10

Income Distribution and Lottery Expenditures in Taiwan: An Analysis Based on Agent-Based Simulation

Shu-Heng Chen[1], Bin-Tzong Chie[1], Hui-Fen Fan[1], and Tina Yu[2]

[1] National Chengchi University, Taipei, Taiwan 116
{chchen,g0258501,92921029}@nccu.edu.tw
[2] Memorial University of Newfoundland, St. John's, NL A1B 3X5 Canada
tinayu@cs.mun.ca

Summary. We investigate the impact of income distribution on lottery expenditures in Taiwan, using an agent-based model developed in [2, 3]. The agents in the model are potential lottery buyers, whose characteristics are described by three features: the percentage of income spent on the lottery, the preferences among lottery numbers selected and the aversion to regret. We used a genetic algorithm to drive the model simulation under agents with different incomes, based on household income data in Taiwan from 1979 to 2003. The simulation results indicated that the impact of income distribution on lottery sales is not significant. This might be due to the Taiwan economy having a minor degree of income variation which has a low effect on lottery expenditures.

10.1 Introduction

Lotteries are pervasive phenomena worldwide. Most modern lottery games are variations of the pari-mutuel "lotto" design: players select a set of numbers in a given range and win prizes according to how many numbers are guessed correctly. The winning prizes come from a proportion of the lottery sales (the pay-out) and the remaining portion (the take-out) is used to administer the game, pay the ticket outlets and cover tax liabilities and other charges. In many countries, lottery taxes have become a major source of charity and educational funds. In order to maximize tax revenues and minimize the negative social impact of gambling, governments consider socio-economic variables and demographic information in regulating lottery policies.

In Taiwan, the first lottery game, Lotto, was launched in January 2002, with a pay-out rate of 60%. The remaining 40% (take-out) is spent in two areas: 13.25% for administration expenses and 26.75% for government tax. Compared to the rest of the world, where lottery take-out rates vary between 68.4% and 40% (Data source: La Fleur's Lottery World, U.S. Lotteries' Unaudited FY00 Sales by Games. http://www.lafleurs.com/), Lotto has a relatively low take-out rate.

As the lottery market continues to grow, the Taiwanese government has become concerned about the determinants of the lottery expenditures, including income and soicoeconomic variables such as age, sex, and education, which might give rise to a "demographic" burden of the implicit lottery tax. As an initial effort to address these issues, this research investigates *income distribution* and its impact on lottery sales. This is an important subject as there have been many studies reporting the regressivity of lottery

A. Brabazon, M. O'Neill (Eds.): Natural Computing in Computational Finance, Vol. 2, SCI 185, pp. 207–223.
springerlink.com

tax: lotteries are disproportionately consumed by the poor, who become the heavy bearers of the implicit lottery tax [4, 5, 15]. The situation will become even worse if economic inequality also plays a role in lottery expenditures. In this research, we adopt an agent-based simulation approach to study this issue.

The lottery market is a dynamic entity whose macro behaviors are influenced by many factors, such as the size of jackpot rollover and the psychology of lottery players. While various mathematical models, such as multiple linear regression, have been used to study the socio-economic variables related to lottery expenditures [14, 15], these models do not capture the dynamic nature of the market. By contrast, agent-based models simulate the lottery market behaviors by designing agents as lottery players and implement various market scenarios in a systematic manner. The simulation results hence allow us to analyze the cause and effect of market phenomena that are difficult to trace using the traditional models. The agent-based simulation approach is becoming a promising alternative in conducting economic analysis [13].

The remainder of this chapter is organized as follows. Sect. 10.2 introduces the concept of income distribution and explains the lottery market's operations. Sect. 10.3 describes our design of a lottery market model using an agent-based system. In Sect. 10.4, the genetic algorithm used to drive the model's simulation is presented. We provide the experimental setups in Sect. 10.5. Based on the simulation results, we analyze the relationship between income distribution and lottery sales in Sect. 10.6. The tax revenues under different tax rates are evaluated in Sect. 10.7. In Sect. 10.8, we discuss rollover and its impact on lottery sales. Finally, Sect. 10.9 concludes the chapter with a brief outline of future work.

10.2 Background

Income distribution provides an indication of the economic inequality of a society. There are several metrics used by economists to measure income equality. This work has adopted two methods: the *5-range ratio* and the *Gini coefficient* [9]. In the 5-range ratio approach, the household income observations are divided into 5 groups from the lowest 20% to the highest 20%. The average income of the 80th percentile is then divided by the average income of the 20th percentile. A larger ratio means a higher degree of income inequality.

The Gini coefficient is derived from the Lorenz curve. To plot a Lorenz curve, the income observations are ranked from the lowest to the highest. The cumulative proportion of the *population* is plotted on the x-axis and the cumulative proportion of the *income* is plotted on the y-axis. Fig. 10.1 gives an example of a Lorenz curve. The diagonal line represents perfect equality. The greater the deviation of the Lorenz curve from this line, the greater the inequality. To calculate the Gini coefficient ratio, we have used the area between the Lorenz curve and the diagonal line as the numerator and the area under the diagonal line as the denominator. Thus, a low Gini coefficient ratio indicates high income equality. When the ratio is 0, it means perfect equality (everyone having exactly the same income). When the ratio is 1, it indicates perfect inequality (where one person has all the income, while everyone else has zero income). The Gini coefficient requires that no one has a negative net income or wealth.

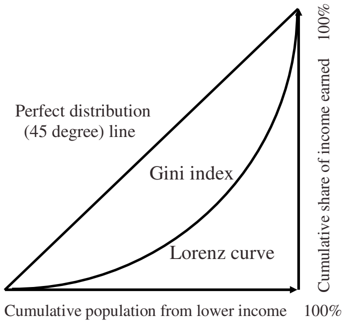

Fig. 10.1. Gini coefficient and Lorenz curve

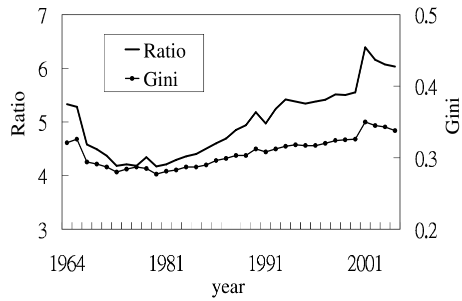

Fig. 10.2. Taiwan's income distribution

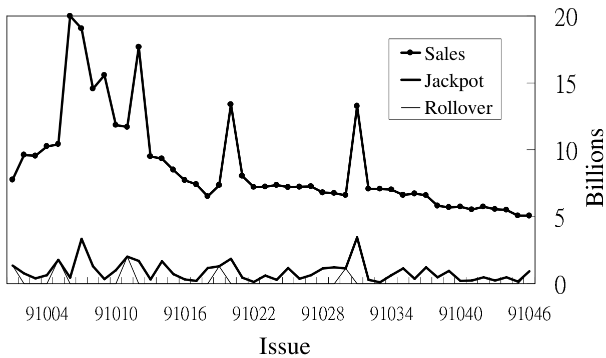

Fig. 10.3. Lotto sales volume (in $NT) in 2002

Fig. 10.2 gives the household income distribution in Taiwan from 1964 to 2004, measured by the *5-range ratio* and the *Gini coefficient* (Data source: `http://fies2.tpg.gov.tw/doc/result/93/211/Year04.doc`). The 5-range ratio from 1964 to 1971 ranged between 5.33 and 4.17. After that, the ratio first increased then started decreasing after 2001. The Gini coefficient index has a similar trend. The highest Gini index was 0.35 in 2001, which is still below the international standard of inequality threshold of 0.4. This indicates that the Taiwan economy has minor income variation. This research will study lottery sales in Taiwan based on this set of income data.

The two major lotteries sold in Taiwan are Lotto and Big Lotto. Lotto was first issued in January 2001. A Lotto player picks 6 out of 42 numbers to match the 6 winning numbers. If the picked numbers match all 6 winning numbers, the player wins the *jackpot*. If nobody wins the jackpot, the prize is rolled over to the next draw. In 2005, the rule was changed to pick 6 out of 38 numbers to promote sales. After the rule change, the odds of winning the jackpot was increased from 1.90629×10^{-7} to 3.62229×10^{-7}.

In 2004, a new game Big Lotto was introduced. This game has the same rules as that of the Lotto game except that the game player picks 6 out of 49 numbers. The odds of winning its jackpot is about 0.715×10^{-7}. When nobody wins the *jackpot*, the rollover frequently increases the sales of the next draw. Fig. 10.3 gives the Lotto sales volume (in NT$) in 2002. It is clear that during the rollover draws, such as issues 91006, 91012, 91020 and 91031, the sales increased dramatically. We will incorporate this market phenomenon to design the agents in our agent-based lottery market model, which is described in the following section.

10.3 An Agent-Based Model of Lottery Markets

An agent-based model typically has two parts: the environment and agent engineering. The environment is described by a set of rules governing the interactions of agents in the model while agent engineering involves the design of agents' representative characteristics. This research used the lottery market model built in [2, 3] to conduct the simulation. In this model, the following x+4-tuple vector is used to describe the lottery market environment:

$$M = (x, X, \tau, s_0, ..., s_x) \tag{10.1}$$

where x is the number of picks that a lottery player has to make from a total of X numbers. Depending on the number of matches between a player's picks and the winning numbers, different prizes are awarded.

Let y denote the number of matches, where $y = 0, 1, ..., x$, and S_y is its prize. The prize with the highest amount is called the *Jackpot*. Each prize, S_y, is shared by all players who picked the numbers that matched the winning numbers. In the event when nobody wins a particular prize S_y, the amount is added to the next draw. When the *jackpot* prize is added to the next draw, it is referred to as a *rollover*. Rollovers usually attract more participants in the next draw, called the *rollover draw*.

The amount of monetary rewards to the lottery winners is governed by the lottery tax rate, τ. Let S be the total lottery sales, the *pay-out rate* is $1-\tau$ and the total amount of monetary rewards is $(1-\tau)\times S$. This amount is distributed among different prizes with rates $s_0 \dots s_x$, such that $\sum_{y=0}^{x} s_y = 1$. In other words, $S_y = s_y(1-\tau)\times S$. Normally, s_y increases as y (the number of matches) increases.

The second part of the lottery market model is agent design. In a lottery market model, agents are potential lottery buyers. What motivates an individual to gamble? How much does one bet? We do not think there is a unique answer to these questions nor a single approach to address these issues. Among many possible ways of designing the agents, we focus on three features that capture the "stylized facts" of lottery markets: *participation level*, *conscious selection* and *aversion to regret*.

10.3.1 Participation Level

Participation level α is defined as the percentage of an individual's income I that is spent on the lottery. When considering the reasons why someone wants to buy lottery tickets, we perceive two types of possible influences: external and internal. The most noticeable external influence is the size of the jackpot. In Section 10.2, we have observed from the empirical data that the lottery sales increased during the rollover draws. We have therefore used the size of the jackpot as a factor that influences an agent's lottery participation. Another possible influence on individuals' decisions to purchase lottery tickets is their own internal subjective belief (probability) that they will win the jackpot.

We have also implemented this subjective belief of lottery winning as an alternative way to determine α. The simulation results from the two different implementations will be compared and discussed.

Lottomania and the Halo Effect

Lottomania refers to the phenomenon that sales following a *rollover* are higher than those of normal draws. Lottomania is mostly created by the media to arouse gamblers' desires to play the lottery. This effect may last for a few draws after the rollover draw [1]. Another related observation is that lottery sales are unexpectedly high right after a large jackpot prize is won. In the lottery industry, this is called the halo effect [7, 11].The following equation defines an agent's lottery participation level based on the jackpot prize

$$\alpha = \rho(J) \tag{10.2}$$

where ρ is the participation function, α is the participation level, and J is the *jackpot* prize. This work assumes that agents base their decisions on some heuristics, and hence uses an if-then rule to represent ρ and approximate Equation (10.2). One example of a rule is "if the jackpot is unusually high, then I will spend 10% of my income to buy lottery tickets." The linguistic term *unusually high* in this rule is a common form of human reasoning. We have used fuzzy sets [16] to define linguistic terms and describe the value range of J in the if-then rule.

In this work, J is mapped into 4 linguistic terms (*Low*, *Medium*, *High* and *Huge*) by 4 different membership functions. A membership function decides the degree of membership of a value to a particular fuzzy set. For example, if the fuzzy set *High* and *Medium* are defined as

$$Medium = \{jackpot|500,000 < jackpot < 1,200,000\}$$
$$High = \{jackpot|1,000,000 < jackpot < 2,000,000\}$$

The jackpot prize of 1,500,000 belongs to *High* 100%, while the prize 900,000 belongs to *High*, maybe 90% and *Medium* 10%. The degree of membership of a value to a particular fuzzy set is decided by the membership function associated with that fuzzy set. In this study, we have adopted triangular-shape membership functions for simplicity. A triangular-shaped membership function is defined by two base points: left leg and right leg. They in turn give the peak of the triangle point (see Fig. 10.4).

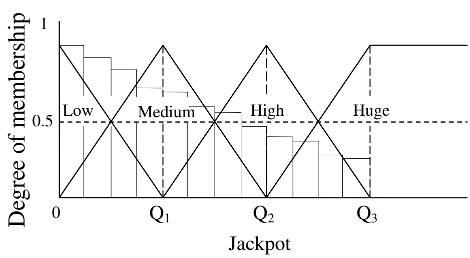

Fig. 10.4. Four triangular-shaped membership functions for the Jackpot prize

The four membership functions are not fixed but change over time depending on the historical jackpot prizes. In other words, the definitions of “Low”, “Medium”, “High” and “Huge” of a jackpot prize differ every day. On day t, the jackpot prizes from day 0 to day *t-1* are analyzed based on their frequency under different prize ranges. The prize ranges with lower frequency normally have high values and are very motivating in terms of lottery purchasing. By contrast, those jackpot prizes with high frequency are normally with low values and are unattractive to lottery buyers. After arranging the prize ranges according to their frequency, we divided the entire jackpot prize ranges evenly into 3 partitions. (Q_1, Q_2, Q_3), which in turn give the base points of the four membership function. Fig. 10.4 shows this scheme of membership function design. The degree of membership to the four fuzzy sets is represented as a vector $\overrightarrow{\mu} = \{\mu_1, \mu_2, \mu_3, \mu_4\}$, where μ_1 is the degree of membership to “Low”, μ_2 is the degree of membership to “Medium”, μ_3 is the degree of membership to “High” and μ_4 is the degree of membership to “Huge”. For example, the jackpot prize 900,000 has $\overrightarrow{\mu} = \{0, 0.1, 0.9, 0.0\}$.

An individual view of the significance of a jackpot prize differs from one person to another. To customize the membership functions for each agent, an individual weight vector $\overrightarrow{a} = \{a_1, a_2, a_3, a_4\}$ is used, where each a_i is between 0 and 1. This vector is applied on the right-hand side of the if-then rule when calculating the participation level α_i for agent i

$$\alpha_i = a_{i1}\mu_1 + a_{i2}\mu_2 + a_{i3}\mu_3 + a_{i4}\mu_4 \tag{10.3}$$

This gives α_i a value between 0 and 1. While the value of $\overrightarrow{\mu}$ changes over time as the jackpot prizes change, each agent’s weight vector $\overrightarrow{a}$ also evolves. The adaptation of $\overrightarrow{a}$ is carried out by a genetic algorithm, which will be described in Sect. 10.4.

Subjective Belief

The second implementation of an agent’s lottery participation level is the agent’s subjective belief [3]. Let p_i be agent i’s *subjective belief* (probability) that he/she will win the jackpot. Assume the agents are *risk-averse* with the log utility function $u(c) = \log c$. It can be shown that the optimal participation level α_i^* is

$$\alpha_i^* = \begin{cases} \frac{I_i - \left(\frac{I_i}{Jp_i}\right)^{\frac{1}{1-p_i}}}{I}, & if \;\; \log^{I_i} \le p_i \log^J + (1-p_i)\log^{(I_i - e)}, \\ 0, & otherwise. \end{cases} \tag{10.4}$$

where I_i is the income of agent i and e is the unit cost of the lottery ticket. Eq. (10.4) shows that both the size of jackpot (J) and the agent’s subjective belief (p_i) have a positive impact on the participation level α^*. However, the impact of income (I_i) on α^*, can not be analyzed easily. Do poor people spend a larger portion of their incomes on the lottery or it is the other way round? Research work on this issue has given rise to mixed reports [6, 14] and the answer to that question is inconclusive.

10.3.2 Conscious Selection

Despite the fact that the lottery winning numbers are generated randomly, lottery players tend to believe that one can predict the future winning numbers by analyzing the winning numbers in the past, and hence choose numbers in a non-random manner. This is called *conscious selection* [8].

We implemented an agent's conscious selection using an X-dimensional vector $\overrightarrow{b}$ whose elements take either "0" or "1" (recall X is the total number of possible selections that a lottery player can make). For a number z, where $1 \le z \le X$, if the zth element in $\overrightarrow{b}$ is "1", the number z is consciously selected by the agent, while "0" indicates the opposite. Therefore, $\overrightarrow{b}$ gives a list of numbers which are consciously selected by the agent. If $\overrightarrow{b}$ has exactly x 1s, there is only one possible combination and that combination is used by the agent to purchase the lottery ticket(s). If $\overrightarrow{b}$ has more than x 1s, there is more than one combination. The agent will randomly select one of them to purchase the ticket(s). Finally, if $\overrightarrow{b}$ has less than x 1s, some random numbers will be generated to make up a total of x numbers for the lottery tickets.

10.3.3 Aversion to Regret

Regret aversion arises because of peoples' desire to avoid the pain of regret resulting from a poor decision. In the case of lottery purchases, the regret refers to the cost of not gambling after knowing somebody has won the lottery. Lottery promoters capitalize on the aversion to regret to encourage lottery buyers to keep on buying [12]. We have incorporated an agent's aversion to regret θ using the following equation

$$r_i = \begin{cases} -\theta_i I_i, & if \;\; \alpha_i^* = 0 \;\; and \;\; N_x > 0, \\ \theta_i I_i, & if \;\; \alpha_i^* = 0 \;\; and \;\; N_x = 0, \\ 0, & otherwise. \end{cases} \tag{10.5}$$

where θ_i denotes how regretful agent i, who did not purchase lottery ($\alpha_i^* = 0$), felt after knowing someone had won the jackpot, i.e., the number of winners of the jackpot is positive ($N_x > 0$). In the case when the jackpot is not drawn ($N_x = 0$), agent i may also derive pleasure from the gamblers' misfortune, and hence has a positive r_i value. If an agent did engage in lottery purchases, however, then there would be no such regretful effect, and r_i is 0. In summary, an agent i in the lottery market model is described by the following 3 characteristics

- *participation level*: α_i under lottomania implementation or α_i^* under subjective belief implementation.
- *conscious selection*: $\overrightarrow{b}_i$;
- *aversion to regret*: θ_i;

These agent properties evolve during the simulations using a genetic algorithm, which is explained in the following section.

10.4 Genetic Algorithms

Genetic algorithms (GAs) [10] start with a population of randomly generated individuals. With a defined fitness criterion, individuals are evaluated and selected for reproduction. The two most commonly used reproduction methods are crossover and mutation.

The generated offspring then form a new generation of population. This process of evaluation, selection and reproduction is repeated many times until a termination criterion is met. We will describe the implementation of our GA in the following sub-sections.

10.4.1 Representation

In the lottery market model, the population consists of a group of agents, each of which has the 3 characteristics, α/α^*, $\overrightarrow{b}$ and θ, as described in the previous section. These three characteristics are represented using a linear chromosome. As given in Eq. (10.3), α is decided by two vectors $\overrightarrow{u}$, which is calculated from the historical jackpot prizes, and $\overrightarrow{a}$, which is an agent's individual weight vector. We encode $\overrightarrow{a}$ in the GA representation using 4 binary bits for each of the four elements in the vector. The total number of bits representing $\overrightarrow{a}$ is therefore 16. Each of the four binary bits is decoded to a real value between 0 and 1 using the following equation

$$a = \frac{\sum_{j=1}^{4} c_j 2^{j-1}}{2^4 - 1} \tag{10.6}$$

where c_j is the jth bit value counted from the right. For example, 0011 is decoded as $\frac{2^0+2^1}{2^4-1} = \frac{3}{15} = 0.2$; 1001 is decoded as 0.6; 1100 is decoded as 0.8 and 1111 is decoded as 1.0. Assume $\overrightarrow{a} = \{0.2, 0.6, 0.8, 1.0\}$ and $\overrightarrow{\mu} = \{0, 0, 0.25, 0.75\}$. According to Eq. (10.3), the agent would invest $\alpha = 0.95$ of his income to purchase lottery tickets.

When the participation level is determined using subjective belief, the probability p is coded as a real number in the GA representation. This only takes one gene space. Based on the p, we can calculate α^* using Eq. (10.4). The coding of $\overrightarrow{b}$ (conscious selection vector) is straightforward: it is a binary string of length X. Fig. 10.5 gives an example of the case where $X = 20$. The consciously selected numbers are 1, 6, 9, 11 and 12.

Aversion to regret, θ, is also a real number between 0 and 1. When the participation level (α) is decided using a fuzzy rule, θ is coded with a 4-bits binary string, so that the entire chromosome is a binary string, which is easier for genetic operation. Eq. (10.6) is used to decode the 4-bits binary string to a real number. In the case where the participation level (α^*) is decided by the agent's subjective belief (p), θ is coded as a real number taking one gene space in the chromosome. The value θ is used to calculate r using Equation (10.5), which is used to calculate the agent's fitness (see Section 10.4.2). In summary, when the agent's participation level is decided by a fuzzy rule, its representation contains $(\overrightarrow{a}, \overrightarrow{b}, \theta)$, which takes 20+$X$ bits. If the agent's participation

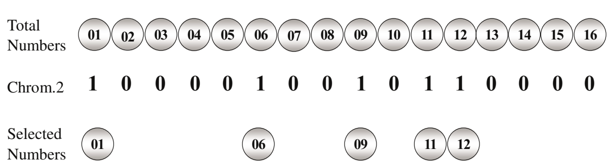

Fig. 10.5. An example of the selected lottery numbers

level is decided by its own subjective belief, its representation contains $(p, \overrightarrow{b}, \theta)$, which has length $2+X$.

10.4.2 Fitness Function

Each agent is provided with different income, based on the Taiwan income data (see Table 10.2). The fitness of agent i is

$$F_i = I_i - G_i + \pi_i + r_i \tag{10.7}$$

where I_i is agent i's income, G_i is the amount of money spent on the lottery (calculated from α_i), π_i is the prize won and r_i is the aversion to regret, per Eq. (10.5).

10.4.3 Genetic Operators

The genetic production starts from the selection of a mating pool. There are several different selection schemes in GA. However, to have a better focus, only tournament selection will be tried in this paper. By tournament selection, each individual in the mating pool is determined as follows. We first randomly select φ chromosomes without replacement, and then take the *best* two of them. The parameter φ is known as the *tournament size*, and it is also the mating pool size.

Given the mating, two genetic operators, *crossover* and *mutation*, are applied to the winning pair to generate two offspring. Since each chromosome contains three characteristics of an agent, crossover is restricted to swapping only the same characteristics between the two parent agents. We first randomly determine which one of the three characteristics of the crossover will take place. If it is on the bit-strings $\overrightarrow{a}$, $\overrightarrow{b}$ or θ, the one-point crossover is applied with probability P_c. If it is on the real-valued p or θ, the arithmetic crossover is applied with the same probability P_c. The arithmetic crossover works by averaging the two parents' gene values as the gene value of the offspring.

After crossover, each of the two offspring has a probability of P_m to be mutated. For bit-strings $\overrightarrow{a}$, $\overrightarrow{b}$ and θ, we apply bit mutation, i.e., 0 is flipped to 1 and 1 is flipped to 0. For the real-valued p and θ, we apply an arithmetic mutation, which is designed to be an equivalent of the bit-mutation shown in Eq. (10.8)

$$v^{new} = v^{old} + \sum_{i=1}^{16} B_{P_m} (\frac{1}{2})^i \cdot (-1)^{B_{\frac{1}{2}}}, \tag{10.8}$$

where v^{old} and v^{new} are the gene values *before* and *after* the mutation. B_{P_m} and $B_{\frac{1}{2}}$ are the Bernoulli random variables with a success probability of P_m, and a mutation rate of one half, respectively. In this way, the arithmetic mutation size, denoted by σ, is $\sum_{i=1}^{16} B_{P_m}(\frac{1}{2})^i \cdot (-1)^{B_{\frac{1}{2}}}$. When all offspring are produced, they replace the entire population and become the new generation for another evolutionary cycle [2].

10.5 Experimental Setup

The agent-based lottery market is defined by two sets of parameters; one is associated with the market (the top half of Table 10.1), and the other is associated with the GA

Table 10.1. Parameter values for the agent-based lottery market

Market parameters	
Pick x from X (x/X)	5/16
Lottery Tax Rate (τ)	10%,... 90%
Prize Distribution Rate (s_0,s_1, ... s_5)	0%,0%,35%,15%,12%,38%
Drawing Period (w)	3 days
Number of Agents (N)	5,000
Agent Income (I)	see Table 10.2
GA parameters	
p_i value range	[0,0.003] (for α^*)
Crossover Rate (P_c)	90%
Mutation Rate (P_m)	0.1%
Arithmetic Mutation Size (σ)	in Equation (10.8)
Population Size	5,000
Tournament Size (φ)	200
Number of Generations	500
Number of runs (lottomania implementation)	25
Number of runs (subjective belief implementation)	50

(the bottom half of Table 10.1). In this lottery game, a player picks 5 out of 16 possible numbers. Players with 0 or 1 matching numbers will not receive any prize. 35% of the total prize pool is given to players with 3 matching numbers. The largest prize is the jackpot which receives 38% of the total prize pool. Note that the 2-matching-number winners receive a larger proportion of the total prize pool than the 3-matching-number and 4-matching-number winners. This is because the number of 2-matching-number winners is large, and hence a larger proportion of the prize pool is allocated to that prize. After the prize is divided among all winners, the individual prize for a 2-matching-number winner is still smaller than that for a 3-matching-number or a 4-matching-number winner.

The prize pool is the total lottery sales after the deduction of tax. Various lottery tax rates (τ) from 0% to 90% are implemented to investigate whether there is an optimal tax rate for the government to receive the maximum lottery tax revenue in this lottery market model.

The drawing period is the number of days between two draws. In this model, each period is 3 days long. On the first day, each agent is assigned with a different income (explained later in this section). During the 3-day period, all agents can purchase lottery tickets as desired. The lottery sales are added to the prize pool. At the end of the 3rd day, the winning numbers are drawn and the prizes are distributed. All prizes that are not won are rolled over to the next period.

After the prize distribution, the fitness of each agent in the population can be calculated using Eq. (10.7). Based on this fitness, selection and reproduction take place to generate a new generation of agents. This ends the current period and all new agents are allocated with new income to start a new period. This also means that each GA generation is equivalent to one period and is 3 days long.

Table 10.2. Average household income of Taiwan from 1979 to 2003

Year	First 20%	Second 20%	Third 20%	Fourth 20%	Highest 20%	The High-Low Ratio	Gini coefficient
1979-1983	87.12	137.55	175.66	227.33	372.34	4.27	0.2201
1984-1988	82.30	135.45	175.21	228.13	378.91	4.60	0.2286
1989-1993	74.81	132.66	175.64	231.83	385.07	5.15	0.2399
1994-1998	72.33	129.37	174.55	232.87	390.89	5.40	0.2469
1999-2003	68.05	124.96	171.82	232.15	403.01	5.92	0.2590
Fix Income	200	200	200	200	200	1.00	0.0000

Each agent is assigned with a different income using the data in Table 10.2. From the Taiwan income data, we selected 5 periods from 1979 to 2003 where income distributions have different degrees of variation. Each period is 5 years long. Their Gini coefficients lie between 0.22 to 0.26 and their 5-range ratios range from 4.27 to 5.92. For each of the periods, the household incomes are partitioned into 5 groups, from the lowest 20% to the highest 20%. The average incomes of each of the 5 groups are assigned to the agent population uniformly. We also conduct an extra set of experiments where all agents have the same income (200). This allows us to compare lottery sales under any income distribution that varies and with perfect distribution. For each of the 6 different income distribution setups, we conduct 25 simulation runs for the lottery market model under lottomania implementation and 50 runs for the lottery market model under subjective belief implementation. The average results are used in the analysis and discussion.

10.6 Lottery Sales vs. Income Distribution Analysis

We apply two statistical tests on the simulation data to determine whether the lottery sales are different under the 6 different income distributions. The first test is the Kolmogorov-Smirnov (KS) test and the second one is the Mann-Whitney-Wilcoxon (MWW) test. Both tests are non-parametric or distribution free methods as they do not assume that the data are drawn from a given probability distribution. The procedures for the two-sample KS-test are as follows

1. Calculate the cumulative frequency for the first data set $S_1(X)$;
2. Calculate the cumulative frequency for the second data set $S_2(X)$;
3. Find the greatest discrepancy between the two frequencies, which is called the "D-statistic", $D_s = Max|S_1(X) - S_2(X)|$.
4. Compare this against the critical D-statistic (D) for that sample size.
5. If $D > D_s$, reject the null hypothesis that the two data sets are distributions of the same form.

The KS-test is more accurate when the sample size is small. Since our data series are collected from GA runs of 500 generations, the data size is considered to be large. We therefore use a second statistical test, MWW, to validate the first test results. The procedures for the MWW-test are as follows:

1. Combine the data from both data sets and rank each value;
2. Take the ranks for the first data set and sum them as W_1
3. Take the ranks for the second data set and sum them as W_2
4. Calculate $u_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1$; $u_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2$, where n_1 is the size of the first data set and n_2 is the size of the second data set.
5. $U = min(u_1, u_2)$;
6. Calculate $Z = \frac{U-E(u)}{\sqrt{V(u)}}$, where $E(u) = \frac{n_1 n_2}{2}$ and $V(u) = \frac{n_1 n_2(n_1+n_2+1)}{12}$;
7. Compare Z against the critical Z-statistic (Z_s) for that sample size.
8. If $Z > Z_s$, reject the null hypothesis that the two data sets are distributions of the same form.

The 6 sets of simulation data from the 6 different income distributions setups under the 50% lottery tax rate ($\tau = 0.5$) are paired to carry out the 2 statistical tests. Table 10.3 gives the test results for the lottery market model under lottomania implementation and Table 10.4 gives the test results for the model with subjective belief implementation. The entries whose p-values are less than 0.05 are marked with an $*$.

The tests show that only income distribution in 1984 under the model with lottomania implementation results in lottery sales that are different from the sales under income with equal distribution (200): both of the $p-values$ of the null hypothesis, based on D_s and Z, are almost nil. All other income distributions produce lottery sales that are not significantly different from each other. There are also a number of entries in the tables whose p-values are < 0.05 (with $*$). However, these entries do not demonstrate any consistent trend, and hence no conclusion can be drawn from them.

Table 10.3. Lottery market model with lottomania implementation

	K-S test statistic ($\tau = 0.5$)					M-W test statistic ($\tau = 0.5$)				
Y	1979	1984	1989	1994	1999	1979	1984	1989	1994	1999
200	0.2370	0.0000*	0.1236	0.0258*	0.0590	0.4971	0.0000*	0.0232*	0.1073	0.1073
1979		0.0013*	0.2370	0.8774	0.4141		0.0041*	0.2216	0.4377	0.4263
1984			0.0258*	0.0258*	0.0104*			0.0625	0.0344*	0.0397*
1989				0.8774	0.6485				0.6554	0.4971
1994					0.8774					0.9536

Table 10.4. Lottery market model with subject belief implementation

	K-S test statistic ($\tau = 0.5$)					M-W test statistic ($\tau = 0.5$)				
Y	1979	1984	1989	1994	1999	1979	1984	1989	1994	1999
200	0.3584	0.5077	0.3584	0.2408	0.0951	0.5649	0.4100	0.7020	0.4380	0.4628
1979		0.8409	0.8409	0.9541	0.6779		0.0076*	0.9862	0.7174	0.6766
1984			0.8409	0.9541	0.8409			0.7801	0.9478	0.9478
1989				0.8409	0.8409				0.7277	0.7485
1994					0.9541					0.9423

One possible reason why agents with different income distributions produce similar lottery sales is the low income variations of the data sets: the Gini indexes are very close to each other, between 0.22 and 0.26. We will perform simulations on data from other countries with a higher income inequality to verify this hypothesis.

10.6.1 Lottery Sales vs. Tax Rates

We also examine the lottery sales under different tax rates (τ) and income distributions. As shown in Figs. 10.6 and 10.7, the sales volume is the lowest when all agents have the same income of 200. In addition, the general trend is that the sales volume decreases as the tax rate (τ) increases. However, there are a couple of exceptions. For example, 1984 data have an increased sales volume when τ increased from 40% to 50%. Another important observation is that, under all income distribution data, the sales volume exhibits a sharp decline when the tax rates are higher than 80%. This indicates that lottery sales are strongly influenced by tax rates. We will analyze lottery tax revenues under different tax rates in the following section.

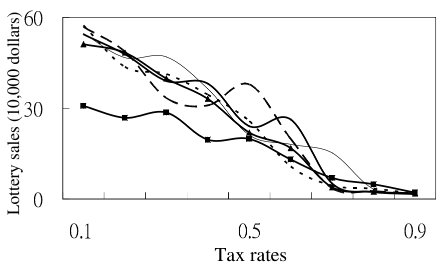

Fig. 10.6. Lottery sales vs. tax rates: lottomania implementation

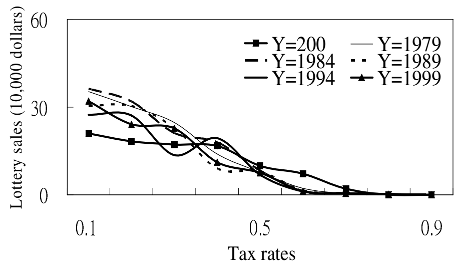

Fig. 10.7. Lottery sales vs. tax rates: subject belief implementation

10.7 Lottery Tax Revenue vs. Tax Rates Analysis

Lottery tax revenue (R) is affected by two factors: lottery sales volume (S) and the tax rate (τ) (see Eq. (10.9)). The two factors generate two counter-balancing forces, which are described in Eq. (10.10).

$$R = \tau S = \tau(\alpha I) \tag{10.9}$$

$$\frac{\partial TR}{\partial \tau} = \underbrace{S}_{+} + \tau \underbrace{\frac{\partial \alpha}{\partial \tau}}_{-} I \tag{10.10}$$

The positive force is characterized by the plus sign in Eq. (10.10), which says that given the sales volume S, the higher the lottery tax rate τ, the higher the tax revenue R. Meanwhile, we also expect a negative co-relation between lottery participation α and the tax rate τ, i.e., $\frac{\partial \alpha}{\partial \tau} < 0$. The second term in Eq. (10.10) hence has a negative value.

To plot the Laffer curves to depict the relationship between tax revenue (R) and the tax rate (τ), we first transform our simulation data in the following ways. For each of the simulation runs, there are 500 generations, i.e., 500 lottery draws. At each draw t, tax revenue R_t is collected. The tax revenue for each simulation run is therefore a time series $\{R_t\}_{t=1}^{500}$. We normalize the revenue series by dividing R by the total income ($\sum_{i=0}^{n} I_i$, where n is the number of agents and I_i is the income of agent i) and call this new series $\{r_t\}_{t=1}^{500}$. Notice that the normalized tax revenue r_t can be interpreted as the *effective tax rate* or the *tax revenue*. To avoid the possible initialization biases, we removed the first 100 data points from the series and calculated the average of the rest of the 400 data points, $\bar{r}$. Since we made 25 runs for each tax rate, the median values are used to plot the Laffer curves in Figs. 10.8 and 10.9.

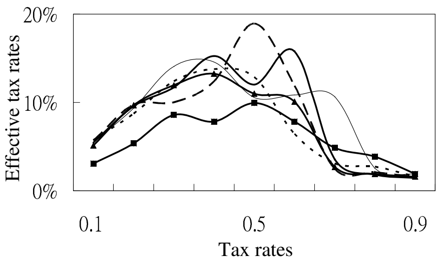

Fig. 10.8. Laffer curves: lottomania implementation

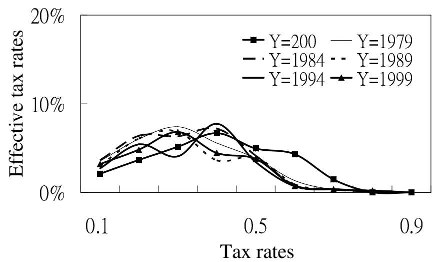

Fig. 10.9. Laffer curves: subjective belief implementation

As analyzed in Eqs. 10.9 and 10.10, the (normalized) tax revenue $\bar{r}$ first increases with the lottery tax rate τ, and then decreases with it. The curves also show that under different income distributions, there are different optimal tax rates τ that give the optimal tax revenue $\bar{r}$. Under lottomania implementation, the optimal τ is between 40% and 60% which generates an $\bar{r}$ of between 10% and 19%. Under subjective belief implementation, the optimal τ is between 30% and 40% which generates an $\bar{r}$ of between 6% and 8%.

We also evaluate the uncertainty of $\bar{r}$ using box-whisker plots. Since the plot pattern is similar for all simulation results from the 6 different income distribution data, we only show two of them, one for the lottomania implementation (Fig. 10.10) and one for the subjective belief implementation (Fig. 10.11). In a box-whisker plot, the box in the middle covers 50% of the simulated tax revenue. The longer the box, the more uncertain the tax revenue is. The two box-whisker plots show that the tax revenue is relatively low and stable when the tax rate is at its two extremes (τ = 10%, 90%). The box starts to inflate when the tax rate is moving toward the center, which signifies the growing uncertainty in tax revenue. The degree of uncertainty is further compounded by the enlarging whiskers, which extend the box to the frontier of the sample distribution. The high degree of uncertainty makes it unclear if there exists an optimum tax rate τ that gives the maximum revenues.

10.8 Discussion on Rollovers and Sales

Many researchers have reported that a large size of rollover would make the lottery more attractive and increase sales [1, 7]. A previous work applying statistical tests to lottery

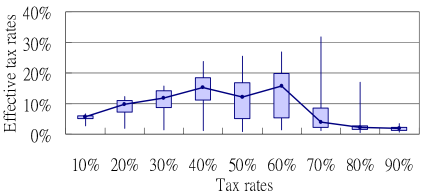

Fig. 10.10. Box-whisker plot: lottomania implementation

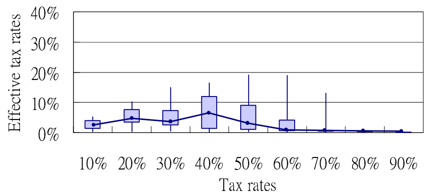

Fig. 10.11. Box-whisker plot: subject belief implementation

sales data from 7 countries also supports this proposition [2]. In this study, we have applied the same statistical tests to the simulation data generated from our experiments to examine if the same phenomenon appears in our lottery market model.

For each experimental run, we collected the sales from each draw. The data that were from rollover draws and from the regular draws were then separated into two data sets. Next, we conducted statistical tests to evaluate whether the two data sets were significantly different one from the other. Fig. 10.12 and Fig. 10.13 provide the t-statistics.

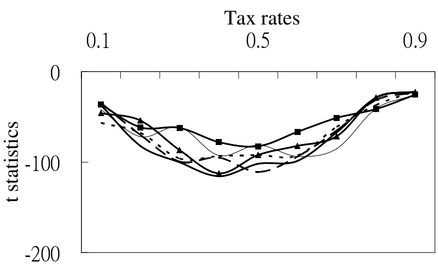

Fig. 10.12. Rollover vs. sales: lottomania implementation

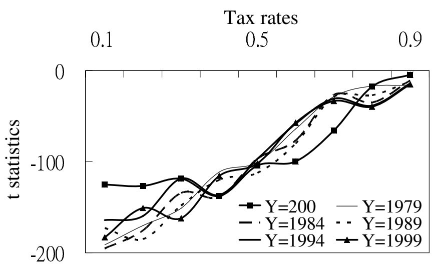

Fig. 10.13. Rollover vs. sales: subjective belief implementation

All t-statistics on data from the simulation runs under 6 different income distributions and 10 different tax rates have negative values. This means that the lottery sales for the rollover draws have a decreased volume, which is the opposite of what has been observed in the real lottery markets. A similar result was reported previously using the same agent-based modeling system under agents with equal income [2, 3]. The explanation presented there can be applied to our case.

In general, GA learning favors agents who win the lottery and who propagate those winning agents' characteristics (α_i, $\overrightarrow{b}_i$, θ_i) to the following generations. However, most gamblers do not win the prizes, and hence will end up with less money than the non-gamblers. In other words, the rank of the agents based on the money they have after the draws is as follows: winning gamblers, non-gamblers, and losing gamblers. During the rollover draw, there was no jackpot winner in the last draw. All winning gamblers won small prizes. The non-gamblers' agents hence have good chances to be propagated from the last draw (generation) to the rollover draw (generation). With these non-gambler agents, the lottery sales in the current draw (rollover draw) are reduced. However, if there is a jackpot winner in the previous draw, the situation is completely reversed. The jackpot winner's characteristics would be greatly propagated to the current generation, since the winner has very high fitness (the amount of money after the draw). Consequently, the lottery sale in the current draw (the draw after the jackpot winning draw) is increased.

10.9 Concluding Remarks

While the lottery has been widely adopted in many countries to raise charity and educational funds, there are concerns about its side-effects, such as the regressivity of the lottery tax and the addiction to gambling. Various studies have been devoted to identifying these side-effects so that the government can regulate related policies.

This chapter presents a study on the impact of income distribution on lottery expenditures in Taiwan based on the simulation of an agent-based model. The simulation results show that the impact is not significant enough for the government to raise any concerns. Although the lottery market model is simple and does not reflect the real market perfectly, it provides a vehicle to study the issues that are difficult to investigate using traditional models such as regression. We will continue this research by improving the model with more sophisticated agent design and incorporating different algorithms to drive the simulation. Meanwhile, we plan to investigate other socio-economic issues related to lottery expenditures based on the newer model.

Acknowledgments

The authors are grateful to two anonymous referees for their very painstaking review of the paper. The research support in the form of NSC grant No. NSC. 95-2415-H-004-002- MY3 is gratefully acknowledged.

References

[1] Beenstock, M., Haitovsky, Y.: Lottomania and other anomalies in the market for lotto. Journal of Economic Psychology 22, 721–744 (2001)
[2] Chen, S.H., Chie, B.T.: Lottery markets design, micro-structure, and macro-behavior: An ACE approach. Journal of Economic Behavior and Organization 67, 463–480 (2008)
[3] Chen, S.H., Chie, B.T.: Agent-based modeling of lottery markets: Policy-making from the bottom up. In: Dennard, L., Richardson, K.A., Morcol, G. (eds.) Complexity and Policy Analysis-Tools and Concepts for Designing Robust Policies in a Complex World. ISCE Publishing (2008)
[4] Clotfelter, C., Cook, P.: Selling Hope: State Lotteries in America. Harvard University Press, Cambridge (1989)
[5] Kitchen, H., Powells, S.: Lottery expenditures in Canada: A regional analysis of determinants and incidence. Applied Economics 23(2), 1845–1852 (1991)
[6] Deboer, L.: Lottery taxes be too high. Journal of Policy Analysis and Management 5(3), 594–596 (1984)

[7] Farrell, L., Morgenroth, E., Walker, I.: A time series analysis of U.K. lottery sales: Long and short run price elasticities. Oxford Bulletin of Economics and Statistics 61(4), 513–526 (1999)
[8] Farrell, L., Hartley, R., Lanot, G., Walker, I.: The demand for lotto: The role of conscious selection. Journal of Business and Economic Statistics 18(2), 228–241 (2000)
[9] Hall, R.E., Lieberman, M.: Economics: Principles and Application, updated 2nd edn., pp. 365–367 (2003)
[10] Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975); MIT Press (2nd edn.)
[11] Matheson, V.A., Grote, K.: Examining the 'Halo Effect' in Lotto Games. Applied Economics Letters 14(4-6), 307–310 (2007)
[12] Statman, M.: Lottery players / stock traders. Financial Analysts Journal 58(1), 17 (2002)
[13] Tesfatsion, L.: Agent-based computational economics: A constructive approach to economic theory. In: Tesfatsion, L., Judd, K.L. (eds.) Handbook of Computational Economics, Agent-Based Computational Economics. Handbooks in Economics Series, vol. 2. North-Holland, Amsterdam (2006)
[14] Vrooman, D.H.: An economic analysis of the New York state lottery. National Tax Journal 4, 482–490 (1976)
[15] Worthington, A.C.: Finance in gambling expenditures: Australian evidence on socioeconomic and demographic tax incidence. Public Finance Review 29(4), 326–342 (2001)
[16] Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)