

## Chapter 2

# Fundamentals

**Abstract** In this chapter, I introduce some fundamental knowledge in the field. My main focus is on how to design both controlled experiments and agent-based models. For showing the validity of the former, I present the Hayek hypothesis in advance; for clarifying the latter, I first present the El Farol bar problem and minority game. In addition, I also present both the information theory (with an emphasis on the Shannon entropy) and a nonparametric regression analysis (Hodrick-Prescott filter), which will be used in some other chapters.

**Keywords** Controlled experiments · Agent-based model · Hayek hypothesis · Minority game · Shannon entropy · Regression analysis

This chapter presents some fundamental knowledge or background that may help to understand the forthcoming chapters.

### 2.1 Hayek Hypothesis

Even nowadays, the typical method to study economic problems is the hypothetical–deductive method. Economists often derive an optimal situation of a system from certain assumptions, such as the complete knowledge of a preference system or information. F. A. Hayek (May 8, 1899–March 23, 1992; 1974 Nobel Prize winner in Economic Sciences) pointed out in his famous thesis “The Use of Knowledge in Society” published in 1945 [27] that this was totally a misunderstanding of social problems because no one could simply acquire the entire data of such assumptions. So despite the allocation problem under certain assumptions, a more important problem was how to obtain and use the decentralized resources and information.

Another thing economists always neglect is the specific knowledge of the individual. Other than scientific knowledge, this specific knowledge only gives its owner a unique benefit due to his/her own understanding of people, environment, and other special circumstances. That is, the exact part of knowledge economists put into the assumptions is equally important as scientific knowledge. As the comparative

stability of the aggregates cannot be accounted for by the “law of large numbers” or the mutual compensation of random changes and the fact that knowledge of this kind by its nature cannot enter into statistics, all plans should be made by the “man on the spot” rather than by any central authorities.

Since central authorities are limited, will the plans made by individuals reach a so-called equilibrium state? Hayek’s answer was “yes.” If all individuals follow the simple regulation of “equivalence of marginal rates of substitution,” which is the basic of microeconomics, the market will indeed be in an equilibrium state finally, without the necessity of the knowledge of the entire market. Under the “magical” market mechanism of price, once someone finds an arbitrage opportunity of a commodity, the price of this commodity will change. Thus the marginal rates of substitution of this commodity to other commodities change, causing another round of price change. This effect spreads to more and more kinds of commodities and gradually covers the whole market, though maybe no one knows why such changes happen. The whole acts as one market, not because any of its members survey the whole field, but because their limited individual fields of vision sufficiently overlap so that through many intermediaries the relevant information is communicated to all.

Actually, the price system is just what A. Smith (June 5, 1723–July 17, 1790) called “the invisible hand” [28], a mechanism for communicating information, and the most significant fact about this system is the economy of knowledge with which it operates, or how little the individual participants need to know to be able to take the right action. Even if people know about all the factors of a commodity, the actual price is not available unless obtained from a market with price system.

The above content is known as the Hayek hypothesis [27], which asserts that markets can work correctly even though the participants have very limited knowledge of their environment or other participants. Certainly, traders have different talents, interests, and abilities, and they may interpret data differently or be swayed by fads. However, there is still room for markets to operate efficiently.

In 1982, “the father of experimental economics” V. L. Smith (2002 Nobel Prize winner in Economic Sciences) [29] tried proving the Hayek hypothesis using 150–200 experiments under different circumstances which he thought the correct method to select a reliable theory. The trading behavior of the market participants led the market to a competitive equilibrium under a double auction regulation without any extra information (the participants, i.e., subjects, only knew their own value of the commodity and the market price), the result of which was contrary to the classical theory of price taking hypothesis and complete knowledge hypothesis.

A key characteristic of controlled experiments was its specific convertible supply and demand condition and the reward system to stimulate the subjects. Once the supply and demand are determined, the equilibrium market price is also determined and whether the market was operated well could be easily observed. Although all the experiments [29] had different subjects and supply and demand conditions, they all ended with the equilibrium state, whether in a stationary or dynamic environment. Although the experiments [29] were still not perfect, a reliable result related to the Hayek hypothesis was that the attainment of competitive equilibrium outcomes is possible under less stringent conditions.

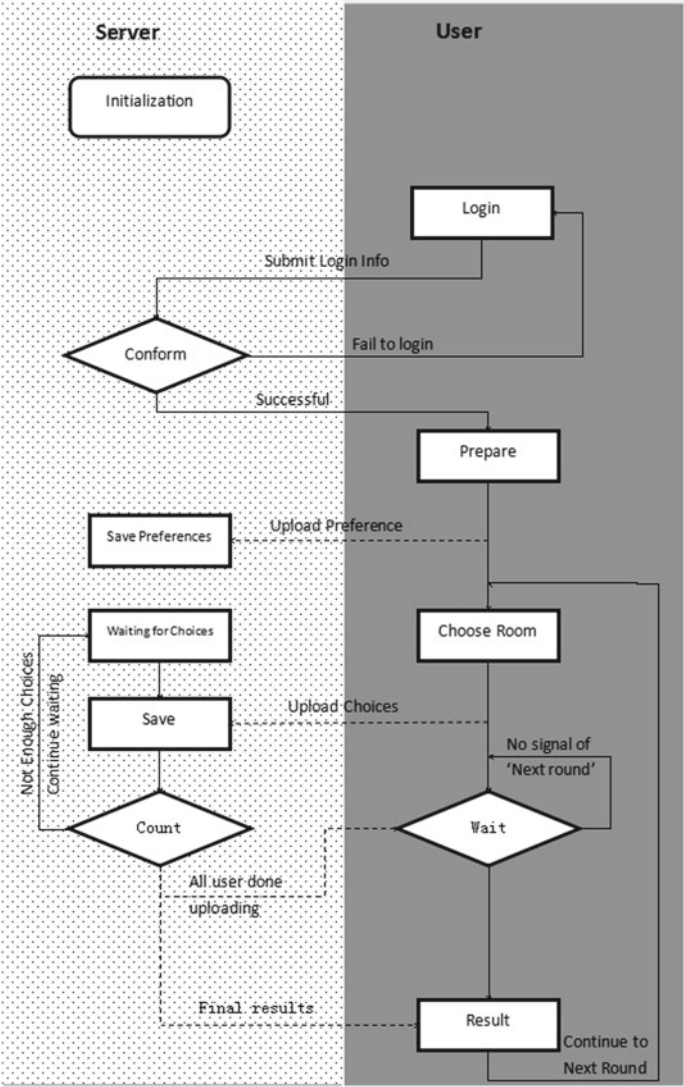
To sum, the Hayek hypothesis helps to account for the reliability of controlled experiments that contain finite subjects. Thus, this Hayek hypothesis may serve as a theoretical foundation for experimental economics [30] and experimental econophysics (introduced in this book), both of which are based on controlled human experiments in the laboratory. Besides, in the field of experimental econophysics, theoretical analysis also helps to validate and generalize results obtained from controlled experiments; details can be found in Sects. 1.1 and 1.5.

## 2.2 How to Design Computer-Aided Controlled Experiments

In the following chapters, I shall introduce the research progress on controlled experiments in the field of experimental econophysics. As a source of econophysics research ideas, the method of controlled experiments has become increasingly important in relative studies. Controlled experiments can be conducted in various ways. For example, in a simple experiment based on the minority game [1], the organizer may require subjects to close their eyes and raise hands to signal their choices between the two rooms in the game. Here, closing eyes prevents communications among the subjects so that it can be guaranteed that they make decisions independently. However, when regulations of an experiment become more complicated, or when the number of subjects becomes larger, computer-aided controlled experiments begin to show their efficiency as they can implement any experimental design easily and also collect experimental data and reveal real-time statistical results quickly.

A computer-aided controlled experiment needs programming of both the server and the client (Fig. 2.1). Two primary missions of the client are to distribute related information to users and also to provide users a channel to upload their own personalized choices. The server's main tasks are to store users' information, process users' uploaded data, and generate new information based on feedback from the behaviors of users. Generally, the client may be designed in the form of web pages to reduce costs and increase scalability. Any computer with a network connection can be easily set as a client. The servers of all the experiments in Chaps. 3–12 were constructed on the architecture of Linux + Apache + MySQL + PHP/Python (Readers can download a source code example from the link: <http://t.cn/zOIkLEk>).

Figure 2.1 shows a general schematic flowchart abstracted from the various experimental systems which will be introduced in Chaps. 3–12. Actually, the detailed designs of various systems are different in many aspects. For example, in the experiment of controlled laboratory stock market (Chap. 3), the server has to process every new order immediately since the time in the experiment is continuous. The herd experiment in Chap. 5 needs to add some robot agents (produced by a computer program) when computing the final outcome. The experiment of risk and return in Chap. 12 requires every subject to set their initial investment ratio, and the final outcome is not simply the “win” or “lose” but the final returns. Hence, we need to make minor revisions of the server and the web page accordingly for each specific system.



**Fig. 2.1** A schematic flowchart showing how to compile the computer program for conducting controlled experiments. Adapted from Ref. [26]

Figure 2.2 depicts some screenshots of web pages in a certain controlled experiment. Figure 2.3 shows a site photograph in a controlled experiment.



**Fig. 2.2** Web pages of a controlled experiment. **a** Background management interface, **b** login, **c** login successful, **d** choose a room, **e** waiting for the result, and **f** result. Adapted from Ref. [26]



**Fig. 2.3** A site photograph in a controlled experiment organized by my group on September 28, 2013

## 2.3 El Farol Bar Problem and Minority Game

Here, we first present the real El Farol bar problem [31], which gave birth to the minority game [1] discussed in Chaps. 5–8. An early summary of minority games can be found in the book by Challet et al. [21].

### 2.3.1 *El Farol Bar Problem*

The essence of formation of human social activities lies in the acquisitiveness for resources. In many social and biological systems, the agents always spontaneously adaptively compete for limited resources, and thus change their environments. In order to effectively describe the system with the complexity, scientists have made a series of attempts. Such a resource competition system is just a kind of complex adaptive system.

For economic systems, the basic issue appears as well. Generally, in an economic market, if the resources are rationally allocated the market is full of vitality. Otherwise, the development will be impeded at least to some extent. Thus, the allocation of resources is the most fundamental economic problem. As is known, most popular economic theories are related to deductive reasoning. According to these economic theories, as long as all individuals are almost smart, everyone will choose the best action, and then each individual can reason his/her best action.

However, people gradually find that in real life, individuals often have no complete rationality and superb deductive reasoning ability when making decisions. Instead,

it is common for them to simply use a feasible method of trial and error. Therefore, it looks like inductive generalization and continuous learning when real individuals make decisions (namely, inductive reasoning).

In the game theory, researchers often use evolutionary games to study the similar dynamic process. However, when using evolutionary game models, economists usually do not take into account the character of limited rationality. Therefore, they cannot convincingly yield interesting phenomena and critical phase transition behaviors. A social human system contains a large number of agents who have the limited ability for inductive reasoning. Even so, the microscopic simplicity can still lead to the complexity of the macroscopic system. Obviously, from a physical point of view, this system has a variety of statistical physical phenomena.

In the past, there were studies on the allocation of resources. For example, in 1994, economist W. B. Arthur put forward a very representative resource allocation problem, the El Farol Bar problem, when he studied the inductive reasoning and bounded rationality. [31] It can be described as follows.

There is the El Farol bar in Santa Fe (a city in New Mexico of United States) which offers Irish music every Thursday night. Each Thursday, 100 persons (here 100 is only set for concreteness) need to decide independently whether to go to this bar for fun or stay at home because there are only 60 seats in the bar. If more than 60 persons are present, the bar is so crowded that the customers get a worse experience than staying at home. If most people choose to stay at home on that day, then the people who go to the bar enjoy the elegant environment and make a wise choice.

In this problem, Arthur assumed no communication in advance among the 100 persons. They only know the historical numbers from the past weeks and have to make decisions independently. In order to make a wise choice, each person needs to possess his own strategies which are used to predict the attendance in the bar this week. People cannot obtain the perfect equilibrium solutions at initial time when making decisions. They must consider others' decisions, and keep learning according to the limited historical experience in their mind. The elements of inductive reasoning and limited rationality in the El Farol bar problem lay the foundation for the further development of modeling in econophysics, as shown in Sect. 2.3.2.

### 2.3.2 *Minority Game*

Inspired by the above El Farol bar problem, physicists D. Challet and Y. C. Zhang in 1997 proposed a minority game to quantitatively describe this problem and statistically analyzed the emerging collective phenomena in complex adaptive systems [1]. In the following years, scientists did extensive research on the minority game and its applications in different fields, which have significantly promoted the development of econophysics. [20, 21] We introduce the minority game model as follows.

There are two rooms (indicated as Room *A* and Room *B*) and  $N$  agents, where  $N$  is an odd number. Each agent chooses independently to enter one of the two rooms. If one room contains fewer agents than the other, then the agents in this



**Table 2.1** A model strategy table in the minority game with memory length  $m = 2$

Information	Choice
00	1
01	0
10	1
11	0

room win. That is to say, the minority wins. The two rooms in the minority game actually correspond to the case of unbiased distribution of two resources. This game is repeated. Each agent can only make a decision next time according to the historical information. As a matter of fact, in daily life, people often face similar choices. Examples include choosing which road to avoid traffic jam during rush hours and choosing a less crowded emergency exit to escape. Although each of us can keep learning from limited historical experiences, it does not guarantee that we will make a correct choice every time.

In the minority game, the decision-making process which is based on historical information is modeled to form strategy tables. For the minority game, one assumes that agents' memory length of the historical information is limited. Each agent can only remember the latest  $m$  rounds. If  $m = 2$ , it can form a strategy as shown in Table 2.1. The historical information in the left column records the attendance in the past two rounds, which is filled with a string of bits of 0 and 1. For example, a string of "10" represents the past two winning rooms, Room A and Room B. The right column is the prediction which is filled with bits of 0 or 1. Bit 1 is linked to the choice of Room A for entrance, while bit 0 to that of Room B. So one can obtain a strategy pool with a size of  $2^{2^m}$ . As  $m$  increases, the total number of strategy tables increases rapidly. In the original minority game model, the designers let each agent randomly select strategy tables. That is, the right column of each strategy table is randomly filled with 0 or 1. These agents are likely to repeat the same selected strategy (namely, the right columns of the strategy tables are the same). However, appropriately increasing the memory length can significantly reduce the repetition probability. Here, it is worth noting a special case: if the right column of a strategy table is all 1 (or 0), this strategy means that the agents are always locked into Room A (or B) no matter what happens.

According to these results, it is not hard to find that the minority game model with such a strategy structure is closely related to memory length  $m$ . And the historical information can only increase with  $2^m$ .

In econophysics, minority game models have been widely used to simulate a special kind of complex adaptive systems, the stock markets [20, 32, 33]. Researchers always hope to generate similar stock market data through the minority game model. The stand or fall of this similarity often needs to be tested to see whether model data have the same stylized facts as the real market data. Besides, the minority game can also be used to study competition problems about an unbiased distribution of resources [3, 34–36].



## 2.4 How to Design Agent-Based Models

Agent-based modeling [37] plays an important role in the progress of complex systems researches. Differing from stochastic equations, agent-based models try to regenerate the evolution of a complex system via a bottom-up approach by means of simulating the behaviors of plentiful homogeneous or heterogeneous agents at the micro scale. There are two general ideas that can guide the design of a particular agent-based model. I shall discuss the two ideas in the following two subsections.

### 2.4.1 *Modeling by Abstracting Real-World Systems*

The minority game is a famous agent-based model in the field of econophysics. As one can see from Sect. 2.3, it originates from the El Farol bar problem. There have been many modifications on the minority game. A particular one is to model on the stock market [20]. It can be seen that the minority game on stock market has many simplifications compared to the real market, such as the adoption of linear relation between excess demand and price change, the neglect of transaction costs, etc. Even under such simplifications, the minority game can still reproduce many statistical characteristics of the stock market successfully [20], which gives a clear illustration of the capabilities of agent-based models to reveal the endogenous mechanisms under financial markets. Hence, one important idea when trying to build an agent-based model is to abstract real-world systems. First, regulations in the associated real-world system should be written down one by one. Second, the importance of each regulation should be evaluated; key regulations should be introduced into the model, while trivial ones can be eliminated to make the model simple and clear. Third, decision-making process for the virtual agents should be carefully designed to mimic the behaviors of real-world humans. Finally, one can complete the designs of an agent-based model by combining the simplified structure and a large number of interacting virtual agents. It can be seen that this idea guides our designs of all the models appearing in Chaps. 5–11.

One more thing I wish to discuss is the simplifications made in the agent-based modeling method. In real financial markets, it can be seen that the price of an asset is the reflection of every market participant's information-collecting and decision-making abilities. Moreover, there also exist communications among participants that may lead to the herding phenomenon. Clearly, if we want to include all these factors into our agent-based model, the model can become much more complicated. So we have to compromise by simplifying the model accordingly. But does our model lose its generality and reasonability? One may refer to the famous Ising model in statistical physics. In the Ising model, the time is discrete, which is clearly an unreal assumption in our real world. However, the model can regenerate many ferromagnetic phenomena very successfully. Hence, we can conclude that a proper simplification can wipe off trivial factors in the real-world system and make the model more

powerful in explanations of the real-world phenomena. But as we know, making a good simplification is not always a simple task.

Next, we discuss the second approach for building agent-based models, that is, building models through physical models.

### 2.4.2 Modeling Through Borrowing from Physical Models

Since many physical models have already been proven proper to explain related natural phenomena, it is worth academic research to extend them to economic or social systems. Here, the Ising model is taken as an example for illustrating this idea.

The Ising model aims at studying the temperature dependence of magnetic susceptibility during the ferromagnetic phase transition. The model contains a large number of interacting spins which form a certain topological structure. It is usually assumed that (1) each spin only has two states, i.e.,  $\sigma_j = +\frac{1}{2}$  for the up direction and  $\sigma_j = -\frac{1}{2}$  for the down direction; (2) the range of spin interactions is limited to the first neighborhood [38].

The fundamental physical picture of the ferromagnetic phase transition is that when increasing temperature, the dominant state (up or down) shown in the overall lattice changes through spin interactions. On one hand, the principle of least action requests that all spins are aligned in the same direction so that the spin interactions are at the lowest level. On the other hand, thermal motions tend to drive the directions of the spins to a random arrangement at which the system's entropy is the largest. The probabilities of spin states obey the Boltzmann distribution. As long as the system's temperature gets higher than the Curie temperature, the thermal motions among the spins become dominant so that a phase transition occurs from ferromagnetic to paramagnetic. For a spin, suppose its energy is  $E_+$  for the up state (i.e.,  $\sigma_j = +\frac{1}{2}$ ), and  $E_-$  for the down state ( $\sigma_j = -\frac{1}{2}$ ). When  $E_- > E_+$ , the probability for the spin to be in the up state  $|\frac{1}{2}\rangle$  in the next time step is denoted as  $p_+$ , and in the opposite state  $|\frac{1}{2}\rangle$  with a probability of  $p_-$ . Then we can write,

$$p_+ = \frac{e^{-\frac{E_+}{kT}}}{e^{-\frac{E_+}{kT}} + e^{-\frac{E_-}{kT}}}, p_- = \frac{e^{-\frac{E_-}{kT}}}{e^{-\frac{E_+}{kT}} + e^{-\frac{E_-}{kT}}}.$$

It can be seen that when the temperature  $T$  gets closer to 0,  $p_+ \rightarrow \frac{E_+}{E_+ + E_-}$ ,  $p_- \rightarrow \frac{E_-}{E_+ + E_-}$ , now all the spins tend to be arranged in the same direction. If  $T$  becomes much large,  $p_+ \rightarrow p_- \rightarrow \frac{1}{2}$ , which means that the directions of spins become totally random. That is, the system transfer from ferromagnetic to paramagnetic as  $T$  increases.

To have a further discussion of the Ising model, here we take the two-dimensional orthogonal cubic lattice for example. The Hamiltonian can be composed of two parts:

$$H(\sigma_j) = - \sum_i J_{i,j} \sigma_i \sigma_j - h_j \sigma_j.$$

Here, the first part  $J_{i,j} \sigma_i \sigma_j$  represents interactions between spins (near-field or local effect), where spin  $i$  belongs to the first neighborhood of spin  $j$ . The second part  $h_j \sigma_j$  is the Hamiltonian of  $\sigma_j$  in the external magnetic field  $h_j$  (global effect). According to the principle of least action, every spin tends to choose states that obey the Boltzmann distribution. Under this distribution, the probability of being the least energy state for  $H(\sigma_j)$  is the largest. The state probability is

$$P(\sigma_j) = \frac{e^{-\beta H}}{\sum_{\sigma_j} e^{-\beta H}},$$

where  $\beta = k_B T$ .

In the simulations, a spin in the lattice is selected randomly and the state of the spin is adjusted on the basis of thermodynamical laws. By repeating this procedure, the system can finally reach equilibrium. At equilibrium, one macroscopic property of the system simply equals the ensemble average of the associated microscopic property among all the spins. Note that  $P(\sigma_j)$  is not independent of  $T$ , thus we can obtain the relation between one observation of the value of  $E$  and the temperature  $T$  or other parameters:

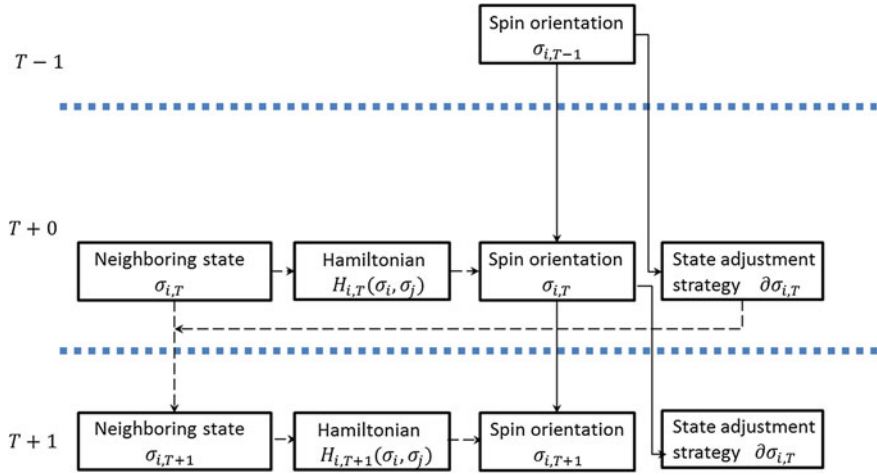
$$E(f) = \sum_j f(\sigma_j) P(\sigma_j).$$

A logical framework of the Ising model is shown in Fig. 2.4.

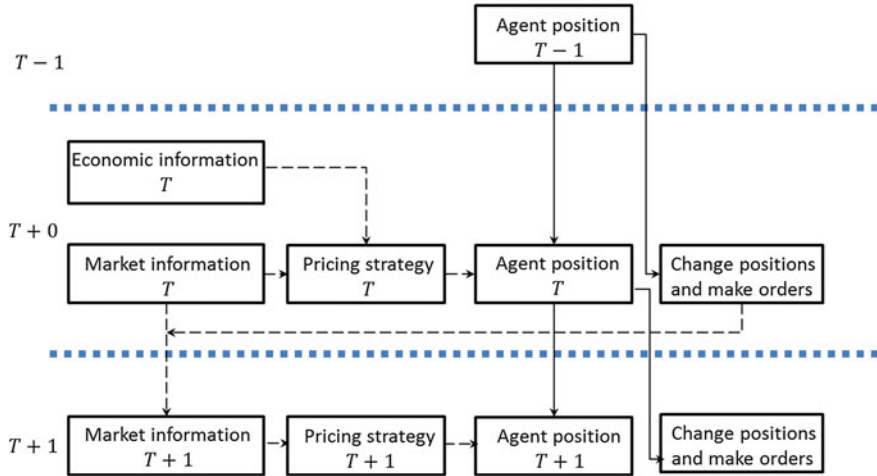
By comparing the similarities between ferromagnetic lattice and financial markets, one can establish an agent-based model for financial markets based on the logical framework of the Ising model (Fig. 2.5). The analogy between financial markets and ferromagnetic lattice is obvious.

First, from the point of interactions, in the Ising model, the spin states depend on the combined result of both global effect from the macro-external magnetic field and local effect from the micro-neighboring spin states. And in the financial markets, investing strategies made by market participants also depend on the combined result of global information such as the price and trading volume of an asset, fundamental market information (like GDP or CPI), and local information from the “neighboring” traders (here “neighboring” means the first neighborhood in the social network appearing in financial markets for a trader). Hence, we may write the associated “Hamiltonian” at one lattice grid for traders as  $H(\sigma_j) = - \sum_i J_{i,j} \sigma_i \sigma_j - h_j \sigma_j$ .

Second, from the point of “strategies,” in the Ising model, spins always tend to change their states to find the best one that confirms the least action and the maximum entropy principles; similarly, in the financial market model, agents should also be able to modify their investing strategies to find the best one that brings in maximum



**Fig. 2.4** Two-dimensional ferromagnetic phase transition in the Ising Model. Adapted from Ref. [26]



**Fig. 2.5** Agent-based financial market model borrowing from the framework of the Ising Model in Fig. 2.4. Adapted from Ref. [26]

returns with minimum risks. Third, from the point of feedback process, in the Ising model, the macromagnetic susceptibility can be obtained by summing up all the microspin states; accordingly, in the financial market model, through the match of orders executed by market makers, price at each time is generated.

Thus, we can see that there exists a deep analogy between physical and economic models.

The financial market models based on the Ising model's framework have received preliminary success [39, 40]. But what strategies should be adopted by agents so that the real-world phenomena in financial markets can be regenerated? What kind of payoff function is able to reflect the investment demands properly? Do the time scales at which agents make decisions impact on global information? All these questions are still waiting to be answered in the future.

### ***2.4.3 How to Test the Reliability of Agent-Based Models***

Whether the agent-based model is successful or not depends on whether it can stand empirical and experimental tests. Despite that it seems reasonable for the two ideas of building an agent-based model introduced in Sects. 2.4.1 and 2.4.2, yet the process does not equal results. The logicality and rationality of the course of establishing a model cannot prove that the final model is perfectly correct and dependable.

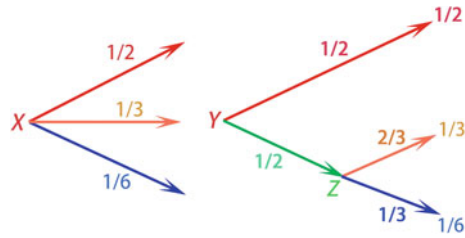
Compared with experimental observations, it is the only way to test whether an agent-based model is reliable. The experimental observation mentioned here can be the induction and analysis of not only the existing economical or financial data but also the controlled experiment data. The former corresponds to empirical econophysics [14], however, the latter corresponds to experimental econophysics (topic of this book). From this, it can be seen that agent-based modeling is a fundamental method and it may have stronger development in the future.

## **2.5 Information Theory**

### ***2.5.1 Initial Remarks***

In books on statistical mechanics, information theory is a chapter that cannot be ignored. At first look, you may wonder how information theory is linked to statistical physics. They seem like two totally different scientific fields. In this section, I give a brief introduction to the theory. Through it, you can see how information theory borrowed the concept of entropy from physics and then how physicists used information theory to reinterpret their statistical researches and further developed information theory. What is more, inspired by information theory, physicists also gave new explanations to the long-hunted Maxwell's demon. Needless to say, in econophysics, information theory can have applications as well; see Chap. 9.

**Fig. 2.6** Illustration of the composition law of  $H$ . Adapted from Ref. [19]



## 2.5.2 Shannon Entropy: Historical Beginning and the Unit of Information

### 2.5.2.1 Historical Beginning

In 1948, C. E. Shannon (April 30, 1916–February 24, 2001) published his classic article “A mathematical theory of communication” [41] in the *Bell System Technical Journal* which established the discipline of information theory. Obviously, the article title showed no relation to the physical world. In this article, Shannon discussed how to quantitatively analyze the amount of uncertainty in a discrete source of information.

Suppose for an experiment, there are  $n$  possible outcomes each of which happens at a rate of  $p_1, p_2, \dots$ , and  $p_n$ , respectively. Under different probability distributions of the outcomes, the experiment contains various amounts of uncertainty. To compare the uncertainty quantitatively, a measure, denoted as  $H(p_1, p_2, \dots, p_n)$ , should be found. Shannon suggested the expression of  $H$  having the following three properties:

1.  $H$  should be a continuous function of the  $p_i$ 's.
2. If all the  $p_i$ 's are equal, i.e.,  $p_i = 1/n$ ,  $H$  should increase monotonically with  $n$ . This is easy to understand, because with equally likely outcomes, there is more uncertainty when there are more possible outcomes.
3. The composition law: if an experiment is broken into two successive experiments, the original  $H$  should be the weighted sum of the two individual values of  $H$ . Figure 2.6 shows an illustration of the composition law. The single experiment  $X$  on the left has three outcomes with probabilities of  $p_1 = 1/2$ ,  $p_2 = 1/3$ , and  $p_3 = 1/6$ . For the two successive experiments on the right, the experiment  $Y$  has two outcomes, each with a probability of  $1/2$ , and if the second occurs, the other experiment  $Z$  takes place with two outcomes under probabilities  $2/3$  and  $1/3$ . It can be seen that the final outcomes of the two successive experiments  $Y$  and  $Z$  have the same probabilities as the ones in the single experiment  $X$ . Then, the composition law requires, in this special case, that  $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$ . Here, the coefficient  $\frac{1}{2}$  in the equation is because the experiment  $Z$  only takes place half the time.

To satisfy the three properties above, Shannon showed that the simplest form for  $H$  is

$$H(p_1, p_2, \dots, p_n) = \sum_{i=1}^n f(p_i), \quad (2.1)$$

where  $f(p_i)$  is a continuous function due to the first property. Considering that the expression of  $H$  is universal for any set of probability distributions  $\{p_1, p_2, \dots, p_n\}$ , we can deduce  $H$  under the special case of equal probabilities, namely,  $p_i = 1/n, \forall i$ . Then, Eq. (2.1) gives

$$H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = nf\left(\frac{1}{n}\right). \quad (2.2)$$

Suppose a single experiment has  $n$  outcomes with equal probabilities of  $1/n$ . It breaks into two successive experiments: the first one has  $r$  evenly happened outcomes and for each of its outcomes, the second occurs sequentially with  $s$  evenly happened outcomes. And there is  $n = rs$ . Then, the composition law requires

$$H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = H\left(\frac{1}{rs}, \frac{1}{rs}, \dots, \frac{1}{rs}\right) = H\left(\frac{1}{r}, \frac{1}{r}, \dots, \frac{1}{r}\right) + H\left(\frac{1}{s}, \frac{1}{s}, \dots, \frac{1}{s}\right). \quad (2.3)$$

Plugging Eq. (2.3) into Eq. (2.2) yields

$$rsf\left(\frac{1}{rs}\right) = rf\left(\frac{1}{r}\right) + sf\left(\frac{1}{s}\right). \quad (2.4)$$

Let

$$g(M) = \frac{1}{M}f(M), \quad (2.5)$$

then Eq. (2.4) becomes

$$g(RS) = g(R) + g(S), \quad (2.6)$$

where  $R = 1/r$ ,  $S = 1/s$ . Differentiating Eq. (2.6) with respect to  $R$  or  $S$  yields the following two equations accordingly:

$$Sg'(RS) = g'(R), \quad (2.7)$$

$$Rg'(RS) = g'(S), \quad (2.8)$$

where  $g'(M)$  means differentiating  $g(M)$  with respect to  $M$ . So we obtain

$$Rg'(R) = Sg'(S). \quad (2.9)$$

Because  $R$  and  $S$  are two independent variables, Eq. (2.9) gives

$$Mg'(M) = Rg'(R) = Sg'(S) = A, \quad (2.10)$$



where  $A$  is a constant. Integrating Eq. (2.10) for the unspecified variable  $M$ , we can get the general expression of function  $g(M)$

$$g(M) = A \ln(M) + C, \quad (2.11)$$

where  $C$  is also a constant and  $\ln(M)$  gives the natural logarithm of  $M$ . Plugging Eq. (2.11) into Eq. (2.5) and letting  $M = \frac{1}{n}$ , we obtain

$$f\left(\frac{1}{n}\right) = -\frac{A}{n} \ln(n) + \frac{C}{n}. \quad (2.12)$$

Now we should try to find the values of the two constants,  $A$  and  $C$ . On the boundary condition  $n = 1$ , the experiment only has one sure outcome. Therefore, now the uncertainty should be zero. From Eqs. (2.2) and (2.12), we have  $H(1) = f(1) = C = 0$ . Hence, for equal probabilities, there is

$$H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = -A \ln(n). \quad (2.13)$$

Now, let us turn to the second property of  $H$ . It is said that  $H$  should increase with  $n$  monotonically, hence  $\frac{dH}{dn} = -\frac{A}{n} \geq 0$ , which means  $A \leq 0$ . It is obvious that  $A \neq 0$ . So, we let  $K = -A$  with  $K$  being a positive constant. Then we obtain the general expression of function  $f(p_i)$ , namely

$$f(p_i) = -K p_i \ln(p_i). \quad (2.14)$$

Back to Eq. (2.1), now we can write the final form of  $H$  that satisfies the three properties,

$$H(p_1, p_2, \dots, p_n) = -K \sum_{i=1}^n p_i \ln(p_i), \quad (2.15)$$

where  $K$  is a positive constant.

Now let us take a break from information theory and turn our attention to the physical world. Consider a system that contains  $N$  distinguishable particles obeying Boltzmann statistics. Assume there are  $n$  nondegenerate quantum states, then we can obtain the entropy of the system expressed as [42],

$$S = -kN \sum_{j=1}^n \left(\frac{N_j}{N}\right) \ln\left(\frac{N_j}{N}\right), \quad (2.16)$$

where  $k$  is the Boltzmann's constant and  $N_j$  stands for the number of particles at the  $j$ th state. If we link  $\frac{N_j}{N}$  to the probability  $p_j$  which means that in average  $N_j$  particles are in the  $j$ th state, Eq. (2.16) becomes

$$S = -kN \sum_{j=1}^n p_j \ln(p_j). \quad (2.17)$$

Compared with Eq. (2.15), we can see the relation between  $H$  and  $S$  is

$$S = \frac{kN}{K} H. \quad (2.18)$$

In Shannon's 1948 article [41], he wrote "The form of  $H$  will be recognized as that of entropy as defined in certain formulations of statistical mechanics... We shall call  $H = -\sum p_i \log p_i$  the entropy of the set of probabilities  $p_1, \dots, p_n$ ." This is where the information theory and Shannon entropy all began.

### 2.5.2.2 The Unit of Information

In the expression of Shannon entropy, i.e., Eq. (2.15),  $K$  is still left as a constant with no specific value. In information theory, a common unit for Shannon entropy is *bit*, which is a contraction of *binary digit*. One bit is typically defined as the uncertainty in one time of coin toss that has equally likely outcomes (i.e., heads or tails). Hence, in this case, there is  $H(\frac{1}{2}, \frac{1}{2}) = K \ln(2) = 1$ , which yields  $K = 1/\ln(2)$ . Then the expression of Shannon entropy becomes

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2(p_i). \quad (2.19)$$

Now we can calculate the Shannon entropy for some specific cases. Here is one example. For a decimal digit, suppose it can choose a value from 0 to 9 with equal probabilities, i.e.,  $p = \frac{1}{10}$ . Then, using Eq. (2.19), it can be calculated that  $H = \sum_{i=1}^{10} \frac{1}{10} \log_2(10) = \log_2(10) = 3.32$ . Therefore, the decimal digit contains 3.32 bits of uncertainty.

## 2.5.3 When Information Meets Physics: The Principle of Maximum Entropy and the Fight with Maxwell's Demon

### 2.5.3.1 The Principle of Maximum Entropy

Now we have seen that Shannon entropy and the entropy of statistical mechanics have similar forms. Is this a coincidence? The answer is no. In physics, entropy is a measure of system's disorder. The word "disorder" means there is a lack of "information" for us to know the exact physical state of the system. Hence, in other words, we can say entropy is the amount of additional information needed to specify

the exact state of a system. This shows a kind of qualitative relationship between the two concepts of entropy. In 1957, E. T. Jaynes [43] further developed information theory by expounding the principle of maximum entropy, and then reinterpreted statistical mechanics through the viewpoints of information theory. Reading this article, we can see a deeper connection between the two fields.

The principle of maximum entropy states that given only partial information of a system, the probability distribution with the largest entropy is the least biased estimate possible for its current state. Suppose for a particular variable  $x$ , it can have  $n$  discrete values  $x_i$ ,  $i = 1, 2, \dots, n$ . If one knows that the mean value of  $x$  is  $\bar{x}$ , the restrictions on the unknown probability distribution  $\{p_1, p_2, \dots, p_n\}$  of  $x$  now are

$$\bar{x} = \sum_{i=1}^n p_i x_i, \quad (2.20)$$

$$\sum_{i=1}^n p_i = 1. \quad (2.21)$$

So, now the principle of maximum entropy tells us that the most proper values of  $\{p_1, p_2, \dots, p_n\}$  are those that maximize the Shannon entropy  $H$  in Eq. (2.15). Using the method of Lagrange multipliers [42], we can finally get the expression for  $\{p_1, p_2, \dots, p_n\}$ , i.e.,

$$p_i = \frac{e^{-\mu x_i}}{Z}, \quad (2.22)$$

where  $Z$  is the partition function,

$$Z \equiv \sum_{i=1}^n e^{-\mu x_i}, \quad (2.23)$$

and  $\mu$  can be derived from the expression

$$\bar{x} = -\frac{\partial}{\partial \mu} \ln(Z). \quad (2.24)$$

Inserting Eq. (2.22) into Eq. (2.15) leads to the maximal Shannon entropy

$$H_{\max} = K\mu\bar{x} + K\ln(Z). \quad (2.25)$$

Now let us turn to statistical mechanics. For the physical system containing  $N$  distinguishable particles described in Sect. 2.5.2, the expression of entropy is given in Eq. (2.16). Here, assume the associated energy level for each state is  $\varepsilon_j$ ,  $j = 1, 2, \dots, n$ , and the average energy of the system is  $U$ . Then, we can follow the same steps above to obtain the most proper particle distribution using the principle of maximum entropy, which is

$$\frac{N_j}{N} = \frac{e^{-\beta \varepsilon_j}}{Z}, \quad (2.26)$$

where the partition function  $Z$  is

$$Z = \sum_{j=1}^n e^{-\beta \varepsilon_j}. \quad (2.27)$$

It is obvious that Eqs. (2.26) and (2.27) fit the Boltzmann distribution for nondegenerate energy states. And  $\beta = 1/kT$  where  $k$  is the Boltzmann constant. Now the entropy of the physical system is maximal, which is

$$S_{\max} = \frac{U}{T} + Nk \ln(Z). \quad (2.28)$$

Note that Eq. (2.28) also has a similar form with Eq. (2.25).

Hence, in Ref. [43], E. T. Jaynes stated that “If one considers statistical mechanics as a form of statistical inference rather than as a physical theory, it is found that the usual computational rules, starting with the determination of the partition function, are an immediate consequence of the maximum-entropy principle.”

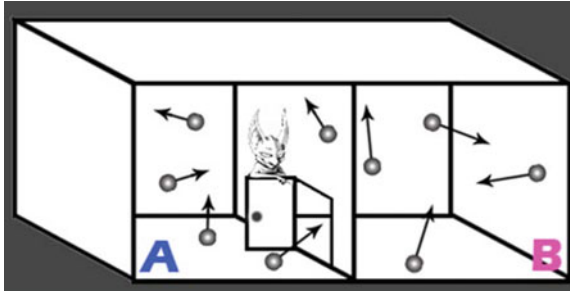
Now we have seen that the principle of maximum entropy can be used in both statistical mechanics and information theory, which means the methodologies in both fields can be connected together. But can we further integrate the two concepts of entropy as a uniform one? The Maxwell’s demon may help to give an answer.

### 2.5.3.2 The Fight with Maxwell’s Demon

There are various descriptions of the second law of thermodynamics. One is called the principle of increasing entropy, i.e., the entropy of an isolated system increases in any irreversible process and is unaltered in any reversible process. Another is the Clausius statement, namely heat can never pass from a colder to a warmer body without external work being performed on the system.

In 1867, J. C. Maxwell (June 13, 1831–November 5, 1879) first proposed a thought experiment that challenged the second law of thermodynamics. This is when the Maxwell’s demon was born. An illustration of the thought experiment is shown in Fig. 2.7. The detailed process is described in the following:

1. Suppose there is a container which is divided by an adiabatic diaphragm into two parts, denoted as part  $A$  and part  $B$ , respectively, as shown in Fig. 2.7. The container is filled with a sort of gas and the gas in part  $A$  is assumed to be hotter than the gas in part  $B$ .
2. Imagine there is a demon. He can know the paths and velocities of every gas molecule by simple inspection. He can do nothing but to open or close a hole in the diaphragm with a zero-mass frictionless slide.



**Fig. 2.7** Illustration of Maxwell's demon who apparently violates the second law of thermodynamics

3. His mission is to open the hole to let a gas molecule in part *A* to enter part *B* when that molecule has a velocity less than the *rms* velocity (*rms* is short for *root mean square*; and the *rms* velocity is expressed by  $v_{\text{rms}} \equiv \sqrt{v^2}$ ) in *B*; meanwhile, a gas molecule from *B* is allowed to pass into *A* through the hole if its velocity exceeds the *rms* velocity in *A*. The two procedures are conducted in such a way that the total number of gas molecules in *A* or *B* is unchanged.

We know that a higher temperature shown in a gas system's macrostate means a higher average kinetic energy of the gas molecules in the microstate, and vice versa. Hence, through the demon's operation, the hotter part *A* gets hotter and the colder part *B* gets colder, and no external work is done. This violates the second law of thermodynamics obviously.

Since the demon was born, it has attracted a huge amount of discussion. Different explanations have been proposed. After Ref. [41], a new explanation framework emerged which combines the demon with information theory and computer science. In 1961, R. Landauer (February 4, 1927–April 28, 1999) proposed a physical principle which was later called Landauer's principle [44]. It states that any logically irreversible manipulation of information, such as the erasure of a bit or the merging of two computation paths, must be accompanied by a corresponding entropy increase in noninformation-bearing degrees of freedom of the information processing apparatus or its environment [45]. In 1982, C. H. Bennett argued that to determine whether to let a molecule pass the hole, the demon must acquire information about the molecule's state and then store it; but no matter how well the demon prepares in advance, he will eventually run out of his information storage space and must begin to erase the previous information he has collected; since erasing information is a thermodynamically irreversible process according to Landauer's principle, an additional part of entropy will be created [46]. Hence, this means no matter how hard the demon works, he still cannot violate the second law of thermodynamics.

Nowadays, the subject of Maxwell's demon with information theory and computer science draws a lot of attention. For example, in Ref. [47], an inanimate device that

mimics the intelligent Maxwell's demon is designed. The device only requires a memory register to store information, which is quite interesting.

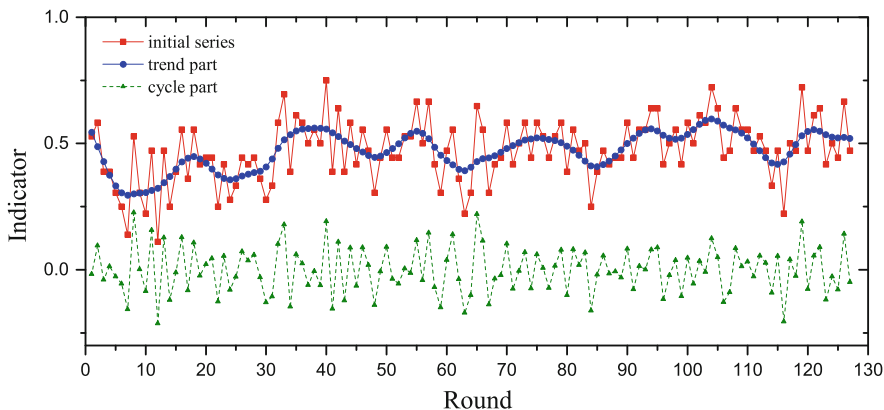
### **2.5.4 Discussion**

So far, we have seen how information theory and statistical mechanics are connected and influence each other. This may give us many inspirations when we deal with econophysics. First, in econophysics, we also need to deal with situations where a lot of uncertainty exists, such as the different kinds of human behavior in some game models or the movement of a stock price. In these situations, the Shannon entropy may have application; see Chap. 9. Second, econophysics has already adopted many physical concepts or methods (such as fractals, chaos, or even quantum mechanics) to tackle economic problems. But now it is still far from mature, lacking fundamental theories or principles. So comparing with the development history of information theory, we can be confident that maybe one day, in textbooks of physics such as statistical mechanics, there will be a chapter on econophysics that cannot be neglected.

## **2.6 Nonparametric Regression Analysis: Hodrick-Prescott Filter**

When talking about the analysis of data (e.g., obtained from controlled experiments or collected from real markets) in econophysics, we often focus on statistical distribution analysis and time correlation analysis for the time series. Commonly, we transform the initial raw time series into the dimension of return. On one hand, we always neglect the time attribute of a series when analyzing the statistical distribution problem. On the other hand, the statistical attribute is also ignored when arguing about the time correlation. So, the critical point is that we need not face the time series directly and primordially but in a side way. It becomes a reduced problem about the linear or other obvious relationship between two physical quantities, which is what econophysicists like to do. Compared with the situation where the relationship is known with an equation, sometimes there is no specific relationship between two quantities. Therefore, except linear regression or other parametric regression, nonparametric regression may be preferred occasionally. For example, for a price time series about one stock, if we want to know its cyclical volatility, one method is to directly consider the relationship between price and time. However, the price curve is possibly without regularity. At this time, for perfect effectiveness when solving real economic problems, economists begin to pay more and more attention to nonparametric regression analysis. For e.g., here I introduce a kind of nonparametric regression analysis method.

As is known, a filter generally refers to a physical tool widely used in signal processing. We can also regard the economic time series as one kind of signal. Since the signal contains too much information and behaves too randomly, naturally we



**Fig. 2.8** An example showing the Hodrick-Prescott filter method

can use filters to decompose it and extract useful information from it. To separate the behavior of a time series into regular and irregular components, several filters have been developed and are already commonly used in studies of macroeconomic and financial phenomena. Such filters are Hodrick-Prescott filter [48, 49], Baxter King filter [50, 51], Christiano-Fitzgerald random walk filter [52, 53], and Butterworth square wave filter [54, 55]. The Hodrick-Prescott filter is one of the most commonly used and known methods. It was first proposed in 1980 by R. J. Hodrick and E. C. Prescott when they tried to analyze postwar United States business cycles [48]. Since economic quantities are changing tardily rather than invariably, they think that economic performance can be reckoned as the combination of two parts, long-term potential growth, and short-term fluctuation. The core thought in this method is to decompose the time series  $X_t$  into trend component  $G_t$  and cyclical component  $C_t$ , that is,  $X_t = G_t + C_t$ . However, economists are more interested in cyclical component, which refers to business cycles. That is why the Hodrick-Prescott filter is called as a detrending method and a high-pass filter. If we denote  $L$  as a linear lag operator,  $L \cdot G_t = G_{t-1}$  and the second difference will be like

$$\Delta^2 G = (1 - L)^2 \cdot G_t = (G_t - G_{t-1}) - (G_{t-1} - G_{t-2}).$$

Suppose that both the cyclical part and the second difference in growth part have a zero mean value and a normally distributed variance. Then, through solving the minimization problem below

$$\min \sum_{t=1}^T \left\{ C_t^2 + \lambda [(G_{t+1} - G_t) - (G_t - G_{t-1})]^2 \right\},$$

the optimal trend and cyclical components can be obtained from the original series. Actually, this optimal result refers to the minimized sum of both variance of cyclical



part and squares of growth part's second difference. The only parameter  $\lambda$  is given as the ratio of the two variances

$$\lambda = \frac{\sigma_C^2}{\sigma_{\Delta^2 G}^2},$$

which determines the smoothness for the trend part. As long as  $\lambda$  gets larger, the trend part  $G_t$  will get smoother. So it becomes an important problem to select an appropriate  $\lambda$  when applying the Hodrick-Prescott filter to economic data of different frequencies. We should have known that the trend component, which represents the potential growth in an area, is always steady and mainly affected by long-term economic policies or others. Economists adopt several distinct  $\lambda$  to fit for different economic cycles. Empirically, most economists use 6.25 for annual data [56],  $\lambda = 1600$  for quarterly economic data, and 1,29,600 for monthly data [56].

Figure 2.8 schematically shows an example of a time series applied with the Hodrick-Prescott filter. The original data come from one human experiment conducted by my group (for clarity, experimental details are neglected herein; relevant experiments can be found in Chap. 9), which aims to study a kind of business cycle. The horizontal axis stands for experimental rounds, and the vertical axis represents one economic indicator, shown in red, which may present cyclical property. After we apply the Hodrick-Prescott filter, the separated trend part and cyclical part are shown above in blue and green, respectively.

Experimental Econophysics

Properties and Mechanisms of Laboratory Markets

Huang, J.-P.

2015, XV, 192 p. 75 illus., 35 illus. in color., Hardcover

ISBN: 978-3-662-44233-3