Extracting Pricing Densities for Weather Derivatives using the Maximum Entropy Method

Antonios K. Alexandridis[a,*] and Henryk Gzyl[b] and Enrique ter Horst[c] and German Molina[d]

[a]Department of Accounting and Finance, University of Macedonia, Greece
Email: alexandridis@uom.edu.gr
[b]Center for Finance, Instituto de Estudios Superiores de Administración, Venezuela
Email: henryk.gzyl@iesa.edu.ve
[c]Universidad de los Andes School of Management, Bogota, Colombia
Email: ea.terhorst@uniandes.edu.co
[d]Quantitative Trading, Idalion Capital Group, U.S.A.
Email: german@alumni.duke.edu

ARTICLE HISTORY
Compiled June 18, 2020

* Corresponding author: A. K. Alexandridis. Email: alexandridis@uom.edu.gr

Abstract
In this paper we propose the use of the maximum entropy method to extract pricing densities directly from the weather market prices. The proposed methodology can overcome the data sparsity problem that governs the weather derivatives market and it is model free, non-parametric, robust and computationally fast. We propose a novel method to infer consistent pricing probabilities, and illustrate the method with a motivating example involving market prices of temperature options. The probabilities inferred from a smaller subset of the data are found to consistently reproduce out-of-sample prices, and can be used to value all other possible derivatives in the market sharing the same underlying asset. We examine two sources of the out-of-sample valuation error. First, we use different sets of possible physical state probabilities that correspond to different temperature models. Then, we apply our methodology under three scenarios where the available information in the market is based on historical data, meteorological forecasts or both. Our results indicate that different levels of expertise can affect the accuracy of the valuation. When there is a mix of information available, non-coherent sets of prices are observed in the market.

KEYWORDS
Extracting probability densities, weather derivatives pricing, temperature options, maximum entropy, out-of-sample valuation

JEL CLASSIFICATION
D52, C44, C61

1. Introduction

Weather fluctuations affect the economy both directly and indirectly. Even minor weather changes often have significant impact on the day-to-day operations and revenues of many businesses in sectors such as agriculture, energy, tourism, transportation and construction, Bertrand et al. (2015), Challis (1999), Hanley (1999). Extreme temperatures significantly impact earnings in over 40% of industries, Addoum et al. (2020). Similarly many businesses are sensitive to even small changes in weather. It is estimated that in 2012, 3.4% of US GDP was affected by routine weather variation, Zhou et al. (2019). A dry period could destroy farmers crops, and warm winters could cost millions to energy companies, caused by the reduced energy consumption for heating. Investors around the world desire products that allow them to hedge against the realizations of climate risk, Engle et al. (2020). Due to the recent economic crisis and the increased weather volatility caused by climate change, the need for efficient and effective weather risk management is evident, Stulec (2017), Weagley (2019), Engle et al. (2020).

In consideration of the attractive characteristics of weather derivatives, it is clear that their use can have significant managerial implications to a firm. As it is demonstrated in Perez-Gonzalez and Yun (2013) the use of weather derivatives can lead to higher firm value, investments, and leverage. In Buxey (1988) the importance of production planning under seasonal demand is highlighted. Similarly, weather derivatives can be used to create profitable investment portfolios for diversification purposes, Alexandridis and Zapranis (2013), Gulpinar and Canakoglu (2017), Jewson et al. (2005). In Buchholz and Musshoff (2014) it is shown that weather derivatives have the potential to substantially affect and alter farm plans. The optimal utilisation of weather derivatives by risk managers can provide a competitive advantage in the marketplace and increased profits, Chen and Yano (2010). Hence, an accurate and fast tool for valuation of weather derivatives and for identification of inconsistent subsets of market prices is essential for efficient and effective weather risk management.

The weather derivatives market is a classic incomplete market, because the underlying weather variables have no value, cannot be traded or stored. Despite the increasing number of studies in weather derivatives pricing Cao and Wei (2004), Davis (2001), Alaton et al. (2002), Benth et al. (2007), Benth and Saltyte-Benth (2013), Zapranis and Alexandridis (2008), Alexandridis and Zapranis (2013) the market still lacks a general accepted pricing framework. Furthermore, the market is characterized by a lack of liquidity, which increases its level of incompleteness, and most observed prices are theoretical market valuations or ranges, rather than traded levels.

In order to derive the option prices, often stochastic differential equations are used to describe the dynamics of the daily average temperatures (DAT), Alaton et al. (2002), Benth and Saltyte-Benth (2007, 2013), Benth et al. (2007), Alexandridis and Zapranis (2013), Zapranis and Alexandridis (2008), Sun and van Kooten (2015), Zong and Ender (2018). Alternatively, regime switching, Elias et al. (2014), Autoregressive Moving Average, Gulpinar and Canakoglu (2017), Campbell and Diebold (2005), Castellano et al. (2020) and Neural Networks, Zapranis and Alexandridis (2008), Cao et al. (2012), have been proposed in the literature to estimate the temperature process. However, modelling the DAT is not a straightforward process and various assumptions about the temperature model and the noise-generating process are made. More precisely, agents face the issue of model risk, since the estimation of the market price of risk depends on the assumed model. Small misspecifications in the DAT model can lead to large mispricing errors, Zapranis and Alexandridis (2008). Furthermore, it is usually difficult to solve the stochastic differential equation in order to price the financial weather products, and often it is impossible to find closed-form solutions for the pricing equations. Additionally, one is faced with the problem of illiquidity and data sparsity that characterizes this market. Finally, as a result of data unavailability, previous studies model the underlying temperature process but do not proceed on testing the accuracy of the forecasted contract prices against real market data.

In this study, we propose an alternative approach where we extract the pricing probabilities directly from the market prices of temperature options, which avoids all the aforementioned drawbacks. Our approach intends to add to the existing literature for pricing and risk management, hence providing the more time-sensitive traders and decision-makers with fast and efficient tools to re-consider these markets for their trading universe. Our proposed methodology is based on the maximum entropy methods, Jaynes (1957), Borwein and Lewis (2000), Zhou et al. (2013). This is a very powerful technique which needs only few market data points in order to extract the pricing probabilities, and can be used for density reconstruction, Borwein et al. (1996), and for the valuation of the prices of other option contracts traded in the market. Hence, the maximum entropy approach naturally overcomes the problem of data sparsity. The advantage of extracting the probabilities directly from the market prices is that volatility and other moments can easily be calculated independently of any particular model, Hardle et al. (2015). Additionally these results can be incorporated to the portfolio management/risk management/systematic trading tools for online decision-making. The concept of entropy has been used with great success in various finance related applications, Abbas (2006), Cozzolino and Zahner (1973), Gulko (1999a,b, 2002), Gzyl and Mayoral (2017), Judge and Mittelhammer (2011), Rajasekera and Yamada (2001), Rouge and El Karoui (2000).

Due to their guaranteed asymptotic exactness Markov chain Monte Carlo (MCMC) methods have become the mainstream followed approach, Gelman et al. (2013). For example in Hardle et al. (2015) the Bayesian quadrature method has been applied in order to derive the state price densities. Since MCMC's can take very long to

run and to converge, Variational Bayes (VB), is also becoming widely used because of its tractability, detectable convergence, and parameter estimation performance in practice, Jordan et al. (1999), Wainwright and Jordan (2008). However, our approach improves over those methods because:

(1) Approaches such as the Bayesian quadrature need to specify prior assumptions on the model parameters to overcome problems with data sparsity, while our approach can provide estimates with very small numbers of observations without imposing strict prior constraints. This is especially relevant as the number of available points observable can oscillate significantly over time.
(2) The processing speed of our method to calculate densities allows for online decision-making.[1] Our method is extremely fast to run and does not need any MCMC tuning nor convergence diagnostics, Gelman et al. (2013). This is a clear contrast with MCMC-based quadrature methods, which require larger amounts of computational time, plus performance of diagnostics of convergence. Speed of processing makes our approach more suitable for automatization and embedment into existing management and trading tools.
(3) Our method allows for range-based definitions of possible states of the underlying, as further discussed in our motivating example - this aligns with weather derivative decision-making, where the impact of different states of the nature may be discretized (for example, when a range of temperatures can have similar effect on crops).

These characteristics make maximum entropy methods highly suitable for online/automatizable-decision making within the weather derivative pricing literature.

The main contributions of our approach are the following: we present a novel model-free, non-parametric and robust approach to determine numerically the state prices, or equivalently the implied pricing probabilities, for any prior physical probabilities. The proposed method overcomes the data sparsity problem that governs the weather derivatives market. Hence, an advantage of the method is that we do not have to calibrate for model parameters, nor to provide prior knowledge about them, which would potentially require constant model feedback. In addition, the maximum entropy based procedure allows us to determine which data points (call or put contracts) are more informative. To numerically implement the maxentropic methodology we propose a market driven, systematic and intuitive discretisation procedure, in which any prior information consisting of the physical probabilities of the market is integrated in a natural way in the methodology. Furthermore, we decompose the out-of-sample valuation error in two components. The first component shows the proportion of the error that arises from the level of available meteorological information in the market while the second one shows the proportion of error that arises from the temperature model used to derive the prior physical probabilities.

The purpose of the paper is to introduce our new methodology, and provide illustrative examples from the motivating weather derivatives problem. The remainder of the paper is organized as follows. Section 2 briefly presents the weather market. In Section 3 the proposed maximum entropy method is presented. The data, numerical examples and results are presented in Section 4. Finally, in section 5 we present our concluding remarks.

[1] Computations can be performed in around 0.15 seconds with a 2.8Ghz dual core PC.

2. The weather market

Weather can disrupt the operations and project schedules of various types of business, Bowers and Mould (1994), Bowers (2001). Retailers have used weather-linked promotions, such as weather rebates, to protect against adverse financial outcomes caused by unfavourable weather, Caliskan Demirag (2013), Chen and Yano (2010). However, the outcome of this strategy is unknown and can lead to significantly volatile returns. The necessity to hedge adverse weather effects and unseasonal weather resulted in the creation of a new class of financial assets called weather derivatives. In general, weather derivatives are designed to cover non-catastrophic weather events, i.e. high probability, low risk events. Non-catastrophic weather risk is gaining importance as climate change becomes more pronounced, Stulec (2017). Weather derivatives were developed to hedge volume or quantity risk, rather than the price risk, Brockett et al. (2005). The payoff of a weather derivative depends on the measurement of the underlying weather index. Clearly, weather derivatives can provide superior hedging opportunities. The majority of weather derivatives are written on temperature, and more precisely on the Heating Degree Day (HDD) index. Hence, in this study we focus on temperature derivatives. The weather market, although more recently stagnant, was a fast developing market greatly affected by liquidity and the ability/willingness of traders to participate. According to the Weather Risk Management Association the market grew by 20% in 2010-2011, to a total notional value of $11.8 billion, and further understanding of the uncertainty about weather trends or increased variability will potentially enhance even further its expansion. The weather derivatives market has undertaken a move toward OTC trading, vastly affected by the inability to maintain sufficient market participation to justify exchange-traded trading.

Temperature derivatives are settled in three main temperature indices: the HDDs, the Cooling Degree Days (CDDs) and the Cumulative Average Temperature (CAT). The CAT index is the sum of the DATs over the contract period. HDD is the number of degrees by which the daily temperature is below a base temperature, T_{base}, and CDD is the number of degrees by which the daily temperature is above the base temperature. The base temperature is usually 65 degrees Fahrenheit (or 18°C). A HDD measures the extra energy needed for heating in a cold day. Similarly, a CDD measures the extra energy that is needed for cooling in a hot day where. It is assumed that no extra energy is needed when the temperature is 65 degrees Fahrenheit (or 18°C). HDDs and CDDs are accumulated over a period, usually over a month or a season. The value of the three indices for a measurement period in the time interval $[\tau_1, \tau_2]$ is given by the following expressions:

$$CAT(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} T(t) \tag{1}$$

$$HDD(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \max\big(T_{base} - T(t), 0\big) \tag{2}$$

$$CDD(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \max\big(T(t) - T_{base}, 0\big) \tag{3}$$

where the DAT, $T(t)$, is the average of the daily maximum and minimum temperature, $T(t) = \big(T_t^{max} + T_t^{min}\big)/2$.

Focusing on a specific temperature index given by equations (1)-(3) we are interested in deriving the value of a temperature derivative. For a measurement period $[\tau_1, \tau_2]$ the value of a forward (or futures) contract can be derived by:

$$e^{-r(\tau_2 - t)} \mathbb{E}_Q \Big[Index - F_{Index}(t, \tau_1, \tau_2) \mid \mathcal{F}_t \Big] = 0$$

where $Index$ is the CAT, HDD or CDD, F_{Index} is the value of a contract written on the specific index, r is the risk-free interest rate, $\mathcal{F}_t$ is the filtration (i.e. all historical information) at time $t \leq \tau_1 < \tau_2$ and Q is the pricing probability. Since F_{Index} is $\mathcal{F}_t$-adapted, we derive the value of the contract to be

$$F_{Index}(t, \tau_1, \tau_2) = \mathbb{E}_Q \Big[Index \mid \mathcal{F}_t \Big]. \tag{4}$$

Consequently, the European temperature call option price written on the futures price with a strike price K_{Index} at exercise time $\tau \leq \tau_1$ is defined as[2]:

$$C_{Index}(t, \tau, \tau_1, \tau_2, K_{Index}) = e^{-r(\tau - t)} \mathbb{E}_Q \Big[\max\big(F_{Index}(t, \tau_1, \tau_2) - K_{Index}, 0\big) \mid \mathcal{F}_t \Big]. \tag{5}$$

The majority of weather derivatives are written on a temperature and specifically on the HDD index. In this study we focus only on this class of weather derivatives, although our methodology can be easily adapted and applied in any weather derivative.

3. Methodology

3.1. The market model

In this section we present the market model. More precisely we focus on temperature derivatives written on the HDD index. The first step is to discretise the temperature index into K non-overlapping intervals:

$$[X_0, X_1), \ldots, [X_{K-2}, X_{K-1}), [X_{K-1}, X_K]$$

There will be a market state associated to each temperature index interval. The values X_0 and X_K can be chosen flexibly using historical data, in such a way that the probability of the temperature index to be greater than X_K or lower than X_0 is effectively zero. Average temperatures will be sufficiently bounded in practice, and

[2]The time τ is the exercise time of the option. At that time, the holder has the option to buy a futures contract with a measurement period $[\tau_1, \tau_2]$. In reality $\tau \leq \tau_2$, however, the most interesting case is when $\tau < \tau_1$. Otherwise, the transaction takes place after the start of the futures contract and the measurement period. Please see Chapter 6 of Alexandridis and Zapranis (2013) for a rigorous treatment of these cases.

our method does not rely on actual estimation of plausible average temperatures, so a practical solution is to determine the lower and upper bounds that would be physically impossible to observe in the geographical area. Note that the above partition still reflects the continuous nature of the underlying weather variable.

Second, we need to estimate the reference (physical) probabilities

$$p_j = P(X_{j-1} \leq X < X_j), \quad j = 1, 2, \ldots, K. \tag{6}$$

where X is the temperature index at the time horizon of interest, i.e. the probability of the temperature index computed using a preferred temperature model[3] to lie between X_{j-1} and X_j. Next, we choose appropriate "levels" $\hat{X}_j$ such that $\hat{X}_j$ occurs when the actual temperature index is $X_{j-1} \leq X < X_j$. As the temperature ranges are homothetically related to the HDD or the CDD, the simplest choice is to take the levels to be the mid points of the ranges as the value of the asset in the corresponding market state. The levels are not to be confused with market prices. Their role is to determine the cash flow associated with the available options.

Fixing "today" as $t = 0$, we consider all temperature options on the HDD index that have the same maturity day, $t = T$. We consider the mid points of the intervals as the possible outcomes of the temperature index (HDD) for the measurement period $[\tau_1, \tau_2]$, and denote as $\hat{S}_j$ the possible values of the underlying in the market.

Table 1 describes the characteristics of the data. The first row lists the market states. In the second row the asset level that characterizes each state is presented, while in the third row we list the corresponding ranges. Finally, in the last row the physical probabilities are listed. As shown in Section 4, we use three different sets of physical probabilities to test the robustness and stability of the performance of the proposed methodology.

[Table 1 about here.]

In this market there exist both European call and put options, where the underlying asset is the price of a futures on the HDD temperature index. Our aim is to determine the pricing probabilities that are used by traders to value these options. For this purpose, the maximum entropy method will be used. Although not of focus in this study, a method that can reconstruct the put and call curves in full, such as the one proposed in this manuscript, automatically unveils the no-arbitrage price, which may not be available either.

Suppose for the sake of definiteness we consider $M < K$ European call options with strike prices $K_m = S_{j_m}$, where $\{j_1, \ldots, j_M\} \subset \{1, \ldots, K\}$ and K is the number of intervals. Observe that we are interested in the case in which there are more market states (K) than market data (M) in order to have an interesting inverse problem to solve.

The observed price is given by the discounted expected payoff. Thus, the problem that we want to solve consists of determining the probabilities $\{q_j, |j = 1, \ldots, K\}$ such that

$$\pi_m = e^{-rT} \sum_{j=1}^{K} q_j O\Big(\hat{S}_j, K_m\Big) \quad m = 1, \ldots, M \tag{7}$$

[3]For example one might use any of the temperature models for weather derivatives pricing proposed in Alexandridis and Zapranis (2013), Benth and Saltyte-Benth (2013) or Alaton et al. (2002)

where M is the number of option prices used, K_m the corresponding strikes, $O\Big(\hat{S}_j, K_m\Big)$ is the payoff of the $m-$th option (which will be either a call or a put of European type), and π_m is its observed price.

3.2. The maximum entropy method

In this section we describe the maximum entropy method that we use in order to solve (7). This method was originally proposed in Jaynes (1957) and the mathematical details are worked out in Section 3.3 of Borwein and Lewis (2000). To continue, first, we rewrite (7) as

$$\pi_m = \sum_{j=1}^{K} \rho_j O\Big(\hat{S}_j, K_m\Big)\, p_j, \quad m = 1, \ldots, M \tag{8}$$

where ρ_j is the density of the pricing probability measure $\boldsymbol{q}$ with respect to the probability measure $\boldsymbol{p}$. Note, that the discount factor in (7) is dropped in (8) for simplicity. If we assume a constant interest rate then we can simplify the notation by making the discount factor part of the option prices by replacing π_m with $e^{rT}\pi_m$.[4]

The constraint for ρ_j, $j = 1, ..., K$ to be a density (or equivalently for q to be a probability) is:

$$\sum_{j=1}^{K} \rho_j p_j = 1 \tag{9}$$

Consider the set $\mathcal{D} = \{\rho_j : j = 1, \ldots, K\}$ satisfying (8) and (9). Then, $\mathcal{D}$ is a closed and convex (in $\mathbb{R}^K$) set. We can then select a point from within that set by solving the following optimization problem

$$\max_{\rho} H(\boldsymbol{\rho}) = -\sum_{j=1}^{K} \rho_j \ln \rho_j p_j \tag{10}$$
$$\text{s.t. constraints (8) and(9) are satisfied}$$

The solution is given by

$$\rho_j^* = \frac{1}{Z(\boldsymbol{\lambda}^*)} \exp\left(-\sum_{m=1}^{M} \lambda_m^* O\left(\hat{S}_j, K_m\right)\right) \tag{11}$$

where the normalization factor $Z(\boldsymbol{\lambda})$ is given by

$$Z(\boldsymbol{\lambda}) = \sum_{j=1}^{K} \exp\left(-\sum_{m=1}^{M} \lambda_m O\left(\hat{S}_j, K_m\right)\right) p_j. \tag{12}$$

[4]Note, that we assume a constant interest rate for simplicity however this is not restrictive for the proposed method.

calculated at $\boldsymbol{\lambda}^*$. The $\boldsymbol{\lambda}^*$ is computed minimizing the "dual entropy"

$$\Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi}) = \ln Z(\boldsymbol{\lambda}) + \langle \boldsymbol{\lambda}, \boldsymbol{\pi} \rangle \tag{13}$$

as explained below.

3.2.1. Heuristics of the duality argument

To understand where (11) comes from and the connection between the maximization of $H(\boldsymbol{\rho})$ and the minimization of (13), we begin by observing that for any two densities $\boldsymbol{\rho}(1)$ and $\boldsymbol{\rho}(2)$, an application of Jensen's inequality yields

$$\sum_{i=1}^{K} \rho_i(1) \ln \left(\frac{\rho_i(1)}{\rho_i(2)} \right) p_i \geq 0,$$

with equality occurring whenever $\boldsymbol{\rho}(1) = \boldsymbol{\rho}(2)$. Now, let $\boldsymbol{\rho}(1) = \boldsymbol{\rho}$ satisfy the constraints (8)-(9), and let $\boldsymbol{\rho}(2) = \boldsymbol{\rho}(\boldsymbol{\lambda})$ be given by (11) computed at an arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^M$. An application of the inequality yields

$$\sum_{i=1}^{K} \rho_i \ln \left(\frac{\rho_i}{\rho_i(\boldsymbol{\lambda})} \right) p_i = -H(\boldsymbol{\rho}) + \Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi}) \geq 0.$$

The entropy $H(\rho)$ of any density ρ satisfying the constraints is always bounded above by the entropy $H\left(\rho(\boldsymbol{\lambda})\right) = \Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi})$ of a density of the type $\rho(\boldsymbol{\lambda})$. The problem is to find a $\boldsymbol{\lambda}$ such that the density $\rho(\boldsymbol{\lambda})$ satisfies the constraints. Note as well that the first order condition on $\boldsymbol{\lambda}^*$ to be a minimizer of $\Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi})$ is for $\rho(\boldsymbol{\lambda}^*)$ to satisfy the constraints, and therefore to maximize the entropy. Furthermore $H(\boldsymbol{\rho}(\boldsymbol{\lambda}^*)) = \Lambda(\boldsymbol{\lambda}^*, \boldsymbol{\pi})$.

The key step in the maxentropic procedure is the minimization of the dual entropy described by (13). A modified Newton algorithm is carried out, Lange (2010). The (fast) algorithm stops when the gradient of the function is less than 10^{-6}. Finding the minimizing vector $\boldsymbol{\lambda}^*$, the pricing density is obtained according to equation (11), and consequently the pricing probabilities can be determined. In all cases the condition $\sum_i \rho_i p_i = 1$ of constraints (9) holds. Each optimization took, on average, around 0.15 seconds in R to derive the pricing probabilities, making this tool suitable for online monitoring/pricing and trading within these markets.

To complete, we remark that equation (11) can be used for the computation of any derivative with payoff $f(\hat{S}_j)$ at time T by computing the expected values

$$\pi(f) = \sum_{j=1}^{K} \rho_j^* f(\hat{S}_j) p_j. \tag{14}$$

3.3. Boundedness of dual entropy function

In this section we present more analytically the case where $\Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi})$ can be unbounded below. To understand this, first we write the dual entropy function as

$$\Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi}) = \ln \left(\sum_{j=1}^{K} e^{-\langle \boldsymbol{\lambda}, (\boldsymbol{O}\boldsymbol{e}_j - \boldsymbol{\pi}) \rangle} \right)$$

where $\boldsymbol{e}_j$ are the standard basis vectors for $\mathbb{R}^K$, and $\boldsymbol{O}$ is the payoff matrix where $O_{m,j}$ denotes the payoff of the $m-$th option in the $j-$th state.

Lemma 3.1. *If there exists a $\boldsymbol{\lambda}_c \in \mathbb{R}^M$ such that $\langle \boldsymbol{\lambda}_c, (\boldsymbol{O}\boldsymbol{e}_j - \boldsymbol{\pi}) \rangle > 0$ for all $j = 1, \ldots, K$, then the dual entropy function is unbounded below.*

Proof. Suppose that such a $\boldsymbol{\lambda}_c$ exists and let $m = \min\{\langle \boldsymbol{\lambda}_c, (\boldsymbol{O}\boldsymbol{e}_j \boldsymbol{\pi}) \rangle \mid j = 1, \ldots, K\} > 0$. Then for $g > 0$,

$$\Lambda(g\boldsymbol{\lambda}_c, \boldsymbol{\pi}) \leq \ln \left(K e^{-gm} \right) = \ln K - gm$$

which tends to $-\infty$ as $g \to \infty$. Therefore, along λ_c the dual entropy is not bounded below and the minimum of $\Lambda(\boldsymbol{\lambda}, \boldsymbol{\pi})$ on $\mathbb{R}^M$ does not exist. □

Remark. *If we think of $\boldsymbol{O}$ as (the matrix of) a linear mapping $\boldsymbol{O} : \mathbb{R}^K \to \mathbb{R}^M$, and any probability $\boldsymbol{p}$ as a point in the simplex $\mathbb{P} := \{\boldsymbol{p} \in \mathbb{R}^K \mid \sum p_j = 1\}$, then the possible price vectors are in the points in $\boldsymbol{O}(\mathbb{P})$. If Lemma 3.1 holds, then $\boldsymbol{\pi} \notin \boldsymbol{O}(\mathbb{P})$. This is because $\boldsymbol{\pi}$ and $\boldsymbol{O}(\mathbb{P})$ are separated by the hyper-plane defined by $\boldsymbol{\lambda}_c$, as indicated in the Lemma 3.1. In this case we can say that some prices have not been consistently assigned.*

4. Results

This section is separated into two parts. In the first part we describe the data utilized to illustrate the methodology, and in the second we present a summary of the results of applying the maxentropic technique for a few stylized examples.

4.1. The market data

We use CME options and futures daily closing prices, written on the temperature HDD index as measured in the meteorological station in New York. We focus on New York since it is the biggest temperature market with volume around 20% of the total market volume. The data availability ranges from 6/3/2006 to 30/11/2010[5].

For this study we are interested in examining whether our maximum entropy approach can extract the pricing probabilities and whether we can use these probabilities in order to value new (unobserved) contracts. Additionally, we want to examine the effectiveness of the method by decomposing the out-of-sample valuation error in two components. There are two ways in which the level of information can affect the pricing.

[5]The high cost involved in obtaining historical weather derivatives prices prohibit us from using more recent and diverse data.

First, meteorological forecasts can be used in order to approximate the temperature index. Second, when temperature derivatives are traded before the measurement period, and meteorological forecasts are not available, advanced statistical models can be used for forecasting the DAT and hence the HDD index. We assume different scenarios in order to examine the performance and robustness of the proposed framework.

First, since meteorological forecasts are not accurate for more than 10 days, Alexandridis and Zapranis (2013), Wilks (2011, 2002), we examine the performance of the maximum entropy method with 3 different remaining times to maturity. More precisely, we apply our method: 1) before the contract's measurement period; 2) at the beginning of the measurement period; and 3) in the middle of the measurement period. Usually weather derivatives are monthly contracts. In general, traders are not very active long before the measurement period. However, as we get closer to the measurement period, trading becomes more active. In the first case, accurate meteorological forecasts are not available for the traders. Therefore, their decisions are based solely on historical data (or seasonal forecasts). In the second case, we set the beginning of the measurement period as the first 1-5 days of the contract's measurement period. Therefore, meteorological forecasts are available for some days during the measurement period but not until the end of the contract. In the last case, we define as middle of the measurement period the last two weeks of the measurement period (middle of the measurement month). Hence, meteorological forecasts are available for all days until the maturity day.

Second, we assume three different scenarios in the computation of the physical probabilities described in section 3.1. The first one corresponds to the scenario where all market states have an equal probability (EQP) to occur. The second set of probabilities corresponds to a simple model often used by practitioners, namely the Historical Burn Analysis (HBA) method. In other words, by studying the possible HDD outcomes in the previous 30 years we compute the corresponding probabilities for each market state[6]. Finally, the last set of probabilities corresponds to a more advanced model. More precisely, a CAR(p) model proposed by Benth et al. (2007) is estimated, and then it is used to forecast the DAT and the HDD index for the corresponding period of interest. This analysis will help us evaluate the robustness of the proposed method and its dependence on the initial assumption of the physical probabilities.

The above scenarios examine roughly all the possible combinations of the availability of meteorological forecasts based on the remaining time to maturity and of the physical probability distributions based on the complexity of the temperature model.

The weather market is characterised by data sparsity and lack of liquidity. We want to examine the number of contracts that are needed in order to extract pricing probabilities. For each day and dataset, we consider all the contracts with different strike prices but with the same maturity. We illustrate our methodology with a sample of fourteen datasets, as considering the both the number of strike prices and the set of probabilities the total number of scenarios that we analyse grows exponentially. More precisely, as it is presented in the next section, for these fourteen datasets we examine 8,214 different scenarios.

In Table 2 the datasets are presented. This includes the day t where our methodology is applied, the maturity date, and the number of available calls and puts. We do not include details of each of the 8,214 optimizations obtained for each subset within those datasets, as reporting that would be prohibitive. Instead we will provide some stylized examples of results. The purpose of these datasets is to provide an illustration of

[6]For an analytical description of the HBA we refer to Alexandridis and Zapranis (2013).

the methodology, rather than a full representation of the weather derivatives market characteristics.

[Table 2 about here.]

4.2. The market states

In this section we will describe the market model for the aforedescribed datasets. More precisely, we will present a description of the underlying market model. As described in Table 1, first we are interested in the number of states, along with the range of values of the HDDs, and the level that characterises each state. Second we need a set of physical probabilities for each state. As explained in the previous section, we will use three different sets of probabilities representing a naive, a semi-advanced and a state of the art temperature model. Third, we need a collection of European options specifying their types, strikes and prices. Due to space limitations we present only two cases that exhibit typical behaviour: using 3 or 4 options we were able to obtain the pricing probabilities that reproduce the market prices of the remaining options in the market not used as input for the proposed maxentropic procedure. All 8,214 cases are available on-line in the supplementary material of this paper[7].

As it is described in Table 2 the difference between some datasets is the time at which the physical probability is known, and the time at which the data about the price of the option was collected, which influence the price considerably. For example when the remaining time to maturity is just a couple of weeks, future weather is more certain.

In Table 3 we describe the data behind the market model for contract 1, which has 10 states and 6 assets. In Table 3 the level and range of each state is given, as well as the three different sets of probabilities. A closer inspection of Table 3 reveals that the three sets of probabilities are significantly different.

[Table 3 about here.]

To calibrate the pricing probability we consider the collection of options of the European type that are available in the market at the day of interest, e.g. 20/11/2007, which is before the measurement period of the contract. The details of the options are presented in Table 4.

[Table 4 about here.]

In Table 4, the types of the options are listed in the first row, and the strikes and prices are defined in the following two rows, respectively. In the next section we will use a subset of the available options to calibrate the pricing probabilities, and then use those probabilities to obtain further option prices of different strike prices, including, but not limited to, those not used during the calibration. To examine the consistency of the prices we compare the predicted price of the options not used for calibration versus their prices as quoted in the market.

Let us label the options from left to right according to the first row of Table 4, and denote their payoff at the exercise time by $O_i(\hat{S}_j, K_i)$, where $O_i(\hat{S}_j, K_i) = (\hat{S}_j - K_i)^+$ if the option is a call or $O_i(\hat{S}_j, K_i) = (K_i - \hat{S}_j)^+$ if it is a put. Table 5 presents the payoffs of all options depending on the strike price and the state of the market.

[7] Full set of results available online as supplementary material.

[Table 5 about here.]

Note that in this case there are only 6 assets while there are 10 states. Thus the (constrained) linear problem (8) is quite undetermined. As it is shown in the next section, we do not even need all the equations to determine the pricing probabilities consistently.

Next we consider the 8^{th} dataset. This refers to the same contract as dataset 7, but with different remaining time to maturity. This contract is very interesting since the final outcome of the HDDs is significantly different from the historical average.

In this case we have nine states and there are nine options available in the market. The dataset for this case is displayed in the next three tables. In Table 6 we specify the underlying market model. The available options traded in the market together with their prices and strikes are presented in Table 7. Finally, in Table 8 the payoffs of the 9 options for the different states are presented.

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

4.3. Summary of results

4.3.1. In-sample

As it is described in Section 4.1 we consider fourteen datasets to illustrate our methods. The number of available options varies from 6 in the first dataset to 9 in the last one. For each dataset we consider all possible subsets of the options and their prices. Using the physical state probabilities as reference, we determine the pricing probabilities for each subset. Next, using the estimated pricing probabilities we compute the option (call and put) prices for a fine grid of strikes. Finally, we examine the predictive power of the methodology by predicting the price of the options not included in the data (sub)set. We repeat our methodology for the three different sets of physical probabilities.

As it was expected the accuracy of the prediction of the option prices varies depending on the number of options used, the subset and the selected options within the subset.

In the cases in which the dual entropy was bounded below, usually 4 options (total of puts and calls observed) were sufficient to completely characterize the pricing measure in a consistent and accurate fashion, and provide a full representation of both put and call curves.

Next, we analyse the results obtained by dataset 1. The data described in Tables 3 – 5 were used. In order to derive the risk pricing probabilties, for this particular set, we consider a subset of two puts and two calls with strike prices of 710, 760, 825 and 850 respectively.

[Table 9 about here.]

We can compute the price of any derivative using those 4 options. First, we want to confirm whether the prices in the market are reproduced.

Figure 1 shows the price of the options computed using the pricing probabilities shown in Table 9. We estimated three sets of prices, one for each set of physical probabilities, and proceeded to reconstruct the prices of the remaining options (marked

by dots), whereas the predicted prices of the options are marked by circles. Figure 1 reveals that the proposed method can reconstruct the option prices and can be used for valuation of the remaining options. The results are similar for the three sets of probabilities indicating the robustness of the proposed methodology.

[Figure 1 about here.]

The second case that we present corresponds to dataset 8. The market data for this case was detailed in Tables 6 – 8. Using the three options with the highest strikes as data, we applied the proposed maxentropic method and obtained the pricing probabilities presented in Table 10.

[Table 10 about here.]

Next, in Figure 2 the curves of the option prices are presented. The description of the panels is as above. From Figure 2 it is clear that using only three options, we can confirm consistency of the valuation of the remaining six options. A closer inspection of Figure 2 reveals that using physical probabilities derived from the CAR model we obtain a more accurate valuation.

[Figure 2 about here.]

4.3.2. Out-of-sample

In Table 11 we present the out-of-sample mean absolute distance (MAD) between the fitted curve and the options observed but not used for fitting, as a function of the probability and the number of options used for fitting across sets. Our results indicate that using only one or two options for fitting produces large out-of-sample pricing errors. For datasets 1-3, 9 and 11 EQP is best for small number of options, but as the number of options used is increasing, the results become similar for all the sets of initial probabilities. On the other hand for datasets 4-8, 10 and 12-14 the CAR and HBA probabilities produce the lower out-of-sample pricing errors. However, the differences between the three methods, especially if three or more options are used, are very close to each other. This highlights our initial claim that our approach can provide accurate estimates even with a very small number of observations and without imposing strict prior constraints. The results in the rest of this section and in the simulation study presented in section 4.5 show that the proposed method is model free, non-parametric, and robust.

The above results are confirmed by Table 12. Table 12 groups the datasets by contract. Hence, we can isolate the error in MAD that arises from a specific contract. For the first contract (datasets 1+2+3) the EQP produces the lowest error followed by CAR and HBA. On the other hand, for the second (datasets 4+5+6), third (datasets 7+8), fourth (datasets 9+10+11) and fifth contract (datasets 12+13+14) CAR and HBA produce lower errors while the highest MAD is obtained when EQP is used. It is worth mentioning that for both the first and the fifth contracts the realised HDDs indices (817.5 and 872 for the first and fifth contract respectively) are very close to the historical average of the last 30 years which are 837.5 and 831 respectively. On the contrary, the realised HDDs index for the second contract is only 467 while the historical average is 505.9. Finally, for both the third and the forth contracts the realised HDDs indices (672.5 and 850 respectively) are significantly lower than the historical average of 764.2 and 1003 respectively.

Finally, Table 13 represents the MAD between the fitted curve and the observed

prices, not used for fitting, across converging sets of the same dimension, i.e. sets with the same number of observations used for fitting. The grouping was done column-wise by combining datasets 1+4+9+12 (first group), datasets 2+5+7+10+13 (second group) and datasets 3+6+8+11+14 (third group). This grouping will help us isolate the time effect and analyse whether it affects the accuracy of the pricing versus other observed prices. In other words, in the first group all contracts are traded before the measurement period when meteorological forecasts are not available. In the second group all contracts are traded at the beginning of the measurement period while the third group contains all the contracts that are traded in the middle and near the end of the measurement period.

A closer inspection of Table 13 reveals that CAR probabilities appear to be more consistent for the reconstruction of the observed option prices not used for fitting, followed by HBA and EQP. Moreover, when using four or more options all methods produce similar out-of-sample MADs. In fact it appears that using 4 or 5 options for fitting is the optimal since more than that, for our dataset, options that are not coherent with each other are included, which increases the likelihood of non-convergence. This aligns with the natural sparsity in the weather derivatives markets, where observing large numbers of market prices is rare. We observe that the MAD when the CAR probabilities are used is decreasing at a much faster rate. Finally, if 3 or more options are used all methods give very similar results indicating that the proposed method can extract the pricing densities without depending strictly on the accuracy of the prior physical probabilities.

Even under the (strong) assumption that prices are coherent between each other, the benchmark MAD under perfect fit would be 0.25, given the integer-based granularity of the premium data in the exchange (any observed price X could be any value between X-0.5 and X+0.5, with an expected absolute error of 0.25 if we assume they are uniformly distributed in that range). For example, the mean absolute distances in the three plots in Figure 2 are 1.40, 0.49 and 0.66 for HBA, CAR and EQP respectively indicating that option prices were, on average, only 1.15, 0.24 and 0.41 away from the benchmark.

Assuming market pricing being coherent, one might expect that as we get closer to the maturity date the accuracy of our pricing will increase, with curves fitting better out-of-sample observed prices. On the contrary, if we expect lack of coherence, it will manifest itself most often during times when there can be disagreement about the inputs and models, probably at the beginning of the measurement period, where there is still a number of potential ways in which the information available is used. Indeed our results in Table 13 confirm this. The first and the third group have smaller errors. For contracts traded in group 1 traders rely on historical data, while for group 3 traders rely mostly on meteorological forecasts. On the other hand for the second group some meteorological information is available but traders also have to rely on historical data. The mix of information in group 2 also means higher room for subjective interpretation/valuation by ad hoc approaches. Uncertainty may be higher when the volatility from all information sources is higher. This scenario is expected in the beginning of the measurement period where there is room for subjective approaches by traders and not before, where most of the information arises from historical data, or at the end of the measurement period, where most of the information arises from meteorological forecasts. Hence the second group is the one where most diverse pricing can occur, while there may be more coherence in groups 1 and 3, i.e. more of a "standard" approach taken by traders versus a higher set of inputs that could be used when pricing during periods of group 2. Inconsistent sets of market prices are a reality of life, and

identifying them opens the gate for arbitrage opportunities.

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

4.4. Discussion on unboundedness

Note that in all cases we were dealing with an undetermined linear system of equations subject to convex constraints. Data will often contain option prices that are inconsistent between them. This induces unboundedness in the dual entropy, but it is not a problem nor does it preclude our method from finding a solution. For example, in all 2,339 (28.5%) sets where the dual entropy was unbounded, we found a bounded solution when a subset of the options was used for fitting. This is an expected outcome and it has the added advantage of helping the agent identify options which are not consistently priced between themselves. For our particular dataset, the percentage of converging sets is negatively correlated to the number of options used for fitting (100%, 97%, 87%, 65%, 33%, 7%, 0%, 0% and 0% for $1, \ldots, 9$ respectively), indicating that multiple options were potentially mispriced. Note that mispricing under exchange constraints can happen even if options are priced correctly. This can be a consequence of the observable granularity of prices in the exchange (for example two options with different strike may have the same observed price; the exchange does not admit decimal granularity in the pricing, and a theoretical price of 1.76 would be observed as the same value as a theoretical price of 1.99). Also, as a result of the same granularity constraint, a drop of 1 unit in the strike can be accompanied by 1 unit drop in the price for every out-of-the-money (OTM) option, which would seem counter-intuitive. Our method is able to extract a solution by simply sampling from subsets of the observed set.

Out of the $\binom{9}{2} = 36$ sets in dataset 8 containing only 2 options, our algorithm converged in 35 sets. The set where convergence was not achieved was the case where only the 640 and 650 puts were used (first and fourth options in dataset 8). In our context, we use the Barzilai and Borwein (1988) algorithm to numerically minimize the Entropy function and by convergence we mean it in the Bayesian sense of the Markov Chain Monte Carlo to the Ergodic distribution. Convergence of the Barzilai and Borwein (1988) algorithm does not depend on the probabilities used, but on the characteristics of the set of options used. For each set of options comprising a given set, if that set did not converge, i.e. the entropy problem didn't have a finite solution, there was a subset of options of that set which would converge. This is achieved by simply sampling within the original set, to capture a smaller subset with consistently-priced options, which will always provide a solution. Additionally note that this problem is most likely to appear with deep OTM options, where mispricing is more often following the aforementioned exchange-imposed granularity. Note, for the aforementioned two options, that a choice of $\lambda = (-2.5, 1)$, for example, aligns with the condition in Lemma 3.1 for the unboundedness of the entropy function when using solely these two options for the optimization. We found that often unboundness occurs when OTM options too close to each other (and of the same nature, whether puts or calls) are included in the optimization. These two options, for example, are the ones with the lowest premium in dataset 8.

In the remaining 5,875 (71.5%)cases, the method provided us directly with a so-

lution. We have mentioned in Section 4.3 that in most cases only 4 options were sufficient to completely characterize the pricing measure in a consistent and accurate fashion. This is an asymptotic result, as described in Gzyl (2017). When the number of observations increases, the entropy value decreases and converges to its true value. This implies that when we are close to the minimum value of the entropy, its value does not change much when we add an additional data point. This result is based on the reported experiment that included all 8,214 different combinations of options over which the model was calculated. As indicated earlier, we achieve convergence in 65% of the times with only 4 options. This is what we also observe in our simulation study which we present in the next section.

Additionally, consistency among prices is a key factor. If more options are available, and the prices are not consistent among them, bootstrapping subsets of options that converge is a possible solution. Then, uncertainty about the curves could be measured with each subset providing a different, valid solution. However, we do not have a theoretical quantification for this relationship, since it will probably be determined by the consistency among prices rather than the absolute number of observed prices. As liquidity increases, we expect prices to be more consistent and therefore higher rates of convergence would be observed. However, the relationship between the amount and type of inconsistency of observed prices and the convergence of the model is not trivial to explore from a theoretical standpoint.

4.5. Simulation study

In the marketplace, different classes of weather derivatives are available. The prior model for the physical probability might be different for each class of weather derivatives and as a result it might have an impact on the extracted pricing probabilities. In the previous section we have shown that the proposed methodology is robust to different physical probabilities. Furthermore, two of the methods are "no method" at all: namely, the EQP and the HBA.

Note that the role of the prior physical is ancillary while we estimate a density (maxentropic probability/physical probability). In the option valuation stage the final probability as it appears in (8) is used. More precisely, the density ρ^* is multiplied by the physical prior p to obtain the maxentropic probability $\rho^* p$ of the state which in turn is used to compute the option prices.

In this section we conduct a simulation study to assess the sensitivity of our framework to the prior probabilities. It is important to note that the purpose of our study is to reconstruct accurately the unknown options curves. Hence, in this simulation study we examine the distance between the observed out-of-sample options prices that were not used for the fitting of our method and the reconstructed option prices. We achieve this through the following simulation:

(1) Choose a dataset i where $i = 1, .., 8$
(2) Choose of subset of observed options
(3) Calculate the MAD for the HBA, CAR and EQP sets of probabilities following the procedure described in the previous sections
(4) Generate $1,000$ simulations of possible physical probabilities P where $P \sim Dirichlet(1/n, ...1/n)$ which is centered around the EQP vector and n is the number of options in the selected dataset.
(5) Calculate the MAD for each simulation.

Note that the variance in the probabilities P is very large and we are able to explore the $n-$dimensional space of physical probabilities for each combination of dataset and subset of observed option prices. However, the above approach induces high levels of noise and it will generate many sets of physical probabilities that are highly non-informative or not possible in reality. We expect to observe the largest errors in the reconstructed curves in these sets.

Figure 3 shows the distribution of the MAD under a non-informative Dirichlet prior on the model probabilities. For illustration reasons we focus on dataset 8 and we select 4 options, 2 calls and 2 puts for fitting. Then we try to value the remaining 5 options. We repeat our method for all three methods, i.e. we compute the MAD for the EQP, HBA and CAR physical probabilities. We choose to present dataset 8 because we observe a large difference between the realised and the historical average HDDs. Hence, we expect the worst performance to occur in this dataset.

The values for MAD for HBA, CAR, and EQP in Figure 3 are represented by a circle, a triangle, and a diamond, respectively, at the x-axis level. The histogram represents the distribution of the MAD for the simulations.

A closer inspection of Figure 3 reveals that the reconstructed curves are somewhat affected by the quality of the physical model behind the derivation. However, it is noteworthy the resilience of the method to the choice of physical probabilities. Even when generating them from a highly non-informative model (centered at EQP and with high levels of noise), the reconstructed curves have similar values of MAD to those coming from more informative physical density choices. In all cases the model converged, meaning that convergence was not dependent on the physical model, but only on the observed prices, confirming our analysis in Section 4.4 that coherence between observed prices is the key factor for convergence of the method.

For this particular example, HBA and EQP appear to provide similar MAD levels, while CAR is slightly larger. However, all three are better than the great majority of random probability draws, many of which led to very extreme physical probability values which are highly implausible caused by the high variance of the proposed Dirichlet distribution.

[Figure 3 about here.]

5. Conclusions

We have provided a computationally fast, non parametric, robust and model-free method for inferring pricing probabilities, along with their densities, with respect to several possible physical state probabilities. Clearly, when presented with a collection of prices of options, a market analyst does not have a way to decide which ones are more informative than the others. The methodology that we have developed can handle that issue in a simple way: a subset of the data is more informative if it can consistently reproduce the data not used to determine the pricing probabilities. When this is the case, we feel more confident for the prices of any other derivative computed using the same pricing probabilities.

Our out-of-sample results indicate that the proposed method can extract the pricing probabilities using only few market data points for in-sample fitting, and it can be used for valuation of other options traded in the market. Hence, the proposed method is ideal to overcome the data sparsity problem that governs the weather market.

It is worth mentioning that we were able to value options accurately even in the

cases where the realized underlying HDDs indices were significantly different from the historical average. Our results show that our methodology produces similar results independently of the initial set of physical probabilities. Furthermore, when the final option's payoff is very close to the historical average even simplistic and naive set of probabilities produce accurate results. In cases where, the final payoff is quite different than the historical average, the accuracy of the valuation can be improved by utilising physical probabilities from more advanced models such as the CAR model. The above results are confirmed from our simulation study where even unrealistic and impossible sets of physical probabilities produce relatively similar values of MAD – albeit higher – to the EQP, CAR and HBA.

In an extensive analysis we examined whether different sets of physical probabilities, indicating different levels of expertise, can affect the accuracy of out-of-sample valuation. The three sets of probabilities are derived by a naive, a semi-advanced and a state of the art modelling procedure. Our results indicate that the CAR and HBA probabilities can provide a better reconstruction of the original option prices, and also produce a lower MAD in the valuation of the options not used in the fitting procedure. As the number of options used for fitting increases the differences between the three methods become smaller. Finally, we observe that regardless the set of probabilities that is used, our methodology is robust.

Our results indicate that when the available information in the market arrives from historical data or from meteorological forecasts pricing is more coherent. However when there is a mix of information in the market, non-consistent sets of market prices are observed. Testing for consistency using different data subsets should be regarded part of the procedure.

While high frequency trading in these types of illiquid assets is unlikely, constant/online monitoring of prices and densities is a must for portfolio management of positions, as well as speedy decision-making within lower frequency trading when opportunities arise. Our approach is especially useful for time-sensitive decision-makers who will need constant feeds of up-to-date pricing densities for every asset in their portfolios, including those of more illiquid nature and with limited information available. By providing fast estimates while overcoming the issue of sparsity, our method allows for more efficient trading and management of illiquid assets in incomplete markets, allowing traders and portfolio/risk managers to more safely expand their trading universes to these assets.

Finally our results show that the proposed method already produces excellent results when as little as 4 options are used for fitting. On the other hand, including OTM options too close to each other usually generates convergence issues, perhaps related to the granularity of the pricing in the exchange, where lower premium options will suffer the most impact as a proportion of the true price, or related to potential incoherent pricing.

An interesting by-product of our methodology becomes apparent when regarding the call and put price curves. If we make use of the put-call parity, we can obtain a no-arbitrage price as the value at which a put and a call of the same strike intersect. This is interesting because these prices are not really available from the market. Hence, an interesting future research strand would be the building on this methodological approach for unveiling prices of unobservable no-arbitrage prices in a wider context. Furthermore, it would be interesting to examine how the extracted price can be used for valuation of 1) weather derivatives on a different temperature index, such as the CAT or CDD, 2) weather derivatives on a different but correlated weather variable such as precipitation, 3) derivatives in correlated markets such as agricultural commodities

or energy contracts, and 4) derivatives in other illiquid asset classes.

Acknowledgement

The authors would like to thank the editor and the anonymous referees for their constructive comments that helped to substantially improve the final version of this paper.

References

Abbas, A. E. (2006). Maximum entropy utility. *Operations Research*, 54(2):277–290.

Addoum, J., Ng, D., and Ortiz-Bobea, A. (2020). Temperature shocks and establishment sales. *The Review of Financial Studies*, 33(3):1331–1366.

Alaton, P., Djehince, B., and Stillberg, D. (2002). On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9:1–20.

Alexandridis, A. and Zapranis, A. (2013). *Weather Derivatives: Modeling and Pricing Weather-Related Risk*. Springer, New York.

Barzilai, J. and Borwein, J. M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148.

Benth, F. E. and Saltyte-Benth, J. (2007). The volatility of temperature and pricing of weather derivatives. *Quantitative Finance*, 7(5):553–561.

Benth, F. E. and Saltyte-Benth, J. (2013). *Modeling and Pricing in Financial Markets for Weather Derivatives*. World Scientific, Singapore.

Benth, F. E., Saltyte-Benth, J., and Koekebakker, S. (2007). Putting a price on temperature. *Scandinavian Journal of Statistics,*, 34:746–767.

Bertrand, J.-L., Brusset, X., and Fortin, M. (2015). Assessing and hedging the cost of unseasonal weather: Case of the apparel sector. *European Journal of Operational Research*, 244(1):261–276.

Borwein, J. and Lewis, A. (2000). *Convex Analysis and Nonlinear Optimization*. CMS Books in Mathematics. Springer-Verlag, New York.

Borwein, J. M., Lewis, A. S., and Noll, D. (1996). Maximum entropy reconstruction using derivative information, part 1: Fisher information and convex duality. *Mathematics of Operations Research*, 21(2):442–468.

Bowers, J. (2001). The cost of weather in a floating oil production system. *Journal of the Operational Research Society*, 52(2):135–142.

Bowers, J. and Mould, G. (1994). Weather risk in offshore projects. *Journal of the Operational Research Society*, 45(4):409–418.

Brockett, P., L., Wang, M., and Yang, C. (2005). Weather derivatives and weather risk management. *Risk Management and Insurance Review*, 8(1):127–140.

Buchholz, M. and Musshoff, O. (2014). The role of weather derivatives and portfolio effects in agricultural water management. *Agricultural Water Management*, 146:34–44.

Buxey, G. (1988). Production planning under seasonal demand: A case study perspective. *Omega*, 16(5):447–455.

Caliskan Demirag, O. (2013). Performance of weather-conditional rebates under different risk preferences. *Omega*, 41(6):1053–1067.

Campbell, S., D. and Diebold, F., X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100:6–16.

Cao, M. and Wei, J. (2004). Weather derivatives valuation and market price of weather risk. *Journal of Future Markets*, 24(11):1065–1089.

Cao, Q., Ewing, B. T., and Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*, 221(1):148–154.

Castellano, R., Cerqueti, R., and Rotundo, G. (2020). Exploring the financial risk of a temperature index: a fractional integrated approach. *Annals of Operations Research*, 284:225–242.

Challis, S. (1999). Bright forecast for profits. *Reactions*, June edition.

Chen, F. Y. and Yano, C. A. (2010). Improving supply chain performance and managing risk under weather-related demand uncertainty. *Management Science*, 56(8):1380–1397.

Cozzolino, J. M. and Zahner, M. J. (1973). The maximum-entropy distribution of the future market price of a stock. *Operations Research*, 21(6):1200–1211.

Davis, M. (2001). Pricing weather derivatives by marginal value. *Quantitative Finance*, 1:1–4.

Elias, R. S., Wahab, M. I. M., and Fang, L. (2014). A comparison of regime-switching temperature modeling approaches for applications in weather derivatives. *European Journal of Operational Research*, 232(3):549–560.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gulko, L. (1999a). The entropic market hypothesis. *International Journal of Theoretical and Applied Finance*, 2(03):293–329.

Gulko, L. (1999b). The entropy theory of stock option pricing. *International Journal of Theoretical and Applied Finance*, 2(03):331–355.

Gulko, L. (2002). The entropy theory of bond option pricing. *International Journal of Theoretical and Applied Finance*, 5(04):355–383.

Gulpinar, N. and Canakoglu, E. (2017). Robust portfolio selection problem under temperature uncertainty. *European Journal of Operational Research*, 256(2):500–523.

Gzyl, H. (2017). Superresolution in the maximum entropy approach to invert laplace transforms. *Inverse problems in Science and Engineering*, 25(10):1536–1545.

Gzyl, H. and Mayoral, S. (2017). Maxentropic solutions to a convex interpolation problem motivated by utility theory. *Entropy*, 19(4):153.

Hanley, M. (1999). Hedging the force of nature. *Risk Professional*, 1:21–25.

Hardle, K. W., Lopez-Cabrera, B., and Teng, H.-W. (2015). State price densities implied from weather derivatives. *Insurance: Mathematics and Economics*, 64:106–125.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106:620–630.

Jewson, S., Brix, A., and Ziehmann, C. (2005). *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*. Campbridge University Press, Cambridge, UK.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233.

Judge, G. G. and Mittelhammer, R. C. (2011). *An Information Theoretic Approach to Econometrics*. Cambridge University Press.

Lange, K. (2010). *Numerical Analysis for Statisticians*, volume 1. Springer.

Perez-Gonzalez, F. and Yun, H. (2013). Risk management and firm value: Evidence from weather derivatives. *The Journal of Finance*, 68(5):2143–2176.

Rajasekera, J. and Yamada, M. (2001). Estimating the firm value distribution function by entropy optimization and geometric programming. *Annals of Operations Research*, 105(1):61–75.

Rouge, R. and El Karoui, N. (2000). Pricing via utility maximization and entropy. *Mathematical Finance*, 10(2):259–276.

Stulec, I. (2017). Effectiveness of weather derivatives as a risk management tool in food retail: The case of Croatia. *International Journal of Financial Studies*, 5(1):2.

Sun, B. and van Kooten, G. C. (2015). Financial weather derivatives for corn production in northern China: A comparison of pricing methods. *Journal of Empirical Finance*.

Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305.

Weagley, D. (2019). Financial sector stress and risk sharing: Evidence from the weather derivatives market. *The Review of Financial Studies*, 32(6):2456–2497.

Wilks, D. S. (2002). Realizations of daily weather in forecast seasonal climate. *Journal of Hydrometeorology*, 3(2):195–207.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*, volume 100 of *International Geophysics Series*. Academic Press, Oxford, UK, 3rd edition.

Zapranis, A. and Alexandridis, A. (2008). Modelling temperature time dependent speed of mean reversion in the context of weather derivetive pricing. *Applied Mathematical Finance*, 15(4):355 – 386.

Zhou, R., Cai, R., and Tong, G. (2013). Applications of entropy in finance: A review. *Entropy*, 15:4909–4931.

Zhou, R., Li, J. S.-H., and Pai, J. (2019). Pricing temperature derivatives with a filtered historical simulation approach. *The European Journal of Finance*, 25(15):1462–1484.

Zong, L. and Ender, M. (2018). Comparison of stochastic and spline models for temperature-based derivatives in China. *Pacific Economic Review*, 23(4):547–589.

Table 1.: Market states for the discrete model

State	ω_1	ω_2	...	ω_K
Level	$\hat{S}_1$	$\hat{S}_2$	...	$\hat{S}_K$
Range	$[S_0, S_1)$	$[S_1, S_2)$	...	$[S_{K-1}, S_K)$
Phys. Prob.	p_1	p_2	...	p_k

Table 2.: Data description

Dataset	t	Maturity	Number of Calls	Number of Puts
1	20/11/2007	31/12/2007	3	3
2	03/12/2007	31/12/2007	3	3
3	14/12/2007	31/12/2007	3	3
4	21/10/2010	30/11/2010	4	2
5	01/11/2010	30/11/2010	4	2
6	15/11/2010	30/11/2010	4	2
7	06/03/2006	03/04/2006	4	3
8	14/03/2006	03/04/2006	6	3
9	20/12/2007	31/01/2008	3	2
10	02/01/2008	31/01/2008	3	2
11	15/01/2008	31/01/2008	5	3
12	24/11/2009	31/12/2009	4	2
13	01/12/2009	31/12/2009	6	2
14	11/12/2009	31/12/2009	6	2

Table 3.: Market model for dataset 1

State	ω_1	ω_2	ω_3	ω_4	ω_5
Level	347.5	715	737.5	765	796
Range	$[0, 695)$	$[695, 725)$	$[725, 750)$	$[750, 780)$	$[780, 812)$
Phys. Prob. HBA	0.1282	0.0769	0.0001	0.1026	0.1537
Phys. Prob. CAR	0.1692	0.0924	0.0910	0.1248	0.1311
Phys. Prob. EQP	0.1000	0.1000	0.1000	0.1000	0.1000
State	ω_6	ω_7	ω_8	ω_9	ω_{10}
Level	824.5	846	861.5	884	1100
Range	$[812, 837)$	$[837, 855)$	$[855, 868)$	$[868, 900)$	$[900, 1300)$
Phys. Prob. HBA	0.1026	0.0769	0.0256	0.1282	0.2051
Phys. Prob. CAR	0.0957	0.0662	0.0391	0.0796	0.1109
Phys. Prob. EQP	0.1000	0.1000	0.1000	0.1000	0.1000

HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities

Table 4.: Call and Put HDD options for dataset 1

Type	Put	Put	Put	Call	Call	Call
Strike	710	740	760	800	825	850
Price	18	27	34	47	36	27

Table 5.: Option payoffs for dataset 1

Option	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
O_1	362.5	0	0	0	0	0	0	0	0	0
O_2	392.5	30	2.5	0	0	0	0	0	0	0
O_3	412.5	50	22.5	0	0	0	0	0	0	0
O_4	0	0	0	0	0	24.5	46	61.5	84	300
O_5	0	0	0	0	0	0	21	36.5	59	275
O_6	0	0	0	0	0	0	0	11.5	34	250

Table 6.: Market model for dataset 8

State	ω_1	ω_2	ω_3	ω_4	ω_5
Level	250	560	632.5	652.5	666.25
Range	$[0, 500)$	$[500, 620)$	$[620, 645)$	$[645, 660)$	$[660, 672.5)$
Phys. Prob. HBA	0.0001	0.0270	0.0270	0.0001	0.0541
Phys. Prob. CAR	0.0064	0.1461	0.0809	0.0574	0.0495
Phys. Prob. EQP	0.1111	0.1111	0.1111	0.1111	0.1111
State	ω_6	ω_7	ω_8	ω_9	
Level	677.5	688.75	747.5	900	
Range	$[672.5, 682.5)$	$[682.5, 695)$	$[695, 800)$	$[800, 1000)$	
Phys. Prob. HBA	0.1350	0.1080	0.2973	0.3514	
Phys. Prob. CAR	0.0452	0.0564	0.4146	0.1435	
Phys. Prob. EQP	0.1111	0.1111	0.1111	0.1111	

HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities

Table 7.: Call and Put HDD options for dataset 8

Type	Put	Call	Call	Put	Call	Call	Put	Call	Call
Strike	640	640	650	650	670	675	690	690	700
Price	1	61	53	3	36	33	14	24	21

Table 8.: Option payoffs for dataset 8

Option	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9
O_1	390	80	7.5	0	0	0	0	0	0
O_2	0	0	0	12.5	26.25	37.5	48.75	107.5	260
O_3	0	0	0	2.5	16.25	27.5	38.75	97.5	250
O_4	400	90	17.5	0	0	0	0	0	0
O_5	0	0	0	0	0	7.5	18.75	77.5	230
O_6	0	0	0	0	0	2.5	13.75	72.5	225
O_7	440	130	57.5	37.5	23.75	12.5	1.25	0	0
O_8	0	0	0	0	0	0	0	57.5	210
O_9	0	0	0	0	0	0	0	47.5	200

Table 9.: Pricing probabilities for dataset 1 using 4 option prices

PP	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}
HBA	0.0497	0.2703	0.0002	0.0868	0.1301	0.0868	0.1012	0.0346	0.1551	0.0853
CAR	0.0497	0.2208	0.1101	0.0865	0.0908	0.0663	0.0988	0.0652	0.1237	0.0882
EQP	0.0497	0.2146	0.1239	0.0790	0.0790	0.0790	0.0931	0.0958	0.0953	0.0906

PP = Pricing Probabilities, HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities

Table 10.: Pricing probabilities for dataset 8 determined from 3 option prices

PP	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9
HBA	0.0001	0.0379	0.0479	0.0002	0.1071	0.2771	0.2299	0.2557	0.0443
CAR	0.0000	0.0137	0.0683	0.0888	0.1165	0.1496	0.2631	0.2557	0.0443
EQP	0.0000	0.0105	0.0595	0.0958	0.1330	0.1739	0.2274	0.2557	0.0443

PP = Pricing Probabilities, HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities

Table 11.: Mean absolute distance of out-of-sample points by dataset, probability, and number of options used for fitting across experiments.

N	Prob	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	HBA	22.9	23.1	24.3	7.9	3.9	18.0	36.9	18.6	37.8	41.0	54.4	5.6	5.9	8.8
	CAR	16.7	16.9	22.6	7.8	4.6	18.4	20.5	18.7	18.1	20.7	40.2	16.3	17.1	7.9
	EQP	7.7	8.1	15.2	23.1	24.5	31.0	33.8	21.5	24.7	26.6	28.2	15.2	10.7	14.5
2	HBA	10.1	10.3	9.8	3.9	2.6	5.4	8.9	3.7	23.7	25.1	30.3	3.5	3.6	5.3
	CAR	7.2	7.3	7.9	3.5	2.9	5.8	5.2	5.1	9.2	10.9	21.5	7.3	7.8	2.6
	EQP	4.8	5.0	5.9	10.9	11.1	11.4	15.4	7.6	13.9	14.9	14.4	9.7	7.0	8.4
3	HBA	5.1	5.0	4.8	2.1	2.1	2.9	4.5	1.6	10.2	10.6	15.3	3.4	3.1	3.0
	CAR	4.3	4.0	4.2	1.8	2.1	2.7	4.7	1.9	4.1	5.3	11.5	4.2	5.0	1.3
	EQP	2.4	2.3	2.8	6.0	5.1	5.4	10.8	3.6	6.2	6.8	7.7	7.8	6.4	5.6
4	HBA	1.9	2.2	2.2	1.1	2.3	1.1	5.1	1.2	2.8	4.0	7.0	2.7	2.6	1.8
	CAR	1.8	1.8	2.0	0.9	2.2	0.9	5.1	1.3	2.6	3.7	5.8	3.1	4.0	0.9
	EQP	1.2	1.1	1.3	1.6	3.1	1.1	6.6	2.2	3.7	4.5	4.1	3.2	5.2	3.7
5	HBA	0.8	1.2	1.3	0.2	1.4	-*	5.4	0.8	2.3	2.7	2.7	1.8	2.4	1.2
	CAR	0.9	0.7	0.9	0.2	1.5	-*	5.4	1.1	2.2	2.6	2.3	1.8	3.5	0.7
	EQP	0.9	0.6	0.7	0.3	2.0	-*	7.6	1.4	3.2	3.8	1.9	1.8	4.3	2.3
6	HBA	-*	-*	-*	-*	-*	-*	-*	1.3	-	-	1.3	-*	2.3	0.7
	CAR	-*	-*	-*	-*	-*	-*	-*	1.0	-	-	1.0	-*	3.1	0.6
	EQP	-*	-*	-*	-*	-*	-*	-*	1.1	-	-	1.3	-*	3.3	1.1
7	HBA	-	-	-	-	-	-	-	0.7	-	-	1.3	-	2.1	0.6
	CAR	-	-	-	-	-	-	-	-*	-	-	1.2	-	2.7	0.7
	EQP	-	-	-	-	-	-	-	-*	-	-	1.5	-	2.1	0.9
8	HBA	-	-	-	-	-	-	-	-*	-	-	0.4	-	-*	-*
	CAR	-	-	-	-	-	-	-	-*	-	-	1.1	-	-*	-*
	EQP	-	-	-	-	-	-	-	-*	-	-	1.7	-	-*	-*

HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities, * = No coherent solution

Table 12.: The contribution of the contract effect to the accuracy of out-of-sample pricing.

	Datasets 1+2+3			Datasets 4+5+6			Datasets 7+8			Datasets 9+10+11			Datasets 12+13+14		
N	HBA	CAR	EQP	HBA	CAR	EQP	HBA	CAR	EQP	HBA	CAR	EQP	HBA	CAR	EQP
1	23.4	18.7	10.3	9.9	10.3	26.2	27.8	19.6	27.7	44.0	26.3	26.5	6.8	13.7	13.5
2	10.1	7.5	5.2	4.0	4.1	11.1	6.3	5.1	11.5	26.4	13.9	14.4	4.1	5.9	8.4
3	5.0	4.2	2.5	2.4	2.2	5.5	3.1	3.3	7.2	12.0	7.0	6.9	3.2	3.5	6.6
4	2.1	1.9	1.2	1.5	1.3	1.9	3.1	3.2	4.4	4.6	4.0	4.1	2.4	2.7	4.0
5	1.1	0.8	0.7	0.8	0.8	1.1	3.1	3.2	4.5	2.5	2.4	3.0	1.8	2.0	2.8
6	-*	-*	-*	-*	-*	-*	1.2	1.0	1.1	1.3	1.0	1.3	1.5	1.9	2.2
7	-	-	-	-	-	-	0.7	-*	-*	1.3	1.2	1.5	1.4	1.7	1.5
8	-	-	-	-	-	-	-*	-*	-*	0.4	1.1	1.7	-*	-*	-*

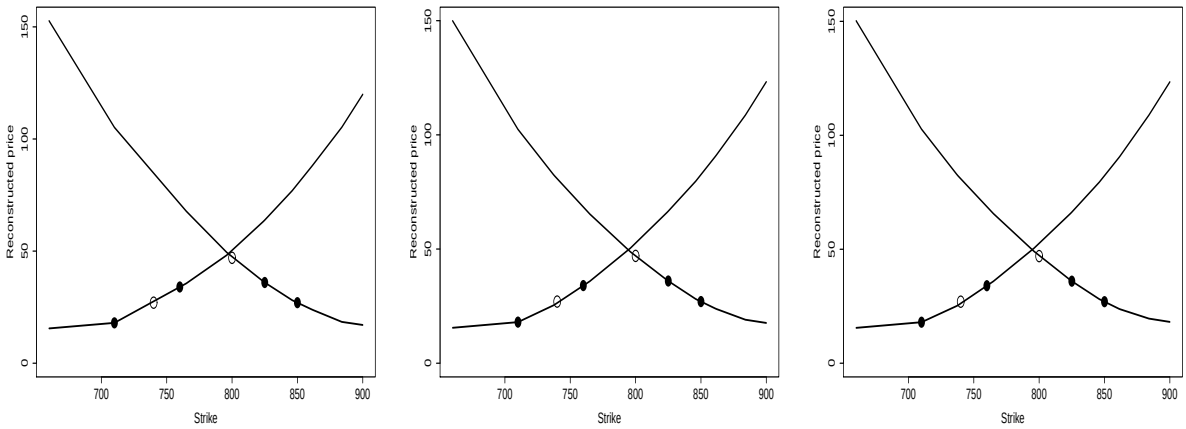
HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities * = No coherent solution

Table 13.: The contribution of the time effect to the accuracy of out-of-sample pricing.

	Datasets 1+4+9+12			Datasets 2+5+7+10+13			Datasets 3+6+8+11+14		
N	HBA	CAR	EQP	HBA	CAR	EQP	HBA	CAR	EQP
1	18.5	14.7	17.7	22.2	16.0	20.8	24.8	21.5	22.1
2	10.3	6.8	9.8	10.1	6.8	10.7	10.9	8.6	9.5
3	5.2	3.6	5.6	5.1	4.2	6.3	5.5	4.3	5.0
4	2.1	2.1	2.4	3.2	3.4	4.1	2.7	2.2	2.5
5	1.3	1.3	1.5	2.6	2.7	3.7	1.5	1.3	1.6
6	-*	-*	-*	2.3	3.1	3.3	1.1	0.9	1.2
7	-	-	-	2.1	2.7	2.1	0.9	0.9	1.2
8	-	-	-	-*	-*	-*	0.4	1.1	1.7

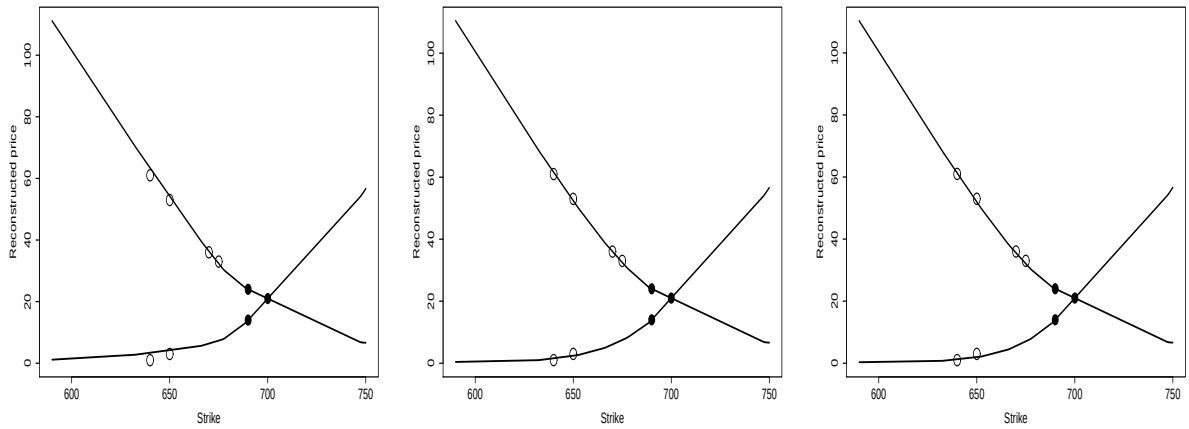
HBA = Historical Burn Analysis, CAR = Continuous Autoregressive, EQP = Equal probabilities * = No coherent solution

List of Figures



(a) HBA (b) CAR (c) EQP

Figure 1.: Option prices across strikes for dataset 1, determined from four options under each of the 3 different probabilities. Dots represent the option prices used for fitting, and circles represent the observed option prices not used for fitting.



(a) HBA (b) CAR (c) EQP

Figure 2.: Option prices across strikes determined for dataset 8, from three options under each of the 3 different probabilities. Dots represent the option prices used for fitting, and circles represent the observed option prices not used for fitting.

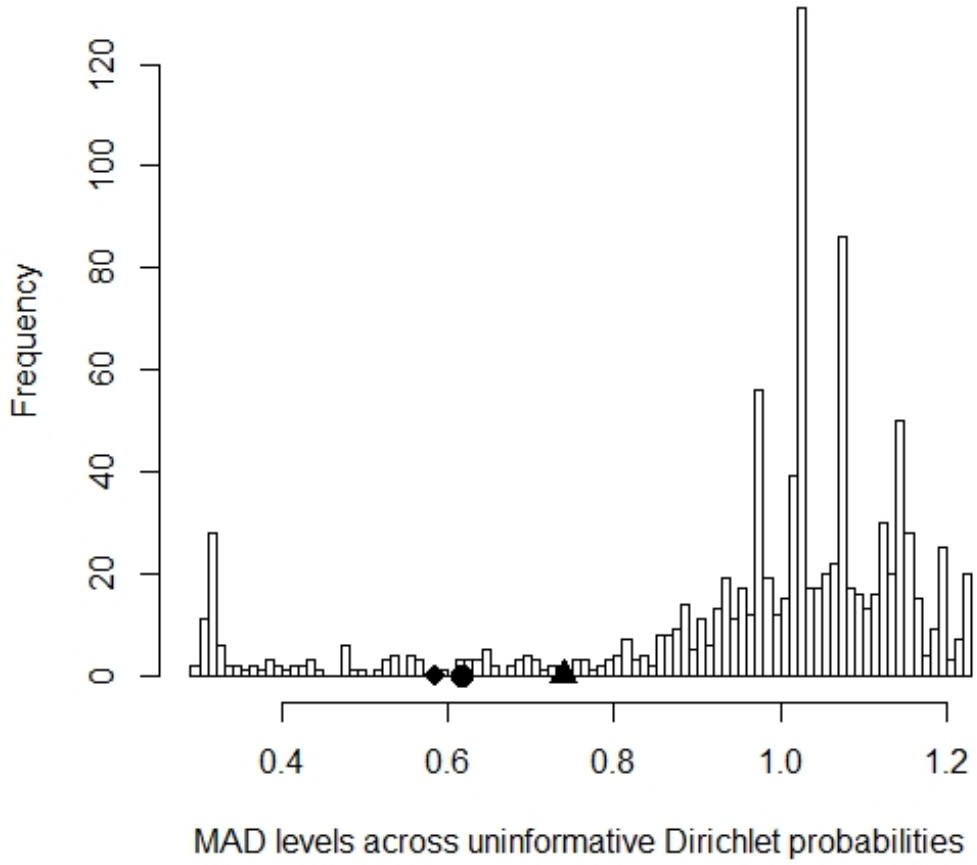


Figure 3.: Distribution of mean absolute differences under an non-informative Dirichlet prior on the model probabilities. The MAD for HBA, CAR and EQP are represented by a circle, a triangle and a diamond respectively at the x-axis level.