.473	0.444	0.164	0.584
07-22	0.596	0.455	0.414	0.02	0.581
09-19	0.608	0.482	0.438	0.118	0.581
12-17	0.584	0.434	0.4	0.046	0.566

Table 2.4: APE for 1 month at the money cliquets with Local Cap

Date	Max Ent.	BGS T2	BGS T1	BGS T0	GBM 20
03-20	0.231	0.044	0.025	0.075	0.217
04-18	0.23	0.041	0.039	0.109	0.21
06-19	0.275	0.11	0.045	0.051	0.268
07-22	0.277	0.128	0.044	0.094	0.276
09-19	0.269	0.094	0.013	0.125	0.25
12-17	0.241	0.045	0.019	0.089	0.225

Table 2.5: APE for 3 month at the money cliquets with Local Cap

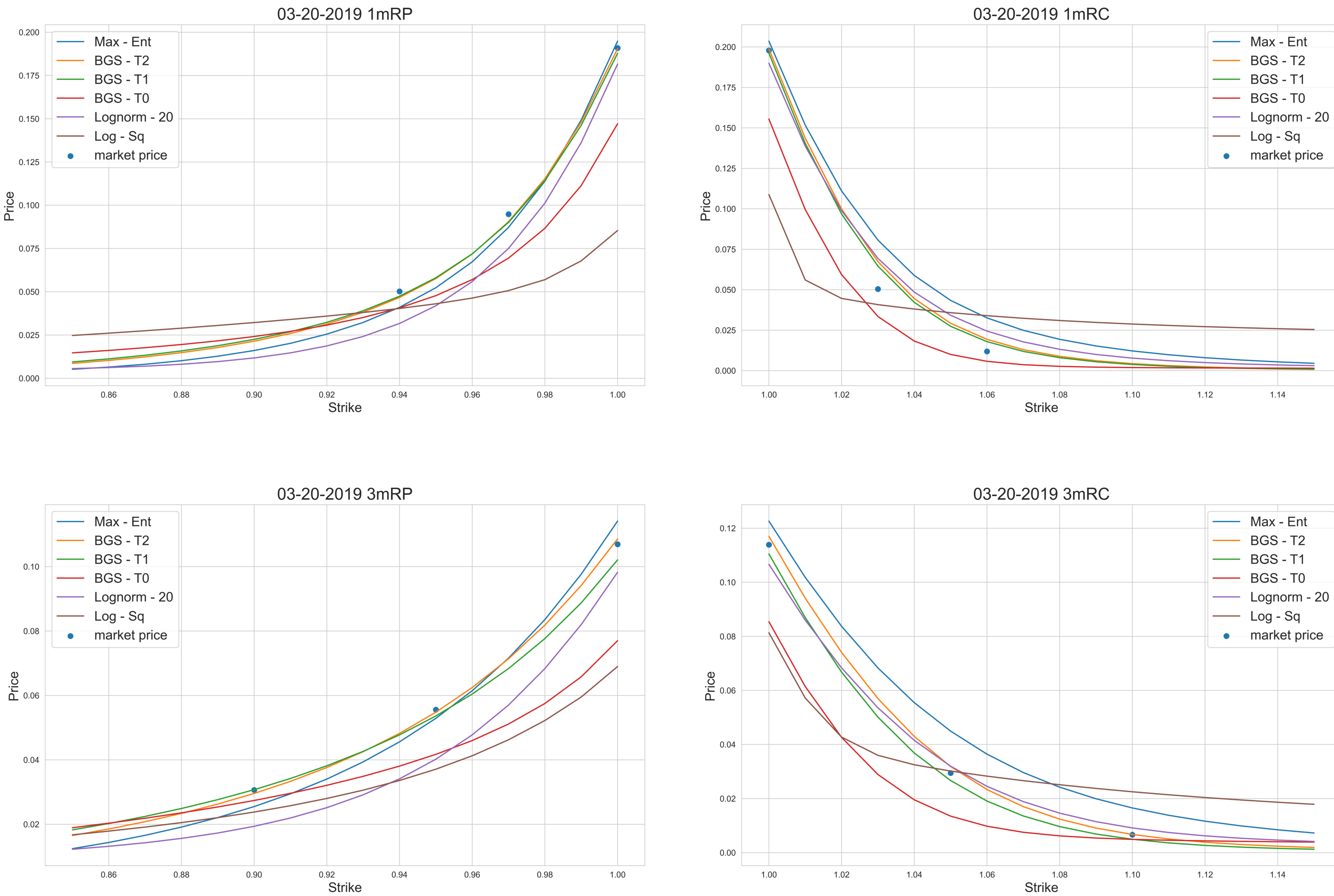


Figure 2.3: Price comparison for 1 month and 3 month rolling cliquets (March)

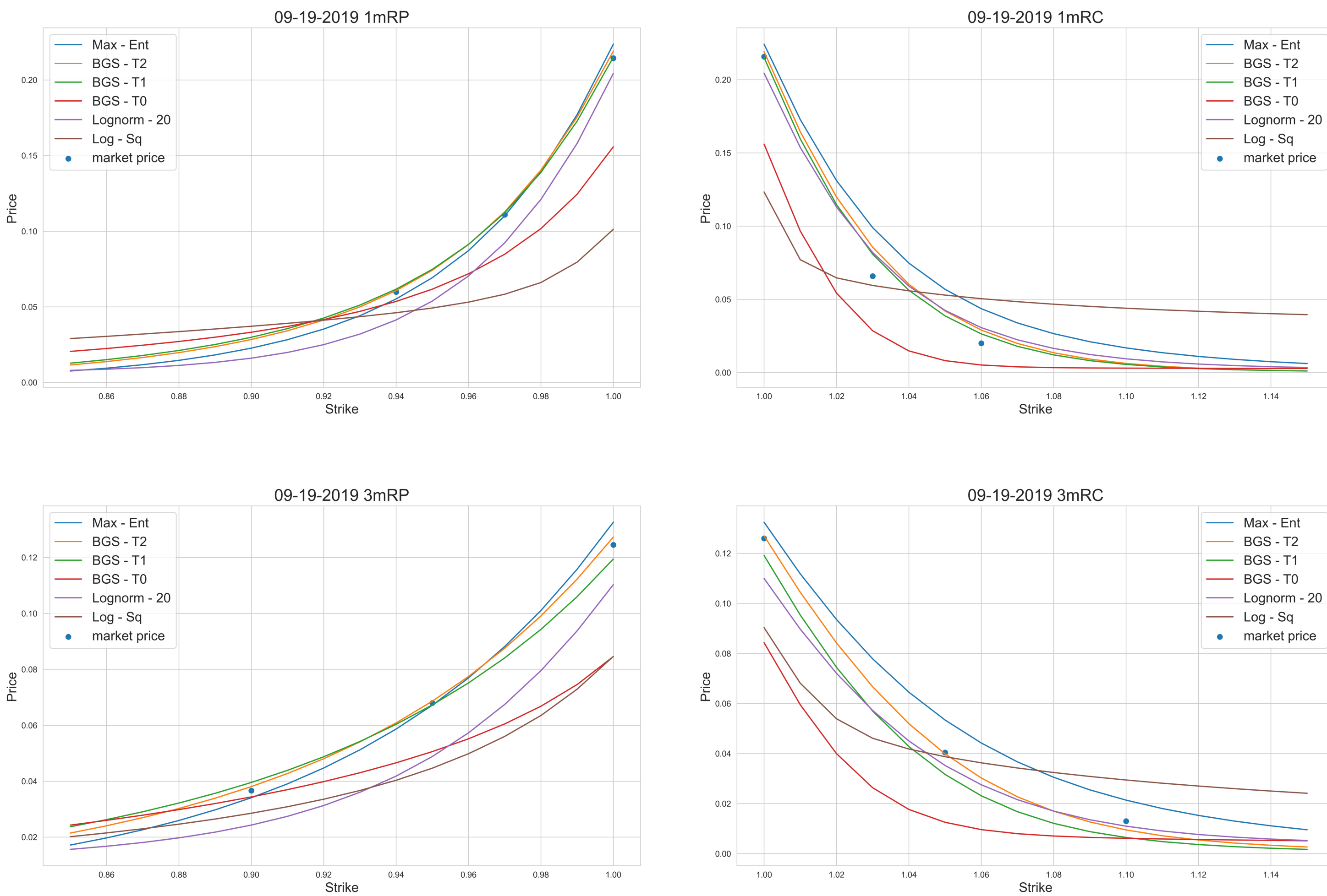


Figure 2.4: Price comparison for 1 month and 3 month rolling cliquets (September)

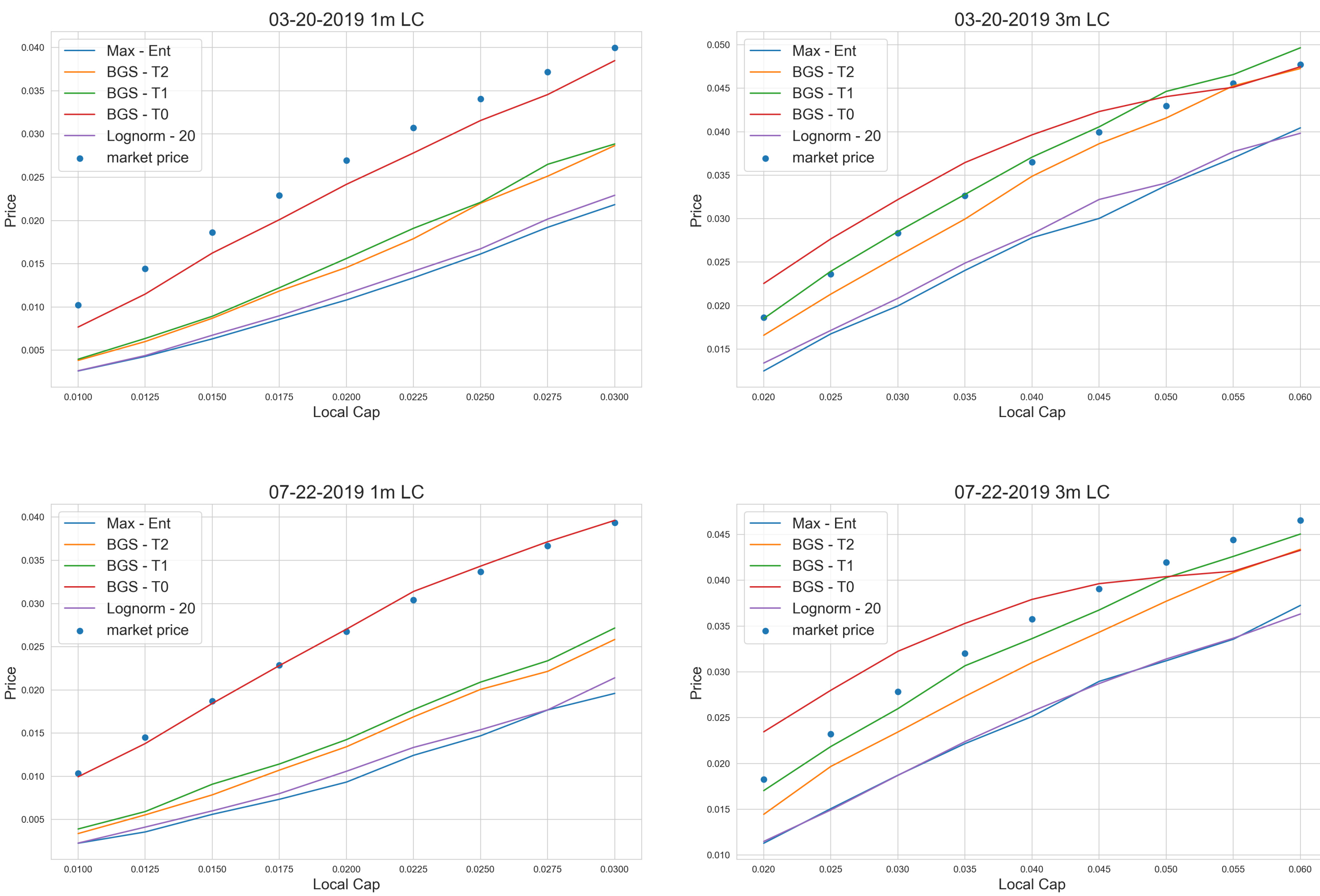


Figure 2.5: Price comparison for 1 month and 3 month cliquets with varying local caps (March and July)

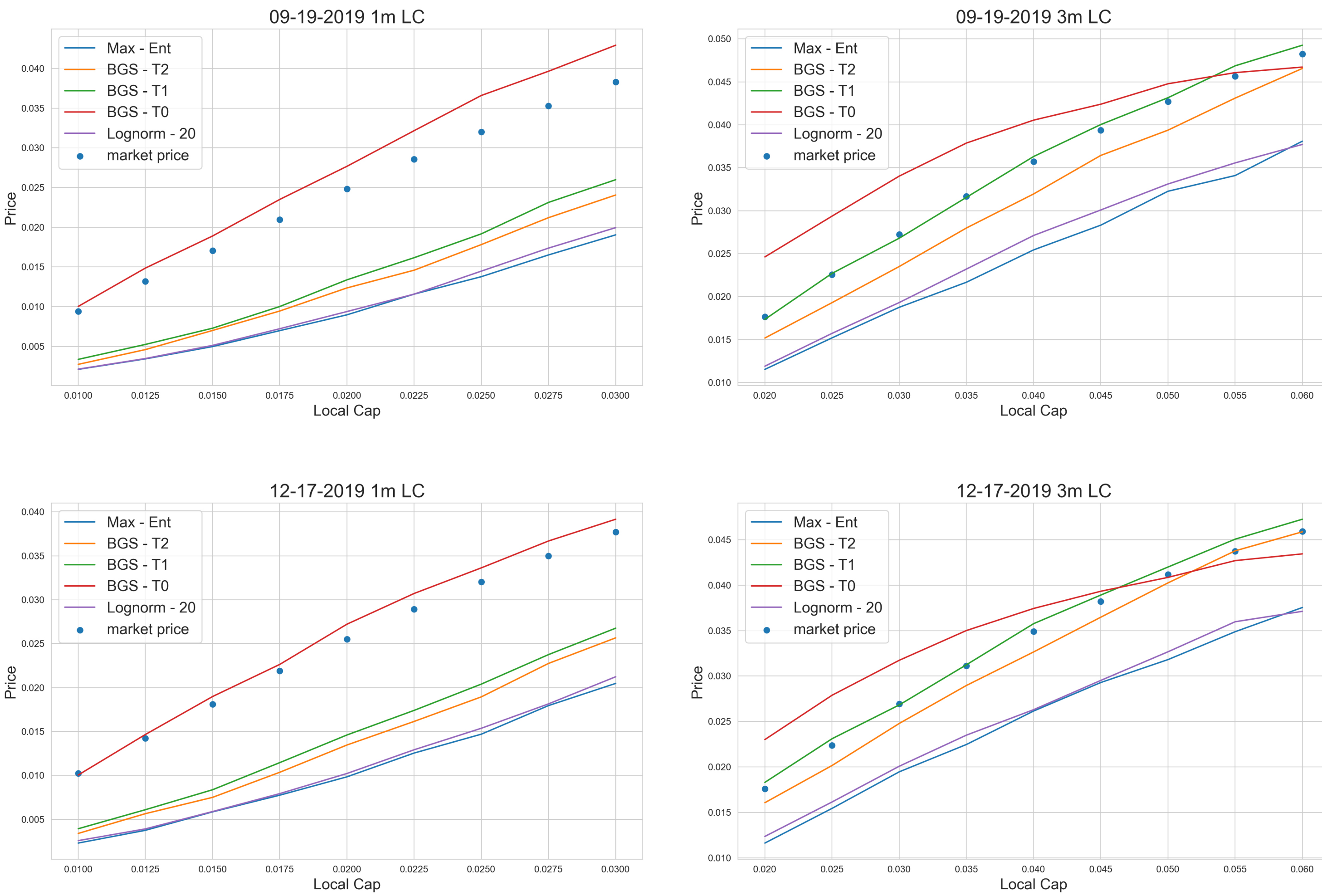


Figure 2.6: Price comparison for 1 month and 3 month cliquets with varying local caps (September and December)

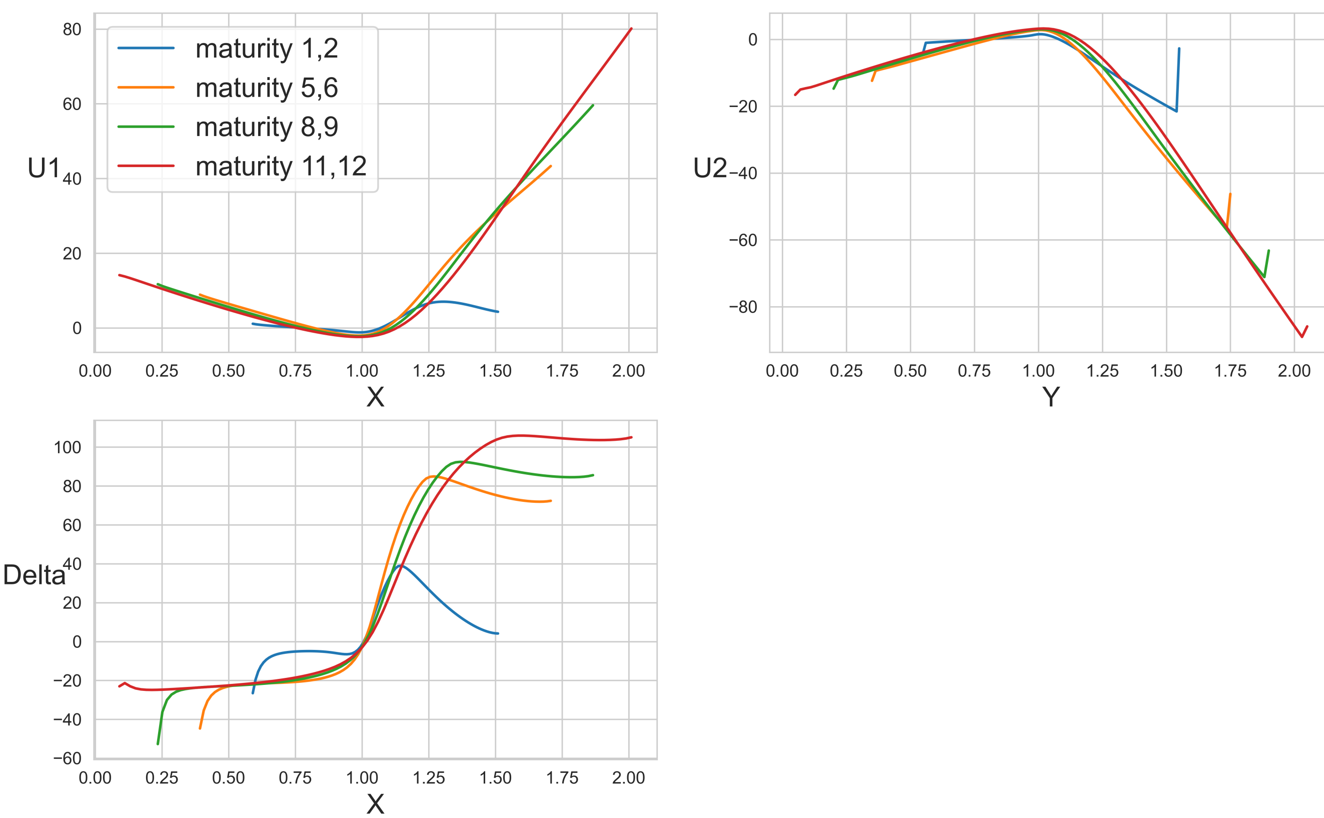


Figure 2.7: Dual variables for 1 month rolling cliquets at different maturities

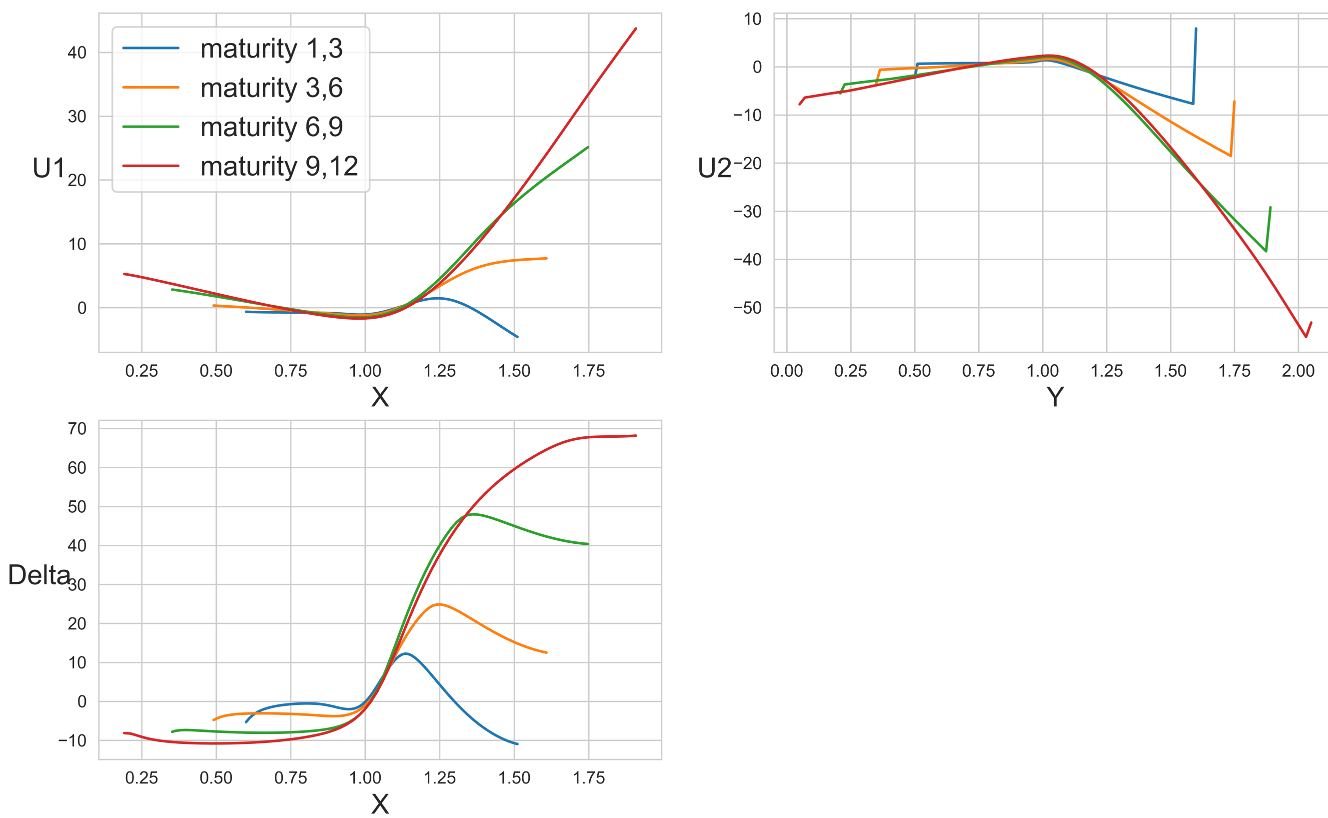


Figure 2.8: Dual variables for 3 month rolling cliquets at different maturities

2.9 Conclusion and Further Research

Cliquets have been priced under a wide variety of models including Heston, local volatility, and numerous Lévy and Sato models with a resulting wide range of prices computed. A basic requirement on a model is that it must fairly accurately price the vanilla options at the traded maturities and this dismisses the Black Scholes model as a candidate. Among the models meeting the requirement of matching the marginals the possible range of cliquet prices is quite large and no one knows which if any of these models is an acceptable candidate for a true model. Hence it is recognized that cliquet prices are not known and there is little interest in their value under specific models.

Our study shows that the optimal joint distribution derived from the reference measures BGS - t_1 and BGS - t_2 are able to provide a close fit to the market data and are clearly better than the other measures. In the case of rolling cliquets, BGS - t_2 provides the best fit for all rolling periods. In the case of 3 month locally capped cliquets BGS - t_1 has the better performance. Our models did not price the 1 month local caps correctly and further work is needed in order to determine the best pricing measure. It is important to note that our models are able to match the price of exotics using information only from the marginal distribution.

We also report the shape of the dual variables that we obtain from using Sinkhorns algorithm. As noted by [21] and [22], the dual variables u_1 and u_2 are similar to taking opposite positions in call options.

A future direction of research involves turning this method into a fully model-free approach. In this study we assumed that the marginals were modelled using Bilateral Gamma Sato process. Alternatively, we can replace the marginal constraints in the primal and dual by constraints using actual option prices. Given a collection of call options $\{C_i^1\}$ and $\{C_j^2\}$ with maturities t_1 and t_2 respectively at different strikes, we replace the marginal constraint with the constraints that

$$\forall i, j, \ \ \mathbb{E}^{\mu}[(S_1 - K_i)^+] = C_i^1 \quad \text{and} \quad \mathbb{E}^{\mu}[(S_2 - K_j)^+] = C_j^2.$$

For further examples using this approach see [22] and [23].

Another related problem that is a useful extension is considering constraints on the correlation of the joint distribution. We have seen that the martingale condition imposes a correlation structure on the optimal distribution. If we choose the reference measure to be $\mu_1 \otimes \mu_2$ then we can obtain an explicit distribution μ_{ent} with maximal entropy. Initial experiments show that correlation constraint must be greater than the correlation for μ_{ent} for the problem to remain feasible. Further work is needed to determine the lower bound on feasible correlation values. In addition we can also change the reference measure and study how the lower correlation bounds changes.

Chapter 3: Martingale Optimal Transport under Acceptability

3.1 Introduction

The price bounds under the Martingale Optimal Transport are based on the no arbitrage principle. The bounds are computed by considering the maximum and minimum values of the sub-replicating and super-replicating portfolios respectively. In complete markets it is possible to obtain reasonable prices for any cash flow using exact replication. However, as we have seen in Chapter 2, in incomplete markets where it is not possible to exactly replicate a payoff, the no arbitrage bounds are far too wide for any practical use. By recognizing that exclusion of arbitrage is too weak of an assumption, we can obtain narrower price bounds by removing additional payoffs that are not arbitrages.

In order to define the payoffs we need to exclude, we use the concept of *acceptable risks* introduced by Artzner et al. [47]. Acceptable risks are defined by a class of risk measures called *coherent risk measures*. The set of acceptable risks is postulated to be a convex cone that contains the set of all arbitrage opportunities. A critical result is that with each cone of acceptable risks, there is a corresponding set of probability measures. A random variable belongs to an acceptable cone if and only if it has non-negative expectation under each of the associated probability measures.

Madan and Cherny [48] provide an explicit representation for the cone of acceptability that can be used for pricing claims. Specifically, they introduce an *acceptability index* with which we can define a family of convex cones each representing payoffs acceptable at different levels. For example, a non-negative random variable accessed at zero costs, represents an arbitrage opportunity and will be acceptable at all levels.

In this chapter we introduce acceptability to the martingale optimal transport in order to get narrower price bounds. Other authors have dealt with this problem by introducing additional constraints. [23] calibrate their model to additional market instruments and impose a penalty term if the model is too far from a prespecified prior distribution. [49] improve the bounds by incorporating information about the variance of the underlying returns. Our approach is different in that, we relax the requirement that the cash flow must always be greater than the sub-replicating portfolio. Instead, we require the gap between the two to be acceptable. This is equivalent to stating that the cash flow is greater than the sub-replicating portfolio on average in some sense.

Our main contribution is a duality theorem similar to Theorem 2 under acceptability. We then use the explicit forms of acceptability given by [48] to provide an algorithm that numerically computes the upper and lower acceptable bounds of cliquets. A similar study was conducted by Madan [46] where they use related methods to compute bounds on forward start contracts.

3.2 Setting

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *risk* is defined as a random variable $X : \Omega \to \mathbb{R}$ that represents possible future values of a portfolio currently held. Let $\mathcal{P}$ denote the set of all probability measures on $\mathcal{F}$ that are absolutely continuous with respect to $\mathbb{P}$. We assume zero interest so that there is no discounting.

A set of risks $\mathcal{A}$ is convex if for any $\alpha \in [0,1]$ and $X, Y \in \mathcal{A}$, it holds that $\alpha X + (1-\alpha) Y \in \mathcal{A}$. Moreover, a set of risks is a cone if for any $c \in (0, \infty)$ we have $cX \in \mathcal{A}$.

3.3 Coherent measures and acceptability

A risk measure $\rho : L^1 \to \mathbb{R}$ is simply a functional that maps a random variable to a real number. If X represents a random payoff that we receive at a future date, then $\rho(X)$ is the minimum cash position that needs to be added today for the payoff to be acceptable. Holding a position with a larger risk corresponds to a larger $\rho(X)$. If on the other hand X is non-negative in all states, such as the case of holding an option, then $\rho(X)$ is negative representing the maximum amount that can be taken out of the future payoff for it to remain acceptable. In the classical sense, this would correspond to the premium of an option.

Artzner et al [47] introduce the notion of *acceptable risks* using a class of risk measures that satisfy a list of axioms. Such risk measures are termed as *coherent risk measures*.

Definition 3.3.1. *A coherent risk measure $\rho : \mathcal{A} \to \mathbb{R}$ is a mapping on a convex cone of risks that satisfies the following four properties:*

1. *(Translational Invariance): For all $X \in \mathcal{A}$ and $c \in \mathbb{R}$ we have $\rho(X + c) = \rho(X) + c$.*

2. *(Subadditivity): For all $X, Y \in \mathcal{A}$ we have $\rho(X + Y) \leq \rho(X) + \rho(Y)$.*

3. *(Positive Homogeneity): For all $X \in \mathcal{A}$ and $c \in (0, \infty)$ we have $\rho(cX) \in \mathcal{A}$.*

4. *(Monotonicity): For all $X, Y \in \mathcal{A}$ such that $X \leq Y$ almost surely we have $\rho(X) \leq \rho(Y)$.*

Moreover, they show that ρ is a coherent measure if and only if there exists a set of nonempty probability measures $\mathcal{D} \subset \mathcal{P}$ such that

$$\rho(X) = -\inf_{Q \in \mathcal{D}} \mathbb{E}^Q[X]. \tag{3.1}$$

The set $\mathcal{D}$ for a given risk measure ρ is not necessarily unique. However, there exists a largest set given by

$$\mathcal{D}^{(\rho)} = \{Q \in \mathcal{P} : \mathbb{E}^Q[X] \geq -\rho(X),\ \forall X \in L^1\}$$

and we term this as the set of *supporting kernels* for ρ.

We say that X is acceptable if $\rho(X) \leq 0$. This is equivalent to X being acceptable if and only if, for all $Q \in \mathcal{D}^{(\rho)}$, we have $\mathbb{E}^Q[X] \geq 0$. We introduce the *cone of acceptability* associated with ρ as the set $\mathcal{A}^{(\rho)}$ defined as

$$\mathcal{A}^{(\rho)} = \{X : \rho(X) \leq 0\}.$$

For any coherent risk measure ρ, $\mathcal{A}^{(\rho)}$ is clearly a convex cone as can be seen by the super-additivity of the infimum. Furthermore it contains the set of all non-negative payoffs i.e., $X \geq 0$ almost surely. This is equivalent to stating that all arbitrages are acceptable to the market.

The set of acceptable zero-cost cash flows does not remain static over a period of time. Indeed, we would expect it to change depending upon the state of the market. Under additional assumptions, Madan and Cherny [50] model the cone of acceptability using options on S&P500 and show that the cone contracted significantly during the 2008 financial crisis and opened up after that. The cone is modelled by introducing, for $x \in [0, \infty]$ an increasing family of coherent risk measures ρ_x (i.e, the map $x \mapsto \rho_x(X)$ is increasing for any X) that represent the market's acceptability of cash-flows under different stress levels. Let $\mathcal{A}_x$ be the associated cone of acceptable zero-cost cash flows and let $\mathcal{D}_x$ be the set of supporting kernels. It is easy to see that for $x' < x$, it holds that $\mathcal{A}_x \subset \mathcal{A}_{x'}$ and $\mathcal{D}_{x'} \subset \mathcal{D}_x$. As mentioned earlier, for each x, $\mathcal{A}_x$ contains the set of all non-negative random variables which represents the set arbitrage opportunities. An explicit form for ρ_x and $\mathcal{D}_x$ is provided in Section 5 that allows for pricing under different levels of acceptability.

3.4 Martingale optimal transport under acceptability

We assume the setting as in Chapter 2 with a measure space $(\Omega, \mathcal{B}(\Omega))$ where $\Omega = \mathbb{R}_{>0} \times \mathbb{R}_{>0}$. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{B}(\Omega))$ representing a physical measure with finite moments and let $\mathcal{P}(\Omega)$ be the set of all probability measures absolutely continuous with respect to $\mathbb{P}$. Let $\mathcal{M}(\mu_1, \mu_2)$ be the set of all martingales with the marginals μ_1 and μ_2. Let $c(s_1, s_2)$ be the terminal payoff whose upper and lower bounds we are interested in computing. In this section we restrict ourselves to the lower bound. Results on the upper bounds can be adapted easily. Let $\mathcal{A} \coloneqq \mathcal{A}_\rho$ denote a the convex cone of acceptable zero-cost cash flows. Then it holds that there exists $\mathcal{D} \subset \mathcal{P}(\Omega)$ such that

$$X \in \mathcal{A} \iff \mathbb{E}^Q[X] \geq 0, \forall Q \in \mathcal{D}.$$

Recall that the lower bound for $c(s_1, s_2)$ under the martingale optimal transport is essentially the highest possible value of a semi-static sub-replicating portfolio. That is,

$$\sup_{(u_1,u_2,\Delta)\in\mathcal{U}} \mathbb{E}^{\mu_1}[u_1(s_2)] + \mathbb{E}^{\mu_2}[u_2(s_2)]$$

subject to

$$\forall (s_1, s_2) \in \Omega, \;\; c(s_1, s_2) \geq u_1(s_1) + u_2(s_2) + (s_2 - s_1)\Delta(s_1).$$

Note that this corresponds to the no arbitrage condition. In order to compute the lowest *acceptable* bound for $c(s_1, s_2)$, we relax this constraint and consider all such (u_1, u_2, Δ) that satisfy

$$\forall (s_1, s_2) \in \Omega, \;\; c(s_1, s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1) \in \mathcal{A}.$$

which is equivalent to

$$\forall Q \in \mathcal{D}, \;\; \mathbb{E}^Q[c(s_1, s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1)] \geq 0.$$

In order to compute the lower bound for $c(s_1, s_2)$ numerically, we must transform the dual to the primal which is given by

$$\inf_{Q \in \mathcal{M}(\mu_1, \mu_2) \cap \mathcal{D}} \mathbb{E}^Q[c(s_1, s_2)] \tag{3.2}$$

The following proposition establishes the relation between the dual and primal formulation. We will assume a weak topology on the set of measures $\mathcal{P}(\Omega)$ that is addressed in the next section.

Proposition 3.4.1. *Let $\mathcal{D} \subset \mathcal{P}(\Omega)$ be a weakly compact set of probability measures under a suitable topology and let $\mathcal{D} \cap \mathcal{M}(\mu_1, \mu_2)$ be non-empty. Let $L^1_s(\mathcal{D}) \subset L^0$ be the set of measurable functions defined by*

$$L^1_s(\mathcal{D}) = \left\{ X \in L^0 : \lim_{n \to \infty} \sup_{Q \in \mathcal{D}} \mathbb{E}^Q[|X| 1_{|X| > n}] = 0 \right\}.$$

Let $c \in L^1_s(\mathcal{D})$, let $\mathcal{U}_{\mathcal{D}} := L^1_s(\mathcal{D}) \times L^1_s(\mathcal{D}) \times \mathcal{C}_b(\mathbb{R}_{>0})$ and let $\mathcal{U}_{c,\mathcal{D}} \subset \mathcal{U}_{\mathcal{D}}$ denote the set of all

$(u_1, u_2, \Delta) \in \mathcal{U}_{\mathcal{D}}$ *such that for all* $Q \in \mathcal{D}$ *it holds that*

$$\mathbb{E}^Q[c(s_1, s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1)] \geq 0.$$

Then it holds that

$$\inf_{Q \in \mathcal{M}(\mu_1, \mu_2) \cap \mathcal{D}} \mathbb{E}^Q[c(s_1, s_2)] = \sup_{(u_1, u_2, \Delta) \in \mathcal{U}_{c,\mathcal{D}}} \mathbb{E}^{\mu_1}[u_1(s_2)] + \mathbb{E}^{\mu_2}[u_2(s_2)]. \tag{3.3}$$

Proof. Observe that

$$\begin{aligned}
&\sup_{(u_1, u_2, \Delta) \in \mathcal{U}_{c,\mathcal{D}}} \mathbb{E}^{\mu_1}[u_1(s_2)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
=& \sup_{(u_1, u_2, \Delta) \in \mathcal{U}_{\mathcal{D}}} \mathbb{E}^{\mu_1}[u_1(s_2)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&+ \inf_{Q \in \mathcal{D}} \mathbb{E}^Q[c(s_1, s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1)]
\end{aligned}$$

Let $f_c : \mathcal{D} \times \mathcal{U} \to \mathbb{R}$ be a function defined by

$$\begin{aligned}
f_c(Q, (u_1, u_2, \Delta)) =& \mathbb{E}^Q[c(s_1, s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1)] \\
&+ \mathbb{E}^{\mu_1}[u_1(s_1)] + \mathbb{E}^{\mu_2}[u_2(s_2)].
\end{aligned}$$

We equip $L^1_s(\mathcal{D})$ with the L^1 norm, $\mathcal{C}_b(\mathbb{R}_{>0})$ with the sup norm and $\mathcal{D}$ is assumed to be compact under some topology. For any $Q \in \mathcal{D}$, $f_c(Q, \cdot)$ is continuous and concave on $\mathcal{U}_{\mathcal{D}}$. Moreover we have from the proof of Theorem 3.4 in [51] that for $X \in L^1_s(\mathcal{D})$ the map $Q \mapsto \mathbb{E}^Q[X]$ is weakly continuous. Therefore it holds that for any $(u_1, u_2, \Delta) \in \mathcal{U}_{\mathcal{D}}$, $f(\cdot, (u_1, u_2, \Delta))$ is convex and continuous in the weak topology. This justifies the use of Sion's minimax theorem to interchange the supremum and infimum. Hence it holds that

$$\begin{aligned}
&\sup_{(u_1,u_2,\Delta)\in\mathcal{U}_{\mathcal{D}}} \mathbb{E}^{\mu_1}[u_1(s_2)] + \mathbb{E}^{\mu_2}[u_2(s_2)] \\
&+ \inf_{Q\in\mathcal{D}} \mathbb{E}^{Q}[c(s_1,s_2) - u_1(s_1) - u_2(s_2) - (s_2 - s_1)\Delta(s_1)] \\
= &\inf_{Q\in\mathcal{D}} \mathbb{E}^{Q}[c(s_1,s_2)] + \sup_{(u_1,u_2,\Delta)\in\mathcal{U}_{\mathcal{D}}} \Big\{\mathbb{E}^{\mu_1}[u_1(s_1)] - \mathbb{E}^{Q}[u_1(s_1)] \\
&+ \mathbb{E}^{\mu_2}[u_2(s_2)] - \mathbb{E}^{Q}[u_2(s_2)] - \mathbb{E}^{Q}[(s_2 - s_1)\Delta(s_1)]\Big\}
\end{aligned}$$

The last equality implies that $Q \in \mathcal{D}$ must satisfy the following three constraints that for any $(u_1, u_2, \Delta) \in \mathcal{U}$ it holds that

$$\mathbb{E}^{Q}[u_1(s_1)] = \mathbb{E}^{\mu_1}[u_1(s_1)], \quad \mathbb{E}^{Q}[u_2(s_2)] = \mathbb{E}^{\mu_2}[u_2(s_2)], \quad \mathbb{E}^{Q}[(s_2 - s_1)\Delta(s_1)] = 0.$$

since otherwise we can scale $(u_1, u_2, \Delta) \in \mathcal{U}_{\mathcal{D}}$ to be arbitrarily large. Hence we conclude that

$$\begin{aligned}
&\inf_{Q\in\mathcal{D}} \mathbb{E}^{Q}[c(s_1,s_2)] + \sup_{(u_1,u_2,\Delta)\in\mathcal{U}_{\mathcal{D}}} \Big\{\mathbb{E}^{\mu_1}[u_1(s_1)] - \mathbb{E}^{Q}[u_1(s_1)] \\
&+ \mathbb{E}^{\mu_2}[u_2(s_2)] - \mathbb{E}^{Q}[u_2(s_2)] - \mathbb{E}^{Q}[(s_2 - s_1)\Delta(s_1)]\Big\} \\
&= \inf_{Q\in\mathcal{M}(\mu_1,\mu_2)\cap\mathcal{D}} \mathbb{E}^{Q}[c(s_1,s_2)].
\end{aligned}$$

□

3.5 Acceptability via distortions

In order to make the problem tractable, we need an explicit formulation for $\mathcal{D}$. Following [50] this can be achieved by making additional assumptions on the risk measure: law invariance and comonotone additivity.

Law invariance implies that acceptability only depends on the probability distribution of the risk. That is, if X and Y and two risks such that $X =^{law} Y$, then either both are acceptable or neither is acceptable.

Secondly, two risks X and Y are said to be comonotone if they are driven by a single risk factor. That is, there exists a random variable U on the unit interval such that

$$X = F_X^{-1}(U) \quad \text{and} \quad Y = F_Y^{-1}(U).$$

We assume that any two comonotone risks X and Y are additive, i.e., $\rho(X) + \rho(Y) = \rho(X + Y)$.

Kusuoka [52] show that under these two additional conditions acceptability can be defined through distorted expectations. That is, there exists an increasing concave distortion function $\Psi : [0, 1] \to [0, 1]$, with the property that $\Psi(0) = 0$ and $\Psi(1) = 1$ such that it holds

$$\inf_{Q \in \mathcal{D}} \mathbb{E}^Q[X] = \int_{-\infty}^{\infty} x d\Psi(F_X) \tag{3.4}$$

where the distorted expectation is given by

$$\int_{-\infty}^{\infty} x d\Psi(F_X) := -\int_{-\infty}^{0} \Psi(F_X)dx + \int_{0}^{\infty} (1 - \Psi(F_X))dx$$

Hence it holds that X is acceptable if and only if it has positive distorted expectation i.e., $\int_{-\infty}^{\infty} x d\Psi(F_X) \geq 0$. Results from Madan and Cherny, [48] show that the determining set of measures $\mathcal{D}$ in (3.4) is given by

$$\mathcal{D} = \left\{ Q \in \mathcal{P} : \mathbb{E}\left[\left(\frac{dQ}{dP} - a \right)^+ \right] \leq \Phi(a) \ \ \forall a \in \mathbb{R}_{>0} \right\},$$

where $\Phi : \mathbb{R}_{>0} \to \mathbb{R}$ is the convex conjugate of Ψ defined as

$$\Phi(a) = \sup_{u \in [0,1]} (\Psi(u) - au)$$

Cherny [51] (see Example 2.10, and references therein) show that $\mathcal{D}$ is a compact subset of $\mathcal{P}(\Omega)$ under the weak topology on L^1 induced by L^∞ functions. Recall that $\mathcal{P}(\Omega)$ can be characterized by Radon-Nikodyn derivatives with respect to $\mathbb{P}$.

3.5.1 Choice of distortion function

We can parametrize different levels of acceptability by introducing a family of distortion functions $\{\Psi^{(\gamma)}\}$. In this study we use the *minmaxvar* distortion function $\Psi^{(\gamma)} : [0,1] \to [0,1]$ introduced in [48] defined for $\gamma \in [0, \infty]$ by

$$\Psi^{(\gamma)}(u) = 1 - (1 - u^{\frac{1}{1+\gamma}})^{1+\gamma}$$

with the corresponding convex conjugate $\Phi^{(\gamma)}(a)$ given by

$$\Phi^{(\gamma)}(a) = 1 - \frac{a}{(1 + a^{\frac{1}{\gamma}})^{\gamma}}.$$

This distortion function allows us to define a family of acceptability cones $\mathcal{A}_\gamma$ and associated set of supporting kernels $\mathcal{D}_\gamma$ that represent the state of the market at different stress levels γ. A higher stress level implies a smaller cone $\mathcal{A}_\gamma$ and a larger set of kernels $\mathcal{D}_\gamma$. As γ increases $\mathcal{A}_\gamma$ shrinks to the set of arbitrages.

3.6 Numerical implementation

By defining acceptability through distortions, we are able to numerically compute the primal formulation via convex programming. In order to solve the numerical problem, we first must re-frame the problem using densities. Let $Z \coloneqq \frac{dQ}{dP}$ denote the density of Q w.r.t the base measure $\mathbb{P}$. Then the lower acceptable bound for c is computed via the following optimization problem

$$\begin{aligned}
& \inf_Z \mathbb{E}[Zc(s_1, s_2)] \\
s.t. \quad & \int Z(s_1, s_2)\mathbb{P}(s_1, s_2)ds_2 = \mu_1(s_1) \;\; \forall s_1 \\
& \int Z(s_1, s_2)\mathbb{P}(s_1, s_2)ds_1 = \mu_2(s_2) \;\; \forall s_2 \\
& \int (s_2 - s_1)Z(s_1, s_2)\mathbb{P}(s_1, s_2)ds_2 = 0 \;\; \forall s_1 \\
& \int (Z(s_1, s_2) - a)^+\mathbb{P}(s_1, s_2)ds_1ds_2 \leq \Phi(a) \;\; \forall a \in \mathbb{R}^+.
\end{aligned}$$

We use the CVXPY ([53], [54]) package to solve the convex program in Python. CVXPY uses *disciplined convex programming*, which is a system for constructing expressions of known curvature from base functions, to ensure that the optimization problems are convex. The optimization is done using the Embedded Conic solver [55] and the Splitting Cone solver [56].

The results of the optimization at different stress levels are presented in the next section.

We require a base measure that delivers prices close to the market quotes. The totem cliquet prices are the market's best candidate for a true price given the wide spreads observed across models. We use BGS - t_2 as the base measure since it provides the best fit to data. As higher stress levels correspond to larger spaces of supporting kernels $\mathcal{D}$ we expect the the bounds to approach the no arbitrage limits as γ increases and shrink to the base measure at smaller values. It is a fairly safe conjecture that if one uses a base measure delivering cliquet prices far from market then at low stress levels market prices will be outside the spread and at high stress levels the spreads will again be too large.

We use the same methods described in Chapter 2 to obtain the discrete marginals. We restrict ourselves to computing bounds on the 6-month rolling cliquets. In addition we discretize $\Phi(a)$ on the grid $[0.1, 2]$ with step size of 0.1.

3.7 Results

The results for the upper and lower price bounds for $\gamma \in \{0.1, 0.25, 4\}$ are reported along with the bounds from martingale optimal transport. We also show plots of the resulting joint distributions at different strikes and stress levels.

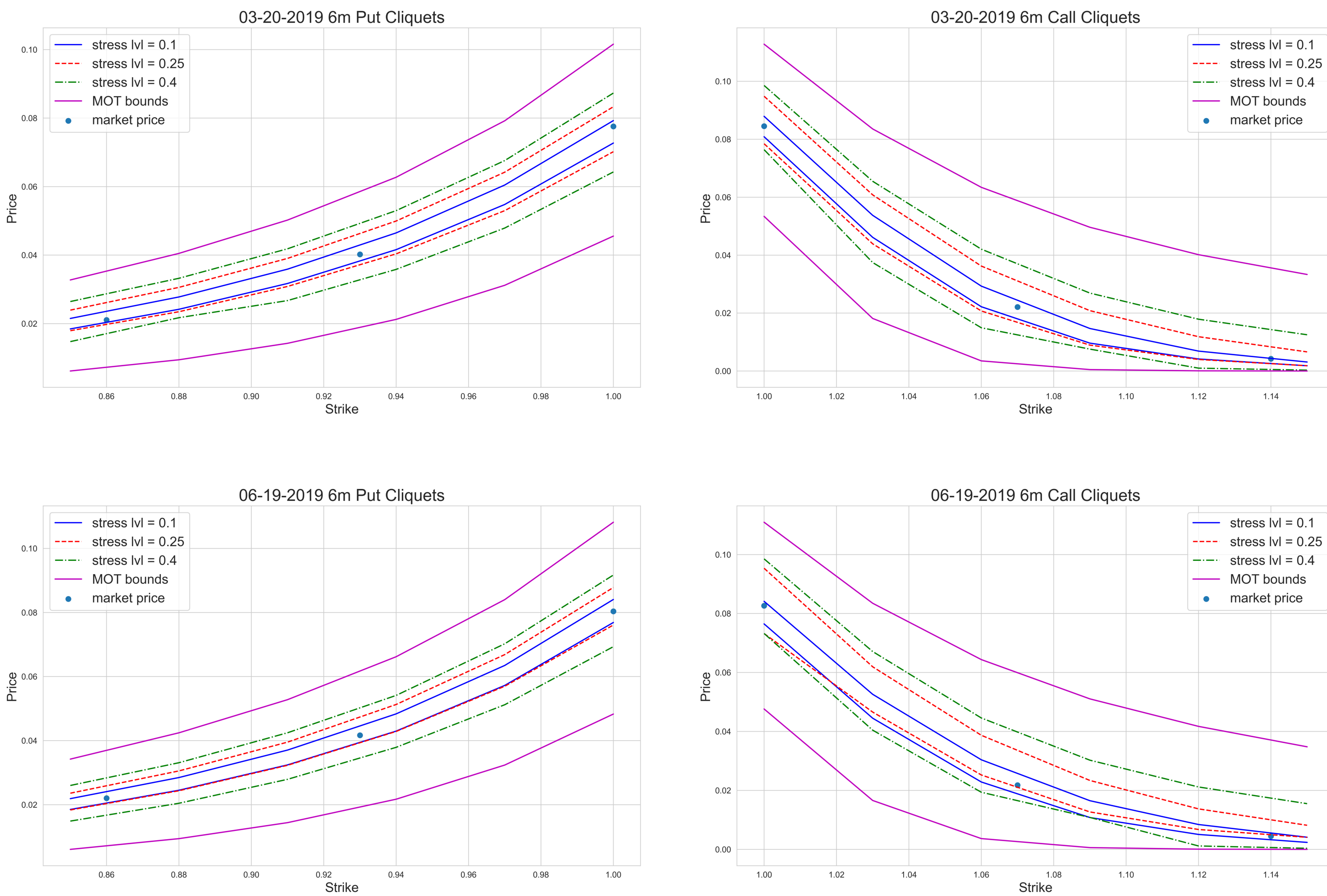


Figure 3.1: Acceptability bounds for 6 month rolling cliquets (March, June)

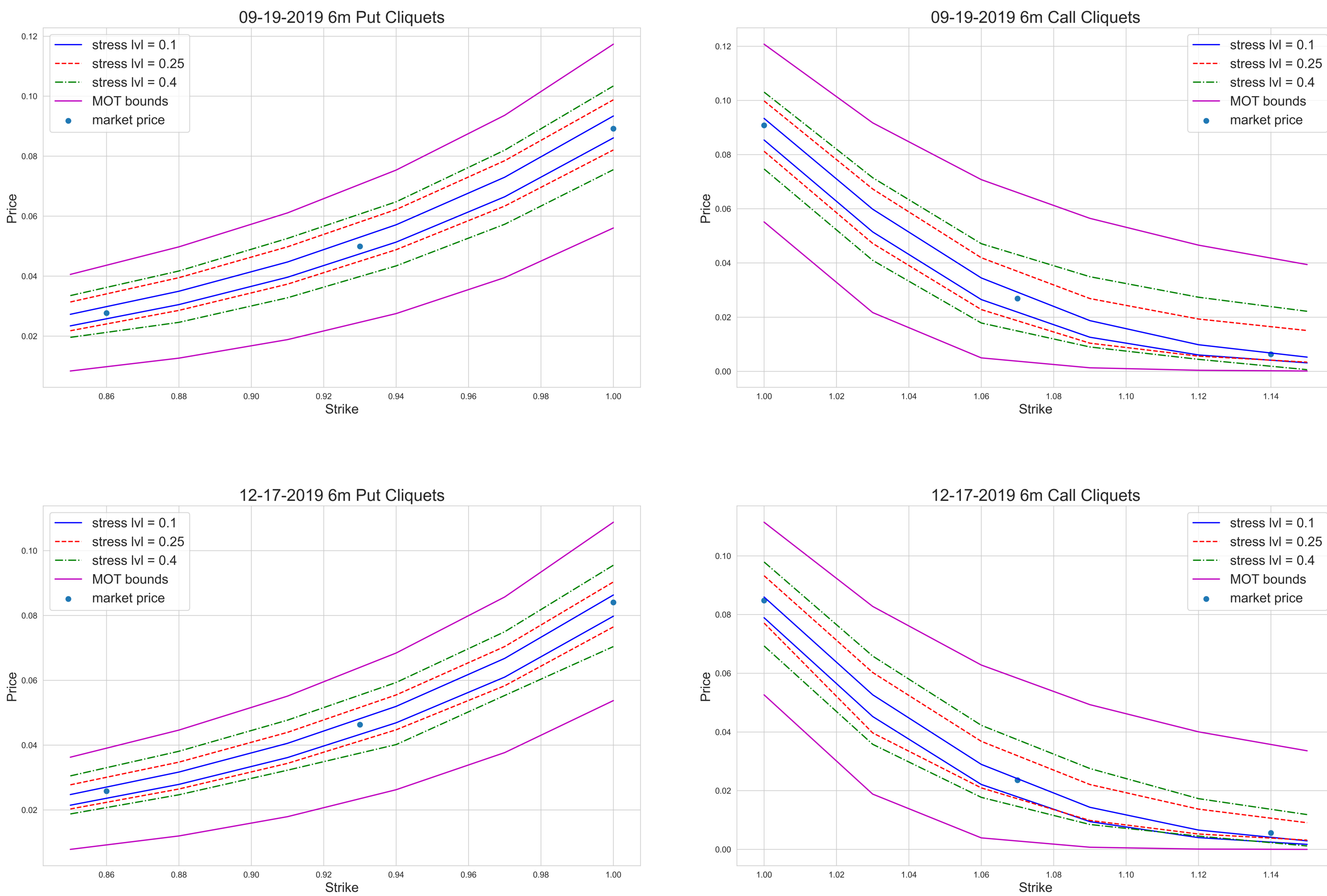


Figure 3.2: Acceptability bounds for 6 month rolling cliquets (September, December)

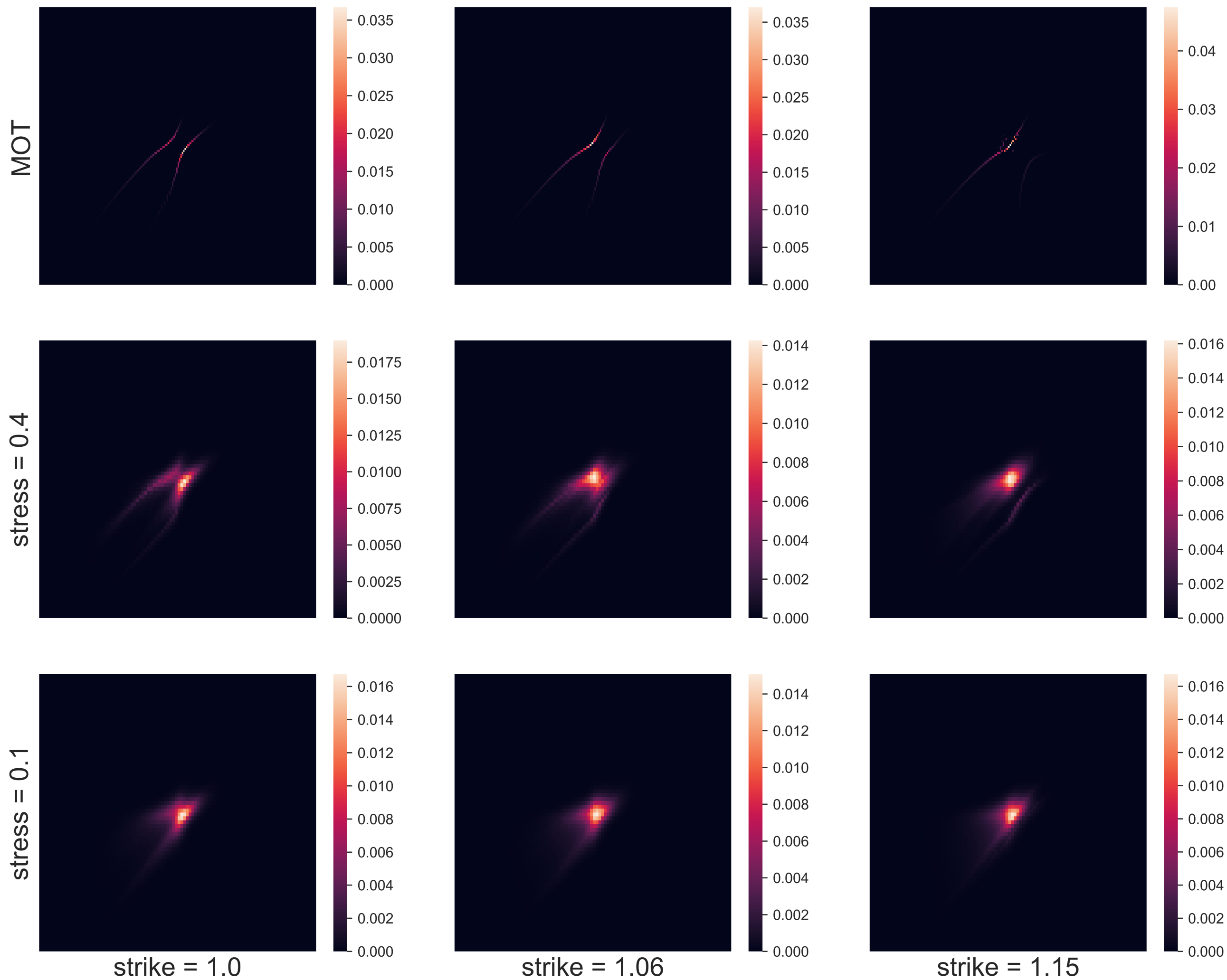


Figure 3.3: Maximizing joint distributions using 6-month and 12-month marginals

3.8 Conclusion

In Chapter 2, cliquets prices were established using no arbitrage arguments assuming the possibility of hedging using traded vanilla options. However, this approach delivers bounds that are too wide. In this study, the bounds are narrowed using concepts of risk acceptability. We see that under higher stress levels, corresponding to a higher γ, we obtain wider bounds. Similarly, as $\gamma \to 0$, the bounds become narrower and converge to the base measure. Our results show that acceptability allows us to obtain narrow spreads for the market cliquet prices. This is contingent on selecting a suitable base measure which was done in Chapter 2.

The plots of the joint distribution also show that the support of the optimal distribution becomes sparser at higher stress levels. This is expected as the set of supporting kernels increases and we are able to access more extreme measures.

The usefulness of the algorithm provided is the ability to deliver sensible spreads on a wider range of cliquet underliers by learning how to reproduce the cliquet prices for the more liquid underliers and then transferring the pricing technology to the markets for the less liquid underliers.

Bibliography

[1] Patrick S Hagan, Deep Kumar, Andrew S Lesniewski, and Diana E Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

[2] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[3] Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[4] Steven G Kou. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101, 2002.

[5] Steven G Kou and Hui Wang. Option pricing under a double exponential jump diffusion model. *Management science*, 50(9):1178–1192, 2004.

[6] Dilip B Madan, Peter P Carr, and Eric C Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

[7] Uwe Küchler and Stefan Tappe. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications*, 118(2):261–283, 2008.

[8] Peter Carr and Dilip Madan. Option valuation using the fast fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

[9] Hélyette Geman, Dilip B Madan, and Marc Yor. Time changes for lévy processes. *Mathematical Finance*, 11(1):79–96, 2001.

[10] Dilip B Madan, Wim Schoutens, and King Wang. Bilateral multiple gamma returns: Their risks and rewards. *International Journal of Financial Engineering*, 7(01):2050008, 2020.

[11] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[12] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[13] Rama Cont and Peter Tankov. Nonparametric calibration of jump-diffusion option pricing models. *The Journal of Computational Finance*, 7:1–49, 2004.

[14] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. In *World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 1: Foundations of CCA and Equity Valuation*, pages 3–21. World Scientific, 2019.

[15] Robert C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[16] Wim Schoutens, Erwin Simons, and Jurgen Tistaert. A perfect calibration! now what? *The best of Wilmott*, page 281, 2003.

[17] David G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2(4):329–347, 1998.

[18] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[19] Leonid V Kantorovich. On a problem of monge. In *CR (Doklady) Acad. Sci. URSS (NS)*, volume 3, pages 225–226, 1948.

[20] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[21] Mathias Beiglböck, Pierre Henry-Labordere, and Friedrich Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[22] Julien Guyon. Dispersion-constrained martingale schrödinger problems and the exact joint s&p 500/vix smile calibration puzzle. *Available at SSRN 3853237*, 2021.

[23] Pierre Henry-Labordere. Automated option pricing: Numerical methods. *International Journal of Theoretical and Applied Finance*, 16(08):1350042, 2013.

[24] Douglas T Breeden and Robert H Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

[25] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[26] Pierre Henry-Labordère and Nizar Touzi. An explicit martingale version of the one-dimensional brenier theorem. *Finance and Stochastics*, 20(3):635–668, 2016.

[27] Hadrien De March. Entropic approximation for multi-dimensional martingale optimal transport. *arXiv preprint arXiv:1812.11104*, 2018.

[28] Erwin Schrödinger. *Über die umkehrung der naturgesetze*. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter, 1931.

[29] Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pages 269–310, 1932.

[30] Michele Pavon, Giulio Trigila, and Esteban G Tabak. The data-driven schrödinger bridge. *Communications on Pure and Applied Mathematics*, 74(7):1545–1573, 2021.

[31] Christian Léonard. A survey of the schr\" odinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

[32] Marcel Nutz. Introduction to entropic optimal transport, 2021.

[33] Maurice Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

[34] O. Dembo A., Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2nd ed edition, 1998.

[35] Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.

[36] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[37] Hadrien De March and Pierre Henry-Labordere. Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. *Available at SSRN 3326486*, 2019.

[38] Peter Carr, Helyette Geman, Dilip B Madan, and Marc Yor. Self-decomposability and option pricing. *Mathematical finance*, 17(1):31–57, 2007.

[39] Sato Ken-Iti. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999.

[40] Ken-Iti Sato. Self-similar processes with independent increments. *Probability Theory and Related Fields*, 89(3):285–300, 1991.

[41] Aurélien Alfonsi, Jacopo Corbetta, and Benjamin Jourdain. Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(03):1950002, 2019.

[42] Gilles Pagès and Benedikt Wilbertz. Intrinsic stationarity for vector quantization: Foundation of dual quantization. *SIAM Journal on Numerical Analysis*, 50(2):747–780, 2012.

[43] David Baker. Martingales with specified marginals. *Theses, Université Pierre et Marie Curie-Paris VI*, 2012.

[44] Gaoyue Guo and Jan Obłój. Computational methods for martingale optimal transport problems. *The Annals of Applied Probability*, 29(6):3311–3347, 2019.

[45] Aleksandar Mijatović and Martijn Pistorius. Continuously monitored barrier options under markov processes. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(1):1–38, 2013.

[46] Dilip B Madan and King Wang. Acceptability bounds for forward starting options using disciplined convex programming. *Journal of Risk, Forthcoming*, 2016.

[47] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[48] Alexander Cherny and Dilip Madan. New measures for performance evaluation. *The Review of Financial Studies*, 22(7):2571–2606, 2009.

[49] Eva Lütkebohmert and Julian Sester. Tightening robust price bounds for exotic derivatives. *Quantitative Finance*, 19(11):1797–1815, 2019.

[50] Dilip B Madan and Alexander Cherny. Markets as a counterparty: an introduction to conic finance. *International Journal of Theoretical and Applied Finance*, 13(08):1149–1177, 2010.

[51] Alexander S Cherny. Pricing with coherent risk. *Theory of Probability & Its Applications*, 52(3):389–415, 2008.

[52] Shigeo Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.

[53] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[54] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

[55] Alexander Domahidi, Eric Chu, and Stephen Boyd. Ecos: An socp solver for embedded systems. In *2013 European Control Conference (ECC)*, pages 3071–3076. IEEE, 2013.

[56] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, June 2016.