# Towards an Agent-Based Foundation of Financial Econometrics: An Approach Based on Genetic-Programming Artificial Markets

**Shu-Heng Chen**
Department of Economics
National Chengchi University
Taipei, Taiwan 11623
E-mail: chchen@nccu.edu.tw

**Tzu-Wen Kuo**
Department of Economics
National Chengchi University
E-mail: g7258502@grad.cc.nccu.edu.tw
Taipei, Taiwan 11623

## Abstract

Using a few nonlinear econometric tools, this paper examines some time-series properties of GP-based artificial markets. We find that GP-based artificial markets are able to replicate several stylized features well documented in financial econometrics. In particular, the time series generated by the GP-based artificial markets are consistent with the efficient market hypothesis in the linear sense. Furthermore, the emergence of these stylized features may be caused by some institutional factors, such as position limits and transaction factors. By introducing the complexity of evolved GP-trees, a bottom-up analysis of the impact of transaction taxes on GP-based artificial markets is also provided.

## 1 Motivation

One of the recent achievements made in *financial econometrics* is to identify several salient features shared by almost all financial markets. Features like *fat tails*, *volatility clusters*, and *nonlinear dependence* have been well documented in Pagan (1996). *Fat tails* concern the fourth moment (kurtosis) of the empirical distribution and refers to the presence of *excess kurtosis*, which is an indicator that the time series under study is *not normally distributed. Volatility clustering* concerns the second moment (variance), more precisely, the dynamics of conditional variance. As Mandelbrot (1963) described, *large changes tend to be followed by large changes-of either sign-and small changes by small changes.* In financial econometrics, this phenomenon is formalized as the GARCH process, where "GARCH" stands for *Generalized AutoRegressive Conditional Heteroskedasticity.* The last one, *nonlinear dependence*, indicates that, while financial time series is not predictable in the linear sense, it may be predictable in the nonlinear sense. Despite their sound statistical basis, a satisfactory economic explanation remains to be established for these *stylized facts.*

In this paper, a time-series econometric study of a GP-based artificial market constructed by Chen and Yeh (1997) is conducted. We attempt to test whether GP-based artificial markets can actually replicate the above-mentioned stylized features. If GP-based artificial markets can, in effect, replicate those patterns, then the expressive power of GP-based markets may help us further explore the possible *institutional connections* for those stylized features. In particular, in this paper, we would like to identify the significance of two institutional factors, namely, *position limits* and *transaction taxes.*

## 2 The Analytical Model

Given the above-mentioned purpose, *GP-based artificial markets* are employed to generate artificial time series of prices of an abstract commodity. The GP-based artificial market used in this paper is based on Chen and Yeh (1996, 1997), which is known as a *cobweb model* in economics. Before proceeding further, let's briefly review this model.[1] Consider a competitive market composed of $n$ firms which produce the same goods by employing the same technology and which face the same cost function described in Equation (1):

$$c_{i,t} = xq_{i,t} + \frac{1}{2}ynq_{i,t}^2 \tag{1}$$

where $q_{i,t}$ is the quantity supplied by firm $i$ at time $t$, and $x$ and $y$ are the parameters of the cost function.

Since at time $t-1$, the price of the goods at time $t$, $P_t$, is not available, the decision about optimal $q_{i,t}$ must be based on the *expectation* (*forecast*) of $P_t$, i.e., $P_{i,t}^e$. Given $P_{i,t}^e$ and the cost function $c_{i,t}$, the expected

[1] One can find details in Chen and Yeh (1996, 1997), which can also be downloaded from the website: http://econo.nccu.edu.tw/ai/staff/csh/vitachen.htm

profit of firm $i$ at time $t$ can be expressed as follows:

$$\pi^e_{i,t} = P^e_{i,t} q_{i,t} - c_{i,t} \tag{2}$$

Given $P^e_{i,t}$, $q_{i,t}$ is chosen at the level such that $\pi^e_{i,t}$ can be maximized and, according to the first order condition, is given by

$$q_{i,t} = \frac{1}{yn}(P^e_{i,t} - x) \tag{3}$$

Once $q_{i,t}$ is decided, the aggregate supply of the goods at time $t$ is fixed and $P_t$, which sets demand equal to supply, is determined by the demand function:

$$P_t = A - B \sum_{i=1}^{n} q_{i,t}, \tag{4}$$

where $A$ and $B$ are parameters of the demand function.

Given $P_t$, the actual profit of firm $i$ at time $t$ is :

$$\pi_{i,t} = P_t q_{i,t} - c_{i,t} \tag{5}$$

In a representative-agent model, it can be shown that the *rational expectations equilibrium price* ($P^*$) and *quantity* ($Q^*$) are (Chen and Yeh, 1996, p.449):

$$P^*_t = \frac{Ay + Bx}{B + y}, \quad Q^*_t = \frac{A - x}{B + y}. \tag{6}$$

To extend the model (Equations (1)-(6)) with *speculation*, the behavior of speculators has to be specified first. Suppose we let $I_{j,t}$ represent the inventory of the $j$th speculator at the end of the $t$th period, then the profit to be realized at the next period $t+1$ is

$$\pi_{j,t} = I_{j,t}(P_{t+1} - P_t). \tag{7}$$

Of course, the actual profit $\pi_{j,t}$ is unknown at the moment when the inventory plan is carried out; therefore, like producers, speculators tend to set the inventory up to the level where speculators' expected utility $Eu_{j,t}$ or expected profit $E\pi_{j,t}$ can be maximized. We shall follow Muth (1961) to assume that the objective function for speculators is to maximize the expected utility rather than the expected profit. Without assuming any specific form of utility function, what Muth (1961) did was to approximate the general utility function $u_{j,t}(\pi_t)$ by taking the second-order Taylor's series expansion about the origin:

$$u_{j,t}(\pi_t) \approx \phi(\pi_t) = \phi(0) + \phi'(0)\pi_{j,t} + \frac{1}{2}\phi''(0)\pi^2_{jt} \tag{8}$$

Based on Equation (8), the approximate utility depends on the moments of the probability distribution of $\pi_t$, i.e.,

$$Eu_{j,t} \approx \phi(0) + \phi'(0)E\pi_{j,t} + \frac{1}{2}\phi''(0)E\pi^2_{j,t} \tag{9}$$

Solving the first and the second moment of Equation (9), we can rewrite the expected utility function as follows.

$$\begin{aligned} Eu_{j,t} \approx\ & \phi(0) + \phi'(0)I_{j,t}(P^e_{j,t+1} - P_t) \\ & + \frac{1}{2}\phi''(0)I^2_{j,t}[\sigma^2_{t,1} + (P^e_{j,t+1} - P_t)^2], \end{aligned} \tag{10}$$

where $P^e_{j,t+1}$ is the conditional expectation $E(P_{t+1} \mid \Omega_t)$ and $\sigma^2_{t,1}$ is the conditional variance $var(P_{t+1} \mid \Omega_t)$ and $\Omega_t$ is the $\sigma$-algebra (the largest information set) generated by $P_t, P_{t-1}, ....$ The optimal position of the inventory can then be derived approximately by solving the first order condition and the optimal position of the inventory $I^*_{j,t}$ is given by

$$I_{j,t} = \alpha(P^e_{j,t+1} - P_t), \tag{11}$$

where $\alpha = -\frac{\phi'(0)}{\phi''(0)\sigma^2_{t,1}}$. Equation (11) explicitly shows that speculators' optimal decision about the level of inventory depends on their expectations of the price in the next period, i.e., $P^e_{j,t+1}$.

Now, if the market is composed of $n$ producers and $m$ speculators, the equilibrium condition is given in Equation (12),

$$\begin{aligned} & \frac{A}{B} - \frac{1}{B}P_t + \sum_{j=1}^{m} \alpha(P^e_{j,t+1} - P_t) \\ = & \sum_{i=1}^{n} \frac{1}{yn}(P^e_{i,t} - x) + \sum_{j=1}^{m} \alpha(P^e_{j,t} - P_{t-1}). \end{aligned} \tag{12}$$

## 3 Experimental Designs

Chen and Yeh (1997) replaced the conditional expectations appearing in Equations (12) by a GP-driven learning processes, and simulated the price dynamics under this new setup. While they showed how speculators may have adverse impacts on market stability, *properties of these price dynamics were largely left unexploited.* In this paper, we shall first resimulate the price dynamics of this market and then conduct a rigorous econometric analysis of the price dynamics. In particular, we would like to see whether our GP-based markets posses the econometric properties widely existing in financial time series. If so, how the emergence of these properties can be possibly accounted for by institutional factors, such as transaction taxes and position limits.

The cobweb markets are composed of two groups of adaptive agents, *producers* and *speculators.* (At this stage, consumption demand is given exogenouly; hence the adaptive behavior of consumers is not explicitly modeled at this moment.) The adaptive behavior addressed here is exclusively restricted to *the formation process of expectations.* At different periods in time, each agent's behavior is characterized by a *model* on which the agent's forecast and decision-making is based. For producer $i$, this model is a forecasting function employed to forecast the next period's price, i.e., $P^e_{i,t}$ in Equation (2). For speculator $j$, this model is an *position function*, which is a function of price history, i.e., $I_{j,t}$ in Equation (11). The *evolving agents* can then be considered as *the evolution of a collection of models*:

$$POP_0 \rightarrow POP_1 \rightarrow POP_2 \rightarrow ... \rightarrow POP_t \rightarrow ..., \quad (13)$$

where $POP_t$ denotes the population of models at time period $t$. A natural approach to implement the evolutionary process depicted above is *genetic programming.*

The end-user supplied control parameters for this study is given in Table 1. Here, we consider a model composed of 300 producers and 100 speculators. These numbers are chosen to roughly mimic a real advanced economy, i.e., 25% of GDP is from the financial industry and 75% of GDP is from the manufacturing industry. The function set defines the set of possible mappings, i.e., the set of all possible forms of $P^e_{i,t}$ and $I^e_{j,t}$. As we may notice, the functions included in our function set are very limited to only $+, -, Sin, Cos$. This choice is based on our calibration described as follows. According to Equation (6), the equilibrium price is determined by four parameters $A, B, x$ and $y$, and is \$1.12 given their values specified in Table 2. Therefore, a simple operation of $Sin$ and $Cos$ is good enough to have a range covering this point, 1.12. In other words, the function set chosen here is a minimal set to satisfy the *closure property.*

The other reason that we have this limited choice is due to *position limits.* For each speculator $j$, $I_{j,t}$ can be both positive (long position) and negative (short position). However, these positions are restricted to a limit $s$, i.e., $-s \leq I_{j,t} \leq s, \forall t$ (See Table 2). So, the inclusion of $Exp$, $RLog$, $\times$ and $\div$ can easily make $I_{j,t}$ beyond this boundary and result in a number being either $s$ or $-s$. Therefore, while speculators can be different in the genotype, but is identical in phenotype, and hence identical in fitness. In this case, the selection process may, in effect, proceed with an almost uniform distribution, which is certainly not a desirable feature.

The terminal set includes the ephemeral random floating-point constant $R$ ranging over the interval [-9.99, 9.99] and the price lagged up to 10 periods $P_{t-1}, ..., P_{t-10}$. While little guidance is available to decide what horizon should speculators use to form their expectations, based on a few pilot experiments, we believe that most of our results presented below would not be sensitive to a longer horizon. The terminal set and the function set together determine inputs of the trees evolved by GP. The output is $P^e_{i,t}$ for producers and $I_{j,t}$ for speculators.

The *selection scheme* is an important operator in genetic programming. When applying genetic programming to *optimization*, the user must notice that different selection schemes may have different implications for the fitness value, selection intensity, selection variance, and loss of diversity. By the same token, when genetic programming is applied to *simulating the evolution and learning of the economic system*, we have to keep in mind that different schemes may have different economic implications. From the viewpoint of *matching processes*, proportionate selection is prone to a *global* network and tournament selection is prone to a *local* network. Since local interaction among speculators plays an extremely important role in finance (Shiller, 1984), tournament selection is more appropriate than proportionate selection.

In the context of economics, profit seems to be a very natural measure for fitness. Here, profit is defined in Equation (5) for the producer and in Equation (7) for the speculator.

Next, the GP-based adaptive agents are placed in ar-

Table 1: Tableau of GP-Based Cobweb Model

| | |
|---|---|
| Number of producers (n) | 300 |
| Number of speculators (m) | 100 |
| Number of trees created by the full method | 30 (P), 10 (S) |
| Number of trees created by the grow method | 30 (P), 10 (S) |
| Function set | $\{+, -, Sin, Cos\}$ |
| Terminal set | $\{P_{t-1}, P_{t-2}, \cdots, P_{t-10}, R\}$ |
| Selection scheme | Tournament selection |
| Tournament size | 2 |
| Number of trees created by reproduction | 30 (P), 10 (S) |
| Number of trees created by crossover | 210 (P), 70 (S) |
| Number of trees created by mutation | 60 (P), 20 (S) |
| Mutation Scheme | Tree Mutation |
| Probability of mutation | 0.2 |
| Maximum depth of tree | 17 |
| Probability of leaf selection under crossover | 0.5 |
| Number of generations ($GEN = t$) | 9000 |
| Maximum number in the domain of Exp | 1700 |
| Criterion of fitness | Profit |

"P" stands for the producers and "S" stands for the speculators. The number of trees created by the full method or grow method is the number of trees initialized in Generation 0 with the depth of tree being 2, 3, 4, 5, and 6. For details, see Koza (1992).

Table 2: Institutional Designs

| Scenario | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Tax Rate ($\tau$) | 0 | | | | 0.01 | | | |
| Position Limit ($s$) | 0.01 | 0.1 | 0.01 | 0.1 | 0.01 | 0.1 | 0.01 | 0.1 |
| Parameter $A$ | 2.296 | | 3.36 | | 2.296 | | 3.36 | |
| Parameter $B$ | 0.0168 | | 0.032 | | 0.0168 | | 0.032 | |
| Parameter $x$ | 0 | | | | | | | |
| Parameter $y$ | 0.016 | | | | | | | |
| Cobweb Ratio ($B/y$) | 1.05 | | 2 | | 1.05 | | 2 | |

For all cases, the time constraint for recovering the short is set to be 20.

tificial markets with the following two particular institutional designs. One concerns *trading restrictions*. In addition to the above-mentioned position limits $s$ and $-s$, for those speculators who hold a short position ($I_{j,t} < 0$), there is a time limit $d$ for recovering the short. In this paper, $d$ is set to be 20 for all simulations. In other words, when the short position has remained for 20 trading days, the speculator is forced to recover the short on the next trading day. The other involves the *transaction tax*. The transaction tax considered in this paper is a proportionate tax and is denoted by the tax rate $\tau$. The tax rate is imputed to speculators only and is imposed on both directions of trading, to buy and to sell. These institutional parameters are summarized in Table 2.

Here, we consider two tax rates (0 and 1%) and two position limits (0.01 and 0). Since transaction taxes and position limits are two major components to affect speculators' potential profits, changing these two parameters will have an influence on speculators' motives and hence their adaptive behavior. It would then be interesting to see how these changes may have further impacts on price dynamics in terms of their econometric properties. Apart from the tax rate and position limit, we also include two different cobweb ratios (1.05 and 2), which is defined as the ratio of the slope of the demand curve to the slope of the supply curve, $B/y$. While, by Equation (6), these two different cobweb ratios have the same rational expectations equilibrium price ($1.12), a higher cobweb ratio tends to be more inherently unstable. (Chen and Yeh, 1996)

Given the description of our adaptive speculators and trading restrictions above, the exact equation to derive $P_t$ is:

$$\begin{aligned} &\frac{A}{B} - \frac{1}{B}P_t + \sum_{j=1}^{m} I_{j,t} \\ = &\sum_{i=1}^{n} \frac{1}{yn}(P^e_{i,t} - x) + \sum_{j=1}^{m} I_{j,t-1}, \end{aligned} \tag{14}$$

where $-s \leq I_{j,t} \leq s$ for all $j, t$, and $0 \leq I_{j,t} \leq s$ if $I_{j,t-k} < 0$ for $k = 1, ...20$.

Everything we have described is also well encapsulated into Equation (13). The evolving targets are $\{P^e_{i,t}\}_{i=1}^{300}$ and $\{I_{j,t}\}_{j=1}^{100}$. At the end of each trading day, $P_t$ is announced, and fitness of $i, j$, $\{\pi_{i,t}\}_{i=1}^{300}$ and $\{\pi_{j,t}\}_{j=1}^{100}$, can be calculated. Genetic operators are then applied to evolve these two populations separately with the tournament selection scheme. The new generation $\{P^e_{i,t+1}\}$ and $\{I_{j,t+1}\}$ is then generated, and the market is open again. The cycle goes on and on until it meets the termination condition, which is the number of generations in this paper.

"Number of Generations" in set to be 9,000 in all simulations. Notice that the number of generations is also the time scale of the simulation, i.e., $GEN = t$. In other words, we are simultaneously evolving the population while deriving the market-clearing price, $P_t$.

Finally, the program to implement all simulations in this paper is called **Speculators**, which is available from the website:
http://econo.nccu.edu.tw/ai/staff/csh/Software.htm

## 4 Time Series Analysis of Price Series

There are totally eight scenarios simulated in this study. For each scenario, we conducted five independent runs, with 9000 periods for each. These resulted in 5 artificial time series for each scenario (40 time series in total). Figures 1.1-1.4 display the time series $\{P_t\}$ for a typical run for some scenarios. Table 3 summarizes the basic statistics of these simulations. Based on these figures and statistics, we can see that $P_t$ basically fluctuates around the rational equilibrium price ($P^* = 1.12$). However, the volatility of $P_t$ ($\sqrt{Var(P)}$) depends on the institutional parameters, in particular, the cobweb ratio and the position limit. Generally speaking, the higher the cobweb ratio and the position limit, the more volatile the price. What seems a little counter-intuitive is that 1 percent tax rate does not stabilize the price movement to a significant degree.

We then examined the econometric properties of these artificial time series. The statistical properties under examination are motivated by the list of stylized features documented by Pagan (1996). To prevent the statistics from being influenced by the initialization process, the first 3000 observations are all

Table 3: Mean and Volatility of Prices

| Scenario | $\overline{P}$ | $\sqrt{Var(P)}$ | Scenario | $\overline{P}$ | $\sqrt{Var(P)}$ |
|---|---|---|---|---|---|
| A | 1.1189 | 0.0191 | E | 1.1291 | 0.0148 |
| B | 1.1181 | 0.0209 | F | 1.1181 | 0.0212 |
| C | 1.1183 | 0.0359 | G | 1.1184 | 0.0717 |
| D | 1.1190 | 0.0481 | H | 1.1197 | 0.0505 |

The $\overline{P}$ and $\sqrt{Var(P)}$ reported here are the average of the mean and the standard deviation of the five runs. For each run, the mean and the standard deviation are calculated from the last 6000 observations only, i.e., $\{P_t\}_{t=3001}^{9000}$.

dropped from the original series. In other words, all statistical properties are examined under the subseries $\{P_t\}_{t=3001}^{9000}$. The first property to examine is *normality*. A common statistic employed to test normality is the *Jarque-Bera statistics*. Based on these statistics, the null hypothesis that the price series is normally distributed is rejected for all 40 series at the 5% significance level. This result is consistent with a well-known result in empirical finance: most financial return series are not normally distributed (the tails are too fat as opposed to the normal distribution).

The next econometric property to examine is the IIDness of the series, i.e., *are price series identically and independently distributed over time?*. The most frequently used test in this area is the celebrated BDS test (Brock, Dechert and Scheinkman, 1996). Due to the page limit, we do not intend to give a full account of the BDS test, the interested reader can find details and a program to run this statistic in the following website:
http://econo.nccu.edu.tw/ai/staff/csh/course/finaecon/lec4/lec4.htm
The BDS test is frequently applied to testing *non-linear dependence*. Hence, the linear process of the time series has to be filtered out before implementing the test. We applied Rissanen's *predictive stochastic complexity* (PSC) to filter out the linear process of each $\{P_t\}_{t=3001}^{9000}$. (Rissanen, 1986). A detailed description of **PSC** with many illustrated examples and a computer program can be found in the website:
http://econo.nccu.edu.tw/ai/staff/csh/course/finaecon/lect8/lect8.htm
One of the by-products of the PSC filter is to inform us of the linear AutoRegressive-MovingAverage process, i.e., the $ARMA(p,q)$ process, extracted from the original series. Among the 40 series examined, 38 have no linear process at all, i.e., they are all identified as $ARMA(0,0)$. The only two exceptions are the one run under Scenario C ($ARMA(0,2)$) and the one run under Scenario E ($ARMA(1,0)$). This result indicates that *the GP-based artificial market is so efficient that there are hardly any linear signals left*. To some extent, this can be considered as a match for the classical version of *the efficient market hypothesis*.

We then applied the BDS test to the *filtered residuals*. The parameter "$\epsilon$" in the BDS test is equal to one standard deviation. (In fact, we also tried other epsilons, but the result is not sensitive to the choice of epsilon.) The embedding dimensions considered are from 2 to 5. Following Barnett et al. (1997), if the absolute value of all BDS statistics under various embedding dimensions are greater than 1.96, the null hypothesis of IIDness is rejected. In this case, nonlinear dependence is detected. If all of them are less than 1.96, then one fails to reject the null hypothesis of IIDness. However, if some are greater than 1.96, and some are less than 1.96, then the result is ambiguous. (Actually, in Barnett's paper, they used the word "strongly reject" and "weakly reject". We do not intend to make such a distinction here.)

Based on this criterion, the null hypothesis of IIDness is rejected 18 times out of 20 when transaction taxes are imposed. However, it is rejected only 14 times out of 20 when there is no transaction tax. Furthermore, when the position limit is relaxed from 0.01 to 0.1, the number of rejection increases from 14 to 18 times. Therefore, imposing transaction taxes and relaxing position limits may weaken the degree of statistical independence of the data. The first half of the result is nothing surprising as transaction cost reduces the chance of arbitraging and hence make the use of information less efficient in terms of statistical independence. What is surprising is the second half of the result, for the relaxation of the position limit should make it more rewarding for speculators to extract information from the price series. The resultant time series is anticipated to be more efficient and is more likely to be IID.[2]

So far, we have examined our simulated time series with a test for non-linear dependence. However, it is well known that most of the non-linearity in financial data seems to be contained in their second comments. The voluminous (G)ARCH (Generalized AutoRegressive Conditional Heteroskedasticity) literature is the outcome of the attempt to capture by appropriate time series models the regularities in the behavior of volatility. In order to proceed further, we carry out the Lagrange multiplier test for the presence of ARCH effects. A detailed description of ARCH and GARCH and an associated SAS program to run the test is available from the website:
http://econo.nccu.edu.tw/ai/staff/csh/course/finaecon/lec6/lec6.htm
If the ARCH effect is rejected, we will further identify the GARCH structure of the series by using the Akaike Information Criterion (AIC). The results are exhibited in Table 4.

[2]However, as we shall see later, relaxing position limits also result in a higher probability of having a GARCH process.

Table 4: GARCH Modeling of Artificial Time Series

| Scenario<br>Run | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | (2,1) | (1,1) | • | (1,1) | (2,1) | (2,2) | (1,2) | (1,1) |
| 2 | (2,2) | (2,2) | (1,1) | (2,1) | (1,1) | (2,1) | (2,2) | (1,1) |
| 3 | (2,1) | (1,1) | • | (1,1) | • | (1,1) | (2,1) | (1,1) |
| 4 | (1,1) | (1,1) | (2,1) | (2,1) | • | (1,1) | (2,1) | (2,2) |
| 5 | (2,1) | • | (2,1) | (1,1) | (1,1) | (2,1) | (2,2) | (1,1) |

The (p,q) within each bracket refers to the model GARCH(p,q), while • means that there is no ARCH effect.

There are couple of points worth noting. First, we can find that *volatility clustering* characterized as the *ARCH effect* is quite ubiquitous. Out of the 40 series, there are only 5 series without the ARCH effect. Not surprisingly, all these five series fail to reject the null hypothesis of the BDS test. However, *a few time series which fail to reject the BDS test still have the ARCH effect.* Second, transaction taxes seem to play no role in accounting for the emergence of the ARCH effect. For example, 17 out of the 20 runs without transaction taxes exhibit the ARCH effect, while 18 out of the 20 runs with transaction taxes have the effect. Nevertheless, position limits may have certain effects. In our simulations, the scenarios with low position limits fail to reject the ARCH effect in 4 out of 20 runs, while the ones with high position limits only fail once. Therefore, *one may hypothesize that position limits may have some connection to the ARCH effect, and relaxing position limits can increase the likelihood of the emergence of the ARCH effect.*

## 5 The Complexity of Evolved Strategies

In addition to the *macro*-phenomenon, i.e., the price series, an equally important thing is the *micro*-phenomenon, i.e., *what happens for the individuals who collectively generate such a complex nonlinear dynamics.* Certainly, one may ask *whether the complex macro-phenomenon is coupled with the complex micro-phenomenon*; in other words, agents with sophisticated strategies collectively generated complex macro-phenomenon.

To give an analysis of the connection between *bottom* and *up*, we give two definitions of the complexity of a GP-tree. The first definition is based on the number of nodes appearing in the tree, while the second one is based on the depth of the tree. At the end of each run, we have a profile of the evolved GP-trees for 300 producers and 100 speculators. The complexity of each GP-tree is computed. We then average the complexity of evolved GP-trees by producers and by speculators. Tables 5 and 6 give the results.

Table 5: Complexity of Evolved Strategies: Number of Nodes

| Average over 300 Producers | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Scenario<br>Run | A | B | C | D | E | F | G | H |
| 1 | 1.05 | 1.21 | 1.01 | 1.01 | 1.00 | 1.01 | 1.02 | 1.02 |
| 2 | 1.02 | 1.01 | 1.01 | 1.03 | 1.02 | 1.05 | 1.01 | 1.00 |
| 3 | 1.10 | 1.01 | 1.03 | 1.05 | 1.02 | 1.02 | 1.06 | 1.03 |
| 4 | 1.03 | 1.06 | 1.02 | 1.01 | 1.01 | 1.06 | 1.06 | 1.05 |
| 5 | 1.02 | 1.03 | 1.02 | 1.05 | 1.03 | 1.01 | 1.02 | 1.09 |
| mean | 1.04 | 1.06 | 1.02 | 1.03 | 1.02 | 1.03 | 1.03 | 1.04 |
| mean | 1.03 | | | | 1.03 | | | |
| Average over 100 Speculators | | | | | | | | |
| Scenario<br>Run | A | B | C | D | E | F | G | H |
| 1 | 1.88 | 4.47 | 1.45 | 3.38 | 1.26 | 2.46 | 3.13 | 5.59 |
| 2 | 2.51 | 2.71 | 4.42 | 4.20 | 1.93 | 1.93 | 2.16 | 3.36 |
| 3 | 4.46 | 3.86 | 4.32 | 2.21 | 2.87 | 1.26 | 3.17 | 4.01 |
| 4 | 1.99 | 1.59 | 1.49 | 1.93 | 2.02 | 1.53 | 2.48 | 3.07 |
| 5 | 1.23 | 1.82 | 23.19 | 8.96 | 2.86 | 1.55 | 2.97 | 1.23 |
| mean | 2.41 | 2.89 | 6.97 | 4.14 | 2.19 | 1.75 | 2.78 | 3.45 |
| mean | 4.10 | | | | 2.54 | | | |

At first sight, it seems to be very remarkable that producers are so *simple*: *many evolved GP-trees consist of only one node.* This seems to indicate that GP is not "at work" at all or the problem is not interesting. A careful examination, however, informs us that this is not the case. In fact, in each stage of evolution, very often the fittest tree belonged to the class of complex nonlinear models. However, this group did not successfully propagate.

The reason is that for producers their profits depend on only the *flow* of the quantity supplied, which is a function of *the first moment of prices*, $E(P_t)$. On the average, this number is constantly around 1.12 (Figures 1.1-1.4 and Table 3). Furthermore, from the previous PSC-filtering results, there is not much *linear signal* left in the first moment of the price series. As a result, statisticians who study our markets may conclude with the following model:

$$P_t = 1.12 + \epsilon_t, \tag{15}$$

where $E(\epsilon_t) = 0$. Given this model, the forecast

$$P^e_{i,t} = 1.12. \tag{16}$$

seems to be very competitive, and this is the most common type of the one-node trees. Although there are other nonlinear models which can outperform Equation (15), but they are only suitable for certain types of nonlinearity and are not *robust* to the general nonlinear properties of $\epsilon_t$. Hence, while complex models can frequently have the championship, they have difficulties to keep it and be prosperous. Eventually, the majority belongs to the robust simple models, such as *taking the simple average.*

While evolving simple strategies may sound strange for GPers, it is a quite popular idea in economics. In economics, simple strategies known as *rules of thumbs*

Table 6: Complexity of Evolved Strategies: Depth of Trees

| Average over 300 Producers | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Scenario | A | B | C | D | E | F | G | H |
| Run | | | | | | | | |
| 1 | 1.03 | 1.10 | 1.01 | 1.01 | 1.00 | 1.01 | 1.02 | 1.01 |
| 2 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.04 | 1.01 | 1.00 |
| 3 | 1.05 | 1.01 | 1.02 | 1.03 | 1.01 | 1.01 | 1.03 | 1.02 |
| 4 | 1.02 | 1.02 | 1.01 | 1.00 | 1.01 | 1.02 | 1.03 | 1.03 |
| 5 | 1.01 | 1.02 | 1.01 | 1.02 | 1.02 | 1.01 | 1.01 | 1.05 |
| mean | 1.02 | 1.03 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 |
| mean | 1.02 | | | | 1.02 | | | |
| Average over 100 Speculators | | | | | | | | |
| Scenario | A | B | C | D | E | F | G | H |
| Run | | | | | | | | |
| 1 | 1.54 | 3.16 | 1.37 | 2.69 | 1.21 | 1.77 | 2.46 | 3.89 |
| 2 | 1.86 | 2.05 | 2.61 | 2.63 | 1.59 | 1.59 | 1.68 | 2.50 |
| 3 | 2.72 | 2.18 | 2.61 | 1.86 | 2.53 | 1.26 | 2.35 | 2.67 |
| 4 | 1.56 | 1.34 | 1.33 | 1.71 | 1.65 | 1.37 | 1.85 | 2.00 |
| 5 | 1.15 | 1.44 | 8.40 | 3.92 | 1.86 | 1.52 | 2.03 | 1.20 |
| mean | 1.77 | 2.03 | 3.26 | 2.56 | 1.77 | 1.50 | 2.07 | 2.45 |
| mean | 2.39 | | | | 1.95 | | | |

are not best (*optimal*) at very point in time, but are very robust for different environments and become the most popular rules on which people rely. One famous example is that 65 to 70 percent of all mutual and pension fund managers fail to beat the simple rule, the market indices over the long run (Malkiel, 1996). Sophisticated strategies can be optimal, but they are not robust and have to be updated quickly. When they are not properly updated, they can perform extremely poor in a very dynamic environment.

On the other hand, speculators have evolved more complex strategies than producers, and this is true for all scenarios. For speculators, their profits depend on both the *flow* and *stock* of the quantity supplied, which is a function of both the first moment and *high-order moments*. From the previous volatility analysis (GARCH analysis), most series have linear structure in the second moment; consequently, it requires speculators to evolve more complex strategies to extract these signals. We plan to put all evolved strategies on the website so that interested reader can have their own analysis.

The other interesting result is *the effect of the transaction tax on the complexity of evolved strategies.* From Table 5, it can been seen that imposing the transaction tax tends to evolve simpler strategies for speculators. This result is consistent with the one found in the BDS test. In the BDS test, imposing the transaction tax increases the probability of rejecting IIDness, i.e., increase the chance of leaving nonlinear signals unexploited. As we conjectured earlier, this is due to a weaker incentive to extract information, as imposing the transaction tax reduce the chance of arbitrage. Here, from the *bottom* part, we actually see that speculators indeed become "*more lazy*" when transaction taxes are imposed.

## 6 Concluding Remarks

This paper provides a thorough time-series analysis of prices generated from genetic-programming artificial markets. Many stylized features well documented in financial econometrics can in principle be replicated from the GP-based artificial markets, which including *leptokutosis*, *non-IIDness* and *volatility clustering.* Moreover, the GP-based artificial markets allow us to search for the behavioral foundation of these stylized features. The two institutional factors, *transaction taxes* and *position limits*, may both contribute to the emergence of those stylized features.

As one may expect that transaction taxes can have adverse effects on speculative trades. Our analysis of GP-based markets partially supports this viewpoint. From the *bottom* part, transaction taxes reduce the chance of arbitrage; hence, speculators have less incentive to search. In particular, the GP-trees evolved get *simpler* when the transaction tax is imposed. Corresponding to the bottom part, what we have experienced on the upper part is a less unexploited or a more nonlinear dependent series. Nevertheless, the emergence of volatility clustering may be a consequence of relaxing position limits and have little to do with transaction taxes.

The empirical evidences accumulated from GP-based markets' simulations are quite limited. At this moment, they can be only useful for the purpose of motivating hypotheses. However, the point of this paper is mainly to show *what* GP-based markets can potentially serve for the advancement of the economic theory. In the future, it is expected that a larger scale of simulation will be conducted for getting more fruitful results.

**Acknowledgement**

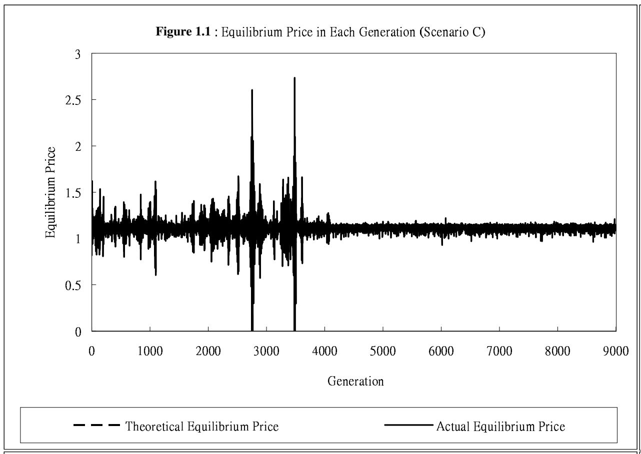
Research support from NSC grants No.88-2415-H-004-008 and No. 87-I-2911-15 is gratefully acknowledged. The authors are also grateful for the helpful comments from three anonymous referees.

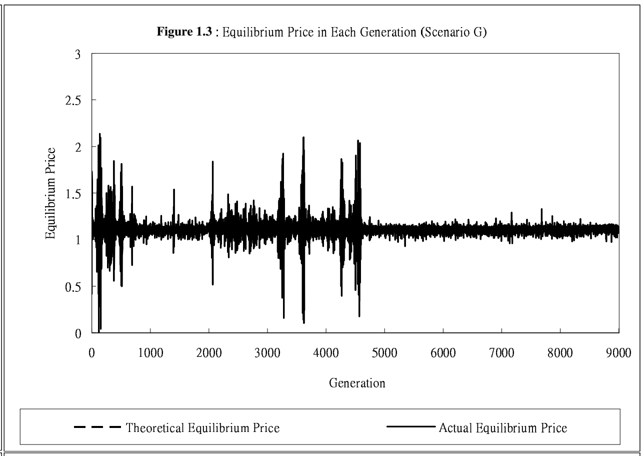
**References**

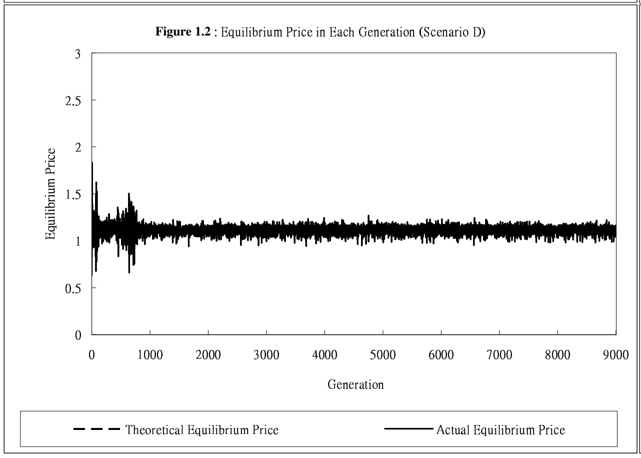
Barnett, W. A., A. R. Gallant, M. J. Hinich, J. A. Jungeilges, D. T. Kaplan and M. J. Jensen (1997), "A Single-Blind Controlled Competition Among Tests for Nonlinearity and Chaos," paper presented at *the 1997 Far Eastern Meeting of the Econometric Society* (**FEMES'97**), Hong Kong, July 24-26, 1997. (**Session 4A**)

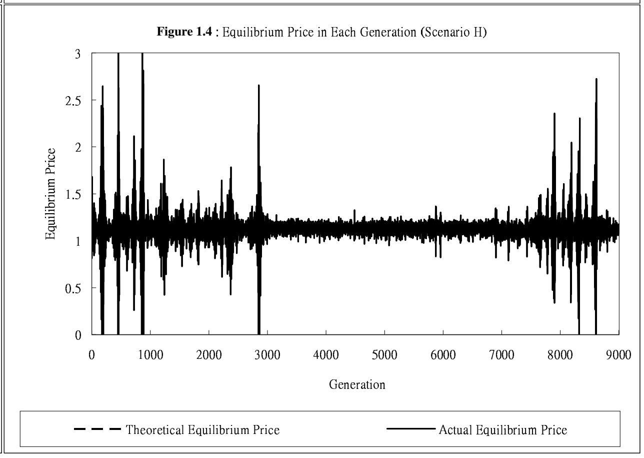
Brock, W. A., W. D. Dechert, B. LeBaron, and J. Scheinkman (1996), A Test for Independence Based on the Correlation Dimension," *Econometric Reviews* 15, pp. 197-235.

Chen, S.-H. and C.-H. Yeh (1996), "Genetic Program-




ming Learning and the Cobweb Model," in Angeline, P. and K. E. Kinnear, Jr. (eds.) *Advances in Genetic Programming 2*, MIT Press, Cambridge, MA. Chapter 22. pp. 443-466.

Chen, S.-H. and C.-H. Yeh (1997), "Modeling Speculators with Genetic Programming," in P. Angeline, R. G. Reynolds, J. R. McDonnell and R. Eberhart (eds.), *Evolutionary Programming* VI, **Lecture Notes in Computer Science**, Vol. 1213, Berlin:Springer-Verlag. pp. 137-147.

Chen, S.-H. and C.-H. Yeh (1998), "Genetic Programming in the Overlapping Generations Model: An Illustration with Dynamics of the Inflation Rate," in V. W. Porto, N. Saravanan, D. Waagen and A. E. Eiben (eds.), *Evolutionary Programming VII*, **Lecture Notes in Computer Science**, Vol. 1447, Berlin:Springer-Verlag. 1998. pp. 829-838.

Koza, J. (1992), *Genetic Programming*, MIT Press, Cambridge.

Malkiel, B. G. (1996), *A Random Walk Down Wall Street*, Norton.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business* 36, pp. 394-419.

Muth, J. F. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrics 29*, pp. 315-335.

Pagan, A. (1996), "The Econometrics of Financial Markets," *Journal of Empirical Finance,* 3, pp.15-102.

Rissanen, J. (1986), "Stochastic Complexity and Modeling," *Annals of Statistics*, Vol. 14, No. 3, pp. 1080-1100.

Shiller, R. J. (1984), "Stock Prices and Social Dynamics," *The Brookings Papers on Economic Activity,* Vol. 2, pp. 457-510.