# Chapter 3

# Methodology

## 3.1 Genetic Algorithm

Genetic algorithm is a mathematical model which imitates the process of natural breeding following Charles Darwin’s rule of natural selection. It begins with two sets of numbers which are called parents. Parents get married and breed children. Children are sets of numbers as well. They receive some pieces of information, the genes, from father and some genes from mother. The process of transferring the information from parents to children is called crossing over. After the genetic transmission, some genes of the children may be mutated. The mutation changes the information of some randomized genes, e.g. changing from number 1 to 0 or from positive to negative value. All the children fight to one another to survive. They try to maximize an objective function, e.g. profit function. The survived children are those two best children who yield the maximum values of the profit function. The rest of the children are dead. In the next generation, the best child will replace the role of father and the second best child may replace the role of mother. However, to avoid the transmission of inferior genes by marrying brothers and sisters, the father may find his wife from outside the family. In this case, only the best child survives. Then the model will generate a wife for him from outside the breeding process. After 1,000 generations, the best child will represent the strongest off-spring of the family. The information written in the best child’s genes is the solution that maximizes the objective function.

To illustrate the mechanism of genetic algorithm, this study constructs a genetic algorithm model with two initial parents (see figure 4 and 5). Each parent contains 30 data indicating the lagged 30 days before the selling day. The selling day is the last day before the XD dates.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Figure 4** A father who has information of 30 days

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 5** A mother who has information of 30 days

There are four families of parents (see figure 6 - 9). For the first family, the father is all one and the mother is all zero. Both father and mother in the second family have randomized numbers as their data. In the third family, the father is still all one but the mother is randomized. The fourth family has the father with randomized numbers whereas the mother is all zero.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 6** Parents who have information of 30 days from the first family

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |

**Figure 7** Parents who have information of 30 days from the second family

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |

**Figure 8** Parents who have information of 30 days from the third family

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 9** Parents who have information of 30 days from the fourth family

Parents get married to each other and their genes will transfer to children by the process of crossing over. The percentage of crossing over is the probability of a child that will be different from his or her father. In the extreme case, the 0 percent of crossing over means that all the children will be as the same as their mother. The 100 percent of crossing over means that all the children will duplicate the father.

This model randomizes the probability of crossing over. If the randomized number is 0.5, then the first half of the father's genes will transfer to the children. The rest of genes will come from the mother (see figure 10). In case that probability of crossing over does not yield an integer for the numbers of gene, then the program will round the number and use that number to be the cut point of the crossing over. Genes after the cut point will come from mother.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 10** Crossing over process with the randomized number of 0.5. The first two lines are the father and mother and the third line is the child.

After the crossing over, the mutation may happen. The mutation rate is the percentage of information that will be mutated. For example, the 0 percent of mutation means that all the genes of a child will remain the same. The 100 percent of mutation means that all the genes will change from 1 to 0 or from 0 to 1. This study allows 30 percent of mutation; nine genes will change from one to zero or vice versa. The positions of mutated genes are randomized (see figure 11).

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

**Figure 11** Mutation process with the mutation rate of 0.3. The first line is the initial child and the second line is the mutated child.

By the process of crossing over and mutation, this study makes each pair of parent produces four different children. Each child will present the information of the buying signals. Number one indicates that an investor should buy the stock on that day. Number zero indicates that the investor should not buy the stock on that day. The first data from the left-hand side and the last data on the right-hand side indicate the lagged 30 days and 1 day before the selling day.

Each child will evaluate its performance against the objective function. In this study, the objective function is the profit function as follows:

$$Profit = (Selling\ price - Average\ price) \times 100 / Average\ Price$$

The selling price is determined by the close price on the last day before the XD dates. The average price is the summation of the price over days that the model suggests the investor to buy the stock divided by the numbers of buying days.

The child that yields the highest value of objective function is the child who maximizes the profit. In the next round, this child will replace the father. The child who yields the second highest profit will replace the mother. The process will repeat 1,000 rounds to find the final best child.

However, the model does not rely on the solution from only the final best child. In every round, the model collects the solution of the best child. This is to ensure that the model must not lose the best solution during the iterative process. Therefore, there are 1,000 solutions for a family. Moreover there are 4,000 solutions from four families. Each family will find the best solution out of those 1,000 solutions. Comparing among the four best solutions from four families, the solution that yields the highest profit will be the best solution (see figure 12).

Finally, for the condition of the best buying strategy, the best solution must yield the highest profit not only in a particular XD round but also in other XD rounds. To find this, the model will produce the best solution of each XD round. A solution will evaluate its performance in the out-of-sample tests. The model with the highest average out-of-sample profit will represent best of the best solution (see figure 13).

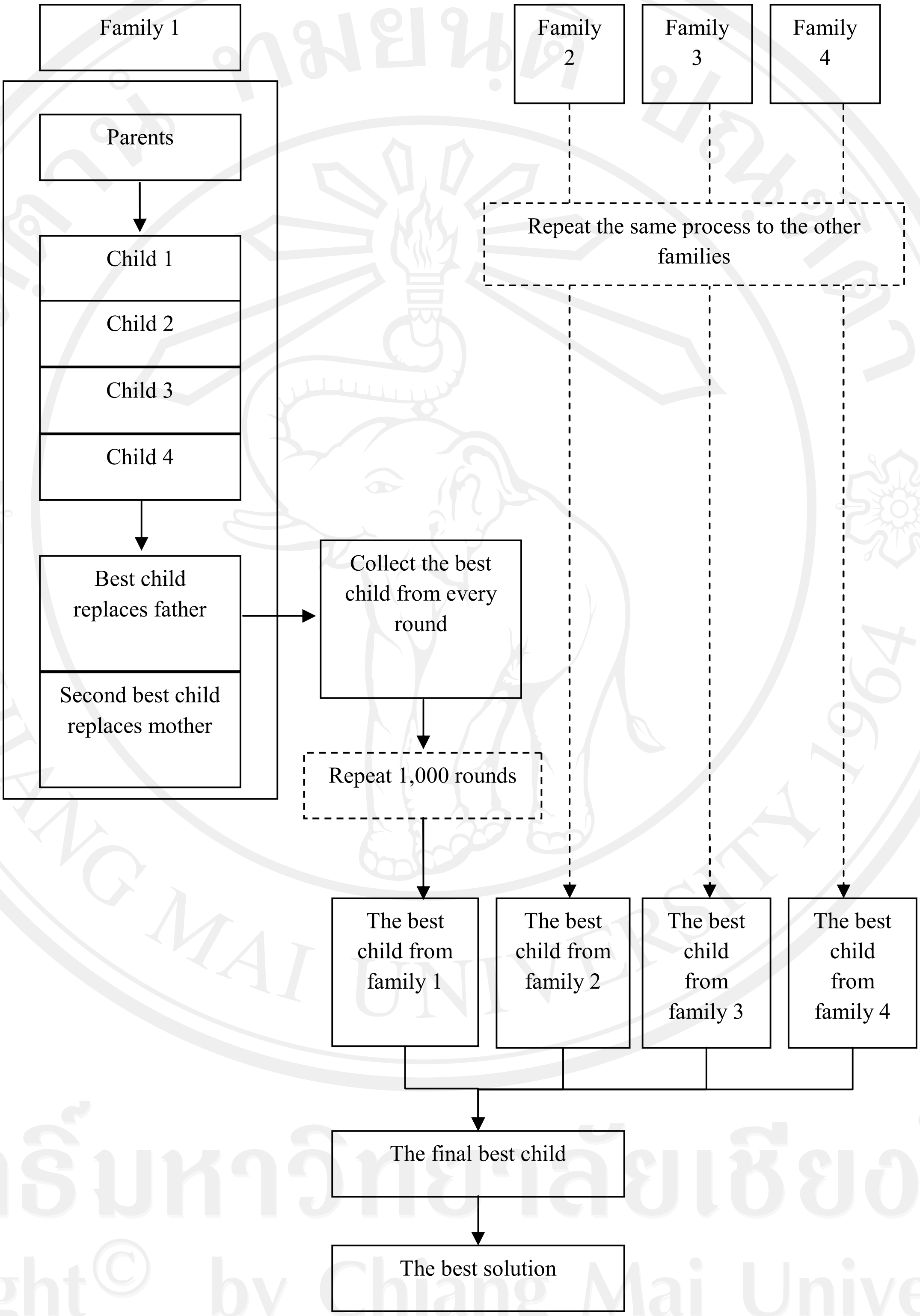


**Figure 12** The process of genetic algorithm in finding the best solution

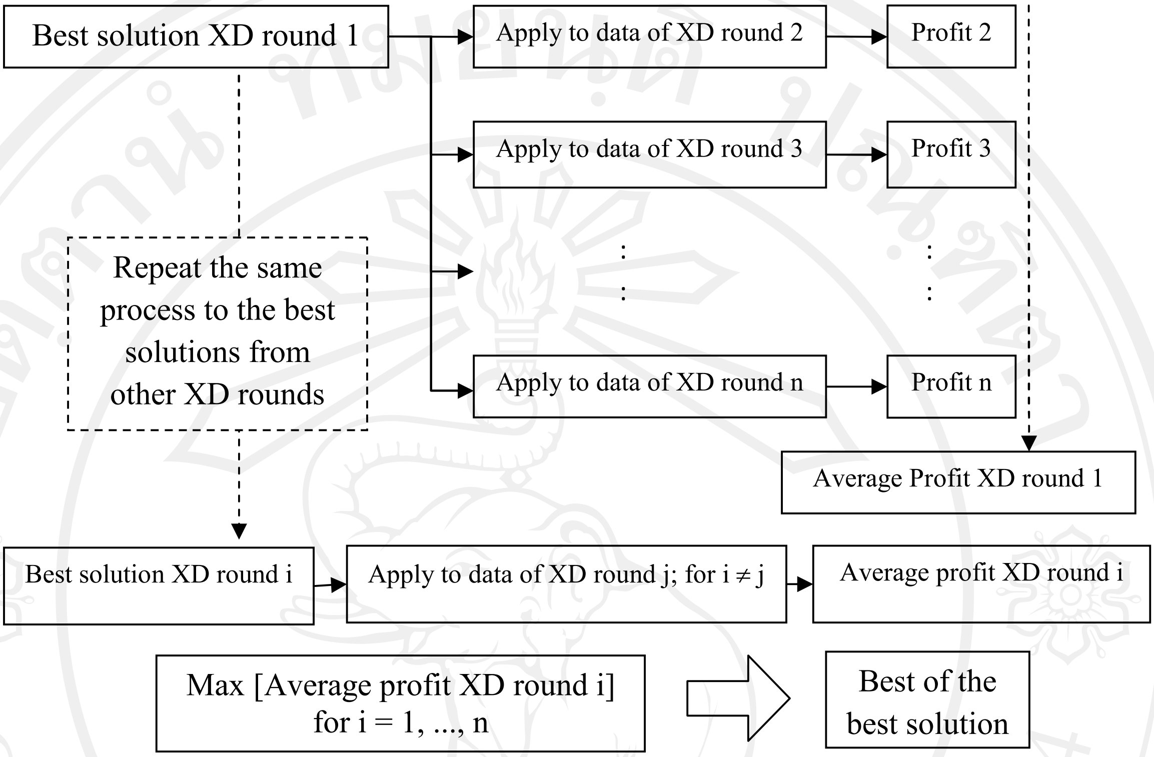


**Figure 13** The process to find best of the best solution (the out-of-sample test)

### 3.2 Criteria for the detection of regime switching

The criteria to detect the regime switching are as follows:

1) The buying signals must appear on at least 4 consecutive days. For example, the buying signals appear on day 40, 39, 38, 37. Then the regime switching takes place on day 37. This rule ensures that the buying signal is strong enough and the days around this period lay on the lower regime. After these days, i.e. day 36 to day 1, the price may rise sharply and lay on the higher regime.

2) In case that the data ends before criteria 1 can be found, the buying signals must appear on at least 2 consecutive days including the end of the data. For example, the buying signals appear on day 30, 29 and 28 for the data of 30 day-investment. In this case, it cannot see what would happen in day 31, 32 and so on. Therefore, it assumes that the buying signal last to those days.

3) When two groups of buying signals with less than 4 consecutive days in each group are separated by only one day of non-buying signal, and when there are at least 2 buying signals in each group, the first day in the first group is the day of regime switching. For example, the buying signals appear on day 12, 11, 9 and 8. In this case day 10 is missing. However, it assumes that the whole period from day 8 to day 12 is on the lower regime. Therefore, regime switching can be counted on day 8.

4) When there is more than one period that meets these criteria, the period that lies nearest to the XD dates is counted as the regime switching. This is to minimize the days of investment. With less days of investment, an investor bears less risk and may earn higher rate of return. For example, two groups of buying signals are day 12, 11, 10, 9 and day 6, 5, 4, 3. The regime switching is detected on day 3 in this case.

5) For the fragmented buying signals, it should be said that there is no signal.

### 3.3 Data

This study examines eight stocks during the period of 2005 – 2011. Seven stocks represent the biggest market capitalized stock of top 7 biggest sectors in the Stock Exchange of Thailand. The eighth stock, TCAP, is a rapidly and strongly growing stock. The author retrieves the data from the Finance and Investment Center (FIC), Chiang Mai University. Names of the stocks are listed in table 1.

**Table 1** List of stocks in the study

| *Abbreviation of stocks (Alphabetical order)* | *Company's Name* | *Sector* | *Market Capitalization in October 2011 (USD Million)* |
|---|---|---|---|
| ADVANC | Advanced Info Service Public Co., Ltd. | Information and communication technology | 12.197 |
| CPALL | CP All Public Co., Ltd. | Commerce | 7,067 |
| CPF | Charoen Pokphand Foods Public Co., Ltd. | Food and Beverage | 7,133 |

**Table 1** (continued)

| *Abbreviation of stocks (Alphabetical order)* | *Company's Name* | *Sector* | *Market Capitalization in October 2011 (USD Million)* |
|---|---|---|---|
| IVL | Indorama Ventures Public Co., Ltd. | Petrochemicals and chemicals | 5,308 |
| KBANK | Kasikornbank Public Co., Ltd. | Banking | 9,081 |
| PTT | PTT Public Co., Ltd. | Energy and utilities | 27,140 |
| SCC | The Siam Cement Public Co., Ltd. | Construction materials | 11,675 |
| TCAP | Tanachart Capital Public Co., Ltd. | Banking | 1,113 |

Note: Exchange rate THB30.8358/USD1.00, Bank of Thailand, October 2011

Source: The Stock Exchange of Thailand, October 2011