IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 27, 2018, accepted June 1, 2018, date of publication June 15, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2846878

Expert Selection in Prediction Markets With Homological Invariants

CARLOS BOUSOÑO-CALZÓN[1], HAROLD MOLINA-BULLA[1], (Member, IEEE), JOSÉ JOAQUÍN ESCUDERO-GARZÁS[2], AND FRANCISCO J. HERRERA-GÁLVEZ[1]

[1]Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain
[2]Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA

Corresponding author: Carlos Bousoño-Calzón (cbousono@tsc.uc3m.es)

This work was supported by the Universidad Carlos III de Madrid under Strategic Action Grant 2013/00199/002.

ABSTRACT Group decision making is a topic of growing interest in today's complex societies. One of the key technologies in this area is the prediction market, where a group of experts plays a fake stock market with assets that represent the outcomes of an uncertain event. The particular problem we address in this paper is the expert selection in these markets to improve their reliability. To aggregate decisions from a particular group of experts, instead of using prices as is typically done, we define a *market deconstruction* considering player portfolios. This decision technology makes the behaviors of experts toward their decisions available through their portfolios evolution. Our main contribution is the identification of two *Persistent Homological Invariants* able to classify experts in groups based on the histories of their portfolios. Interestingly, this translates into the definition of essentially two dominant groups. A simulation of the Prediction Market with artificial agents allow us to interpret these two classes as *rational* and *irrational* players, following the Microeconomic jargon. Four experiments with experts in the insurance sector help us to illustrate the relationship between these two player types with the prediction reliability of the market.

INDEX TERMS Artificial markets, behavioral classification, Betti numbers, group decision-making, expert selection, insurance sector, market deconstruction, market efficiency, market evolution, persistent homology, prediction market, prediction reliability, rational player, Wasserstain distances.

I. INTRODUCTION

The subject of Decision Making has a long history and, because of its practical importance in many aspects of human life, it has also been studied from many different perspectives [1]. The Delphi Method (DM) by the RAND corporation in the 1950s and the Surowiecki's Wisdom of Crowds (WoC) in 2004 have vigorously defended the view that the aggregation of opinions of many people may render better decisions [2], [3]. As examples of recent technology-oriented applications of the DM, we can mention the study of state requirements of cyberdefence [4], the identification of core concepts of cybersecurity [5] or the estimation of software effort and the knowledge management the Scrum methodology [6]; as for the WoC, discovery of categories of images [7], distributed spectrum access assisted by social recommendation [8], mobile sensing for multimedia applications [9], the processing of distributed signals [10] and the determination of course-offering [11] can also be offered as examples. Some other technologies have been added to these two milestones, such as optimization techniques to obtain consensus among decision groups [12]–[14], probability aggregation methods [15]–[17] that are somehow parallel to traditional data fusion approaches [18], or variants of the DM [19].

We use a technology known as Prediction Markets (PM), that works similarity to real Stock Exchanges, but experts trade on shares that represent the outcomes of an uncertain event [20]–[24]. This technology is a kind of WoC approach [25] and shares with DM the fact that it feeds the aggregated decision continuously back to the experts, allowing them to modify their positions accordingly. The advantages of PM we exploit in this paper are that the entire process in PM is carried out intrinsically on line and that the decisional status of all the experts can easily be expressed as a vector of real numbers ready for numerical processing, as shown below. In particular, our PM usually involves a very small number of participants. This fact has a great influence on the design of our PMs and, when it is necessary to emphasize this characteristic, we will refer to them as *Small Prediction Markets* (SPM).

The particular problem we address in this paper is the selection of experts in PMs. The idea of selecting experts or the more relaxed approach of adjusting their contributions have been investigated for practically all the group decision methods mentioned above [26]–[29]. Interestingly, there are no definite general conclusions, not only on the best way to proceed [16], [27] but even on its own relevance [2]. It is worth noting that the entire selection process of experts is independent of the number of players, so we will adhere to the PM term, instead of the SPM, for this particular discussion.

Our approach is inspired by the financial concept of *market efficiency*, which characterizes how well the market is functioning as an allocation mechanism [30], [31]. According to this type of efficiency, players in the market can be classified as *rational* or *irrational* depending on whether or not they project their private knowledge on prices, respectively [32]. However, for this much debated topic to be used to classify experts, it has to be conveniently formalized and tested. We provide a mathematical formalization in terms of topological invariants calculated from the PM trace data by using Persistent Homology tools [33], and we test it on a controlled experiment consisting of four parallel PMs with real experts in the field of insurance. Obviously, these results are not enough to draw definitive conclusions about this selection. Therefore, to add further support to this classification, we developed an artificial market model for the PM [34], [35]. This artificial model allows us to provide an interpretation of the models of the players in terms of the selected homological invariants. Note that although this approach can be considered quite formal in the definition of classification types, their final usefulness can only be illustrated through real PM experiments.

The proposed expert selection has an additional feature that is worth observing: it only depends on the behavior of the participants in each implementation of the PM. This allows to abstract the classes of the players from the real persona although, to some extent, people tend to comply with a particular class, as shown in our experiments.

The main contribution of this paper is to show how two homological invariants (the 0^{th} and 1^{st} *persistent Betti numbers* [36], [37], see also Section III) are sufficient to characterize the experts from their behavior in the decision-making process. Although this fact is illustrated here in the PM field, it may well be able to translate to other group decision-making schemes where data on how the experts behave towards their decisions are available, for instance the DM and its variations [19]. Additionally, as for the PM technology, we provide a novel way of constructing predictions by allowing the aggregation of decisions of arbitriarily chosen groups of experts. Furthermore, the interpretation of the homological classes of players in the market, and their influence in the *efficiency* of the market, as *rational* and *irrational* players, allows a novel approach for experimentation in real markets [38].

The structure of this paper is as follows: first we present both our real SPM and its artificial model in the next section. In Section III, the application of the Persistent Homology tools is analyzed in the artificial market model to define the types of player. The experiments in four real SPMs are given in Section IV to test the usefulness of our classification of real experts. Finally, some conclusions are drawn.

II. SMALL PREDICTION MARKETS: THE REAL AND ITS ARTIFICIAL MODEL

In many PM applications, the actual number of players involved is quite small, either because the motivations are not sufficiently enticing or simply because the forecasting configuration is naturally limited [39]. As already mentioned in the introductory section, we refer to these as SPM.

Although the forecast capacity of SPM does not decrease compared to a standard PM [39], the small number of trading players together with the lack of information about them, due to the protection of personal data and the limited account of market records they generate, make it difficult to test hypotheses to interpret the types of players. To overcome these limitations, we have developed an *Artificial Prediction Market* (APM) that reproduces the same mechanism as in the real SPM. With this APM, different player models are checked to get some insights about the behavior types of player. The coincidence between the Homological invariants in the SPM and the APM makes these interpretations possible, in regard to their *loopy* behavior, which is the behavior characterized by these invariants, as we argue in the next section. Here we only provide some description of how the SPM and APM are designed.

A. THE REAL SMALL PREDICTION MARKET

One of the relevant issues to consider in the design of PMs is the price formation mechanism [40]; even more if the markets are expected to have a very small number of players, that is if they are SPM. Within the two classic options for the market mechanism, double side auction [41] [42] or the use of a market maker [41], our configuration clearly requires the second option to provide liquidity and a live experience for the players. This choice is also in line with the suggestions of simplicity as in [43].

The shares traded in the market embody the different outcomes of the event we want to predict. Market prices reflect the probabilities assigned by the market to each of the possible outcomes. The market closes before the target event occurs, and the last market prices are generally interpreted as the final forecast for the outcome.

Players play with fake money (f$s) to buy and sell the different stocks and thus build a personal portfolio which represents her particular forecast for the PM. The use of fake money would not have a great impact on the accuracy of their predictions [44]. At the beginning of the game, all players are allocated the same amount of fake money and an initial stock portfolio. At any time during the evolution of the market, prices are fed back to the experts who can

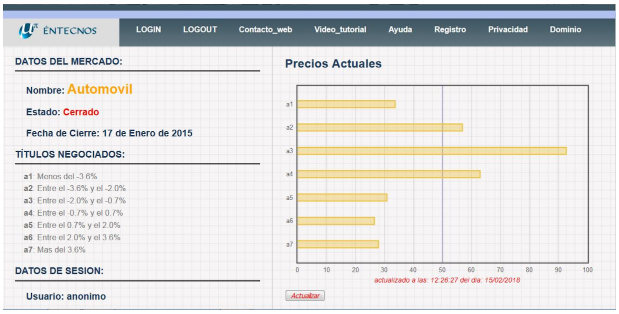

FIGURE 1. Screenshot of our SPM client seen by the players. The most important item is the horizontal bar graph showing the prices for each share. These prices are fed back to the experts to elaborate on their decisions. Note that since prices shown graphically, the market can be easily played by experts in other languages.

update their portfolios. Fig. 1 shows a screen shot of this feedback scheme. Note that the prices of each share are displayed graphically. This fact makes this technology available to experts who speak different languages: in this particular example, even if the PM is in Spanish, it could easily be played. This is a particularly interesting feature of this decision-making technology. As a rewarding scheme for the players, the winner-take-all strategy has been chosen: once the market is closed and the event is revealed, the stock that represents the real outcome will be valued at 100 f\$, while the rest will have no value. The final position of each player is calculated by adding her available money plus the number of the winning shares in her portfolio multiplied by 100. These positions allow us to rank the players and then translate either their positions or their ranks into a real reward.

The pseudo-code for SPM is given in Algorithm 1. The reader can get an idea of market dynamics from the evolution of prices, as illustrated in Figs. 2, see Section IV for additional details about these markets. These figures show the evolution for the daily closing prices: the initial price is set at 50 f\$ and, given that the maximum reward for the winning asset is 100 f\$, it is expected that the price of any asset does not exceed 100 f\$, although this is not enforced by the system. The latest prices in the final stage of the market are then converted into the probability estimate for the outcomes to the event associated with the assets through a simple normalization so that they add up to one.

Algorithm 1 Real Market Dynamics

```
Initialize the market state:
for each player do
  | Player_state = initial_ portfolio
end
Market_state = initial_state
Run the market:
while the market is open do
  | Wait until Mov = Player (Market_state, Player_state)
  | /* Wait for a player to make a
  |    movement, that is to issue a buy
  |    or sell order.                    */
  | Market_state = Markey_Maker (Market_state,
  | Player_state, Mov)
  | /* Update the states of the player
  |    and the market and publish the
  |    prices, see Fig. 1.               */
end
Wrap up the results:
Estimate the probability for each outcome based in the
  final prices.
Calculate the final position for each player and rank
  them.
End.
```

B. PORTFOLIOS AND PREDICTION DECONSTRUCTION

Typically in PMs, the aggregation of information comes in form of the price obtained by the different shares. However, the controlled manipulation of this aggregate may yield important improvements in prediction capacity. This fact is especially relevant in SPM, since the consistency of the aggregation of prices cannot be supported by the law of large numbers.

The first step towards *deconstructing* PM prices can easily be done using the players' portfolios. These portfolios reflect both the expectations of the players and the dynamics of pricing. Given that the evolution of player portfolios is the source of data for the topological study in the next section, it is clarifying to obtain some insight about them: Figures 3(a-c) show them for players no. 1, 4, 10 and 15 in the *Automobile market*, see Section IV.

These portfolios play a similar role to the betting functions on artificial models [34], although their final composition reflects the strategies and capabilities of different players, as well as their shortcomings in the history of the game of a particular market. New prediction aggregations of these final portfolios can be made by first selecting some players, so disregarding the other portfolios, and second by defining a function to aggregate the selected portfolios for a final decision.

Given that in this paper we are mainly interested in the selecting procedure, a basic linear aggregation of the selected portfolios will be used. We elaborate on this topic in Subsection IV.C.

C. THE ARTIFICIAL PREDICTION MARKET

Our main objective when establishing the APM is to check patterns of the players to get insight into the *loopy* classes found by Homological invariants in the real SPMs. This issue is explained in more detail in the next section, after introducing these invariants. Here we outline its pseudocode, see Algorithm2, which mimics our real SPM as an expeditious comparison with Algorithm 1 can reveal.

For the market dynamics to be similar, the main parameters that need to be transferred from the SPM to the APM are: the *tick size*, the *bid-ask spread* and the *depth of the market* [45].

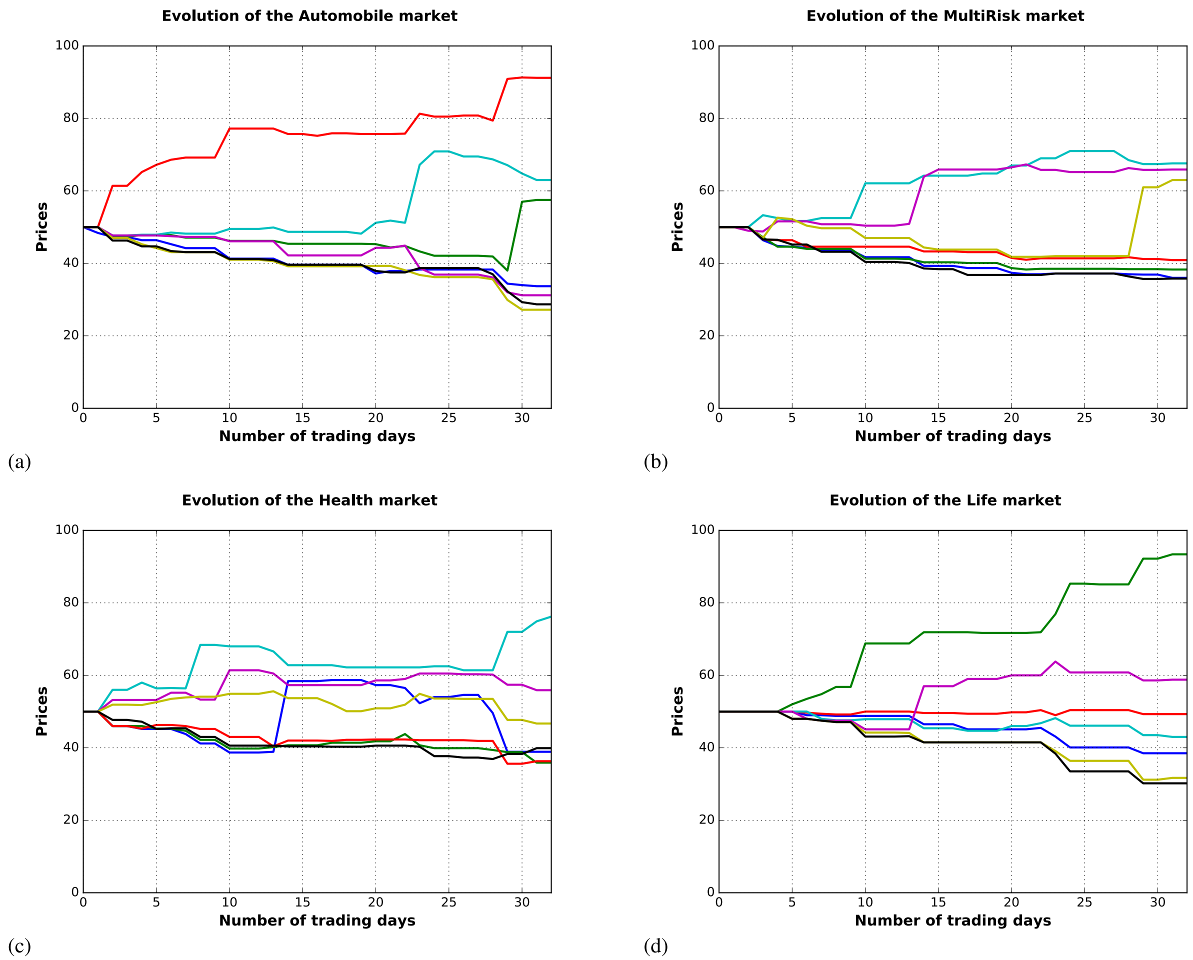

FIGURE 2. Real dynamics of the SPM, given by the evolution of the prices in the four prediction experiments. Each color represents a different asset.

Therefore, the values of these parameters, as well as the initial allocation of money and portfolios, are set with the same values in the SPM and APM. Additionally, we have set a number of artificial players close to the participants in our real markets. The total number of runs is set to allow each of the artificial players to reach a state of equilibrium, that is, the state in which these players no longer trade.

As the market develops, random turns are assigned to random players who select their best move according to their profiles. Once the movement that a player must make is selected, the market maker updates the state of the market, and therefore the prices, in a similar way to the real SPM. Fig. 4 provides an illustration of evolution of the prices in an execution of the APM. When comparing this figure with the evolution of prices in real SPM (Figure 2) it can be observed that the time scale is different, as is the information shown: in the real SPM only the daily closing prices are drawn, while for the APM each individual price movement is collected.

Regarding the evolution of portfolios in the APM, Figs. 5 illustrate them for two artificial players, each representative of their type, as discussed in the next section.

III. PERSISTENT HOMOLOGY: INTERPRETATION AND PLAYER CLASSIFICATION

A. PERSISTENT HOMOLOGY, BARCODE DIAGRAMS

Algebraic Homology is a theory where the ideas from Topology meet the computational tools of Algebra. Its main output for our purposes is a set of features of a topological space which are related to the existence of *holes* at different dimensions, or *loops* when the space dimensionality is low [46]. Persistent Homology is an adaptation of these tools for a set of data instead of an abstract topological space. There is excellent material to delve into these topics: see [47] for a rigorous introduction to the Algebraic aspects, [48], [49] for a gentle introduction to the idea of persistency and [33] for a categorical oriented interpretation in processing data. In this paper, we are mainly interested in its black-box application to the traces left by players in order to classify them in a meaningful way. Therefore, we only review the basic concepts for the interested reader without much familiarity with this theory, and use the interface for libraries GUDHI, Dionysus, and PHAT in the R Language [50].

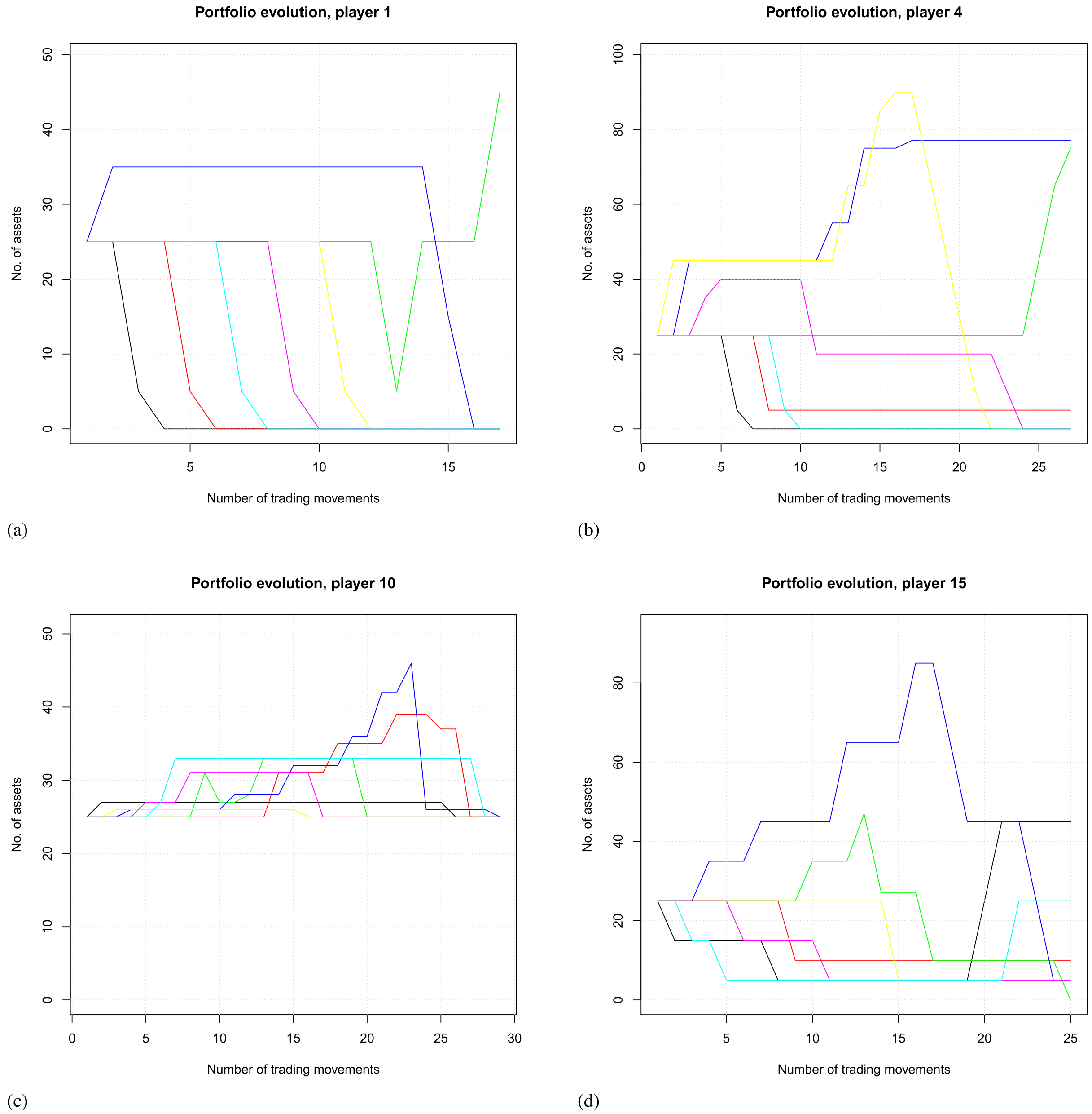

FIGURE 3. Evolution of a sample of portfolios of experts in the real SPM (automobile). Each graph corresponds to different experts and each color to different shares.

As data, we use the history of the portfolios of a particular player, see Figs. 3 and 5. The dimension of these data is, therefore, the number of shares in a particular market. Given that the market evolves asynchronously, we only consider those points in which the player performs an operation, whether sale or buy, of any stock. This selection of data is not only sensible but important since the typical number of movements in a small market is fortunately adapted to the complexity of the tool and no statistical treatment of these data is needed, which might blur the topological content of the data. At this point, it is worth remembering that the complexity of the algorithms for the calculation of invariants comes mainly from the number of points rather than their dimensionality. In this sense, our computer facilities can cope with a history of 40 or 50 points. Beyond this number, the last 50 points are selected as representative of the evolution or a statistical processing is in order. This is not an essential limitation of our IT infrastructure; a much larger computer cloud and infinite patience would hardly increase this limit in a few tens. However, new advances in quantum computing may offer the possibility of a much more extended histories in real time, opening the opportunity to extend this type of analysis to real stock markets instead of the prediction markets [51].

Algorithm 2 Artificial Market Dynamics

```
Initialize the market state:
for each player do
    Player_state = initial_ portfolio
    if Player is predictor or speculator then
        Ideal_portfolio = random_porfolio
    end
end
Market_state = initial_state
Run the market:
for run= 1 to Total_number_of_runs do
    Player = sample(1 to N_players)
    if Player is speculator then
        Mov = Speculator(Market_state, Player_state,
        Ideal_portfolio)
    end
    if Player is predictor then
        Mov = Predictor(Market_state, Player_state,
        Ideal_portfolio)
    end
    if Player is random then
        Mov = Random_player(Market_state,
        Player_state)
    end
    /* Sell or buy movements are issued
    with the corresponding player'
    profile.                              */
    Market_state = Marker_Maker(Market_state,
    Player_state, Mov)
end
/* Since we only need to record the
portfolio dynamics, see Fig. 5, there
is no need of additional processing.
*/
End.
```

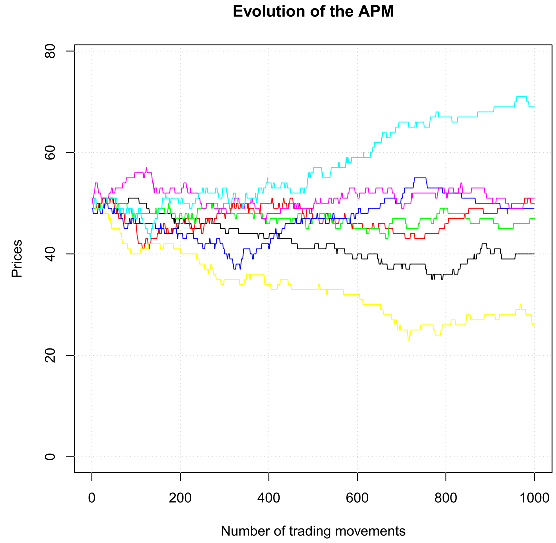

FIGURE 4. **Artificial market dynamics, as given by the evolution of prices in a typical run. Different colors code different shares. In artificial markets, time is measured by the number of trading movements.**

The result of the persistent homology tool for each player is a kind of *topological signature* for him, which can be depicted either as a *barcode* or as a *rotated diagram* [36], see Fig. 6. Technically speaking, the calculated invariants in these representations are the so called *p-th persistent Betti numbers*, an extension of the traditional Betti numbers in the framework of *persistency*, see [37] and [47], respectively.

The barcode summarizes the essential idea behind the persistence homology: first, it shows the homological invariants which can be related to *loopy* structures in the data at different dimensions. In this study, we will show that the only use of invariants at dimension 0, 0^{th} *persistent Betti number* or *components*, and dimension 1, 1^{st} *persistent Betti number* or *loops*, is sufficient for our objectives, see [50] for this terminology. Second, given that the data do not form by themselves a topological space to search for loops, the idea of persistency associates data to different combinatorial spaces, at different scales, to do so. The combinatorial spaces we use here are the so called *Vietoris-Rips complex*, see [37] for a mathematical definition and [50] for its practical application.

As there is no defined scale to characterize the data, it is therefore mandatory to observe the complete pattern left by the invariants in many of them simultaneously, namely the scale at which an invariant appears, the scale at which it disappears and then ponder the persistency of this invariant through all of them. These scales can be thought essentially in those between the minimum and the maximum distances of any two points in the data set.

All this information is displayed in the barcode: the horizontal axis showing the scale and the different invariants drawn as different lines present at the scales where they appear. The color in these lines codes for the dimension of the invariant: black is for zero dimensional components, red is for loops [50]. Rotated diagrams depict the same information but emphasizing the scales where the invariant appears and disappears as the axes of the representation.

However, the information transmitted by these diagrams cannot be used straightforwardly for the intended classification, as discussed in Subsection III-C.

B. PLAYER MODELS IN THE APM AND THEIR TOPOLOGICAL SIGNATURES

In this subsection we address the player types corresponding to the structure of the APM given in II-B. Following insights provided by the Microeconomic Theory [31], we propose three different profiles for the players: *Predictor, Speculator and Random*. Their pseudocodes are reproduced below (Algorithm 3).

The Predictor models the ideal player for a PM according to a Bayesian view [52]: this player model has a *apriori* expectation about the outcomes of the event which is reflected in its ideal portfolio. Its moves tend to minimize the difference

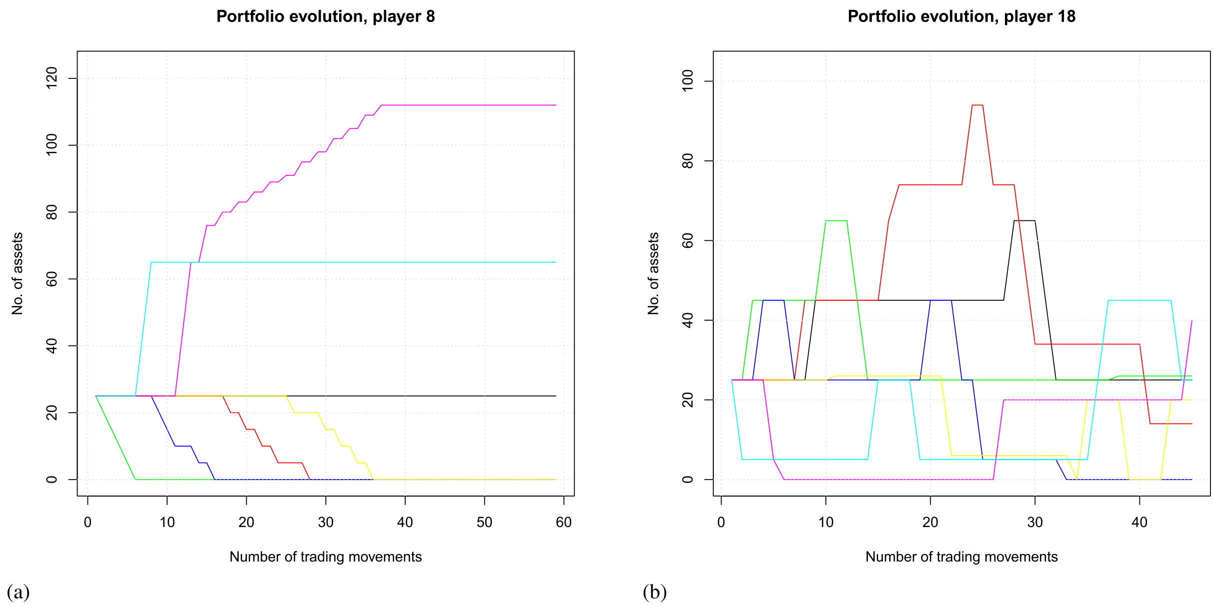

FIGURE 5. **Portfolios evolution in the artificial market for (a) a *predictor* agent and (b) a *random* agent.**

between its actual portfolio and its ideal portfolio. The Speculator model also has a reference for the outcomes and a corresponding ideal portfolio. However, it does not intent to reproduce such a portfolio, but simply plays the market according to the differences between real prices and its expectations, that is, it's ideal portfolio. Finally, the Random model has no reference, so the actions are a random walk on the feasible movements.

In Fig 6, we provide some barcodes and rotated diagrams for Players 1 (a,b), 2 (c,d), 12(e,f) and 13(g,h), respectively. The Player no.1 is a Predictor, no.2 a Speculator, and no.12 and no.13, Random players.

C. WASSERSTAIN DISTANCES: THE KEY TO CLUSTERING.

The diagrams in Fig. 6 for players 1,2,12 and 13 in the APM indicate that only Player 2 has a loopy behavior as shown by a red spot, while the rest only exhibits zero dimensional components. A simplistic attempt to conclude with this tool that speculators are separated from the rest, while predictors and random players come together, is deceptive. A careful observation of the size of the persistence of these diagrams suggests that a more careful treatment of these results is necessary. The solution comes from the concept of *stability*, as it is formulated through the Wasserstain's distances. The precise definition of Wasserstain's distances (W-d) and its relationship to stability can be found in [37]. Here we simply emphasize the interpretation of these results in the context of the selection of players through their topological signatures.

Since Persistent Homology can be seen as a measuring instrument for player data that delivers a topological signature, small changes in the data set are expected to cause small changes in the signature. W-d calculates a distance between two topological signatures in such a way that allows a precise formalization of this concept: small changes in the results of the data in nearby firms, as measured by W-d.

Therefore, a small W-d between two signatures corresponding to two different data sets indicates that these two sets are very similar in terms of looping behavior. The Wasserstain distance matrix for players 1, 2, 12 y 13 in the APM [see Table 1], can be used to gain some insight into this measure. Although there is a dispersion of distance values, players can essentially be grouped into two classes separated by an approximate distance of 0.25. Note that this matrix corresponds to the same players whose topological signatures are given in Fig. 6.

An illustration with a larger set of players in the APM is shown as a cluster dendrogram in Fig.7. Players with numbers from 1 to 11 are either predictors or speculators while players from 12 to 20 belong to the random class. For the APM, the W-d clearly separates players into two clusters: random players and the rest.

Although the final interpretation of the roles of the players is deferred until the introduction of real experts in the next section, some interesting theoretical conclusions can already be drawn at this stage. The predictor model is one related to the ideal player for a prediction market: it essentially provides information to the market in accordance with predefined and immutable expectations. However, in a real market the rational player is much closer to the speculative model [30], [31]. Even for a prediction market, the main hypothesis is that a player is embedded in the game as if he were playing in the real stock market. So, the model closest to the predictor we can reasonably expect from a player in a prediction market is actually a speculator. The good news from these results is that both models behave the same,

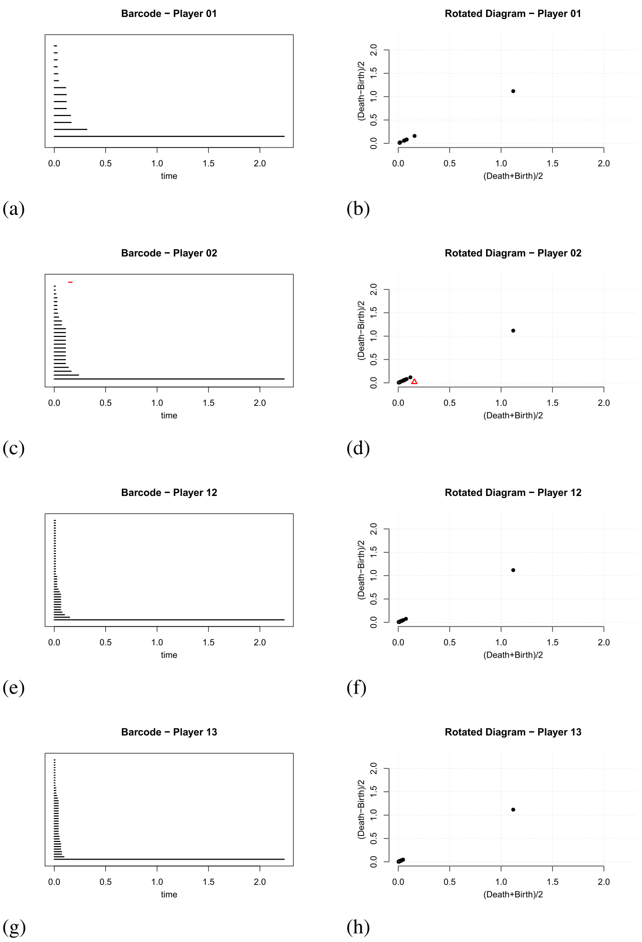

FIGURE 6. Barcodes and rotated diagrams for the APM player models: (a-b) Predictor, (c-d) Speculator, and (c-f) Random players.

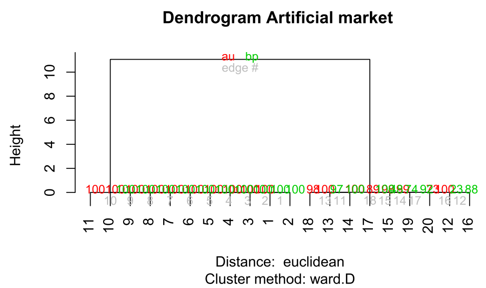

FIGURE 7. Classification dendrogram for the three player models in the artificial market. Players 1 to 11 are predictors or speculators, while players 12 to 18 belong to the random model.

insofar as their *loopy behavior* is considered, so there is no need to make further difference between them.

IV. EXPERIMENTS AND RESULTS

A. FORECASTING SETTING: ACTUARIAL PREMIUM RATES

Four PMs were implemented, in cooperation with actuarial experts, with the objective of forecasting the annual variation of insurance premium rates for four different sectors in the Spanish insurance market: Automobile, Life, Health

Algorithm 3 Pseudo Code for the Three Player Models

```
Function Predictor(Market_state, Player_state,
Ideal_portfolio) is
    Order the potential actions (sell or buy) and stocks
    according to the distance(ideal_portfolio, the actual
    portfolio). The maximum first.
    Mov= Select the first feasible action and stock.
    The number of stocks is such that
    distance(ideal_portfolio, the actual portfolio) is
    minimized
    return(Mov)
end
Function Speculator (Market_state, Player_state,
Ideal_portfolio) is
    Order the potential actions (sell or buy) and stocks
    according to the distance(ideal_portfolio, prices).
    The maximum first.
    Mov= Select the first feasible action and stock.
    The number of stocks is the maximum allowed.
    return(Mov)
end
Function Random_player (Market_state,Player_state)
is
    Mov=randomly select a feasible action (sell or buy)
    and a stock
    return (Mov)
end
```

TABLE 1. Wasserstain distance matrix for players 1,2, 12 and 13 in the APM.

	[,p1]	[,p2]	[,p12]	[,p13]
[p1,]	0.0000	0.0000	0.2516	0.2500
[p2,]	0.0000	0.0000	0.2516	0.2500
[p12,]	0.2516	0.2516	0.0000	0.0016
[p13,]	0.2500	0.2500	0.0016	0.0000

and Multi-Risk. The final data were published by the ICEA Association after all of our SPMs were closed.

For each market, several actuarial experts were invited to play for a month before the ICEA issued the actual data of premium rates. In order to design the stocks to be traded in each market, seven ranges of potential variation of premium rates were calculated from data of previous years together with additional suggestions provided by the experts, to obtain approximately equiprobable intervals. These intervals appear as different steps in the corresponding figures of the outcomes probability estimation. They are also referred to as the *potential decisions* made by the market. The dynamics of the markets, given by the evolution of their prices, is shown in Fig. 2.

Fig. 8 shows the estimated probabilities for each market according to their prices. Actual results, given by the ICEA, are represented by red vertical lines. As it can be observed from this figure, even with a limited number of participants,

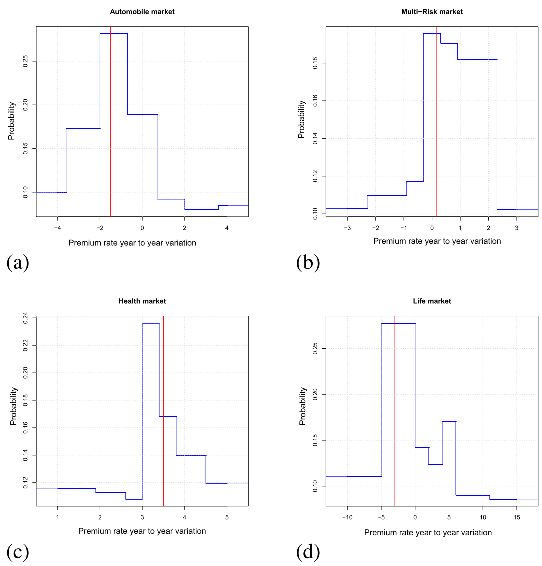

FIGURE 8. SPM results for the four prediction experiments in the Insurance sector. Final prices are normalized to provide the probability estimation for the different quantized levels of premium year-to- year variations.

the obtained results are quite accurate, selecting the right interval in three of the markets and slightly missing the Health market. However the forecast risk, gauged by the distribution of the assigned probabilities, is only correct in the Automobile and Life markets, while it is quite deficient for the other two.

B. HOMOLOGICAL PLAYER CLASSIFICATION

The application of the homological tools to the classification of players has been discussed for the APM in Section III. Here we follow a parallel discussion to approach the classification of human players.

It is worth observing some of the topological signatures for real people first. Figs. 9(a)-(h) show a random sampling of the topological signatures of players in the Life market, given by their barcodes and rotated diagrams. The first characteristic that should be noted is that real people tend to offer more complicated diagrams than artificial models, as shown by both the distribution of persistence bars for 0-dimensional invariants and a larger number of loops (in red).

Nevertheless, we have to resort to the W-d to make a meaningful classification. Table 2 illustrates the distances between Players 1,2,3 and 4 in the Life market. Their topological profiles are those in Figs. 9. In addition, Fig. 10 shows the dendrogram that summarizes the clustering between all the players in the Life market. In total, these figures display how the players are essentially grouped into two classes, as well as in the APM.

Comparing the definition of these two groups in the Life market and in the APM, and examining their W-d matrices, we observe the following pattern: There is a group of players, which corresponds to the predictor model in the APM,

TABLE 2. Wasserstain distance matrix for players 1,2, 3 and 4 in the life market.

	[,p1]	[,p2]	[,p3]	[,p4]
[p1,]	0.0000	0.2583	0.0090	0.2583
[p2,]	0.2583	0.0000	0.2507	0.0000
[p3,]	0.0090	0.2507	0.0000	0.2506
[p4,]	0.2583	0.0000	0.2506	0.0000

whose distances between them are essentially zero. When the real SPM is examined, the equivalent group tends to offer the simplest topological signatures, with essentially zero distances between the players in this group. We call it, the *predictor* group. The justification for this terminology is derived not only from the corresponding model in the APM, but also for its predictive capacity, as seen below.

For the other group, made up of random players in the APM, their W-d matrix shows how these players are essentially at a similar distance from the main group, and very close each other, but not as homogeneously as the players in the group of predictors. Their topological signatures also tend to be a little more complicated. In real prediction markets, this second group players also have very similar distances to the group of predictors, but among them they are even more dispersed than those of the APM, see Tables 1 and 2. Their topological signatures are the most complicated found in this study, see for example Figs. 9 (a) -(b).

Table 3 summarizes the classes found for all participants in the four markets. Each row represents a player. The first four columns stand for the prediction markets: 1 means that the player belongs to the main group, while -1 assigns the player to the second group; a 0 means that the player did not participate in a market. The last column simply counts the number of markets in which a player has contributed.

The number of participants in a market ranges from thirteen, in the Life market, to eighteen in the Auto market. These figures assert for the smallness of the markets in use. Approximately 40% of the players have participated in at least two markets, which allow us to obtain a preliminary idea of the player's perseverance to be in a particular group. Fig. 11 shows the number of markets that each expert has played in his/her typical class minus the number of markets where he/she changed, in percentage. In this figures, 50% means that the expert has played the same number of markets behaving *rationally* than *irrationally*. Only Players 1, 4 and 9, of the thirteen people that played at least two markets, adopt any kind of behavior with equal probability; while eight experts, out of thirteen, have always played as *rational* or *irrational*, but not as both.

C. PREDICTION IMPROVEMENT

Although the APM helps us to interpret the two groups of players according to their loopy behavior, the mere existence of these groups would not make sense from the engineering point of view if this classification does not turn into prediction improvements. This section examines this issue.

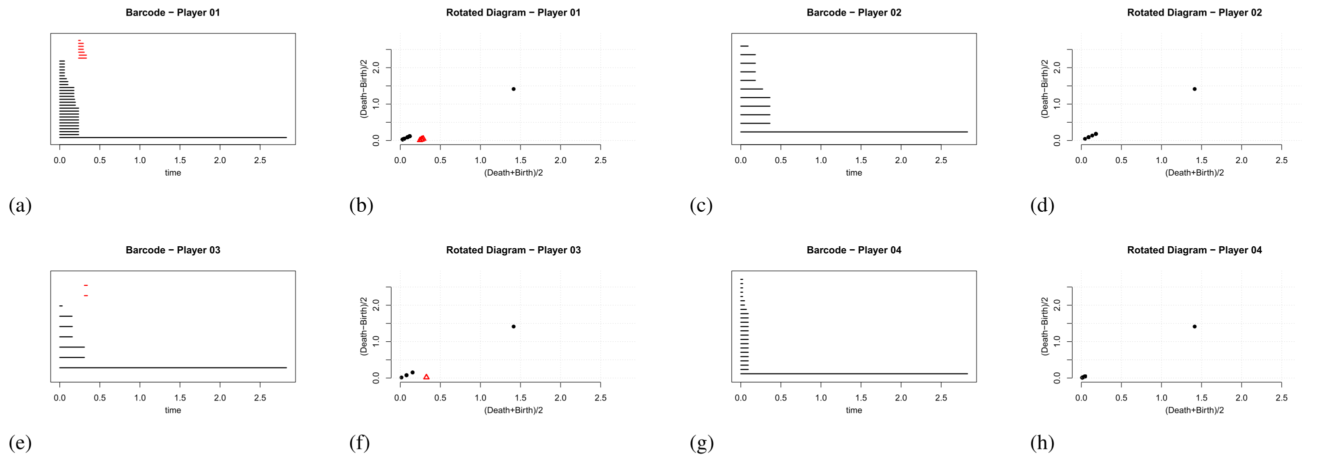

FIGURE 9. Barcodes and rotated diagrams for real experts in the Life market. Players 1 and 3 are classified as *irrational*, while Players 2 and 4 belong to the *rational* class.

TABLE 3. Player classification into class '1' or *rational*, and '−1' or *irrational*. '0' means that the expert has not played. The total number of markets where the expert has played is shown in the last column.

	Auto	Multi	Health	Live	Markets
1	-1	1	1	-1	4
2	-1	1	1	1	4
3	1	-1	-1	-1	4
4	-1	-1	1	1	4
5	1	1	1	1	4
6	0	0	0	-1	1
7	1	1	1	1	4
8	1	1	1	1	4
9	-1	-1	1	1	4
10	1	1	1	1	4
11	0	0	1	1	2
12	0	0	0	1	1
13	0	0	0	1	1
14	1	0	1	0	2
15	1	0	1	0	2
16	1	0	0	0	1
17	1	0	0	0	1
18	1	0	0	0	1
19	1	0	0	0	1
20	1	0	0	0	1
21	-1	0	0	0	1
22	1	1	0	0	2
23	0	1	0	0	1
24	0	1	0	0	1
25	0	1	0	0	1
26	0	-1	0	0	1
27	0	0	1	0	1
28	0	0	-1	0	1
29	0	0	-1	0	1
30	0	0	-1	0	1

Once a player's classification scheme is available, the forecast made by the PM can be adjusted using a partial aggregation of the player's estimates instead of the total market price. As proxies for these estimates, we use their normalized portfolios, interpreted as Bayesian probabilities in the outcomes. The aggregation of portfolios of a selected group of experts is carried out through a simple and effective linear rule [27].

Although accuracy in PMs has been discussed in the literature [53], the influence of market design, the stock

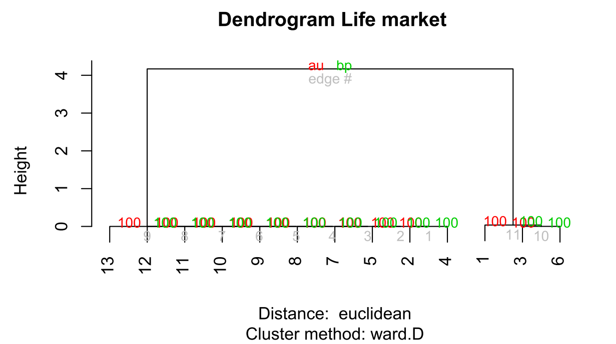

FIGURE 10. Dendrogram for the players in the Life market, showing the emerging of only two classes. The large group corresponds to the *rational* class.

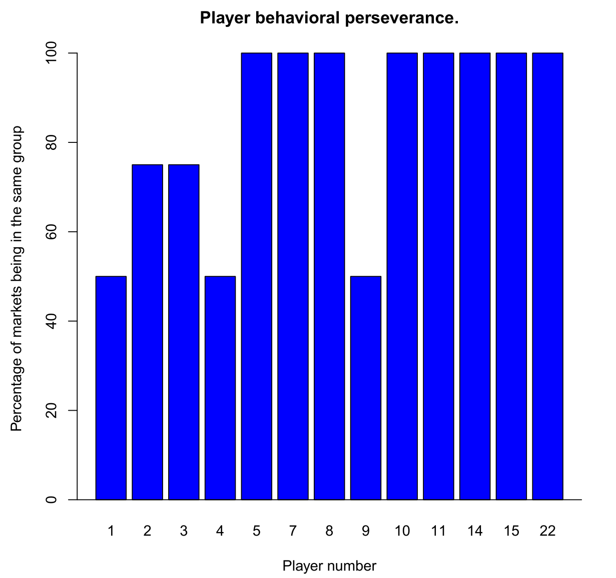

FIGURE 11. The behavioral perseverance of the experts playing in at least two of the four markets.

definition and even the forecasting objectives make the precise measurement of prediction errors a subject still debated [54], and beyond the scope of this paper. Since

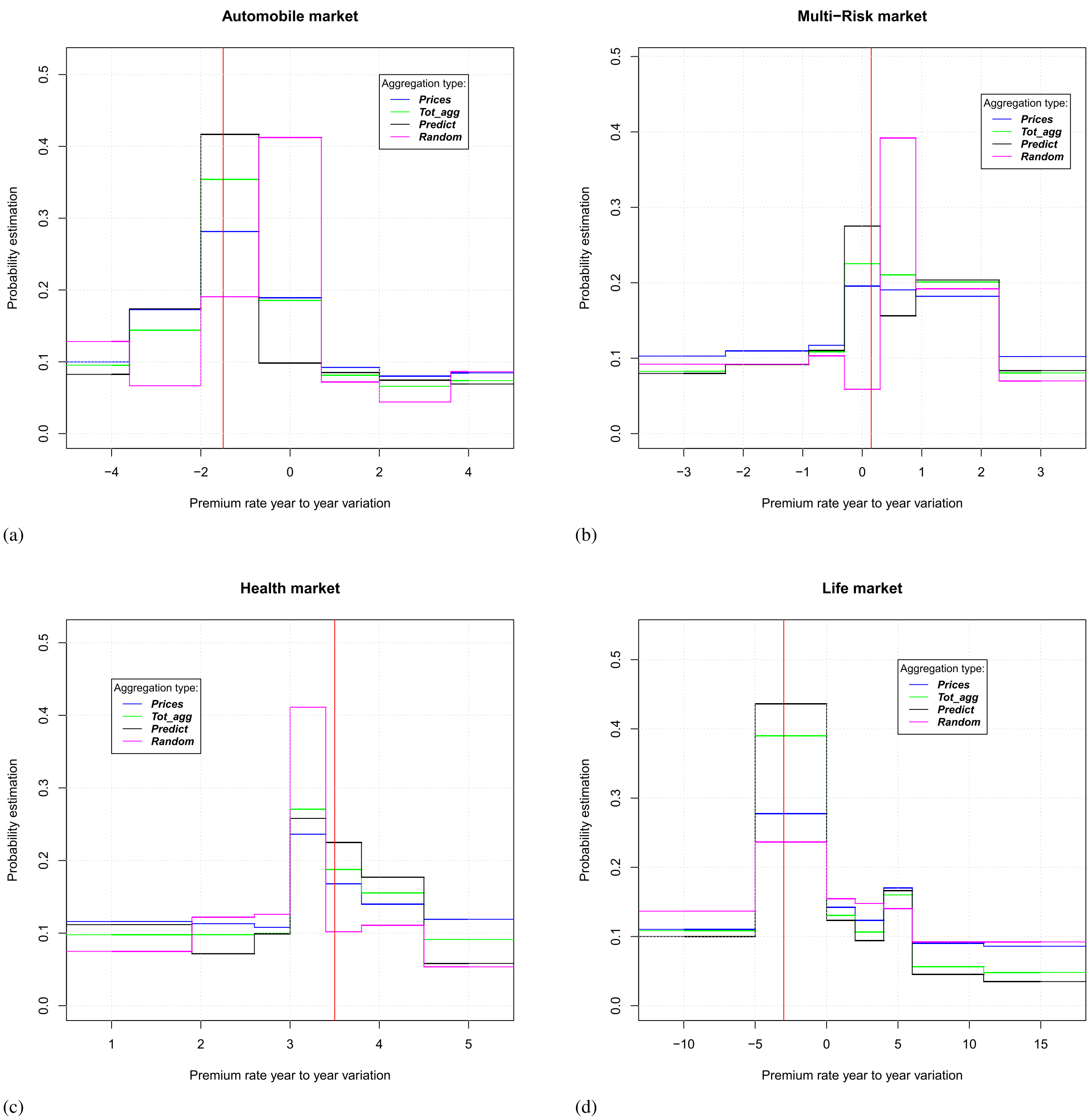

FIGURE 12. **Prediction performance in the four markets given by the prices and three aggregates of portfolios: the whole of the players, the *rational* players, and the *irrational* players.**

TABLE 4. **Prediction mean square errors for the *rational* and *irrational* players in the four SPM. The total aggregation and the prices are shown for comparison.**

Forecaster	Auto	Multi	Health	Live
Prices	5.40	1.92	0.70	44.91
Tot_agg.	4.72	1.68	0.61	32.03
Predict.	4.60	1.67	0.58	27.94
Random	5.02	1.70	0.54	45.64

our main objective is to evaluate the improvement of the prediction by selecting a particular group of experts, we have used typical error measurements: Table 4 shows the *Mean Squared Error* (MSE), and Table 5 shows the *Mean Absolute Error* (MAE). Both measurements assume an error vector defined as the difference of the mean values for each interval used in the definition of the stocks minus the actual data provided by the ICEA. The extreme intervals in the histograms are represented by their internal limit values. To calculate the MSE table, the squared values of the error vector are weighted successively by the probability estimates provided by the market prices and by the different aggregation of the groups of experts. The MAE table is calculated in a similar way but changing the square values of the error vector by their absolute values.

TABLE 5. Prediction mean absolute errors for the the *rational* and *irrational* players in the four SPM. The total aggregation and the prices are shown for comparison.

Forecaster	Auto	Multi	Health	Live
Prices	1.75	1.13	0.69	5.11
Tot_agg.	1.55	1.04	0.63	4.01
Predict.	1.47	1.02	0.60	3.65
Random	1.77	1.08	0.59	5.21

These tables together show the prediction advantage obtained first by considering the portfolios, instead of relying on the prices, and second by selecting the appropriate group. However, as the error for the Health market shows, they do not clearly reveal the gains of the different aggregations as a PM user can perceive them. Therefore, we prefer to discuss them directly from the histograms.

Fig. 12 displays the aggregates for markets and different groups of players: the black color codes for the main group, the predictors; the magenta for the second one, the participants in the random class; and in green we aggregate all of them linearly. The typical estimation directly from the prices is added in blue for comparison.

As can be seen in Fig. 12, the main group clearly improves the estimated probability assigned to the different outcomes compared to the second group. This becomes the main message of this paper: how the segmentation generated by their loopy behaviors selects the players with greater accuracy than the rest. This also motivates the name of *Predictors* for the first group, as it has been used throughout this paper. Although the data provided is quite limited, the comparison between the probabilities assigned by these two groups, 'Predictors' or 'Rational' vs. 'Random' or 'Irrational', suggests some behavioral conclusions about the players in PMs that can be exploited in their design and in future research [38].

Regarding the overall performance improvement due to the group of predictors, in all cases the probability estimation assigned to the correct answer is better for this group than for the entire market, represented by the final prices or even by the total linear aggregation of the portfolios; although no decision was changed, including the incorrect result for the Health market. However, a significant improvement is made in terms of the reliability of the forecast: in the markets where the answer is correct, the probability estimation associated with the intervals surrounding the correct one is quite diminished. Even in the Health market where the incorrect result is maintained, the difference in the probability estimation between the selected and the correct intervals decreases clearly with the Predictor group, as if to rise suspicions about the correctness of the chosen interval.

It is also worth comparing the estimate made by the total linear aggregation with the price estimate, in green and blue, respectively. In all markets, the decision made from the portfolio aggregation is better than that made from prices, which suggests that, for prediction purposes, prices are noisier signals than portfolios.

V. CONCLUSION AND SOME FURTHER RESEARCH SUGGESTIONS

In this paper we have identified two Homological invariants, namely the 0^{th} and 1^{st} *persistent Betti numbers*, which can classify experts into two groups based on their behavior in the decision-making process of a Prediction Market. A simulation of this market by artificial agents clarifies the nature of this classification: in terms of microeconomic jargon, they correspond to *rational* and *irrational* players. Finally, we experiment with four real SPMs with experts in the insurance sector. These markets allows us to relate theses classes of players with their prediction accuracy and to emphasize in some way how this classification depends on the behavior of the expert rather than on his personality. An additional contribution comes from the deconstruction of the market: here we open the way in which PM is aggregating expert information and shows how the prices seem to be nosier signals that the portfolios.

The obtained results suggest some applied research extensions: first, the application of this classification approach to other group decision-making processes where the individual history of the experts' behavior regarding their decisions may have a numerical representation. Second, the use of these categories in the simulation of markets in the field of Experimental Economics. Third, its translation to artificial devices for information fusion where these classes, if they exist, may be connected to malfunctioning or defective designs.

Finally, a theoretically oriented question raises from the apparent superior performance of the portfolios aggregation over the prices. The use of players' private information and the new aggregation mechanism fits well with the idea behind the *hidden profile effect*. A careful investigation of this relationship may open up new avenues for the deconstructed prediction markets.

REFERENCES

[1] L. Buchanan and A. O. Connell, "A brief history of decision making," *Harvard Bus. Rev.*, vol. 84, no. 1, p. 32, 2006.

[2] H. Olsson and J. Loveday, "A comparison of small crowd selection methods," in *Proc. 37th Annu. Meeting Cogn. Sci. Soc.*, D. C. Noelle *et al.*, Eds. Austin, TX, USA: Cognitive Science Society, Jul. 2015, pp. 1769–1774.

[3] M. Yousefnezhad, S.-J. Huang, and D. Zhang, "WoCE: A framework for clustering ensemble by exploiting the wisdom of crowds theory," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 486–499, Feb. 2018.

[4] Y. Nugraha, I. Brown, and A. S. Sastrosubroto, "An adaptive wideband Delphi method to study state cyber-defence requirements," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 1, pp. 47–59, Jan. 2016.

[5] G. Parekh *et al.*, "Identifying core concepts of cybersecurity: Results of two Delphi processes," *IEEE Trans. Educ.*, vol. 61, no. 1, pp. 11–20, Feb. 2018.

[6] M. Adnan and M. Afzal, "Ontology based multiagent effort estimation system for scrum agile method," *IEEE Access*, vol. 5, pp. 25993–26005, 2017.

[7] A. Faktor and M. Irani, "'Clustering by composition'—Unsupervised discovery of image categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1092–1106, Jun. 2014.

[8] X. Chen, X. Gong, L. Yang, and J. Zhang, "Amazon in the white space: Social recommendation aided distributed spectrum access," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 536–549, Oct. 2017.

[9] P.-Y. Chen, S.-M. Cheng, P.-S. Ting, C.-W. Lien, and F.-J. Chu, "When crowdsourcing meets mobile sensing: A social network perspective," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 157–163, Oct. 2015.

[10] F. Rosas, J.-H. Hsiao, and K.-C. Chen, "A technological perspective on information cascades via social learning," *IEEE Access*, vol. 5, pp. 22605–22633, 2017.
[11] B. Wu, X. Zhou, Q. Jin, F. Lin, and H. Leung, "Analyzing social roles based on a hierarchical model and data mining for collective decision-making support," *IEEE Syst. J.*, vol. 11, no. 1, pp. 356–365, Mar. 2017.
[12] J. Ma, Z.-P. Fan, Y.-P. Jiang, and J.-Y. Mao, "An optimization approach to multiperson decision making based on different formats of preference information," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 36, no. 5, pp. 876–889, Sep. 2006.
[13] G. Zhang, Y. Dong, Y. Xu, and H. Li, "Minimum-cost consensus models under aggregation operators," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 6, pp. 1253–1261, Nov. 2011.
[14] P. K. Kwok and H. Y. Lau, "A modified consensus-building methodology for reaching a group decision using minimum costs," *IEEE Access*, vol. 6, pp. 3509–3523, 2018.
[15] A. Timmermann, "Forecast combinations," in *Handbook of Economic Forecasting*, vol. 1, Kidlington, U.K.: Elsevier, 2006, pp. 135–196.
[16] P. Ernst, R. Pemantle, V. Satopää, and L. Ungar, "Bayesian aggregation of two forecasts in the partial information framework," *Statist. Probability Lett.*, vol. 119, pp. 170–180, Dec. 2016.
[17] M. Cai, Y. Lin, B. Han, C. Liu, and W. Zhang, "On a simple and efficient approach to probability distribution function aggregation," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 9, pp. 2444–2453, Sep. 2017.
[18] A. Sinha *et al.*, "Estimation and decision fusion: A survey," *Signal Process.*, vol. 71, nos. 13–15, pp. 2650–2656, Aug. 2008.
[19] S. E. Seker, "Computerized argument delphi technique," *IEEE Access*, vol. 3, pp. 368–380, 2015.
[20] J. E. Berg and T. A. Rietz, "Prediction markets as decision support systems," *Inf. Syst. Frontiers*, vol. 5, no. 1, pp. 79–93, 2003.
[21] M. Spann and B. Skiera, "Internet-based virtual stock markets for business forecasting," *Manage. Sci.*, vol. 49, no. 10, pp. 1310–1326, 2003.
[22] J. Wolfers and E. Zitzewitz, "Prediction markets," *J. Econ. Perspectives*, vol. 18, no. 2, pp. 107–126, 2004.
[23] E. Snowberg, J. Wolfers, and E. Zitzewitz, "Prediction markets for economic forecasting," in *Handbook of Economic Forecasting*, vol. 2. New York, NY, USA: Elsevier, 2013, pp. 657–687.
[24] P. Atanasov *et al.*, "Distilling the wisdom of crowds: Prediction markets vs. prediction polls," *Manage. Sci.*, vol. 63, no. 3, pp. 691–706, 2016.
[25] E. Y. Li, C.-Y. Tung, and S.-H. Chang, "User adoption of wisdom of crowd: Usage and performance of prediction market system," *Int. J. Electron. Bus.*, vol. 12, no. 2, pp. 185–214, 2015.
[26] G. Welty, "Problems of selecting experts for Delphi exercises," *Acad. Manage. J.*, vol. 15, no. 1, pp. 121–124, 1972.
[27] V. Genre, G. Kenny, A. Meyler, and A. Timmermann, "Combining expert forecasts: Can anything beat the simple average?" *Int. J. Forecasting*, vol. 29, no. 1, pp. 108–121, 2013.
[28] Z. Guo *et al.*, "Fine-grained recommendation mechanism to curb astroturfing in crowdsourcing systems," *IEEE Access*, vol. 5, pp. 15529–15541, 2017.
[29] Y. Luo, G. Iyengar, and V. Venkatasubramanian, "Social influence makes self-interested crowds smarter: An optimal control perspective," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 200–209, Mar. 2018.
[30] R. J. Gilson and R. Kraakman, "The mechanisms of market efficiency twenty years later: The hindsight bias," *J. Corp. Law*, vol. 28, pp. 715–742, 2002.
[31] T. Cowen and A. Tabarrok, *Modern Principles of Microeconomics*. Basingstoke, U.K.: Palgrave Macmillan, 2015.
[32] J. Patel, R. Zeckhauser, and D. Hendricks, "The rationality struggle: Illustrations from financial markets," *Amer. Econ. Rev.*, vol. 81, no. 2, pp. 232–236, 1991.
[33] G. Carlsson, "Topology and data," *Bull. Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.
[34] A. Barbu and N. Lay, "An introduction to artificial prediction markets for classification," *J. Mach. Learn. Res.*, vol. 13, pp. 2177–2204, Jul. 2012.
[35] F. Jahedpari *et al.*, "Online prediction via continuous artificial prediction markets," *IEEE Intell. Syst.*, vol. 32, no. 1, pp. 61–68, Jan. 2017.
[36] R. Ghrist, "Barcodes: The persistent topology of data," *Bull. Amer. Math. Soc.*, vol. 45, no. 1, pp. 61–75, 2008.
[37] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. Providence, RI, USA: AMS, 2010.
[38] G. R. Fréchette and A. Schotter, *Handbook of Experimental Economic Methodology* (Handbooks in Economic Methodologies Series). London, U.K.: Oxford Univ. Press, 2015.
[39] J. D. Christiansen, "Prediction markets: Practical experiments in small markets and behaviours observed," *J. Prediction Markets*, vol. 1, no. 1, pp. 17–41, 2012.
[40] D. Grainger, S. Sun, F. Watkin-Lui, and P. Case, "A simple decision market model," *J. Prediction Markets*, vol. 9, no. 3, pp. 41–63, 2015.
[41] P. R. Wurman, W. E. Walsh, and M. P. Wellman, "Flexible double auctions for electronic commerce: Theory and implementation," *Decis. Support Syst.*, vol. 24, no. 1, pp. 17–27, 1998.
[42] R. Hanson, "On market maker functions," *J. Prediction Markets*, vol. 3, no. 1, pp. 61–63, 2009.
[43] W. Chen, X. Li, and D. D. Zeng, "Simple is beautiful: Toward light prediction markets," *IEEE Intell. Syst.*, vol. 30, no. 3, pp. 76–80, May 2015.
[44] E. S. Rosenbloom and W. Notz, "Statistical tests of real-money versus play-money prediction markets," *Electron. Markets*, vol. 16, no. 1, pp. 63–69, 2006.
[45] R. D. Huang and H. R. Stoll, "Tick size, bid-ask spreads, and market structure," *J. Financial Quantitative Anal.*, vol. 36, no. 4, pp. 503–522, 2001.
[46] P. J. Giblin, *Graphs, Surfaces and Homology: An Introduction to Algebraic Topology* (Chapman and Hall Mathematics Series). Dordrecht, The Netherlands: Springer, 2013.
[47] S. Lang, *Algebra* (Graduate Texts in Mathematics), vol. 211. New York, NY, USA: Springer-Verlag, 2002.
[48] H. Edelsbrunner and J. Harer, "Persistent homology—A survey," *Contemporary Math.*, vol. 453, pp. 257–282, Feb. 2008.
[49] H. Edelsbrunner and D. Morozov, "Persistent homology: Theory and practice," Ernest Orlando Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep., 2012. [Online]. Available: https://escholarship.org/uc/item/2h33d90r
[50] B. T. Fasy, J. Kim, F. Lecci, and C. Maria. (2014). "Introduction to the R package TDA." [Online]. Available: https://arxiv.org/abs/1411.1830
[51] S. Lloyd, S. Garnerone, and P. Zanardi, "Quantum algorithms for topological and geometric analysis of data," *Nature Commun.*, vol. 7, p. 10138, Jan. 2016.
[52] P. P. Wakker, *Prospect Theory: For Risk and Ambiguity*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[53] J. Berg, F. Nelson, and T. Rietz, "Accuracy and forecast standard error of prediction markets," Tippie College Bus. Admin., Univ. Iowa, Iowa City, IA, USA, Tech. Rep., 2003.
[54] A. Emrouznejad, B. Rostami-Tabar, and K. Petridis, "A novel ranking procedure for forecasting approaches using data envelopment analysis," *Technol. Forecasting Social Change*, vol. 111, pp. 235–243, Oct. 2016.

CARLOS BOUSOÑO-CALZÓN received the Engineering and Ph.D. degrees in telecommunications from the Universidad Carlos III de Madrid, Madrid, Spain.

He was with Telefónica I+D, Madrid, Spain, and BT Research, Ipswich, U.K., while working for the Ph.D. He was an Assistant Professor with the University of Sevilla and then joined the Universidad Carlos III de Madrid, where he is currently an Associate Professor with the Department of Signal Theory and Communications. From 2007 to 2009, he served as a Seconded National Expert for the FET-Proactive Unit, European Commission. His research interests include information processing for complex systems and their social applications.

HAROLD MOLINA-BULLA (S'90–M'99) received the degree in electronic engineering from Pontificia Universidad Javeriana, Bogotá, Colombia, in 1994, and the Ph.D. degree (Hons.) in telecommunication technologies from the Universidad Carlos III de Madrid, Madrid, Spain, in 1999.

He is currently a Visiting Professor of electronic communications with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid. His current research interests include high performance computing for signal processing, advanced machine learning techniques, and their applications in big data.

JOSÉ JOAQUÍN ESCUDERO-GARZÁS received the Ph.D. degree in electrical engineering from the Universidad Carlos III de Madrid (UC3M).

For five years, he worked for Spanish telcos as a Provisioning Engineer and as the Head of the Communications Network Maintenance Department. He joined the Department of Signal Theory and Communications, UC3M, as a Lecturer in 2002, where he also held positions as Assistant and Visiting Professor. He was also a Post-Doctoral Fellow with Universitat Autònoma de Barcelona, a Research Associate with the University of Virginia, and a Visiting Scholar with the University of Florida. During these years, he has participated in over 20 research projects and taught over 15 undergraduate and graduate courses in the mobile and telecommunication fields. His research interests include optimization of wireless communication systems and resource management for 5G networks.

FRANCISCO J. HERRERA-GÁLVEZ received the M.Eng. degree in telecom engineering from the Universidad Politécnica de Madrid in 1981.

For over 25 years, he worked for Spanish telcos as different positions (Research and Development, strategic planning, mobile networks, and engineering). During these years, he has participated in different research and deployment projects. He joined the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, as a part time Professor Lecturer in 2005. His research interest is focused in the information markets.