



Time-Dynamic Markov Random Fields for price outcome prediction in the presence of lobbying

The case of olive oil in Andalusia

Julia García Cabello¹

Accepted: 5 June 2021
© The Author(s) 2021

Abstract

This paper presents a mathematical/Artificial Intelligence (AI) model for the prediction of price outcomes in markets with the presence of lobbying, whose outputs are pricing trends that aggregate the opinions of lobbies on future prices. Our proposal succeeds in unraveling this complex real-world problem by reducing the solution to straightforward probability computations. We tested our method on real olive oil prices (Andalusia, Spain) with encouraging results in a challenging sector, where opacity in the entry of oil shipments which are stored while waiting for the price to rise, makes it very difficult to forecast the prices. Specifically, understanding by minimum price that the price level is at least reached, specific formulas for computing the likelihood of both the aggregate and the minimum market price are provided. These formulas are based on the price levels that lobbies expect which in turn, can be calculated using the probability that each lobby gives to market prices. An innovative quantitative study of the lobbies is also carried out by explicitly computing the weight of each lobby in the process thus solving a problem for which there were only qualitative references up until now. The structural model is based on Time Dynamic Markov random fields (TD-MRFs). This model requires significantly less information to produce an output and enjoys transparency during the process when compared with other approaches, such as neural networks (known as black boxes). Transparency also ensures that the internal structures can be fine tuned to fit to each context as well as possible.

Keywords Price outcomes · Aggregate and minimum market price · Networks · Time Dynamic Markov Random Fields · Olive Oil sector

1 Introduction and literature review

Price is what a consumer must pay in order to receive a product. It is also known as a *charge* or *fee* when referring to services. Although price is often related to the maximization of a company's profits, it can also compete for or change market share. In any case, knowing the price of a product (or as close to this as possible) confers an advantageous position on companies that trade with it in one way or another.

Price prediction is a complicated task when it relates to products in markets where there are pressure groups with intricate interrelated interests. Indeed, price is the output

of a complex process which combines both factors and conflicting interests amongst related agents. While factors influencing price are known (production costs, profit-margins, market demand, consumer purchasing power, competitors and regulation amongst others), the interests (interrelationships) between lobbies are not. And even if these interests were known, it would not be easy to quantify them to know exactly what *weight* (i.e., how much influence) each lobby has in this process.

This paper presents a mathematical/artificial intelligence (AI) model that calculates price outcomes in markets with lobbies by capturing prices using a time-varying Markov random field over an underlying graph which represents the market. Our main contribution is to use this method to successfully solve the complex problem of what kinds of market prices emerge when there are multiple agents pushing for their prices. Our contribution is not only to solve the problem but also to do it in a clear and direct way through a set of straightforward formulas, based on the price levels expected by lobbies which in turn, can be

✉ Julia García Cabello
cabello@ugr.es

¹ Department of Applied Mathematics, University of Granada, Granada, Spain

calculated by using the probability that each lobby gives to market prices. More specifically, if the minimum price is understood as being the price level at least reached, specific formulas for computing the likelihood of both the aggregate and minimum market price of a product are provided.

An additional contribution is to give an expression to compute the *weight* of each lobby. A quantitative study of this type on the importance of lobbies in the process of pricing outcomes sheds light on a problem for which there were only qualitative references up until now. Actually, to the best of our knowledge, this is the first quantitative study of the lobby's importance in processes of pricing outcomes. This would complete the information necessary to develop adequate pricing strategies that allow price negotiation processes to be conducted from a vantage point. Such help in decision-making would allow practitioners to anticipate and seize market opportunities.

In the final section, this paper also explores the connections between the key problem (what kind of price outcome appears in a market with multiple agents that push for its price) and thermoeconomics, an emerging discipline (involving concepts such as ensemble average of machine learning, see [10]) which provides a novel approach to the problem being studied.

The structural model is based on Markov random fields (MRFs) considered in their dynamic version (Time-Dynamic Markov Random Fields, TD-MRFs) to adjust to reality. Our approach has several advantages over previous ones: a model of the interrelationships and objectives of the agents (as pricing factors, stakeholder's interests, interactions amongst agents) is provided allowing the complex process to be visualized using a suitable dynamic graphical model (DGM). Moreover, a joint probability distribution (associated to a TD-MRF) is given that gathers the perspectives of all the related agents and translate them into expressions with simple and clear formulations. In this way, the use of TD-MRFs clearly shows how the variables involved interact by displaying the relationships amongst them. This is in contrast with other approaches that are classified as opaque because their internal structures remain hidden [18], [24]. As a result, all the processes can be fine-tuned by varying the variables of the model using suitable sensitivity tests to better suit the changing dynamics of the sector. Thanks to the notable computational properties of DGMs, our approach can also be easily translated into coding-terms thereby providing extra support. Furthermore, the use of TD-MRFs for building the structural model has two additional advantages: first, TD-MRFs require significantly less information to produce an output (only the information from cliques is required) compared with existing approaches such as Bayesian or neural networks, which need large amounts of data (one of their major disadvantages). This

also means a higher computing speed and lower storage capacity requirements. And second, the use of TD-MRFs means that there is no need to use a *given* distribution function, as required when using neural networks and related statistical methods, which strongly rely on a given distribution.

Generally speaking, interest groups which compete for influence are called lobbies whether they are fully organized or not. They are pressure groups which try to exercise their influence on policymaking processes, government decisions or the economic course of action, particularly on pricing processes. We illustrate our analysis of price outcomes under the influence of lobbying with the olive oil sector, where lobby pressure along with a lack of transparency in certain links of the value chain, make price prediction difficult. It is important to note that, in this sector, determining the weight of each lobby (i.e., how much influence they come to exert) is challenging since each group exerts pressure in a different way. Moreover, some of the lobbies have extra support that adds to their specific weight. We are referring to agricultural producers -at the lowest link in the chain- who apply pressure with public displays of discontent, in contrast with the more private actions of other parts of the olive oil value chain. Furthermore, the demonstrations held by agricultural producers enjoy public support. The consequence of this extra weight is enormous influence as evidenced by the ability of the producers to impact on the process of reforming the EU(European Union)'s agricultural policy (Agriculture is the EU's largest policy area in budgetary terms, see [2] and [23]). Therefore, our approach clarifies this panorama by providing explicit computation of the weights of the lobbies in price outcomes. Importantly, both the model designed and its findings can be applied, with minor adjustments to other lobbying. As mentioned, our analysis of price outcomes in the presence of lobbying uses Markov random fields, which are stochastic models that generalize the well-known Markov chains. While both stochastic models and Markov decision processes have been previously used for studies on price outcomes, see [4] for instance, to the best of our knowledge, our approach is the first study of price outcomes under the influence of lobbying that explicitly aggregates the interests of pressure groups in a single final output.

In the literature review, we shall address two points: price prediction and those papers which specifically deal with the olive oil sector. There are several studies which address price outcomes. Instead of trying to mention them all, we shall make a brief summary regarding the different techniques used. Let us start with *probabilistic forecasting*: autoregression (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated

moving-average (SARIMA), which are all known as *time series analysis*. While time series techniques can help to clean data by filtering out the noise, one of their major drawbacks in price prediction is that they are strongly dependent on past history: they are unable to forecast anything that has not previously happened (see [25], which contains a systematic literature review on financial time series forecasting, covering the period 2005–2019). For these reasons, these are valuable tools when combined with other methodologies, see the paper [20] which examines performance of traditional time series models, ROBUST models and ARIMA models together with generalized regression neural networks (GRNNs) and multi-layer feedforward networks (MLFNs) in forecasting prices.

According to [21], probabilistic forecasting has been one of the areas that has made significant advances in the last decade against all odds. Our proposal can also be considered to be one of the probabilistic forecasting methodologies inasmuch as it employs joint probability distributions. It has a high degree of novelty as it relies on MRFs for predicting purposes, see [11], [12], while MRFs are tools known in AI as instruments that have traditionally been used in image processing or pattern recognition and, more recently in activity detection (see [8] for an application of MRFs for the detection of spammers).

Artificial neural networks (ANN) are widely used in price prediction, see [6] and [17] to mention more recent papers. The latter includes an ANN model which tests the model effectiveness for the prediction of perishable vegetable price series. *Vector Support Machines (VSM)* are neural network techniques used in solving nonlinear regression estimation problems. They also have been successfully used to predict price, see [28] for instance, where the authors combine VSM with the firefly algorithm (FA) to enhance the global convergence speed. Another example is the more recent paper [13], where time series forecasting of agricultural product prices are used together with a Wavelet-Support Vector Machine (W-SVM) model in forecasting palm oil price. In this type of model, wavelet transform is used to decompose data series into two parts which are then used as the input for the SVM model to forecast the palm oil price, thereby enhancing the forecasting performance (compared with the traditional ANN). A mixture of ANN or VSM with other methodologies are also used, as in [26] which combines an ensemble empirical mode decomposition (EEMD) into a wavelet neural network with a random time effective (WNNRT) to design a hybrid neural network prediction model for energy prices. A further example is the paper [27], where a novel random deep bidirectional

gated recurrent unit neural network is constructed in order to obtain accurate forecasts of crude oil prices, which also integrates historical data into the training process of the model.

While effective in predicting prices, these techniques have some limitations regarding issues related to price outcomes due to the complex dimensionality of this problem, not only because of the large number of factors that affect the final output but also, above all, for the intricate relationships between them. Another limitation of these techniques is that they are data-intensive requiring huge amounts of data. Nonetheless, in the opinion of some authors, [18] or [24], their main drawback is that they implicitly involve a high levels of opacity in their reasoning (they are called black boxes). This makes them difficult to fine tune, hindering the incorporation of the characteristics of each scenario into the models.

Unlike with qualitative analysis, there is very little literature on olive oil sector-specific papers that study the subject from a quantitative perspective. Amongst the most recent ones is the paper [16], referred to as adaptive learning forecasting, where forecast revisions are made from previous forecasting errors. The paper also applies this to agricultural prices. In [22] the interrelationship between olive oil price dynamics and pollen emission as a price determinant are studied by using a pollen monitoring methodology to predict olive yields. Moreover this paper explores the difficulties of defining marketing strategies for oil prices due to the complex nature of the olive oil market. Furthermore, the paper [5], which explores the price determinants of extra-virgin olive oil using a mathematical model based on three price levels (production, intermediation and sales) and states the relationships amongst price factors such as the purchase cost of olives and the production cost of olive oil. Finally, the paper [1] econometrically considers olive oil price formation aimed at identifying the trader' behaviour in the olive oil market. In this paper, for prediction purposes, the artificial neural network (ANN) approach is used.

The remainder of the paper is organized as follows. Section 2 contains the fundamentals on TD-MRFs. In Section 3 the structural model of price outcomes in the presence of lobbying is designed. Section 4 is devoted to developing the procedure for price prediction. In Section 5, an analysis of energy functions is performed. The application of the model to the olive oil market is fully detailed in Section 6, including both a model validation with real data and a complementary study from the perspective of thermodynamics. Finally, conclusions are stated in Section 7.

2 Dynamic graphical models. Time-Dynamic Markov random fields

Any group of random variables $X = \{X_{s_i} | s_i \in S, i \in \mathbb{N}\}$ can be viewed as a spatial stochastic process inasmuch as they may take concrete values $X_{s_i} = x_{s_i}$ for each site s_i of a given space S . For them, terms like node, site or vertice are used interchangeably throughout the literature. These processes are also known as *graphical models* and their main feature is that they express the probabilistic conditional dependence between random variables. For them, let us pay attention to notions of neighbourhood and clique. On one hand, two vertices s_i and s_j are said to be *adjacent*, denoted by $s_i \sim s_j$, if there is an edge connecting them. There is an equivalence between defining edges and the defining *neighbourhood* of a node s_i , $N(s_i)$, in the following sense: for two nodes s_i, s_j there is an edge connecting them $s_i \sim s_j$ if and only if $s_j \in N(s_i)$. On the other hand, *cliques* are maximally connected subgraphs of the underlying graph.

Graphical models are commonly categorized into Bayesian and Markov random fields (MRFs) where Markov random fields are particular graphical models that use an undirected graph to represent a distribution, see (Kiefermann 1980) [15]. These have been (only) used in the image-processing scenario (denoising and vision), where only a static model is needed. The local Markov property which characterizes MRFs is defined as follows: if $\mathbb{P}[X] = \mathbb{P}[\{X_{s_i} | s_i \in S, i \in \mathbb{N}\}] = \{\mathbb{P}[X_{s_i} = x_{s_i}] | s_i \in S, i \in \mathbb{N}\}$ denotes the joint distribution of random variables $X = \{X_{s_i} | s_i \in S, i \in \mathbb{N}\}$, thus

$$\mathbb{P}[X_{s_i} = x_{s_i} | X_{S-\{s_i\}} = x] = \mathbb{P}[X_{s_i} = x_{s_i} | X_{N(s_i)} = x],$$

where \mathbb{P} denotes probability in order to avoid confusion with the P reserved for the variable “price”. The Hammersley-Clifford theorem establishes that an MRF of non-negative random variables (“positivity condition”) has an associated *joint* distribution function which can be described in terms of some functions f on the cliques $C \in \mathcal{C} \subset S$ (i.e., in terms of some functions of the given random variable X taking values only on the cliques). These functions f are called *energy functions*¹ or *clique potentials* when considered as product factors. Thus, the joint distribution function is of the form $\mathbb{P}[X] = \frac{1}{Z} \exp[-\sum_{C \in \mathcal{C}} f(X_C)]$, where Z is a normalizing constant² which ensures that the distribution sums to 1. This is known as a Gibbs distribution.

Time-dynamic (or time-varying) Markov Random Fields (TD-MRFs) are the plain extension of MRFs for

dynamic scenarios (see [19]). These are a special sort of dynamic graphical models (DGMs) which explicitly model the correlations in space and in time as dependencies amongst the random variables such that the corresponding joint distribution function can be written by means of functions of the corresponding time-varying random variable over the cliques $C \in \mathcal{C}$:

$$\mathbb{P}[X] = \frac{1}{Z} \exp \left[- \sum_{C \in \mathcal{C}} f(X_C^t) \right] \text{ where } t \text{ represents the time step.} \tag{1}$$

As mentioned before, TD-MRFs have been chosen as the theoretical structure for our model in order to both reflect the dynamic nature of the reality of price outcomes and to reduce the amount of information needed.

3 The structural model of price outcomes using TD-MRFs (the TD-MRF model)

This section is devoted to developing the fundamentals of price outcomes in the presence of lobbying. Hereafter “price” will be denoted by P while \mathbb{P} will stand for “probability”. There are many agents who are pushing to obtain their price: agents will be denoted by a_i while P_{a_i} is the price they are pushing for. Let A be the set $A = \{a_i, i \in \mathbb{N}\}$ which gathers all the agents pushing for a certain price, no matter whether they agree on the price they want to reach or not (which we will refer to as *target price*). Agents are interconnected by common relationships (commercial, agreement on the target price etc.) which interconnect them. All this information can be visualized jointly with a graph (A, E) whose nodes are the agents a_i and whose edges are represented by such links amongst agents.

Price is a variable that can take positive values in all cases up to a upper bound max_P : i.e., the range of P is $[0, max_P]$. Since P_{a_i} is the target price that each agent a_i pushes for, each P_{a_i} can be viewed as a random variable as it can take different numerical values *depending on each agent* $a_i \in A$. Thus, the previous graph (A, E) shall be the underlying graph of a graphical model with nodes the random variables $P = \{P_{a_i}, a_i \in A\}$ whose values vary depending on each agent $a_i \in A$. Once the edges between the nodes have been specified (which we will do shortly), we shall have a joint graphical representation (and therefore a graphical model) of statistical relationships between non-negative random variables $P = \{P_{a_i}, a_i \in A\}$.

Importantly, price outcomes are the result of a time-varying process. For this reason, we make the transition

¹In a term borrowed from physics.

²In literature Z is also known as “the partition function”.

from a static scenario (graphical models) to *dynamic graphical models* (DGMs) by incorporating the time step as superscript t . Hence, nodes in a dynamic graphical model DGM are the random variables $P^t = \{P_{a_i}^t, a_i \in A\}$ whose values vary depending on both each agent $a_i \in A$ and time t . Additionally, in order to capture the huge variety of factors that influences price outcomes, we shall describe the random variable “ a_i ’s target price”, $P_{a_i}^t$, as an array (or column) of features $p_{a_{ki}}^t, k = 1, \dots, n$ that gathers all the factors associated with price: $P_{a_i}^t \simeq (p_{a_{1i}}^t, p_{a_{2i}}^t, \dots, p_{a_{ni}}^t)^t$ (here, superscript t means transpose). For the sake of simplicity we will identify each agent a_i with their target price $P_{a_i}^t$, i.e., $a_i \simeq P_{a_i}^t \simeq (p_{a_{1i}}^t, p_{a_{2i}}^t, \dots, p_{a_{ni}}^t)$.

Once the nodes have been fully described, the edges amongst the nodes must be defined. To this end, it is important to note that there is an equivalence between the edges and the neighbourhoods of a graphical model,³ i.e., there is an edge connecting two nodes $P_{a_i}^t, P_{a_j}^t$, denoted by $P_{a_i}^t \sim P_{a_j}^t$, if and only if $P_{a_i}^t \in N(P_{a_j}^t)^t$. Thus, we define the neighbourhood of a node as follows:

$$N(P_{a_i}^t) = \{P_{a_j}^t \text{ such that the random variables } P_{a_i}^t, P_{a_j}^t \text{ are equivalent}\},$$

in the usual sense: random variables $P_{a_i}^t, P_{a_j}^t$ are equivalent if and only if their distribution functions are the same: $\mathbb{P}[P_{a_i}^t \leq p] = \mathbb{P}[P_{a_j}^t \leq p]$ for a numerical price p . From this definition it should be noted that, in particular, agents in the same neighbourhood have the same marginal distribution. Thus, the neighbourhood of a node may be easily identified as follows:

Lemma 1 *All agents in the same neighbourhood have an equal conditional distribution.*

Proof In order to prove this result, let $P_{a_i}^t, P_{a_j}^t$ be two nodes in the same neighbourhood. The definition taken implies that their marginal distributions are equal, $\mathbb{P}[P_{a_i}^t] = \mathbb{P}[P_{a_j}^t]$. The result follows from applying Bayes’s theorem:

$$\mathbb{P}[P_{a_i}^t | P_{a_j}^t] = \frac{\mathbb{P}[P_{a_j}^t | P_{a_i}^t] \cdot \mathbb{P}[P_{a_i}^t]}{\mathbb{P}[P_{a_j}^t]} = \mathbb{P}[P_{a_j}^t | P_{a_i}^t]. \quad \square$$

Proposition 1 *The neighbourhood of a node consists of all the agents who have the same pricing interests. That is, the neighbourhoods are the lobbies.*

³This way of describing edges is particularly suitable when there is no order on the set of nodes as is the case with the set of agents $A = \{a_i, i \in \mathbb{N}\}$.

Proof On one hand, from the definition of neighbourhood of a node a_i , agents a_j who belong to this set have the same marginal distribution function, that is, their corresponding achievement prices shall take the same values. On the other hand, previous lemma states that their corresponding conditional distributions are equal which means that their mutual interests are the same on their target price. This implies that agents in a neighbourhood shall jointly push for the same price. This means that the neighbourhoods are the lobbies. \square

Once the nodes and edges have been defined (and considering former identification $a_i \simeq P_{a_i}^t$), previous structures enable us to switch from graph (A, E) to the dynamic graphical model (P^t, N^t) . In other words, price outcomes are the result of a process like the one described in Fig. 1 that uses a DGM. In this way, this complex process of relationships between agents is displayed as a solid graphical representation which clearly shows both the target price of each agent and the affinity-in-price (or lack of) between them.

It is important to remember that for any random variable, the transition from the random variable to a particular value is known as the *realization of the random variable*. This procedure can be carried out using several approaches, known as rating/scoring processes. As mentioned before, price outcomes are the result of a dynamical process, which is time-varying and depend on agents a_i , which has been fully covered by its representation as a former DGM (P^t, N^t) . From this general representation, two kinds of realizations may be carried out. First, if price outcomes for a concrete time instant t_0 are required, the corresponding DGM, formed by

- nodes $P(t_0) = \{P_{a_i}(t_0), a_i \in A\}$ and
- edges $P_{a_i}(t_0) \sim P_{a_j}(t_0)$ if agents a_i, a_j belong to the same lobby,

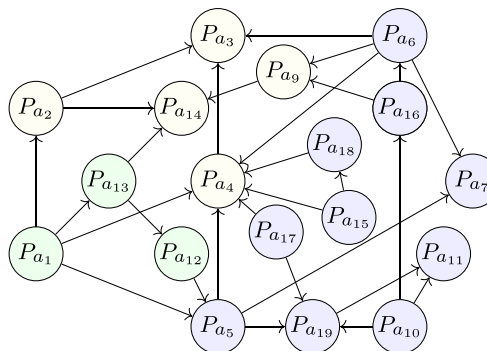


Fig. 1 Price outcomes as a TD-DGM

should display the process as time stopped in t_0 . A second realization of price outcomes may be performed in (P^t, N^t) when the random variables involved take the same value simultaneously. Once a random variable decomposed in a collection of features $P_{a_i}^t \simeq (p_{a_{1i}}^t, p_{a_{2i}}^t, \dots, p_{a_{ni}}^t)$ has taken a concrete numerical value, it is renamed as *feature vector*, i.e., a vector formed by n numerical scores which represents an object according to the definition commonly used in AI literature.

Let us now pay attention to the former DGM. Before addressing the next theorem, there are some points that need to be clarified. In Gibbs distribution contexts, the statement “energy functions f are functions of the random variable taking values only on the cliques” can have more than one interpretation: the functions f are equal for all cliques but they can have different domains (input values of the price) depending on each clique or the f 's can be different depending on cliques with a common domain for all cliques. Since from the perspective of mathematical modelling both interpretations are equivalent, we consider the second one to be:⁴ thus f_{C_i} will stand for the energy function corresponding to the i th-clique. Thus, the following theorem is the most important result in the paper as it proves that price outcomes are a TD-MRF with an explicit joint probability distribution expressed in terms of lobbies. It should be noted that the positivity condition on the set of variables “price” is always satisfied.

Theorem 1 (Price outcomes are a TD-MRF) *Price outcomes viewed as a DGM are a TD-MRF. Its corresponding joint distribution $\mathbb{P}[P^t]$ is equal to*

$$\mathbb{P}[P^t] = \frac{1}{Z} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)} = \frac{1}{Z} \prod_{i \in \mathbb{N}} e^{-f_{L_i}(P^t)},$$

where L_i stands for the i th-lobby. (2)

Proof To prove the theorem is equivalent to proving that the local dynamic Markov property holds. To do so, consider that the target price of a given agent a_i takes a concrete numerical value $P_{a_i}^t = p_{a_i}$ while the target price of the rest of the agents $(A - \{a_i\})$ takes a value p . Since the set of agents that are different from a_i may be expressed as $A - \{a_i\} = N(a_i) - \{a_i\} \cup \text{Not}N(a_i)$, by the definition of

neighbourhood $N(a_i)$ (where $\text{Not}N(a_i)$ represents the set of agents who do not belong to $N(a_i)$) it follows that

$$\begin{aligned} \mathbb{P}[P_{a_i}^t = p_{a_i} | P_{A-\{a_i\}}^t = p] &= \frac{\mathbb{P}[(P_{a_i}^t = p_{a_i}) \cap (P_{A-\{a_i\}}^t = p)]}{\mathbb{P}[P_{A-\{a_i\}}^t = p]} \\ &= \frac{\mathbb{P}[(P_{a_i}^t = p_{a_i}) \cap (P_{N(P_{a_i}^t)}^t = p)]}{\mathbb{P}[P_{N(P_{a_i}^t)}^t = p]} \\ &= \mathbb{P}[P_{a_i}^t = p_{a_i} | P_{N(a_i)}^t = p]. \end{aligned}$$

Hence, price outcomes are a TD-MRF at any time instant. In consequence, the price outcomes $P^t = \{P_{a_i}^t, i \in \mathbb{N}\}$ obey a Gibbs distribution (see (1)) with a joint probability function depending on the price P over the cliques.

By and large, cliques are fully connected subsets of the neighbourhood. However, from our definition of a neighbouring relationship, cliques are equal to the neighbourhood itself. Hence, the price outcome joint distribution can be expressed in terms of clique energy functions, different for each clique but taking the same input prices (i.e., having the same domain). That is to say, the price outcome joint distribution is of the form $\mathbb{P}[P^t] = \frac{1}{Z} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)}$ with energy functions f_{L_i} where L_i is the i th-looby, according to proposition 1. □

4 Deriving the model of price prediction

The aim of this section is to develop a price prediction procedure from the previous structural model. Explicit formulae for computing the likelihood of both aggregate and minimum market price of a product are provided to this end. Such information offers as future trends an output that aggregates the opinions of lobbies on future prices.

Therefore, we will focus on the neighbourhoods of the TD-MRF model. It is important to remember that, from the previous proposition 1, the neighborhoods are the lobbies. Taking into account the fact that the positivity condition on the variable “price” is always satisfied, the following result is achieved:

Theorem 2 *Any lobby in the market, considered as a DGM, is a TD-MRF.*

Proof It is important to remember that, according to the definition of neighbourhood all nodes which belong to this set have the same distribution function. This implies that (Markov) local property holds and hence, the

⁴Once this premise has been established, the analysis of the normalizing constant Z in sections 4 and 6 (under the denomination of “partition function”) will benefit.

neighbourhood of each node is a TD-MRF. By proposition 1 the result holds. \square

As mentioned before (see the proof of theorem 1) from the definition of neighbouring relationship, cliques equal the neighbourhood. Hence the terms “clique”, “neighbourhood” or (by proposition 1) “lobby” shall be used interchangeably. $\mathbb{P}_{L_i}[P^t]$ denotes the distribution which expresses the probability of reaching a (feasible) market price⁵, according to the opinion of the L_i lobby. Thus, the set of all the functions $\mathbb{P}_{L_i}[P^t]$, $i = 1, \dots, n$ (= number of lobbies) condenses the opinions of lobbies on future prices.

The next result provides an explicit expression of the likelihood distribution for the i -th lobby.

Theorem 3 (Joint probability distribution of each lobby) *The joint (i.e., for all agents belonging to the lobby) likelihood distribution for the i th-lobby is given by the formula $\frac{1}{Z}e^{-f_{L_i}(P^t)}$ where Z is a normalizing constant and f_{L_i} is the energy function of the i th-lobby. We refers to this as*

$$\mathbb{P}_{L_i}[P^t] = \frac{1}{Z}e^{-f_{L_i}(P^t)}.$$

Proof On one hand, recall that cliques in a graphical model are connected subgraphs where no node may be added and still be connected (i.e., they are maximally connected). Moreover, there is only one clique which coincides with the whole. On the other hand, an specific joint probability function of the neighbourhood may be derived by applying the Hammersley-Clifford theorem: $\mathbb{P}_{L_i}[P^t]$, which can be expressed in terms of energy functions f which are clique-dependent. Hence, the results follows. \square

A clarification should be made on the notation of the normalizing constant Z .

Remark 1 For simplicity, in both Theorems 1 and 3 the normalizing constant Z has been denoted with the same letter (and continues in this way). However, *it is not the same Z* . Let us examine this in greater detail. From both theorems 1 and 3, Z can be isolated from the fact that the corresponding probability distributions sums to 1:

$$\begin{aligned} 1 &= \sum_{\text{prices } p_j} \mathbb{P}[P^t] = & 1 &= \sum_{\text{prices } p_j} \mathbb{P}_{L_i}[P^t] = \\ &= \sum_{\text{prices } p_j} \frac{1}{Z} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)} = & &= \sum_{\text{prices } p_j} \frac{1}{Z} e^{-f_{L_i}(P^t)} \\ &= \frac{1}{Z} \sum_{\text{prices } p_j} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)} \Rightarrow & &= \frac{1}{Z} \sum_{\text{prices } p_j} e^{-f_{L_i}(P^t)} \Rightarrow \\ &\Rightarrow Z = \sum_{\text{prices } p_j} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)}. & &\Rightarrow Z = \sum_{\text{prices } p_j} e^{-f_{L_i}(P^t)}. \end{aligned}$$

⁵Particularly, it would allow the likelihood of achieving the i th-lobby’s target price price to be computed.

It should be noticed that in both cases the sub-indices set is the same. This is the result of considering that the energy functions f can be different depending on cliques but they have a common domain.

As a consequence of theorems 1 and 3, the following corollaries shall finally provide explicit formulae to compute i) likelihood of an aggregate market price and ii) likelihood of a minimum market price, understanding that the minimum price is that price level at least reached. The corresponding probabilities from the perspective of the lobbies are also given.

Corollary 1 (Likelihood of reaching a market price) *Let us consider a market with lobbies L_i and a given market price p at a time instant t_0 . Thus,*

- i) *the likelihood of reaching an aggregate market price p is $\mathbb{P}[P^{t_0} = p] = \frac{1}{Z}e^{-\sum_{i \in \mathbb{N}} f_{L_i}(p)} = \frac{1}{Z} \prod_{i \in \mathbb{N}} e^{-f_{L_i}(p)}$.*
- ii) *From the perspective of the i th-lobby, the likelihood of reaching p is given by $\mathbb{P}_{L_i}[P^{t_0} = p] = \frac{1}{Z}e^{-f_{L_i}(p)}$.*

Finally, by using the cumulative probability, a second result is derived:

Corollary 2 (Likelihood of reaching a minimum market price) *For a market price p at time instant t_0 ,*

- i) *The likelihood of reaching at least p is given by $\mathbb{P}[P^{t_0} \leq p] = \sum_{p_i \leq p} \mathbb{P}[P^{t_0} = p_i]$.*
- ii) *From the perspective of the i th-lobby, the likelihood of reaching at least p is given by*

$$\mathbb{P}_{L_i}[P^{t_0} \leq p] = \sum_{p_i \leq p} \mathbb{P}_{L_i}[P^{t_0} = p_i].$$

The fact that price outcomes are TD-MRFs provides extra-information: it is possible to weight each lobby. Actually, an instantaneous weight $w_{L_i}^t$ may be attached to each lobby L_i by computing the similarity index between the joint distribution of a given lobby and the distribution of the whole process: $w_{L_i}^t = \frac{\mathbb{P}_{L_i}[P^t]}{\mathbb{P}[P^t]}$

Let us assume that there are n lobbies in the market. Thus, the following proposition gives an explicit formula for computing the weight of the i -th lobby L_i :

Theorem 4 $w_{L_i}^t = K \cdot e^{f_{L_1}(P^t)} \dots e^{f_{L_{i-1}}(P^t)} e^{f_{L_{i+1}}(P^t)} \dots e^{f_{L_n}(P^t)}$ for a constant K .

Proof By substituting the explicit form of $\mathbb{P}[P^t]$ as a product of clique potentials in the definition of weight and

thus applying theorem 3, the weight of the i -th lobby is equal to

$$\begin{aligned}
 w_{L_i}^t &= \frac{\mathbb{P}_{L_i}[P^t]}{\mathbb{P}[P^t]} = \frac{\frac{1}{Z} e^{-f_{L_i}(P^t)}}{\frac{1}{Z} \prod_{i \in \mathbb{N}} e^{-f_{L_i}(P^t)}} = \\
 &= \frac{\frac{1}{Z} \left(e^{f_{L_i}(P^t)} \right)^{-1}}{\frac{1}{Z} \left(\prod_{i \in \mathbb{N}} e^{f_{L_i}(P^t)} \right)^{-1}} = \\
 &= \frac{\frac{1}{Z} \frac{1}{e^{f_{L_i}(P^t)}}}{\frac{1}{Z} \frac{1}{\left(\prod_{i \in \mathbb{N}} e^{f_{L_i}(P^t)} \right)}} = \\
 &= \frac{\frac{1}{Z} \left(\prod_{i \in \mathbb{N}} e^{f_{L_i}(P^t)} \right)}{\frac{1}{Z} e^{f_{L_i}(P^t)}} = \\
 &= K \cdot \frac{e^{f(P_{L_1}^t)} \dots e^{f(P_{L_{i-1}}^t)} e^{f(P_{L_i}^t)} e^{f(P_{L_{i+1}}^t)} \dots e^{f(P_{L_n}^t)}}{e^{f(P_{L_i}^t)}} = \\
 &= K \cdot \frac{e^{f(P_{L_1}^t)} \dots e^{f(P_{L_{i-1}}^t)} \cancel{e^{f(P_{L_i}^t)}} e^{f(P_{L_{i+1}}^t)} \dots e^{f(P_{L_n}^t)}}{\cancel{e^{f(P_{L_i}^t)}}} = \\
 &= K \cdot e^{f(P_{L_1}^t)} \dots e^{f(P_{L_{i-1}}^t)} e^{f(P_{L_{i+1}}^t)} \dots e^{f(P_{L_n}^t)},
 \end{aligned}$$

where, according to remark 1, the quotient of two normalizing constants (Z 's) is a constant (K). \square

In real life, it is necessary to compute the value of this weight as accurately as possible. To this end, the constant K shall be completely determined, which will be performed by following the instructions of the remark 1. For this, the variable price should be regarded as a discrete variable, as follows:

- first, a finite range of values for price should be considered, $[p_{min}, p_{max}]$.
- Second, exact approximations of the variable price within the range should be taken (this step reflects the normal real-life practice of rounding prices up or down thus avoiding the infinite decimal series).

Once price has been viewed as a discrete variable within a range of values, using the following expression considered in the corresponding sub-indices makes sense:

Proposition 2 *In the expression $w_{L_i}^t = K \cdot e^{f_{L_1}(P^t)} \dots e^{f_{L_{i-1}}(P^t)} e^{f_{L_{i+1}}(P^t)} \dots e^{f_{L_n}(P^t)}$, the constant K corresponding to the i th-lobby weight with energy functions f_{L_i} , is*

$$K = \frac{\sum_{p_j=p_{min}}^{p_{max}} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(p_j)}}{\sum_{p_j=p_{min}}^{p_{max}} e^{-f_{L_i}(p_j)}} = \sum_{p_j=p_{min}}^{p_{max}} \frac{e^{-\sum_{i \in \mathbb{N}} f_{L_i}(p_j)}}{\sum_{p_j=p_{min}}^{p_{max}} e^{-f_{L_i}(p_j)}}.$$

Proof The constant K is a quotient of the corresponding normalizing constants Z , as shown in the proof of theorem 4. The exact values for Z 's are provided by remark 1. \square

5 Analysis on energy functions/cliue potentials

The significance of theorem 1 as well as that of previous results is that they delimit a complex problem within the bounds of some probability computations as long as the energy functions f are specified. Since energy functions/cliue potentials have a crucial bearing on our model, we shall examine their significance in depth.

Previous theorems 1 and 3 provide explicit formulae on how to compute the likelihood of reaching an aggregate price for the whole market and the corresponding estimation from the perspective of the lobbies. In both cases, the corresponding joint probability distributions are transcribed in terms of energy functions $f_{L_i}(P^t)$ (it is important to remember that, when expressed as factors, energy functions $f_{L_i}(P^t)$ are known as cliue potentials $e^{-f_{L_i}(P^t)}$), which play a central role in computations, see Table 1. More importantly, they strongly determine the performance of the whole probability distributions.

We will focus here on the construction of the energy functions. It is very important to make it clear that, to the best of our knowledge, there is no information in the literature about energy functions, nor on how to define them

Table 1 Energy functions and cliue potentials

	Price outcomes	Lobbies
In terms of energy functions	$\mathbb{P}[P^t] = \frac{1}{Z} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)}$	$\mathbb{P}_{L_i}[P^t] = \frac{1}{Z} e^{-f_{L_i}(P^t)}$
In terms of cliue potentials	$\mathbb{P}[P^t] = \frac{1}{Z} \prod_{i \in \mathbb{N}} e^{-f_{L_i}(P^t)}$	$\mathbb{P}_{L_i}[P^t] = \frac{1}{Z} \left(e^{f_{L_i}(P^t)} \right)^{-1}$

Table 2 Most commonly used energy functions

Gaussian	Log Linear
$f_{L_i}(P^t) = \frac{(P^t - \mu_i)^2}{\sigma_i^2}$	$f_{L_i}(P^t) = \exp(w_0 + \sum_k w_{ik} P_{ik}^t)$

nor why some have been chosen and others have not. In any case, the specific clique functionals are given without providing any further explanations. Therefore it is worth paying attention here to how they are established and which appear most frequently in literature.

On how they can be defined, the functionals f can be arbitrarily taken as long as they meet the “energy constrain” which sets that the energy functions must be decreasing functions as for the matching of the features of the cliques with a given template. That is, if the features in a clique match the features in a given template, the energy function should decrease. Otherwise it should increase. Bearing this in mind, assuming the clique template is 1, a routine for constructing energy functions f on a clique C could be $f(c) = 1 - d(c)$, $c \in C$, where d is a distance methodology which measures the similarities between the features of the clique and those of a given template. SSD (Sum of Squared Differences) is a further example of functions that meet the idea of measuring the distance from a given template. This pattern for defining energy functions is so wide, that it allows many options. The energy functions that appear most frequently in literature are Gaussian and log linear distributions (Table 2):

where μ_i and σ_i^2 are the mean and the variance of the i -th lobby respectively. It is well-known that the log linear functional takes its name from the fact that

$$f_{L_i}(P^t) = \exp(w_0 + \sum_k w_{ik} P_{ik}^t) \Leftrightarrow \log(f_{L_i}(P^t)) = w_0 + \sum_k w_{ik} P_{ik}^t \text{ (i.e., its log is linear).}$$

6 Model application: The olive oil sector in Andalusia

The olive, the core product in the olive oil industry, is classified under Agribusiness Classification Terms as a non-perishable and long-life agricultural product. However, olive oil can also be stored for long periods without its main qualities suffering (there are advanced techniques to preserve it from deterioration -perishing-⁶ in long periods

⁶This refers to the acidic taste which appears in olive oil after long periods of storage.

of storage). This quality, shared by both the olive and the olive oil, makes it more difficult to accurately foresee the price of olive oil since, as a result of a widespread practice in the sector, part of the oil consignments (which are stored waiting for the optimal moment for sale) go to market depending on unexpected decisions that private individuals (producers/olive oil industry and even distributors) make based solely on their own particular interests. This reality is just one example of the complexity of the olive oil sector, see [9], [22].

The aim of this section is to apply previous results to the olive oil market. This is an extremely complex sector and, therefore, it is especially important to develop price prediction techniques which provide inside information that can help to achieve an advantageous position in the price negotiation processes which take place in the olive oil value chain (e.g., between producers and the oil industry, between the oil industry and distributors and even directly between producers and distributors). Our proposal presents an encouraging solution for such a challenging problem. Specifically, our proposal succeeds in developing effective price prediction techniques which reduce the problem to straightforward probability computations.

Section 6.1 is devoted to the translation of the theoretical TD-MRF model to the context of olive oil. In Section 6.2, energy functions for the olive oil sector are analyzed. This is a necessary intermediate step considering that (as mentioned before) there is no detailed information in the literature about how to define/select the energy functions. Section 6.3 contains an example of how to use the TD-MRF model designed to produce price outcome predictions in real-life (model validation with real data) and finally, a complementary study from the perspective of thermodynamics is provided in Section 6.4.

6.1 Applying the TD-MRF model to the olive oil market

The term “olive oil lobby” includes all the agents in the agricultural value chain, from producers to distributors. Figure 2 displays graphically the olive oil value chain as a linear structure which consists of three categories in which agents of different nature participate with a high degree of specialization, from the first level (producers) to the third one (distributors). It allows also to visualize that the main categories are divided into sub-categories (e.g., the olive oil industry is formed by oil mills and bottling plants). In closer detail, we can see from Fig. 2 that the different pressure groups that appear in this sector make up the following panorama: let us use L_{Pr} , L_{OOI} and L_D for the lobbies corresponding to producers, the olive oil industry and distributors respectively with weights that reflect their

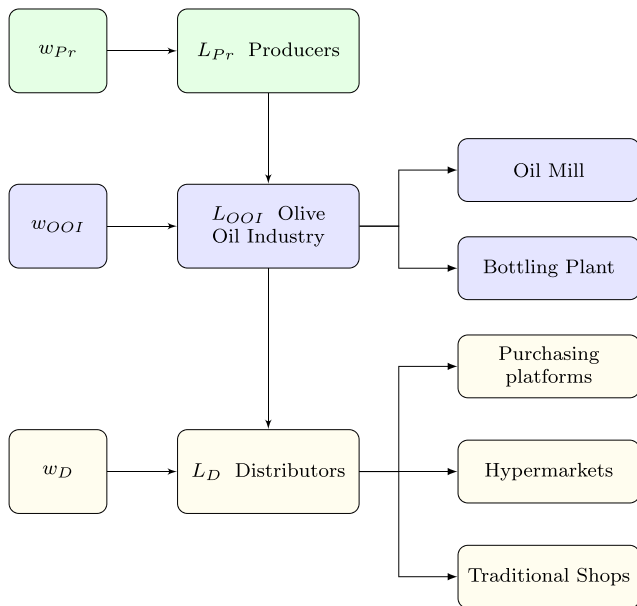


Fig. 2 Olive oil value chain: lobbies and their corresponding weights

importance within the value chain, w_{Pr} , w_{OOI} and w_D respectively.

Another factor of complexity in this sector that prevents the accurate calibration of the weight of each of the pressure groups is that each lobby uses different weapons to further their interests. For instance, agricultural lobbies, at the bottom of the value chain, apply pressure with public displays of discontent whereas other olive oil lobbies apply pressure more privately. For reasons of public well-being, the protests staged by these types of lobbies (agricultural producers) enjoy public support in contrast to what happens with other types of lobbying. Additionally, the power of intervention held by both oil producers and the oil industry, is evidenced in every process present in the reform of the EU’s agricultural policy (see [2]).

Now, we are going to apply the TD-MRF model and the results obtained in previous sections, to the olive oil market. From previous results, the olive oil price outcomes can be viewed as a dynamic graphical model (see Fig. 1) where the neighbourhoods are the lobbies:

Proposition 3 *The neighbourhood of any agent a_i is its lobby: $N(a_i) = L_{Pr}$ or L_{OOI} or L_D .*

Moreover, our model provides the joint probability distribution of each lobby, thereby expressing the expectation of each lobby with regard to the price of olive oil.

Theorem 5 *Each of the three lobbies L_{Pr} , L_{OOI} , L_D considered to be a DGM, is a TD-MRF. Thus, the corresponding joint distribution function which gives the*

likelihood of reaching a price from the perspective of each lobby is

$$\begin{aligned} \mathbb{P}_{L_{Pr}}[P^t] &= \frac{1}{Z} e^{-f_{L_{Pr}}(P^t)}, \quad \mathbb{P}_{L_{OOI}}[P^t] \\ &= \frac{1}{Z} e^{-f_{L_{OOI}}(P^t)} \quad \text{and} \quad \mathbb{P}_{L_D}[P^t] = \frac{1}{Z} e^{-f_{L_D}(P^t)}, \end{aligned}$$

where Z denotes a normalizing constant and the f ’s are the corresponding energy functions of each lobby.

Concrete energy functions for lobbies shall be studied in short (see next subsection). Additionally, the likelihood of reaching an aggregate olive oil price (corresponding to the price outcomes) is stated by the following theorem:

Theorem 6 (The olive oil price outcomes are a TD-MRF) *The olive oil price outcomes, viewed as a DGM, are a TD-MRF. Its corresponding joint distribution $\mathbb{P}[P^t]$ is equal to*

$$\begin{aligned} \mathbb{P}[P^t] &= \frac{1}{Z} e^{-[(f_{L_{Pr}}(P^t)) + f_{L_{OOI}}(P^t) + f_{L_D}(P^t)]} = \mathbb{P}[P^t] \\ &= \frac{1}{Z} e^{-f_{L_{Pr}}(P^t)} e^{-f_{L_{OOI}}(P^t)} e^{-f_{L_D}(P^t)} \end{aligned}$$

as the product of the expectations of lobbies with regard to the olive oil price up to a constant.

The importance of each stakeholder involved in the olive oil price outcomes shall be detailed here from a quantitative perspective using of the specific weigh w . Although in general terms, the olive oil industry has more weight than the rest of the agents (the olive oil industry has a greater ability to influence the institutional course of actions than other agents) the producers enjoy implicit public support to further increase their specific weight. As an application of theoretical developments of Sections 3 and 4, a precise form of calculating these specific weights w is provided by theorem 4 (together with proposition 4) which, in short, states that a weight may be defined for each lobby L_{Pr} , L_{OOI} , L_D by comparing its joint distribution to the distribution of the price outcomes:

$$w_{L_{Pr}}^t = \frac{\mathbb{P}_{L_{Pr}}[P^t]}{\mathbb{P}[P^t]}, \quad w_{L_{OOI}}^t = \frac{\mathbb{P}_{L_{OOI}}[P^t]}{\mathbb{P}[P^t]}, \quad w_{L_D}^t = \frac{\mathbb{P}_{L_D}[P^t]}{\mathbb{P}[P^t]}. \quad (3)$$

These properties, derived from the DGM model, have been expressed in terms of a Gibbs distribution but they can be expressed in terms of a slightly more general function: the Boltzmann distribution, which is a key tool in thermodynamics. Hence, in order to reach more conclusions about the price outcomes in the presence of lobbying, we will proceed by studying this considering thermodynamics where notions and results shall be properly applied to our scenario with a specific economic meaning. This analysis will be carried out in Section 6.3, after the appropriate

energy functions for the influence groups of the oil sector are studied in depth, in Section 6.2.

6.2 Energy functions in the study of the olive oil

Gaussian distribution describes those data sets in which most values cluster around the centre of the range. Compared with the previously described pattern that energy functions should follow, Gaussians measure the distance between price and average price (the template). These are suitable for those products whose price takes values within a small range, in an interval $[p_{min}, p_{max}]$ of short length $p_{max} - p_{min}$. That is, products whose price takes approximately the same values. Further, a Gaussian

— EVOO — VOO — LOO — ROO

In Southern Spain (Andalusia), one of the fundamental pillars of the Andalusian economy is the agri-food sector (olive oil and its derivatives, wine products, meat, fruits and vegetables and seafood) which accounts for 10% of the Andalusian Gross Domestic Product (GDP) by contributing 10% of the total labour force. It is not surprising that olive oil prices are periodically displayed on official platforms of the Government of Spain. Under normal economic and environmental conditions, olive oil in all its varieties is an example of product whose price has very few fluctuations as shown in Fig. 2.

Another argument that supports the choice of Gaussians as energy functions and not the log linear ones, is that, despite the fact that several interconnected factors affect the price of all the varieties of olive oil, in a simplification of the model made by experts in olive oil production, it is only the weather conditions that are considered to be really determining factors: the water index and to a lesser degree, temperature. Hence, Gaussian distribution shall be used (as energy functions) for predicting the prices from the perspective of the lobbies and, consequently, for forecasting the aggregate price.

These Gaussian functions will be different for each lobby, that is, with a mean and a specific variance for each one. In these types of distributions, in general the standard deviation represents the degree of deviation with regard to the mean (the higher the standard deviation is, the further from the mean it is and thus, the more spread out it is). In the context we are using (with a mean and a specific variance for each lobby) it represents how much the olive oil price deviates from the expected mean in the opinion of the lobby. These opinions about the price given by the lobbies are often a combination of economic reality (the price they

distribution fits well with those products whose price suffers from small, smooth variations. Meanwhile, the *log-linear distribution* is used to manage output-contexts where an analysis of the simultaneous effects of several inputs is needed. This could be useful for describing the associations and interrelationships amongst numerous inputs and, therefore, also useful for the price of those products which depends on many interrelated factors.

The olive oil varieties most commonly used for human consumption are (in decreasing order of quality) extra virgin olive oil EVOO, virgin olive oil VOO, lampante olive oil LOO and refined olive oil ROO. They are categorized according to their acidity levels and other characteristics which are regulated by the corresponding competent body:

actually expect to reach under the economic circumstances) and political aspirations (the target price for which they push).

As mentioned, there are various platforms that reflect the variations in the price of olive oil varieties. In Andalusia (Southern Spain), one of the most popular amongst olive producers and oil producers is PoolRed (see <http://www.poolred.com/>) which is used to predict the right moment (that is, when prices are rising) to sell the harvests to the oil companies. Both olive and oil producers use this platform to try to answer the following two questions, which are key when conducting price negotiation processes:

A “which should be the *aggregate* market price of olive oil in the following campaign”

B “which should be the *minimum* market price of olive oil in the following campaign”

Thus, Table 3 provides an overview for computing the price outcome prediction by using the TD-MRF model.

6.3 An example of price outcome prediction by using the TD-MRF model

The aim of this section is to develop an example in order to illustrate how our model of price prediction works. For this, data have been extracted from <https://www.olimerca.com/precios/tipoInforme/3> in a time period from 2017 to 2020 consisting of the monthly average price of certain variety of olive oil (extra virgin EVOO) for 7 major specialist retailers/hypermarkets. Thus, we assume that there are 7 lobbies $L_1, L_2, L_3, L_4, L_5, L_6, L_7$ corresponding to these 7 big stores with data displayed in Table 4:

Table 3 Aggregate and minimum market and lobby's price in the Gaussian case

	Aggregate price p_0 at time t_0	Minimum price p_0 at time t_0
Price outcomes	$\mathbb{P}[P^{t_0} = p] = \frac{1}{Z} \prod_{i \in \mathbb{N}} \left(e^{-\frac{(p - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$	$\mathbb{P}[P^{t_0} \leq p] = \frac{1}{Z} \sum_{p_j \leq p} \prod_{i \in \mathbb{N}} \left(e^{-\frac{(p_j - \mu_i)^2}{\sigma_i^2}} \right)^{-1} =$ $= \frac{1}{Z} \prod_{i \in \mathbb{N}} \left(e^{-\frac{(p_1 - \mu_i)^2}{\sigma_i^2}} \right)^{-1} + \dots + \frac{1}{Z} \prod_{i \in \mathbb{N}} \left(e^{-\frac{(p_r - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$ <p>for prices $p_1, \dots, p_r \leq p$</p>
Lobbies	$\mathbb{P}_{L_i}[P^{t_0} = p] = \frac{1}{Z} \left(e^{-\frac{(p - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$	$\mathbb{P}_{L_i}[P^{t_0} \leq p] = \frac{1}{Z} \sum_{p_j \leq p} \left(e^{-\frac{(p_j - \mu_i)^2}{\sigma_i^2}} \right)^{-1} =$ $= \frac{1}{Z} \left(e^{-\frac{(p_1 - \mu_i)^2}{\sigma_i^2}} \right)^{-1} + \dots + \frac{1}{Z} \left(e^{-\frac{(p_r - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$ <p>for prices $p_1, \dots, p_r \leq p$</p>

On one hand, the probability distribution for each lobby L_i is given by

$$\mathbb{P}_{L_i}[P^{t_0} = p] = \frac{1}{Z} \left(e^{-\frac{(p - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$$

with p representing a certain price at a time instant t_0 . We know that the probability distribution \mathbb{P}_{L_i} provides the price prediction from the perspective of the lobbies and this prediction is based on the mean μ_i and the variance σ_i^2 , which are specific for each lobby. These parameters can be derived from Table 4.

On the other hand, the price of each variety of olive oil oscillates within a specific range (see Fig. 3). From the perspective of each lobby, there are minimum and maximum thresholds for prices as part of the price wars between the supermarkets. These are p_i^{min} and p_i^{Max} . In order to reach a common domain, let $\underline{p} = \inf\{p_i^{min} | i = 1, \dots, 7\}$ and $\bar{p} = \sup\{p_i^{Max} | i = 1, \dots, 7\}$ be the infimum and supremum respectively so that $\underline{p} \leq p \leq \bar{p}$. For each lobby L_i , the probability of EVOO price reaching values p which exceed the upper threshold, $p \in [\bar{p} - p_i^{Max}, \bar{p}]$ must be close to zero, as actually happens when computed by using Gaussian energy functions since the further the deviation, the greater the exponent and the lower the quotient (which is the probability). The mean and variance of each lobby as well as other related parameters are displayed in Table 5:

In order to carry out the necessary computations, let us apply some properties of Gaussian functions to the joint

probability whose output are the minimum price p_0 at time t_0 :

$$\mathbb{P}[P^{t_0} \leq p] = \frac{1}{Z} \sum_{p_j \leq p} \prod_{i \in \mathbb{N}} \left(e^{-\frac{(p_j - \mu_i)^2}{\sigma_i^2}} \right)^{-1}$$

$$= \frac{1}{Z} \sum_{p_j \leq p} e^{-\sum_{i \in \mathbb{N}} \frac{(p_j - \mu_i)^2}{\sigma_i^2}} = \frac{1}{Z} \sum_{p_j \leq p} e^{-\frac{(p_j - \mu)^2}{\sigma^2}},$$

for $\mu = \sum_{i=1}^n \mu_i, \sigma^2 = \sum_{i=1}^n \sigma_i^2$. This property can also be applied to the partition function Z , whose explicit expression was given at remark 1 as $Z = \sum_{\text{prices } p_j} e^{-\sum_{i \in \mathbb{N}} f_{L_i}(P^t)}$. According to the selected energy functions $f_{L_i}(P^t)$, Z is

$$Z = \sum_{\text{prices } p_j} e^{-\sum_{i \in \mathbb{N}} \frac{(p_j - \mu_i)^2}{\sigma_i^2}} = \sum_{\text{prices } p_j} e^{-\frac{(p_j - \mu)^2}{\sigma^2}},$$

$$\mu = \sum_{i=1}^n \mu_i, \sigma^2 = \sum_{i=1}^n \sigma_i^2$$

The calculation of Z entails certain difficulties in practice. This example is intended to be illustrative, in addition to the use of our TD-MRF model prediction, of how to proceed in the calculation of the partition function.

Under certain circumstances, Z can be considered equal to 1. For this, let us remind the reproductive property (see [14]):

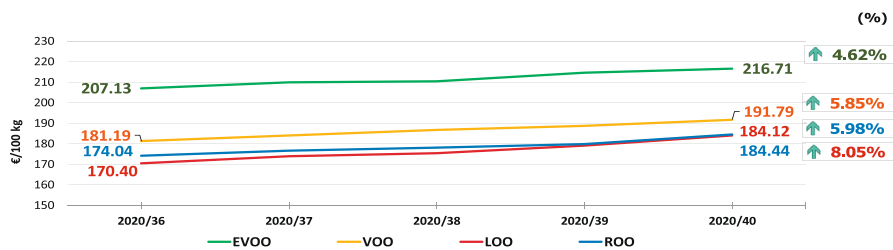
Table 4 Monthly average price of EVOO for specialist major retailers

Date	Ahorramas	Alcampo	Carrefour	Dia	Hipercor	Lidl	Mercadona
Jan17	3.99	4.29	3.99	4.25	4.99	3.99	3.99
Feb17	3.99	4.29	3.99	4.25	4.99	3.99	3.99
Mar17	3.99	4.24	3.95	4.25	4.75	3.99	3.99
Apr17	4.35	4.34	4.35	4.25	5.49	4.35	4.35
May17	4.35	4.34	4.35	4.25	5.49	4.35	4.35
Jun17		4.34	4.35	4.35	5.49	4.35	4.35
Jul17	4.5	4.34	4.35	4.35	5.49	4.49	4.5
Aug17		4.34	3.99	4.35	4.59	4.49	4.35
Sep17	4.35	4.29	4.3	4.35		4.35	4.35
Oct17	4.35	4.29	4.3	4.35	5.59	4.49	4.35
Nov17	4.35	4.29	3.99	4.35	5.59	4.29	4.35
Dic17	3.99	4.29	3.99	4.35	4.69	4.29	4.35
Jan18	3.99	3.98	4.3	4.35	4.69	4.29	4.35
Feb18	4.35	4.29	4.3	4.35	4.29		4.35
Mar18	4.35	4.29	4.3	4.35	5.59	4.29	4.35
Apr18	4.35	4.49	4.3	4.35	4.69	4.29	4.35
May18	4.35	4.24	4.05	4.35	4.95	4.25	4.1
Jun18	4.35	4.24	4.24	3.99	4.59		4.1
Jul18	3.9	4.24	3.85	3.99	3.75	3.89	3.9
Augo18	4.1	4.24	3.85	3.99	3.75	3.89	3.9
Sep18	3.9	3.84	3.85	3.99	4.09	4.09	3.9
Oct18	3.9	4.24	3.85	3.99	4.09	4.09	3.9
Nov18	3.9	4.2	3.85	3.99	4.09	4.09	3.9
Dic18	3.9	3.84	3.59	3.99	4.09	3.89	3.9
Jan19	3.9	4.24	3.7	3.9	4.09	3.89	3.75
Feb19	3.75	4.24	3.7	3.99	4.09	3.75	3.75
Mar19	3.75	3.74	3.39	3.75	3.69	3.69	3.75
Apr19	3.75	3.74	3.7	3.75	3.69	3.69	3.75
May19	3.59	3.74	3.55	3.75	3.69	3.69	3.59
Jun19	2.99	3.54	3.3	3.59	3.69	3.69	3.59
Jul19	3.59	3.54	3.55	3.59	3.89	3.59	3.59
Aug19	3.59	2.98	3.55	3.59	3.75	3.59	3.59
Sep19	3.59	3.54	3.55	3.59	3.75	3.59	3.59
Oct19	3.59	3.54	3.23	3.69	3.75	3.59	3.59
Nov19	3.59	3.54	3.15	3.59	3.75	3.55	3.59
Dic19	2.99	3.54	3.45	3.59	3.75	3.49	3.5
Jan20	3.5	3.54	3.45	3.49	3.75	3.49	3.5
Feb20	3.5	3.54	3.45	3.49	3.75	3.45	3.5
Mar20	3.5	3.54	3.45	3.5	3.75	3.45	3.5
Apr20	3.5	3.54	3.45	3.5	3.75	3.49	3.5
May20	3.5	3.19	3.45	3.5	3.71	3.49	3.5
Jun20	2.99	3.44	3.45	3.99	3.5	3.49	3.5
Jul20	3.5	3.44	3.45	3.5	3.75	3.49	3.5

Proposition 4 *If two independent random variables which follow a certain distribution are added together, the resulting random variable has a distribution of the same type as that of the summands.*

By applying this property to Z itself, we can conclude that Z has an exponential distribution, $Z = e^{-\lambda}$ for some parameter λ . Therefore, Z can be considered equal to 1 when $\lambda = 0$.

Fig. 3 Olive oil price in Spanish market. Source: Government of Spain, at https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/aceite-oliva-y-aceituna-mesa/Evolucion_precios_AO_vegetales.aspx



Source: MINISTERIO DE AGRICULTURA, PESCA Y ALIMENTACIÓN
 Dirección General de Producciones y Mercados Agrarios
 Subdirección General de Cultivos Herbáceos e Industriales y Aceite de Oliva

In this example however, we will consider that this is not the case and a detailed calculation of Z shall be performed.

The necessary information for computing $\mathbb{P}[P^{t_0} \leq p] = \frac{1}{Z} \sum_{p_j \leq p} e^{-\frac{(p_j - \mu)^2}{\sigma^2}}$ is given in Table 6:

According to the values $\underline{p} = 2.98$ and $\bar{p} = 5.59$, let us consider the sample $S = \{2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7\}$. Thus, the probability of reaching a minimum market price $p = 5.7$ is given by the formula

$$\mathbb{P}[P^{t_0} \leq p] = \frac{1}{Z} \sum_{p_j \leq p} e^{-\frac{(p_j - \mu)^2}{\sigma^2}} = \frac{1}{Z} \sum_{p_j \in S} e^{-\frac{(p_j - 27.718)^2}{1.302}} =$$

$$= \frac{1}{Z} \left(e^{-\frac{(2.9 - 27.718)^2}{1.302}} + e^{-\frac{(3 - 27.718)^2}{1.302}} + e^{-\frac{(3.1 - 27.718)^2}{1.302}} + e^{-\frac{(3.2 - 27.718)^2}{1.302}} + e^{-\frac{(3.3 - 27.718)^2}{1.302}} + \right.$$

$$+ e^{-\frac{(3.4 - 27.718)^2}{1.302}} + e^{-\frac{(3.5 - 27.718)^2}{1.302}} + e^{-\frac{(3.6 - 27.718)^2}{1.302}} + e^{-\frac{(3.7 - 27.718)^2}{1.302}} + e^{-\frac{(3.8 - 27.718)^2}{1.302}} +$$

$$+ e^{-\frac{(3.9 - 27.718)^2}{1.302}} + e^{-\frac{(4 - 27.718)^2}{1.302}} + e^{-\frac{(4.1 - 27.718)^2}{1.302}} + e^{-\frac{(4.2 - 27.718)^2}{1.302}} + e^{-\frac{(4.3 - 27.718)^2}{1.302}} +$$

$$+ e^{-\frac{(4.4 - 27.718)^2}{1.302}} + e^{-\frac{(4.5 - 27.718)^2}{1.302}} + e^{-\frac{(4.6 - 27.718)^2}{1.302}} + e^{-\frac{(4.7 - 27.718)^2}{1.302}} + e^{-\frac{(4.8 - 27.718)^2}{1.302}} +$$

$$+ e^{-\frac{(4.9 - 27.718)^2}{1.302}} + e^{-\frac{(5 - 27.718)^2}{1.302}} + e^{-\frac{(5.1 - 27.718)^2}{1.302}} + e^{-\frac{(5.2 - 27.718)^2}{1.302}} + e^{-\frac{(5.3 - 27.718)^2}{1.302}} +$$

$$+ e^{-\frac{(5.4 - 27.718)^2}{1.302}} + e^{-\frac{(5.5 - 27.718)^2}{1.302}} + e^{-\frac{(5.6 - 27.718)^2}{1.302}} + e^{-\frac{(5.7 - 27.718)^2}{1.302}} \left. \right) = \frac{1}{Z} 2.0314E - 162,$$

supported by the information given in Table 7:

Table 5 Mean and variance for each lobby

Mean, Variance	L_1 Ahorramas	L_2 Alcampo	L_3 Carrefour	L_4 Dia	L_5 Hipercor	L_6 Lidl	L_7 Mercadona	\underline{p}	\bar{p}
μ_i	3.988	3.884	3.851	3.858	4.343	3.878	3.916		
σ_i^2	0.172	0.156	0.143	0.108	0.472	0.134	0.117		
p_i^{min}	2.99	2.98	3.15	3.49	3.5	3.45	3.5	2.98	
p_i^{Max}	4.5	4.49	4.35	4.35	5.59	4.49	4.5		5.59

Table 6 Mean and variance for each lobby

Mean, Variance	L_1	L_2	L_3	L_4	L_5	L_6	L_7	p	\bar{p}	$\mu = \sum_{i=1}^n \mu_i$	$\sigma^2 = \sum_{i=1}^n \sigma_i^2$
μ_i	3.988	3.884	3.851	3.858	4.343	3.878	3.916			27.718	
σ_i^2	0.172	0.156	0.143	0.108	0.472	0.134	0.117				1.302
p_i^{min}	2.99	2.98	3.15	3.49	3.5	3.45	3.5	2.98			
p_i^{Max}	4.5	4.49	4.35	4.35	5.59	4.49	4.5		5.59		

Similar computations must be accomplished in order to determine the “partition function” Z , for which the main difference is the range of prices p_j considered. In general terms, *all* possible values for the EVOO price should be

considered ($p_j \geq 0$). To be more realistic, by using a record of market EVOO price data, the range could be further scoped: in this way, we shall consider that $2 \leq p_j \leq 6.5$. Thus, Table 8 gathers all the information required

Table 7 Calculation of the numerator of $\mathbb{P}[P^{t_0} \leq p]$ for $p = 5.7$

p_j	$(p_j - 27.718)$	$(p_j - 27.718)^2$	$\frac{(p_j - 27.718)^2}{1.302}$	$-\frac{(p_j - 27.718)^2}{1.302}$	$e^{-\frac{(p_j - 27.718)^2}{1.302}}$
2.9	-24.818	615.933124	473.0669155	-473.0669155	3.5453E-206
3	-24.718	610.979524	469.2623072	-469.2623072	1.5921E-204
3.1	-24.618	606.045924	465.4730599	-465.4730599	7.0407E-203
3.2	-24.518	601.132324	461.6991736	-461.6991736	3.0662E-201
3.3	-24.418	596.238724	457.9406482	-457.9406482	1.3149E-199
3.4	-24.318	591.365124	454.1974839	-454.1974839	5.5531E-198
3.5	-24.218	586.511524	450.4696805	-450.4696805	2.3094E-196
3.6	-24.118	581.677924	446.7572381	-446.7572381	9.4579E-195
3.7	-24.018	576.864324	443.0601567	-443.0601567	3.8143E-193
3.8	-23.918	572.070724	439.3784363	-439.3784363	1.5149E-191
3.9	-23.818	567.297124	435.7120768	-435.7120768	5.9245E-190
4	-23.718	562.543524	432.0610783	-432.0610783	2.2817E-188
4.1	-23.618	557.809924	428.4254409	-428.4254409	8.6536E-187
4.2	-23.518	553.096324	424.8051644	-424.8051644	3.2319E-185
4.3	-23.418	548.402724	421.2002488	-421.2002488	1.1887E-183
4.4	-23.318	543.729124	417.6106943	-417.6106943	4.3051E-182
4.5	-23.218	539.075524	414.0365008	-414.0365008	1.5354E-180
4.6	-23.118	534.441924	410.4776682	-410.4776682	5.3928E-179
4.7	-23.018	529.828324	406.9341966	-406.9341966	1.8652E-177
4.8	-22.918	525.234724	403.406086	-403.406086	6.3528E-176
4.9	-22.818	520.661124	399.8933364	-399.8933364	2.1307E-174
5	-22.718	516.107524	396.3959478	-396.3959478	7.0376E-173
5.1	-22.618	511.573924	392.9139201	-392.9139201	2.289E-171
5.2	-22.518	507.060324	389.4472535	-389.4472535	7.3317E-170
5.3	-22.418	502.566724	385.9959478	-385.9959478	2.3125E-168
5.4	-22.318	498.093124	382.5600031	-382.5600031	7.1829E-167
5.5	-22.218	493.639524	379.1394194	-379.1394194	2.1971E-165
5.6	-22.118	489.205924	375.7341966	-375.7341966	6.6178E-164
5.7	-22.018	484.792324	372.3443349	-372.3443349	1.963E-162
					2.0314E-162

Table 8 Computing Z

p_j	$(p_j - 27.718)$	$(p_j - 27.718)^2$	$\frac{(p_j - 27.718)^2}{1.302}$	$-\frac{(p_j - 27.718)^2}{1.302}$	$e^{-\frac{(p_j - 27.718)^2}{1.302}}$
2	-25.718	661.415524	507.9996344	-507.9996344	2.3909E-221
2.1	-25.618	656.281924	504.0567773	-504.0567773	1.2329E-219
2.2	-25.518	651.168324	500.1292811	-500.1292811	6.2606E-218
2.3	-25.418	646.074724	496.2171459	-496.2171459	3.1306E-216
2.4	-25.318	641.001124	492.3203717	-492.3203717	1.5416E-214
2.5	-25.218	635.947524	488.4389585	-488.4389585	7.4758E-213
2.6	-25.118	630.913924	484.5729063	-484.5729063	3.5699E-211
2.7	-25.018	625.900324	480.7222151	-480.7222151	1.6788E-209
2.8	-24.918	620.906724	476.8868848	-476.8868848	7.7742E-208
2.9	-24.818	615.933124	473.0669155	-473.0669155	3.5453E-206
3	-24.718	610.979524	469.2623072	-469.2623072	1.5921E-204
3.1	-24.618	606.045924	465.4730599	-465.4730599	7.0407E-203
3.2	-24.518	601.132324	461.6991736	-461.6991736	3.0662E-201
3.3	-24.418	596.238724	457.9406482	-457.9406482	1.3149E-199
3.4	-24.318	591.365124	454.1974839	-454.1974839	5.5531E-198
3.5	-24.218	586.511524	450.4696805	-450.4696805	2.3094E-196
3.6	-24.118	581.677924	446.7572381	-446.7572381	9.4579E-195
3.7	-24.018	576.864324	443.0601567	-443.0601567	3.8143E-193
3.8	-23.918	572.070724	439.3784363	-439.3784363	1.5149E-191
3.9	-23.818	567.297124	435.7120768	-435.7120768	5.9245E-190
4	-23.718	562.543524	432.0610783	-432.0610783	2.2817E-188
4.1	-23.618	557.809924	428.4254409	-428.4254409	8.6536E-187
4.2	-23.518	553.096324	424.8051644	-424.8051644	3.2319E-185
4.3	-23.418	548.402724	421.2002488	-421.2002488	1.1887E-183
4.4	-23.318	543.729124	417.6106943	-417.6106943	4.3051E-182
4.5	-23.218	539.075524	414.0365008	-414.0365008	1.5354E-180
4.6	-23.118	534.441924	410.4776682	-410.4776682	5.3928E-179
4.7	-23.018	529.828324	406.9341966	-406.9341966	1.8652E-177
4.8	-22.918	525.234724	403.406086	-403.406086	6.3528E-176
4.9	-22.818	520.661124	399.8933364	-399.8933364	2.1307E-174
5	-22.718	516.107524	396.3959478	-396.3959478	7.0376E-173
5.1	-22.618	511.573924	392.9139201	-392.9139201	2.289E-171
5.2	-22.518	507.060324	389.4472535	-389.4472535	7.3317E-170
5.3	-22.418	502.566724	385.9959478	-385.9959478	2.3125E-168
5.4	-22.318	498.093124	382.5600031	-382.5600031	7.1829E-167
5.5	-22.218	493.639524	379.1394194	-379.1394194	2.1971E-165
5.6	-22.118	489.205924	375.7341966	-375.7341966	6.6178E-164
5.7	-22.018	484.792324	372.3443349	-372.3443349	1.963E-162
5.8	-21.918	480.398724	368.9698341	-368.9698341	5.7337E-161
5.9	-21.818	476.025124	365.6106943	-365.6106943	1.6493E-159
6	-21.718	471.671524	362.2669155	-362.2669155	4.6717E-158
6.1	-21.618	467.337924	358.9384977	-358.9384977	1.3031E-156
6.2	-21.518	463.024324	355.6254409	-355.6254409	3.5796E-155
6.3	-21.418	458.730724	352.327745	-352.327745	9.6829E-154
6.4	-21.318	454.457124	349.0454101	-349.0454101	2.5793E-152
6.5	-21.218	450.203524	345.7784363	-345.7784363	6.766E-151
					7.034E-151

for the computation of $Z = \sum_{\text{prices } p_j} e^{-\frac{(p_j - \mu)^2}{\sigma^2}} = \sum_{2 \leq p_j \leq 6.5} e^{-\frac{(p_j - 27.718)^2}{1.302}}$. Hence, $\mathbb{P}[P^{t_0} \leq p] = \frac{1}{Z} 2.0314E - 162 = \frac{2.0314E - 162}{7.034E - 151} = 2.88797E - 12$.

6.4 Analysis from thermoeconomics

Econophysics is a novel discipline (from the physicist E. Stanley in 1996) which interrelates economics and physics by applying statistical and mathematical techniques, particularly from thermodynamics (with methodologies around the Boltzmann-Gibbs distributions amongst others, see (Dimitrijevic 2015) . This mixture of principles and methods is called *thermoeconomics*, where physical standards like energy and temperature find an economic analogy. Actually, from this perspective, thermodynamic and economic processes run in parallel. This section is divided into two main parts. The first one is devoted to key notions and results from thermodynamics. The second part is an application of thermodynamic principles to our proposal of price outcomes from which significant market properties shall be derived.

6.4.1 Key notions from thermodynamics

The concepts and results in this paragraph may be found in [3] for instance.

- *The partition function Z.* First, we pay attention to the normalizing constant Z whose basic role has so far been to ensure that the probability distribution sums to 1. In thermodynamics, Z is called the “partition function”. Far from simply being a constant (we shall see in short how a *constant* can also be a *function*), the analysis of Z shall allow us to reach conclusions about some properties of our economic process.

Before giving the definition of Z , we must refer to the Boltzmann distribution, see [3]. While the most common form of joint probability distribution in MRF contexts is the Gibbs distribution (which has been used in this paper up until now) the one used in thermodynamics is called the Boltzmann distribution and it is simply a Gibbs distribution depending on one parameter β with $\beta = \frac{1}{k_\beta T}$ where k_β is the Boltzmann constant and T is the temperature. In most contexts, β is considered to be a Lagrange multiplier. Hereinafter, the Boltzmann constant k_β will be taken as being equal to 1 and β shall be referred to as the inverse temperature.

According to this, the Boltzmann distribution of olive oil price outcomes is

$$\mathbb{P}[P^t] = \frac{1}{Z} e^{-\beta[(f_{LPr}(P^t) + f_{LOOl}(P^t) + f_{LD}(P^t))]}$$

for β the inverse of temperature, $\beta = \frac{1}{T}$,

and that of lobbies is

$$\begin{aligned} \mathbb{P}_{LPr}[P] &= \frac{1}{Z} e^{-\beta f_{LPr}(P^t)}, \quad \mathbb{P}_{LOOl}[P] \\ &= \frac{1}{Z} e^{-\beta f_{LOOl}(P^t)}, \quad \mathbb{P}_{LD}[P] = \frac{1}{Z} e^{-\beta f_{LD}(P^t)}. \end{aligned}$$

In general, the partition function is defined as $Z(\beta) = \sum e^{-\beta H(x_1, x_2, \dots)}$ where $H(x_1, x_2, \dots)$ is the Hamiltonian operator associated to random variables x_i (i.e., it is the Legendre transformation of the corresponding Lagrangian operator). In the context of our work, the normalizing constant Z was explicitly isolated in remark 1. In this way, the partition function associated to a Boltzmann distribution is

$$Z = e^{-\beta f_{LPr}(P^t)} + e^{-\beta f_{LOOl}(P^t)} + e^{-\beta f_{LD}(P^t)},$$

or equivalently, in its log-form, as

$$\log Z = \log \left(e^{-\beta f_{LPr}(P^t)} + e^{-\beta f_{LOOl}(P^t)} + e^{-\beta f_{LD}(P^t)} \right).$$

As can be seen, Z may be considered as function of β , $Z(\beta)$ and function of the temperature, $Z(T)$.

- *Expected value and ensemble average.* Let A be any magnitude (mainly energy in thermodynamics) and let A_i be some A -microstates. Both the expected value and the ensemble average (these are actually the same concept) are tools for studying global properties from local (micro)states. The expected value (also termed as expectation value) of a magnitude A with microstates A_i is defined as usual:

$$\langle A \rangle = \sum_j \mathbb{P}[A = A_j] \cdot A_j, \text{ where } \mathbb{P}[A = A_j]$$

is the likelihood of A to take the microstate A_j .

On its part, the ensemble average $\langle A \rangle$ is defined as $\langle A \rangle = \frac{\sum_i e^{-\beta A_i} \cdot A_i}{\sum_i e^{-\beta A_i}} = \frac{\sum_i e^{-\beta A_i} \cdot A_i}{Z}$ where $Z = \sum_i e^{-\beta A_i}$ is the partition function as stated before.

- *Relationship with partition function.* The relationship between expected energy and the partition function Z may be obtained from the $\log Z$ -expression by first-order partially deriving $\log Z$ with respect to the inverse temperature β : $\langle A \rangle = -\frac{\partial Z}{Z \partial \beta} = -\frac{\partial \log Z}{\partial \beta}$

6.4.2 Our contribution

The insights from thermodynamics could offer a new perspective on price outcomes in the presence of lobbying inasmuch as notions and results (see theorem 3 for instance) have certain parallels with them. Therefore, we proceed by analyzing the price outcomes using *suitable microstates* related to the lobbies. For later interpretations it is important to remember that the energy E , no matter which scenario it is referring to, symbolizes a flow rate that changes the magnitudes. According to some authors ([29]), the ensemble average of some daily repeated stochastic process consists of the daily variations in the process. Therefore here, regarding price outcomes, the *energy of price* shall represent the volatility of price as a rate of change.

Specifically, for our analysis we consider the magnitude *price variations at time instant t* which shall be termed as *energy of price* and denoted as $E(P^t)$ in order to follow the thermodynamical notation as closely as possible. Moreover, while price outcomes refer to the whole market price (which is the global process), their study shall be based on the

price estimates made by lobbies, which shall be taken as the micro levels. The price variations at the micro-level states are given by the energy functions of each lobby f_{L_i} . Particularly, in the olive oil market the energy functions of the lobbies are $f_{L_{Pr}}$, $f_{L_{OOI}}$, f_{L_D} . In consequence, the total energy of price at a time instant t is

$$E(P^t) = f_{L_{Pr}}(P^t) + f_{L_{OOI}}(P^t) + f_{L_D}(P^t).$$

Next, both the expected value and the ensemble average (same concept, different perspectives) of the energy of price are examined. On one hand, the expected value is

$$\begin{aligned} \langle E(P^t) \rangle &= \sum_i \mathbb{P}[E(P^t) = e_i] \cdot \overbrace{f_{L_i}(P^t)}^{e_i} = \\ &= \mathbb{P}_{L_{Pr}}[E(P^t) = f_{L_{Pr}}(P^t)] \cdot f_{L_{Pr}}(P^t) + \\ &\quad \mathbb{P}_{L_{OOI}}[E(P^t) = f_{L_{OOI}}(P^t)] \cdot f_{L_{OOI}}(P^t) + \\ &\quad \mathbb{P}_{L_D}[E(P^t) = f_{L_D}(P^t)] \cdot f_{L_D}(P^t). \end{aligned}$$

On the other hand, the concept known in thermodynamics as “ensemble average” is more easily identified by transforming the former expression into the following one:

$$\begin{aligned} \langle E(P^t) \rangle &= \sum_i \mathbb{P}[E(P^t) = e_i] \cdot \overbrace{f_{L_i}(P^t)}^{e_i} \\ &= \frac{e^{-\beta f_{L_{Pr}}(P^t)}}{Z} f_{L_{Pr}}(P^t) + \frac{e^{-\beta f_{L_{OOI}}(P^t)}}{Z} f_{L_{OOI}}(P^t) + \frac{e^{-\beta f_{L_D}(P^t)}}{Z} f_{L_D}(P^t) \\ &= \frac{e^{-\beta f_{L_{Pr}}(P^t)} f_{L_{Pr}}(P^t) + e^{-\beta f_{L_{OOI}}(P^t)} f_{L_{OOI}}(P^t) + e^{-\beta f_{L_D}(P^t)} f_{L_D}(P^t)}{Z} \\ &= \frac{e^{-\beta f_{L_{Pr}}(P^t)} f_{L_{Pr}}(P^t) + e^{-\beta f_{L_{OOI}}(P^t)} f_{L_{OOI}}(P^t) + e^{-\beta f_{L_D}(P^t)} f_{L_D}(P^t)}{e^{-\beta f_{L_{Pr}}(P^t)} + e^{-\beta f_{L_{OOI}}(P^t)} + e^{-\beta f_{L_D}(P^t)}}. \end{aligned}$$

Now, we shall prove some useful properties of the expected energy of price.

Proposition 5 1. *The expected energy of price is the sum of all the expected energy of these lobbies:*

$$\langle E(P^t) \rangle = \langle E_{Pr}(P^t) \rangle + \langle E_{OOI}(P^t) \rangle + \langle E_D(P^t) \rangle$$

2. $\langle E(P^t) \rangle$ always takes positive values regardless of the values of the temperature.
3. The expected energy of price is increasing function of the temperature.

Proof The result holds from the properties of the expectation values. Moreover, it has been shown in the former development since

$$\begin{aligned} \langle E(P^t) \rangle &= \langle E_{Pr}(P^t) \rangle + \langle E_{OOI}(P^t) \rangle + \langle E_D(P^t) \rangle \\ &= \underbrace{\mathbb{P}_{L_{Pr}}[P^t] f_{L_{Pr}}(P^t)}_{\langle E_{Pr}(P^t) \rangle} + \underbrace{\mathbb{P}_{L_{OOI}}[P^t] f_{L_{OOI}}(P^t)}_{\langle E_{OOI}(P^t) \rangle} \\ &\quad + \underbrace{\mathbb{P}_{L_D}[P^t] f_{L_D}(P^t)}_{\langle E_D(P^t) \rangle}. \end{aligned}$$

The function Z is a decreasing function of β so it is $\log Z$. Hence, the first partial derivative $\frac{\partial \log Z}{\partial \beta}$ should always be

negative. In consequence, $\langle E(P^t) \rangle = -\frac{\partial \log Z}{\partial \beta}$ always takes positive values regardless of the values of β (or of the temperature).

Using the above property and the corresponding Boltzmann distribution from theorem 5, the first derivative of $\langle E \rangle$ with respect to β is

$$\begin{aligned} \frac{\partial \langle E \rangle}{\partial \beta} &= \frac{\partial \langle E_{Pr} \rangle}{\partial \beta} + \frac{\partial \langle E_{OOI} \rangle}{\partial \beta} + \frac{\partial \langle E_D \rangle}{\partial \beta} \\ &= \frac{\partial \mathbb{P}_{L_{Pr}}[P^t]}{\partial \beta} f_{L_{Pr}}(P^t) + \frac{\partial \mathbb{P}_{L_{OOI}}[P^t]}{\partial \beta} f_{L_{OOI}}(P^t) + \frac{\partial \mathbb{P}_{L_D}[P^t]}{\partial \beta} f_{L_D}(P^t) \\ &= \frac{\partial \left(\frac{e^{-\beta f_{L_{Pr}}(P^t)}}{Z} \right)}{\partial \beta} f_{L_{Pr}}(P^t) + \frac{\partial \left(\frac{e^{-\beta f_{L_{OOI}}(P^t)}}{Z} \right)}{\partial \beta} f_{L_{OOI}}(P^t) \\ &\quad + \frac{\partial \left(\frac{e^{-\beta f_{L_D}(P^t)}}{Z} \right)}{\partial \beta} f_{L_D}(P^t) = \\ &= K \left(-f_{L_{Pr}}(P^t)^2 \mathbb{P}_{L_{Pr}}[P^t] - f_{L_{OOI}}(P^t)^2 \mathbb{P}_{L_{OOI}}[P^t] - f_{L_D}(P^t)^2 \mathbb{P}_{L_D}[P^t] \right) < 0, \end{aligned}$$

with K a positive constant (as it is the result of algebraic operations of the normalizing constant Z). This proves that the expected energy is decreasing function of the inverse temperature, and the result follows. \square

We address the economic interpretation of our former proposition 5. Regarding price volatility, the next theorem provides a main market property: the variations that the market price suffers may be computed separately (as the algebraic sum) from the variations in the target price of the lobbies.

Theorem 7 *The expected market price volatility is equal to the algebraic sum of that of lobbies: $\langle E(P^t) \rangle = \langle E_{Pr}(P^t) \rangle + \langle E_{OOI}(P^t) \rangle + \langle E_D(P^t) \rangle$. Moreover, each of these may be explicitly computed once the corresponding energy functions have been selected:*

$$\begin{cases} \langle E_{Pr}(P^t) \rangle = \frac{e^{-\beta f_{L_{Pr}}(P^t)}}{Z} f_{L_{Pr}}(P^t) \\ \langle E_{OOI}(P^t) \rangle = \frac{e^{-\beta f_{L_{OOI}}(P^t)}}{Z} f_{L_{OOI}}(P^t) \\ \langle E_D(P^t) \rangle = \frac{e^{-\beta f_{L_D}(P^t)}}{Z} f_{L_D}(P^t) \end{cases}$$

Before concluding the volatility issue, it should be noticed that, according to the second property of proposition 5, mean volatility is always positive, i.e., market prices move upwards.

In econophysics, the temperature of an economic system is linked with money: the income of involved agents, wages

or monetary value of goods and services (particularly with Gross Domestic Product GDP at a global scale, as a measure of the system efficiency), see [7]. Thus, we shall assume here that the *temperature of an economic process is its financial solvency (soundness or economic health)*. In this line, higher temperatures are more desirable and higher differences of temperatures should generate more profit. Thus, the third property of proposition 5 has the following economic meaning:

Corollary 3 *The expected market price volatility is an increasing function of the solvency of the olive oil market. That is, the higher the market solvency, the higher the volatility of the price.*

7 Conclusions

This paper presents a structural model for explaining what kind of market price emerges when there are multiple agents pushing for their prices (i.e., markets with the presence of lobbying). Apart from offering a mathematical/AI model that successfully unravels a complex real-world problem, the significance of our approach is that it delimits an intricate problem within the bounds of some probability computations providing formulas that allow the calculation of the price estimates of a product. The model is also endowed with a quantitative study which determines the specific weight of each lobby in the process of pricing outcomes thus solving a problem for which up until

now, there have only been qualitative references. This study has been complemented with an analysis from thermoeconomics (econophysics), a relatively young science that offers a novel perspective from which our model can offer additional results.

The originality of our model lies in the use of time dynamic Markov random fields (TD-MRF model) since this problem has been addressed from the perspective of Bayesian approaches, neural networks and related techniques. The use of TD-MRFs means that significantly less information is required (only from cliques) than for the aforementioned methodologies. The consequence is a higher computing speed and a lesser need for storage capacity. Using TD-MRFs also means breaking free from the constraint of a given distribution function.

Our TD-MRF model has been tested on real olive oil prices (the olive oil market in Andalusia, Spain) with encouraging results for a challenging sector in which opacity in the entry of oil shipments throughout the season, with olives/olive oil being stored waiting for the price to rise, makes it very difficult to forecast the prices. In the model-validation process, both the base functions (energy functions) and the partition function have been chosen and computed in detail when, to the best of our knowledge, there is a notable absence of discussion on either how to select the energy functions (to best fit the context) or how to compute the partition function Z in the literature. As we mean the value of Z , not an approximation, this could be considered as a limitation of the model since the final output depends on how good the estimate of Z is (as previously mentioned, the literature gives no indication on how to delimit the range considered in the calculation of Z in reality).

As mentioned in the Introduction section, our TD-MRF model can be applied to any lobbying context although it is possible that fine-tunings of the structural model might be necessary if the lobbies considered have their own characteristics. Furthermore, understanding the “lobbying context” in a broad manner would allow the application of the model to all those scenarios in which the final goal is to obtain a single output as the result of unifying different perspectives (such as reaching a consensus). Thus, the possible extensions of the TD-MRF model should focus on replacing the numerical datasets with wording ones (or integrating both types of datasets).

Acknowledgements Support by the Spanish State Research Agency under Project PID2019-103880RB-I00/AEI/10.13039/501100011033, and Junta de Andalucía “Excellence Groups” (P12.SEJ.2463), and Junta de Andalucía (SEJ340) is gratefully acknowledged. Funding for open access charge: Universidad de Granada / CBUA.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abid F, Kaffel B (2018) The extent of virgin olive-oil prices’ distribution revealing the behavior of market speculators. *Rev Quant Finan Acc* 50:561–590. <https://doi.org/10.1007/s11156-017-0638-9>
2. Bednarikova Z (2012) Why is the agricultural lobby in the European Union member states so effective? *Ekonomie a Management* 15(2):26–37
3. Boltzmann L (1995) *Lectures on Gas Theory*. Dover Publications New York (translated by S.G. Brush). <https://doi.org/10.1029/2004WR003567>
4. Chen N, Kou S, Wang Ch (2018) A partitioning algorithm for markov decision processes with applications to market microstructure. *Manage Sci* 64(2):v-x, 495–981. 10.1287/mnsc.2016.2639
5. DAdamo I, Falcone PM, Gastaldi M (2019) Price analysis of extra virgin olive oil. *Brit Food J* 121(8):1899–1911. <https://doi.org/10.1108/BFJ-03-2019-0186>
6. Deng ZF, Liu CX, Zhu ZL (2021) Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network. *Int J Elec Power* 125:106499. <https://doi.org/10.1016/j.ijepes.2020.106499>
7. Dimitrijevic B, Lovre I (2015) The Role of Temperature in Economic Exchange - An Empirical Analysis. *Journal of Central Banking Theory and Practice* 3:65–89. <https://doi.org/10.1515/jcbtp-2015-0012>
8. El-Mawass N, Honeine P, Vercouter L (2020) Similcatch: Enhanced social spammers detection on Twitter using Markov Random Fields. *Comm Com Inf Sc* 57(6):102317. <https://doi.org/10.1016/j.ipm.2020.102317>
9. Espadas-Aldana G, Vialle C, Belaud JP, Vaca-Garcia C, Sablayrolles C (2019) Analysis and trends for Life Cycle Assessment of olive oil production. *Sustainable Production and Consumption* 19:216–230. <https://doi.org/10.1016/j.spc.2019.04.003>
10. Feng X, Xiao Z, Zhong B et al (2019) Dynamic weighted ensemble classification for credit scoring using Markov Chain. *Appl Intell* 49:555–568. <https://doi.org/10.1007/s10489-018-1253-8>
11. García Cabello J, Herrera-Viedma E (2018) A universal decision making model for restructuring networks based on Markov Random Fields. *Advances in Artificial Intelligence*. CAEPIA 2018 Lect Notes Comput Sc 11160:164–173. Springer, Cham
12. García Cabello J (2019) A model towards global demographics: an application, a universal bank branch geolocator based on branch size. *Soft Comput* 23(16):7193–7205. <https://doi.org/10.1007/s00500-018-3362-z>
13. Kurumatani K (2020) Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *Sn Applied Sciences*, 2(8). <https://doi.org/10.1007/s42452-020-03225-9>

14. Kyar N, Soe Kyaw S, Oo Htike N (2019) Moment-Generating Functions and reproductive properties of distributions. *International Journal of Scientific Research and Engineering* 2(5):654–660
15. Kindermann R, Snell JL (1980) *Markov Random Fields and Their Applications*. American Mathematical Society. ISBN 0-8218-5001-6
16. Kyriazi F, Thomako D, Guerard JB (2019) Adaptive learning forecasting, with applications in forecasting agricultural prices. *Int J Forecasting* 35:1356–1369. <https://doi.org/10.1016/j.ijforecast.2019.03.031>, 0169-2070
17. Li B, Ding JQ, Yin ZQ, Li KY, Zhao X, Zhang LX (2021) Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting. *Expert Syst Appl* 168:114232. <https://doi.org/10.1016/j.eswa.2020.114232>
18. Liang Y, Li SG, Yan CG, Li MZ, Jiang CJ (2021) Explaining the black-box model: a survey of local interpretation methods for deep neural networks. *Neurocomputing* 419:168–182. <https://doi.org/10.1016/j.neucom.2020.08.011>
19. Liu ZY, Kodamana H, Afacan A, Huang B (2019) Dynamic prediction of interface level using spatial temporal markov random field. *Comput Chem Eng* 128:301–311. <https://doi.org/10.1016/j.compchemeng.2019.05.035>
20. Matyjaszek M, Fernandez PR, Krzemien A, Wodarski K, Valverde GF (2019) Forecasting Coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory. *Resour Policy* 61:283–292. <https://doi.org/10.1016/j.resourpol.2019.02.017>
21. Nowotarski J, Weron R (2018) Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew Sust Energy Rev* 81:1548–1568. <https://doi.org/10.1016/j.rser.2017.05.234>
22. Orlandi F, Aguilera F, Galan C, Msallem M, Fornaciari M (2017) Olive yields forecasts and oil price trends in mediterranean areas: a comprehensive analysis of the last two decades. *Exp Agr* 53(1):71–83. <https://doi.org/10.1017/S0014479716000077>
23. Pe'er G, Dicks LV, Visconti P (2014) Agriculture Policy EU agricultural reform fails on biodiversity. *Science* 344(6188):1090–1092. <https://doi.org/10.1126/science.1253425>
24. Shao Y, Cheng Y, Shah RU, Weir CR, Bray BE, Zeng-Treitler Q (2021) Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcomes. *J Med Syst* 45(1):5. <https://doi.org/10.1007/s10916-020-01701-8>
25. Sezer OB, Gudelek MU, Ozbayoglu AM (2020) Financial time series forecasting with deep learning : a systematic literature review: 2005-2019. *Appl Soft Comput* 90:106181. <https://doi.org/10.1016/j.asoc.2020.106181>
26. Yang Y, Wang J (2021) Forecasting wavelet neural hybrid network with financial ensemble empirical mode decomposition and MCID evaluation. *Expert Syst Appl* 166:114097. <https://doi.org/10.1016/j.eswa.2020.114097>
27. Wang B, Wang J (2019) Energy futures price prediction and evaluation model with deep bidirectional gated recurrent unit neural network and RIF-based algorithm. *Energy* 216:119299. <https://doi.org/10.1016/j.energy.2020.119299>
28. Zhang J, Teng YF, Chen W (2019) Support vector regression with modified firefly algorithm for stock price forecasting. *Appl Intell* 49:1658–1674. <https://doi.org/10.1007/s10489-018-1351-7>
29. Zhang DZ (2021) Ensemble average and nearest particle statistics in disperse multiphase flows. *J Fluid Mech* 910:A16. <https://doi.org/10.1017/jfm.2020.974>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Julia García Cabello held a PhD in Pure and Applied Mathematics from the University of Granada where she has been teaching since 1990. Prior to arriving at the world of Applied Mathematics, she developed a successful career in Pure Algebra (known as JG Cabello), publishing in flagship journals on Algebraic Homotopy Theory. Today, she is fully tenured professor and full researcher at the Applied Mathematics Department, Faculty of Business and

Economics of the University of Granada (Spain), where she teaches undergraduate, MBA and Executive MBA courses and conducts seminars on a wide range of mathematical business-related topics. She is also Academic Secretary of the Department of Applied Mathematics.

Her current research interests include Mathematical Modeling / Intelligent Systems for the resolution of real problems including applications in Business, Finance and Economics as well as Decision Making. To this regard, her mathematical baggage (from the Pure Algebra to Applied Mathematics) makes Dr. García Cabello's research characterized by using a wide range of mathematical tools, from stochastic processes to dynamic systems, with the sole purpose of solving the stated real problem. Dr. Julia García Cabello is also reviewer at related journals.