Using genetic algorithms and linear regression analysis for private housing demand forecast

Thomas Ng, S.; Skitmore, Martin; Wong, Keung Fai

Published in:
Building and Environment

DOI:
10.1016/j.buildenv.2007.02.017

Revised Paper submitted to

Building and Environment

Using Genetic Algorithms and Linear Regression Analysis for Private Housing Demand Forecast

(Reference No: 05/1859)

S. Thomas Ng [1]
Martin Skitmore [2]
Keung Fai Wong [3]

Correspondence Address:

Dr. Thomas Ng
Department of Civil Engineering,
The University of Hong Kong,
Pokfulam Road, Hong Kong.

Tel: Int+ (852) 2857 8556
Fax: Int+ (852) 2559 5337

Email: tstng@hkucc.hku.hk

(Version 2a) 6th January 2007

[1] Associate Professor, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. *E-mail: tstng@hkucc.hku.hk*

[2] Professor, School of Urban Development, Queensland University of Technology, GPO Box 2434, Brisbane Q4001, Australia. *E-mail: rm.skitmore@qut.edu.au*

[3] Student, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. *E-mail: kfwkf@graduate.hku.hk*

USING GENETIC ALGORITHMS AND LINEAR REGRESSION ANALYSIS FOR PRIVATE HOUSING DEMAND FORECAST

Abstract

An accurate prediction of prospective construction supply and demand, especially the private residential market, is paramount important to policy makers, as it could help formulate strategies to cultivate/stabilize the economy and satisfy the social needs (at macro level). Despite that, a realistic prediction of future private residential demand is never an easy task, as it is governed by a number of social and economic factors. In this paper, four leading indicator models are developed and compared for directly forecasting Hong Kong private sector residential demand. These comprise a (i) Linear Regression Analysis (LRA) model; (ii) Genetic Algorithms (GA) model; (iii) GA-LRA model, where LRA is used to select the indicator variables; and (iv) GA-LRA model with Adaptive Mutation Rate (AMR) to reduce the likelihood of local optima. The findings indicate that the GA-LRA model with AMR provides the most accurate forecasts and over a longer time horizon. In providing a range of possible forecasts, the model also provides an opportunity for the decision-maker to exercise judgment in selecting the most appropriate forecasts.

Keywords: Forecasting, housing, demand, supply, private sector, models, genetic algorithm

INTRODUCTION

Precise estimation of demand for new residential properties is never a simple task, as it could be influenced by a number of dynamic factors *viz*. demographic change, economic pattern, government policy and external environment (Raftery, 1998). While many major cities are confronted with a shortage of public housing and a soaring private property price (Tse *et al*, 1999), it is usually the government's responsibility to formulate suitable long-term housing strategies and policies to regulate and accommodate the housing needs of different sectors such that a sufficient amount of land and housing units are available to satisfy the demand. In order to make housing policy decisions, it is first necessary to estimate both short-term and long-term future housing demand.

As the housing stock is relatively inelastic in the short run, an overly conservative prediction in the housing demand could result in a shortage of residential supply. However, no one would ever imagine an overly optimistic housing forecast could also lead to profound effects to the locality especially on the overall economy. Recent example in Hong Kong (HK) have illustrated that a surplus supply of residential units had an inverse relation to the price of real estates (the property price in HK plummeted by almost 60% between 1998 and 2003). Reliable estimation of new residential property not only concerns policy makers, planners and home purchasers/tenants, but could also determine the survival of many companies related to the construction sector (Lansley *et al*, 1980).

Despite its strategic significance, little research has been carried out to enhance the methods for predicting the residential demand. In some cases, estimations are made according to a projection of flats required for new households (e.g. new marriage, divorce, new immigrant, etc.) and existing families (e.g. those affected by redevelopment programs). Surely, demographic change would have significant implication to the housing demand, yet one should not ignore the impacts of economic change on the desire of property purchase (Lavender, 1990). According to Hillebrandt (1985), the effects of economy on construction occur at all level and in all aspects of economic life, hinting that the economy (e.g. income, interest rate, etc.) may somehow influence the demand for residential properties, especially on private housing.

This paper reports on a comparison of four leading indicator models for forecasting Hong Kong private sector housing supply (as a proxy for demand) directly. These comprise a (i) Linear Regression Analysis (LRA) model, (ii) Genetic Algorithms (GA) model, (iii) GA-LRA model, where LRA is used to select the indicator variables; and (iv) GA-LRA model with Adaptive Mutation Rate (AMR) to reduce the possibility of local optima. The findings suggest that the GA-LRA model with AMR provides the most accurate forecasts and over a longer time horizon. In providing a range of possible forecasts, the model also provides opportunity for the decision-maker to exercise some judgment in selecting the most appropriate forecasts.

ECONOMIC INDICATORS

The findings of previous research studies (e.g. Killingsworth, 1990; Akintoye and Skitmore, 1994) realized a close relationship exists between the construction and economic cycles, and thereby swings in the economy can be treated as indicators of the prospective movement in the construction industry and *vice versa*. The cyclical indicator technique can be used to exploit this for forecasting purposes. Although not without its shortcomings (i.e. its apparent lack of theoretical basis and inability to explain transmission processes), this technique can be used in any market-oriented economy (Kanaengnid, 1992).

Having been used for various aspects of construction forecasting (Akintoye and Sktimore, 1994; Goh, 1996, 1999; Ng *et al*, 2000), the suitability of economic indicators for forecasting the private residential demand should be explored. In developing a similar system for the private sector, it is clear that some aspects of the public sector model are useful. Indicator variables such as marriage, divorce, etc., are likely to be relevant to both sectors. In addition, public housing demand can also be treated as an indicator for future public sector housing supply (as a proxy for demand) and the government forecasts for these are expected to be reasonably accurate.

By observing the economic indicators used in similar topics of other countries, together with those of the Hong Kong government, and considering the availability and consistence of measurement of data in Hong Kong, a list of candidate economic leading indicators as shown in Table 1 were selected for building the forecasting model. These economic indicators have been used in comparable studies such as Goh (1996, 1999) and Killingsworth (1990) and they should therefore be appropriate for model development. Time series data for the indicators are available from the *"Hong Kong Monthly Digest of Statistics"*, which is one of the general statistical digests compiled by the Census and Statistics Department in HK – with historical records dating from the early 1980s to the present time. To ensure a sufficient amount of data is available for model estimation, 20 years of quarterly records were used for all the time series data relating to construction output and other economic indicators.

< Table 1 >

Where the indicators recorded in the digest were not exactly in quarterly form (e.g. the Hang Seng and marriage indices, which are in monthly form, or housing stock, which are in yearly form), it was necessary to estimate the quarterly figures by either aggregation or interpolation of the figures involved.

MODELING TECHNIQUES

Linear Regressing Analysis

As suggested by Hanke (1989), the main statistical forecasting techniques available are Linear Regressing Analysis (LRA) and Autoregressive Integrated Moving-average (ARIMA) techniques. In the construction industry, LRA and ARIMA models have often been used often to model and forecast construction variables such as demand and price owing to their relative simplicity in both concept and application (Killingsworth, 1983; Thomas and Stekler, 1983; Akintoye and Skitmore, 1991a, 1991b, 1994; Goh, 1999; Kenny 1999; Macpherson and Sirmans, 1999; Tse *et al*, 1999; Mills *et al*, 2003; Ng *et al*, 2004), sometimes in conjunction with other techniques such as artificial neural networks (Majid and Yahya, 2002), decision support systems (Forgionne, 1996) and geographic information systems (Bell *et al*, 2000).

In using LRA for private sector housing forecasts, an obvious starting point is the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots + \beta_n x_n \qquad [1]$$

where the dependent variable y represents the value of private housing supply predicted; and x_1 x_2 ... x_n are economic indicators with the coefficients β_0 β_1 ... β_n to be estimated from the data.

Genetic Algorithms

Genetic Algorithms (GA) are stochastic search and optimization algorithms based on the principles of natural evolution (Holland, 1975) and their ease of use has enabled many applications to be identified in solving business, scientific and engineering optimization problems. Forecasting problems have come in for particular treatment, with GAs being used to: estimate forecasting model parameters (e.g. Chiraphadhanakul *et al*, 1997; Ju *et al*, 1997; Kim and Kim 1997; Jeong *et al*, 2002) and as a part of hybrid algorithms with other heuristics such as neural networks, simulated annealing, taboo search and application-specific heuristics (e.g. Kai and Wenhua, 1997; Yip *et al*, 1997; Kung *et al*, 1998). Readers are referred to Zheng *et al* (2004) for a more detailed description of the basic features of a GA model.

For the purposes of the research, the GA was applied to search the coefficients of the economic indicators in the model. A computer program written in Microsoft Visual Basic™ and Excel™ was developed to handle the calculations, and the main function of which is to generate models that can approximate the actual private housing supply. In the program, each string of chromosomes represents a list of coefficients. If there are ten economic indicators used, then there would be eleven genes in each chromosome – the additional one being for the constant term. The boundary of each gene is different and is determined in advance. When the GA program is running, the gene may randomly change its value when required by the program. For each generation of the program, the solution of the equation is therefore evaluated and larger fitness values assigned to those with a solution closer to the actual private housing supply for the corresponding quarter of year. After a few generations, the program builds a model approximating the actual supply trend. Predictions can then be generated over 3 or 4 years (i.e. 1996 to 1998 and 1999) and compared with the actual values.

Since this is the first time GA has been applied to the topic, the search spaces, or boundaries of search, of each coefficient are not known, so the technique of 'all possible regressions' can be applied to simulate the possible regions involved. This technique involves fitting all the regression equations with one indicator, two indicators, and so on. Therefore, if there are K candidate indicators, there are 2^K total equations to be estimated and examined. The value of the coefficients should then lie between the maximum and minimum value of the coefficients of the indicators in all equations. Since 10 indicators are used, a total of 2^{11} = 1024 equations are generated by the approach to obtain the range of values of the coefficients of all the indicators (Table 2).

< Table 2 >

Other than the search space, several GA parameters need to be assigned to the program. These include the population size (*pop_size*), convergent ratio (c), crossover ratio (P_c), mutation ratio (P_m) and the stopping criteria. To date, no general methodology is available to optimize the selection of these parameters. Therefore, they were selected by trial and error, the values ultimately adopted being:

pop_size	50
c	0.6
P_c	0.6
P_m	0.1

with the stopping criterion being when G, the number of generations, is equal to 300.

MODEL BUILDING PROCEDURE

Data Transformation

Two main transformations were carried out on the data after collection from the *Monthly Digest*:

Turning point analysis: Before running the raw data by GA, a turning point analysis was carried out, as suggest by Levenbach and Cleary (1984), to identify the leading characteristic of the indicators before their selection. Thus involved a four-step procedure:

Step 1: Plot the time series.

Step 2: Remove any trend/seasonality (seasonal adjustment, differencing).

Step 3: If necessary, remove irregularities by low-order moving averages.

Step 4: Fit a trend line to the series in Step 3 and plot the deviations from the trend (this is the cycle).

Correlation analysis: After the turning point analysis, a Pearson correlation analysis was carried out to obtain the correlation values between the time series data of the housing supply and candidate indicators (Table 3). The quarter of each candidate indicator with the largest correlation value was selected as the quarter lead for fitting into the model. For example, the public housing supply values (*PUBH*) has a 5-quarter lead, while the property price index (*PROIN*) has an 8-quarter lead. These are all summarized in Table 4. This produced the following equation:

$$PRI = f\,(PUBH_{t=5},\ PROIN_{t=4},\ HIS_{\,t=8},\ GDP_{\,t=2},\ GCE_{\,t=1},\ HSTOCK_{\,t=1},$$
$$LAND_{\,t=1},\ GOCN_{\,t=8},\ HCP_{\,t=6},\ UER_{\,t=8}) \quad [2]$$

< Table 3 >

< Table 4 >

Where *PRI* is the quarterly fluctuation in completed private new housing and *PUBH*, *PROIN* … *UER* are the leading indicators selected, with *t* being the quarter lead of each indicator.

Multi-Objective Problem

The genetic algorithms are normally used for solving single objective problems. To forecast housing supply, a total of 60 sets of *PRI* data, together with the leading indicators, were used to build the model – with each set of data being treated as one objective. This is therefore a multi-objective problem. As suggested by Coello (2000), for this type of problem, there is no clear definition of an ‘optimum’ as is the case with single-objective problems; neither does there necessarily have to be an absolutely superior solution corresponding to all the objectives due to incommensurability and conflict among the objectives. Since the solutions cannot be simply compared with each other, the ‘best’ solution generated from the optimization process corresponds to the human decision-makers' subjective selection from the potential solution pool.

However, this principle is slightly different here. In this case, although 60 objectives are considered, the aim is to build a model with the best forecasting power, i.e. the highest accuracy and longest prediction period. Thus, the degree of conflict is lower and the subjectivity of the decision-maker is less significant.

Zadeh (1963) has proposed the Adaptive Weight Approach (AWA) for multi-objective problems. This involves assigning weights to the objective functions and then combining them into a single-objective function. The weights assigned should satisfy the following conditions:

$$Z = \sum_{l=1}^{k} w_l f_l(x) \quad [3]$$

$$\sum_{l=1}^{k} w_l = 1 \quad [4]$$

where Z is the combined function; w_l is the weight assigned to the l^{th} function; and k is the number of objectives in the problem.

As no further information indicating how the objectives influence the forecasting power, three types of weight were tested:

Type 1 – equal weight: This assumes that all the objectives have a similar, if not equal, effect on the forecasting power. Therefore, the weight w is equal to $1/k$.

Type 2 – special weight: This assumes a 'recency effect', i.e. the effect of recent years on forecasting power is stronger than the effect of earlier years. Thus, if the model is to predict the *PRI* for year 1996, then the data for year 1995 would have a larger weight than for the year 1994, and that for year 1994 would be larger than for year 1993 and so on. The weight is formulated as:

$$\lambda_l = \frac{1}{(k-l)} \qquad [5]$$

$$w_l = \frac{\lambda_l}{\sum_l \lambda_l} \qquad [6]$$

Type 3 – calculated weight: This is calculated by a payoff matrix as suggested by Belenson and Kapur (1973). The *k* by *k* payoff matrix is shown in Figure 1. In the matrix, the *x** is the ideal solution of each objective. From this (with the values in the shaded boxes), a set of optimal weight can be obtained.

< *Figure 1* >

RESULTS OF INDIVDUAL MODEL

LRA models

Predictions of housing supply were made for the years 1996 to 1998. LRA model 1 was constructed by using the data from 1983 to 1995, and LRA model 2 is the updated version of Model 1 with year 1996 data added (Figure 2). This provides a visual indication of the performance of the models in terms of accuracy of prediction. LRA model 1 is clearly not very satisfactory, while LRA model 2 seems a little better. When attempting a longer forecasting time horizon, however, both models failed to make good predictions.

< Figure 2 >

GA models

With the data transformed and the sets of assigned weights assumed, the GA program was used to generate the model forecasts. As no previous study in this topic has determined the amount of information required to build a good model, 5-years of data and 10-years of data were used.

5-year model: The data for years 1991 to 1995, totaling 20 sets, were used. The results for the three weighting regimes are shown in Figures 3a–c. The thick lines are the actual *PRI* in years of 1991 to 1995, with the dotted lines being the predictions generated by the model. Clearly, the performance of the model is better if the dotted lines lie closer to, and have a similar trend with, the thick lines. As can be seen, however, no matter which type of weighting regimes are used, the generated models are not particularly good.

< Figure 3 >

10-year model: Data from 1986 to 1995, totaling 40 sets, were use to construct the 10-year model. The results after running the GA program are shown in Figure 4, divided into the 3 types of weighting regimes as before. This shows the performance of the model to be much improved over the 5-year model, with some of the dotted lines having a similar trend to the thick lines. But still the pattern of the dotted lines does not seem to be properly related to the assigned weight, and the fluctuations of even the best-dotted lines are still very large.

< Figure 4 >

GA-LRA MODELS

In both the 5-year and 10-year GA models, all the candidate indicators were used. It is possible that some of the indicators, although having leading characteristics with housing supply, are not significant enough to perform well in the model. In this case, LRA can help select the indicators for use by the GA. Methods of LRA model building have been developed for evaluating subsets of independent (indicator) variables wholly ('best subset' regression), by adding variables one at a time ('forward' selection), deleting one at a time ('backward' elimination) or a combination of these latter two ('stepwise' regression). The stepwise regression procedure was selected. After running the program with 10 years of data from 1986 to 1995, 6 out of the 10 significant indicators remained in the model – $PUBH_{(5)}$, $PROIN_{(4)}$, $HIS_{(8)}$, $GCON_{(8)}$, $HCPI_{(6)}$ and $UER_{(8)}$ (Table 5 summarizes the results).

< Table 5 >

10-year model: The GA program was run again with the 10-year data – this time with only the six indicators selected by the LRA method. As Figures 5a–c show, there is much improvement over the basic GA model and some good solutions are available. Comparing the patterns of the dotted lines in the three graphs, it is still difficult to draw any conclusions on the effect of the various weighting regimes – perhaps due to the large number of objectives involved, as these reduce the effects of the differences in weights between each objective.

< Figure 5 >

Regeneration of the 5-year model: Having shown that the 10-year GA-LRA model provides a significant improvement, the LRA stepwise model was built again for the 5-year model using stepwise regression. This produced the reduced variable subset of $PUBH_{(5)}$, $HIS_{(8)}$ and $UER_{(8)}$. The data from 1991 to 1995 for these three economic indicators were processed by the GA program. The performance of the GA-LRA model again improved, with some of the dotted lines being quite similar to the actual supply line.

EFFECTS OF AMR ON THE GA-LRA MODEL

One problem associated with GA models is the occurrence of local optima. This can be caused by the mutation ratio being too small. If the ratio is too large, however, the good chromosome may be damaged. The Adaptive Mutation Rate (AMR) approach was therefore used in an attempt to overcome this problem. As explained by Li *et al* (1999), since a GA is an intrinsically dynamic and adaptive tool, the use of a constant mutation rate is thus contrary to the general evolutionary spirit. At the initial stage, a relative high mutation probability helps prevent premature convergence, while the mutation rate needs to be reduced at a later stage so that good solutions will not be excessively disrupted. As a result of these considerations, a mutation probability formulation as shown in Eqn. 7 was applied:

$$P_m = P_{mi} - 0.3 \times t/G \quad [7]$$

where t is the current generation number and **G** is the maximum generation number; P_m is the mutation probability for current generation; and P_{mi} is the initial mutation probability set by user.

10-year model: Two sets of 10-year models were generated; one with a maximum generation number of 300 and the other with 1000 – the P_{mi} values of both programs being set at 0.6. Figure 6 gives the results (the results from the different weighting regimes being grouped together, since their effect is not significant). As Figures 6a–b show, most of the model predictions have a very similar trend to the actual supply. They are now distributed more evenly, which should reduce the chance of local optima. The solutions provided by the models, however, do not have any obvious improvement over GA-LRA models. This may be because these do not suffer from the presence of local optima, or the AMR is not performing as well as expected.

< Figure 6 >

Figure 7 shows some good models selected from those generated by the 10 year GA-LRA (with AMR) models. The dotted lines in the end section are the predictions for the out sample years 1996 to 1998.

< *Figure 7* >

DISCUSSION

Two of the models from the solution pool generated by the GA-LRA (with AMR) are:

GA-LRA(AMR) model 1:

$$PRI = 500 + 0.366386\ PUBH_{t=5} + 4.789877\ PROIN_{t=4} - 0.37443\ HIS_{t=8} - 4177.53\ GCE_{t=1} + 2204.436\ HCPI_{t=6} + 664.7909\ UER_{t=8} \quad [8]$$

GA-LRA(AMR) model 2:

$$PRI = 10{,}600 + 0.3197\ PUBH_{t=5} + 26.4616\ PROIN_{t=4} - 2.27509\ HIS_{t=8} + 1634.839\ GCE_{t=1} + 143078.7\ HCPI_{t=6} - 3450.52\ UER_{t=8} \quad [9]$$

The predictions of these two models over the year 1996 to 1998 are shown in Figure 8. Table 6 compares the results of these two models with the LRA models. The index shown in Table 6 is derived according to the following equation:

$$I_i^k = \frac{\left|D_i^k\right|}{\sum\left|E_i\right|} \times 100 \quad [10]$$

where E_i is the actual value at the i^{th} quarter; D_i^k is the difference between the predicted and actual values, i=1,2,3,…,12 denotes the twelve quarters being examined; and k=1,2,3,4 is the label of the model.

From the prediction index in Table 6, it is clear that the predictive accuracy of the GA-LRA(AMR) models has improved significantly over the LRA models.

< Figure 8 >

< Table 6 >

The application of the GA also helped to prolong the prediction period. As Figure 9 shows, all the models' prediction powers decline after mid-1998. From the solution pool generated by GA, however, another model can be found that predicts the trend after 1999, although the accuracy does decrease a little. This demonstrates another advantage of using the GA in that more than one solution, or model, can be generated – providing the decision maker with more options from which to choose.

< Figure 9 >

CONCLUSIONS

In comparing four models for forecasting private sector housing demand, it was found that:

i) the LRA method is easier to operate and the amount of time required to build a model is shorter;

ii) the GA method involves some tedious calculations and the aid of a computer is necessary when the problem is complex or the search space is large;

iii) the GA method allows the decision-maker a larger involvement, such as in assigning the GA parameters and choosing the appropriate results from the pool of solutions;

iv) the GA method can generate more than one solution each time, e.g. in this case 50 solutions were generated each time since the population size is 50;

v) the GA method has a better accuracy; and

vi) the GA method has a longer prediction period.

An important point is that the GA program needs to be properly parameterized to avoid reaching local optima and converge to a global optimum with a high degree of consistency, regardless of the specification of the initial population. On the other hand, the program can spend a considerable amount of time without showing improvement, and then suddenly produce a jump. It still not yet clear, however, quite how to do this parameterization, what kind of problems the GA is most suited for, what controls its convergence rate, and what precisely are the roles of crossover, mutation, etc., in the overall search in progress. There is growing evidence that the "optimum" parameters values may be problem-specific – no general methodology being presently available to optimize the selection of these parameters. Only general experience shows that the value of the crossover ratio (P_c) is usually 0.6-0.8; while for the mutation rate (P_m) the expected number of bits mutated per chromosome should be kept less than one. Similarly, setting the convergent ratio c at 0.6 has been found to avoid either reaching a local optima or taking too long to converge.

In addition, as with any form of prediction or forecast, many uncertainties and errors exist. In this case, they may be due to:

i) *Lack of significance of economic indicators:* For this study, a total of 10 economic indicators were available for constructing the model but the GA-LRA model benefited from the selection of only six of these. This is presumably because the omitted indicators fail to, or spuriously, represent the economic factors involved. For example, the reason the real wage index was not selected by the LRA for use in the GA-LRA model may be that this index does not really reflect the income and purchasing power of Hong Kong people, as the actual income for many high-income people is mainly from returns on their investments rather than their salary.

ii) *Interdependence of variables:* In constructing the forecasting model, it is assumed that the economic indicators, which serve as the variables, are independent. In fact, all these indicators are related to different aspect of the economic conditions of Hong Kong and are therefore, by their very nature, likely to be highly interdependent.

iii) *Change in economic indicators:* In making predictions using the leading characteristics of the chosen economic indicators, it is assumed that these indicators will follow a similar pattern or trend in the whole period under consideration. If there is an abrupt change in the indicators, the prediction may fail. Abrupt changes to the indicators can easily happen due to:

iv) *Policy:* Changes in economic policy often have a significance effect on economic conditions, even in the construction industry.

v) *Housing habits:* An increasing number of HK people are now buying houses in People Republic of China – a trend that is very difficult to be shown by an economic indicator and therefore reflected in the models constructed.

vi) *Economic structure:* The HK economic structure has changed very rapidly in recent years. This has resulted in significant changes in land use and redistribution of property and therefore the general economic cycle. There is no guarantee, therefore, that the economic indicators will follow the same cycle.

Finally, it should be noted that only some basic aspects have been explored here, however, and there is considerable potential for future study by:

- carrying out more systematic tests to optimize the parameters – particularly those of population size, the fitness evaluation function, the crossover and mutation rate;
- using 'best subset' instead of step-wise regression to identify suitable indicator variables;
- further tests on the AMR;
- using the niche formation and modified adaptive weight approaches instead of Roulette wheel selection;
- further investigation on the search space beyond that of stimulation from the 'all possible repressor' method; and
- using further stopping criteria, such as when the program does not show significant improvement for certain number of generations, or when the results generated have achieved a certain satisfactory level, instead of just 300 iterations or 1000-AMR iterations.

Developing the GA method beyond that of merely replacing the *R*-square method in constructing the mathematical model – it could be used in other aspects of the problem as well, such as to determine which set of economic indicators would produce the best result or to study the leading characteristics of the candidate economic indicators. Although there is no theory yet to support such a replacement, some empirical tests could be made to gauge the usefulness of this approach.

REFERENCES

Akintoye, A.S. and Skitmore, R.M. (1991a) The supply of construction services: an econometric analysis of UK data, *Proceedings: 4th Yugoslav Symposium on Construction Project Modelling and Productivity*, April, Zagreb, Yugoslavia, M. Horner and M. Zaja (eds), Faculty of Civil Engineering, University of Zagreb, Yugoslavia, 109-115.

Akintoye, A.S. and Skitmore, R.M. (1991b) Dynamics of investment in new housing and other new construction work, *Transactions: the European Symposium on Management, Quality and Economics in Housing and other Building Sectors*, A. Bezelga & P. Brandon (eds), E & FN Spon, 1623-1635.

Akintoye, A.S. and Skitmore, R.M. (1994) Models of UK private sector quarterly construction demand, *Construction Management and Economics*, **12**(1), 3-13.

Belenson, S.M. and Kapur, K.C. (1973) An algorithm for solving multicriterion linear programming problems with examples, *Operations Research Quarterly*, **24**(1), 65-77.

Bell, M., Dean, C. and Blake, M. (2000) Forecasting the pattern of urban growth with PUP: a web-based model interfaced with GIS and 3D animation, *Computers, Environment and Urban Systems*, **24**(6), 559-581.

Chiraphadhanakul, S., Dangprasert, P. and Avatchanakorn, V. (1997) Genetic algorithms in forecasting commercial banks deposits, *Proceedings: IEEE International Conference on Intelligent Processing Systems*, October 28-31, Central Garden Hotel, Beijing, IEEE Press, 116-121.

Coello, A. (2000) Handling preferences in evolutionary multiobjective optimization: survey, *Proceedings: 2000 Congress on Evolutionary Computation*, July 16-19, La Jolla Marriott, San Diego, IEEE Press, 30-37.

Forgionne, G.A. (1996) Forecasting army housing supply with a DSS-delivered econometric model, *Omega*, **24**(5), 561-576.

Goh, B.H. (1996) Residential construction demand forecasting using economic indicators: a comparative study of artificial neural networks and multiple regression, *Construction Management and Economics*, **14**, 25-34.

Goh, B.H. (1999) An evaluation of the accuracy of the multiple regression approach in forecasting sectoral construction demand in Singapore, *Construction Management and Economics*, **17**(2), 231-241.

Hanke, J. (1989) Forecasting in business schools: a follow-up survey, *International Journal of Forecasting*, **5**(2), 259-262.

Hillebrandt, P.M. (1985) *Economic Theory and the Construction Industry*, Macmillan.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

Jeong, B., Jung, H.S. and Park, N.K. (2002) A computerized causal forecasting system using genetic algorithms in supply chain management, *Journal of Systems and Software*, **60**(3), 223-237

Ju, Y.K., Kim, C. and Shim, J.C. (1997) Genetic based fuzzy models: interest rate forecasting problems, *Computers and Industrial Engineering*, **33**, 561–564.

Kai, F. and Wenhua, X. (1997) Training neural networks with genetic algorithms for forecasting the stock price index, *Proceedings: IEEE International Conference on Intelligent Processing Systems*, October 28-31, Central Garden Hotel, Beijing, IEEE Press, 401-403.

Kanaengnid, T.Q. (1992) *Leading indicators of economic growth in Thailand: 1970-1991*, The South East Asian Central Banks (SEACEN) Research and Training Centre, Lorong Universiti A, Kuala Lumpur.

Kenny, G. (1999) Modelling the demand and supply sides of the housing market: evidence from Ireland, *Economic Modelling*, **16**(3), 389-409.

Killingsworth, M.R. (1983) *Labor Supply*, New York: Cambridge University Press.

Kim, D. and Kim, C. (1997) Forecasting time series with genetic fuzzy predictor ensemble, *IEEE Transactions on Fuzzy Systems*, **5**, 523–535.

Kung, C.H., Devaney, M.J., Huang, C.M. and Kung, C.H. (1998) An adaptive power system load forecasting scheme using a genetic algorithms embedded neural network, *Proceedings:*

IEEE Instrumentation and Measurement Technology Conference, May 18-20, Venice, Italy, IEEE Press, 308–311.

Lansley, P., Quince, T. and Lea, E. (1980) Flexibility and efficiency in construction management, *Building Technology and Management*, December, 42-43.

Lavender, S.D. (1990) *Economics for Builders and Surveyors: Principles and Applications of Economic Decision Making in Developing the Built Environment*, Longman.

Levenbach, H. and Cleary, J.P. (1984) *The Modern forecaster,* Van Nostrand Reinhold Company Inc., New York.

Li, M., Qiu, L. and Tong, T. (1999**)** Analysis on searching efficiency of partheno – genetic algorithm, *Journal – Changsha University of Electric Power Natural Science*, **14**(1), 48-50.

Macpherson, D.A. and Sirmans, G.S. (1999) Forecasting seniors housing demand in Florida *Journal of Real Estate Portfolio Management*, **5**(3), 259-274.

Majid, M.Z.A. and Yahya, K. (2002) Forecasting of low cost housing demand in urban areas: artificial neural network and ARIMA model approach, *Proceedings: 2nd International Conference on Systems Thinking in Management,* April 3-5,University of Salford, Manchester, University of Salford Press, A5:14-19.

Mills, A., Harris, D. and Skitmore, R.M. (2003) The accuracy of housing forecasting in Australia, *Engineering Construction and Architectural Management*, **10**(4), 245-253.

Ng, S.T., Cheung, S.O., Skitmore, R.M., Lam, K.C. and Wong, L.Y. (2000) Prediction of tender price index directional changes, *Construction Management and Economics*, **18**(7), 843-852.

Ng, S.T., Cheung, S.O., Skitmore, R.M. and Wong, T.C.Y. (2004) An integrated model for construction tender price index forecasting, *Construction Management and Economics*, (in press).

Raftery, J.J. (1998) *Principles of Building Economics: An Introduction*, Blackwell Science.

Thomas, R.W. and Stekler, H.O. (1983) A regional forecasting model for construction activity, *Regional Science and Urban Economics*, **13**(4), 557.

Tse, R.Y.C., Ho, C.W. and Ganesan, S. (1999) Matching housing supply and demand: an empirical study of Hong Kong's market, *Construction Management and Economics*, **17**(5), 625-634.

Yip, D.H.F., Hines, E.L. and Yu, W.W.H. (1997) Application of artificial neural networks in sales forecasting, *Proceedings: IEEE Instrumentation and Measurement Technology Conference*, May 19-21, Hotel Laurier, Ottawa, R.S. Turgel (ed), IEEE Press, 308–311.

Zadeh, L.A. (1963) *Linear System Theory: The State Space Approach,* McGraw-Hill, New York.

Zheng, D.X.M., Ng, S.T. and Kumaraswamy, M.M. (2004) Applying a GA-based multiobjective approach for time-cost optimization, *Journal of Construction Engineering and Management*, ASCE, **130**(2), 168-176.

CAPTIONS TO FIGURES AND TABLES

Figure 1: *Payoff matrix for the Belenson and Kapur method*

Figure 2: *Results of LRA models*

Figure 3: *Result of 5-year GA model*

Figure 4: *Result of 10-year GA model*

Figure 5: *Results of 10-year GA-LRA model*

Figure 6: *Results of 10-year GA-LRA(AMR) model*

Figure 7: *Forecast made by the 10-year GA-LRA(AMR) model*

Figure 8: *Forecast made by the 2 selected GA-LRA(AMR) models*

Figure 9: *Prediction between 1996 and 1999*

Table 1: *List of leading indicator variables*

Table 2: *Univariate statistics*

Table 3: *Pearson Correlation value with different quarter lead (significant value with P value < 0.05)*

Table 4: *Candidate Indicators with their corresponding quarter lead over housing supply*

Table 5: *Step-wise selection*

Table 6: *The actual and predicted values and prediction index*

	x^{1*}	x^{2*}	$\cdots$	x^{h*}	$\cdots$	x^{k*}
Z_1	$f_1(x^{1*})$	$f_1(x^{2*})$	$\cdots$	$f_1(x^{h*})$	$\cdots$	$f_1(x^{k*})$
Z_2	$f_2(x^{1*})$	$f_2(x^{2*})$	$\cdots$	$f_2(x^{h*})$	$\cdots$	$f_2(x^{k*})$
$\vdots$	$\vdots$	$\vdots$	$\cdots$ $\vdots$	$\vdots$	$\cdots$ $\vdots$	$\vdots$
Z_f	$f_f(x^{1*})$	$f_f(x^{2*})$	$\cdots$	$f_f(x^{h*})$	$\cdots$	$f_f(x^{k*})$
$\vdots$	$\vdots$	$\vdots$	$\cdots$ $\vdots$	$\vdots$	$\cdots$ $\vdots$	$\vdots$
Z_k	$f_k(x^{1*})$	$f_k(x^{2*})$	$\cdots$	$f_k(x^{h*})$	$\cdots$	$f_k(x^{k*})$

Figure 1: Payoff matrix for the Beleson and Kapur method

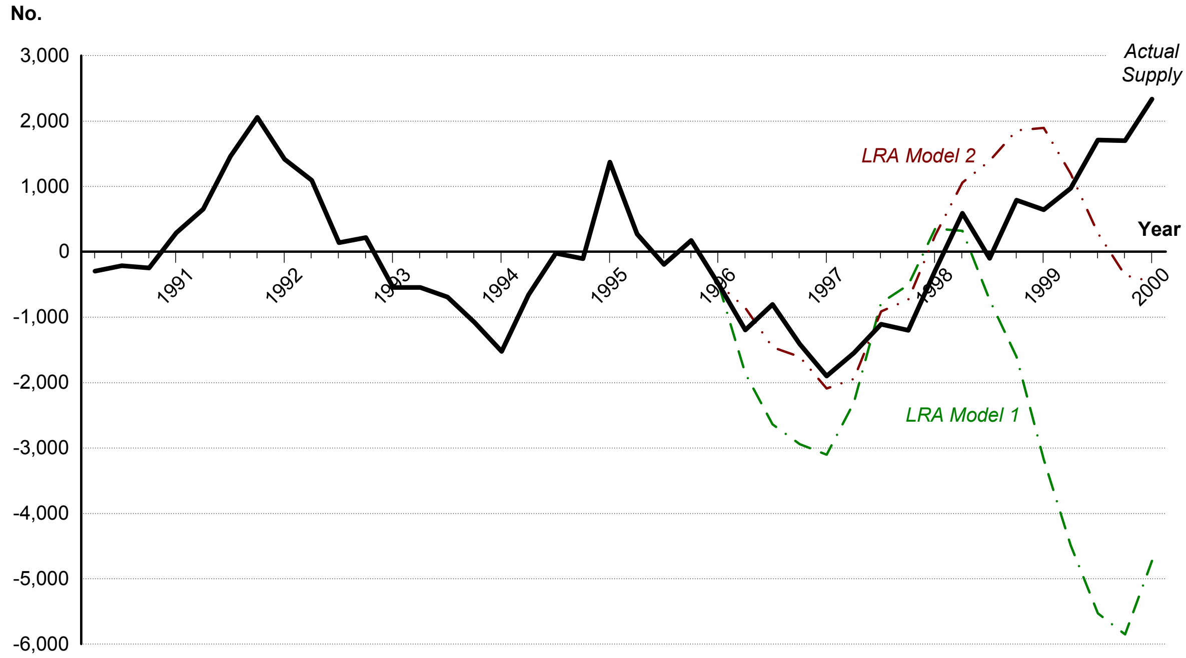
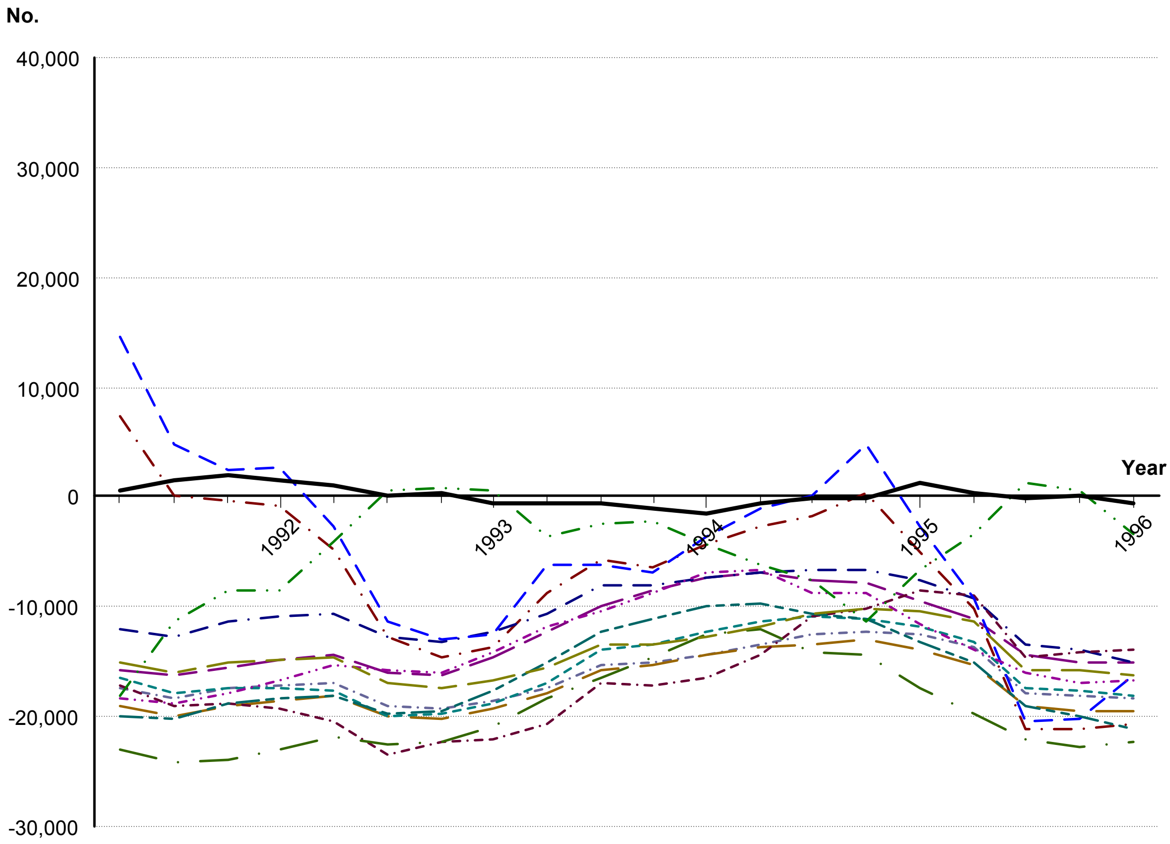
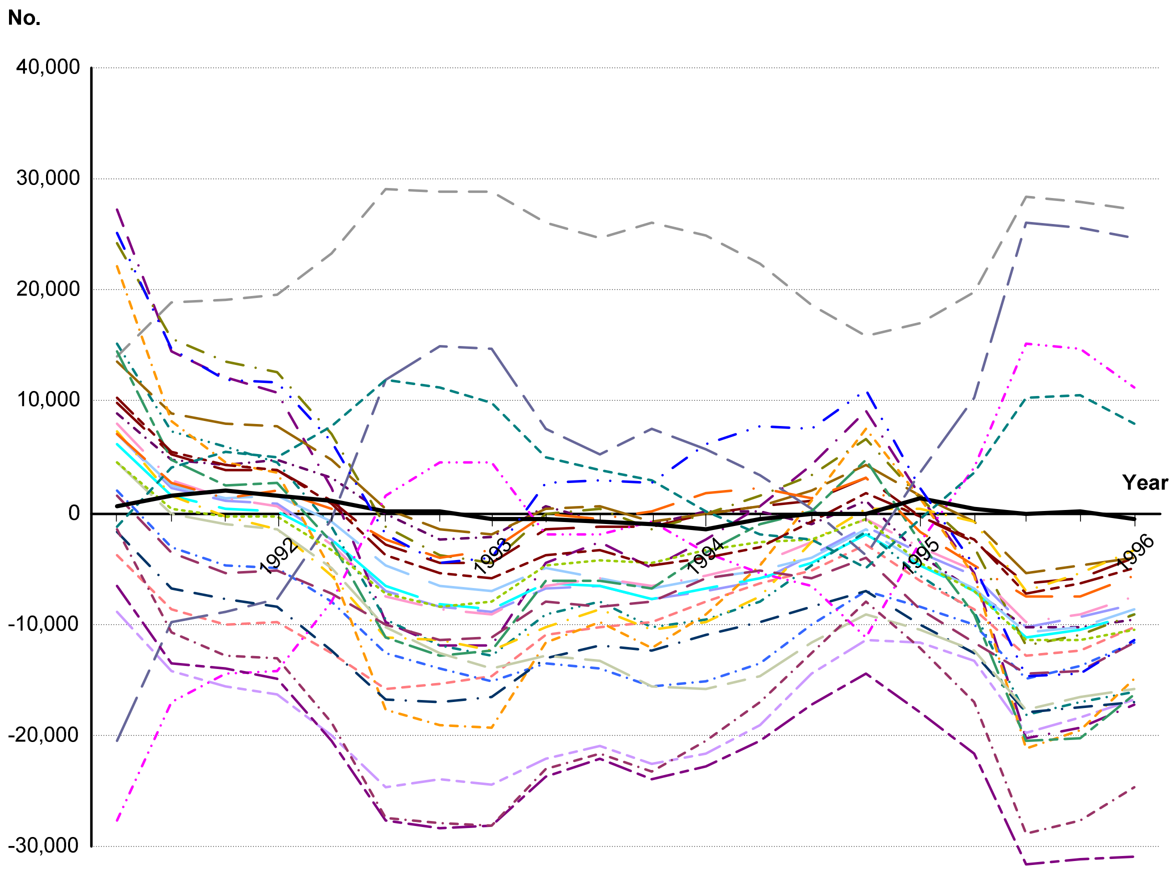


Figure 2: Results of LRA models



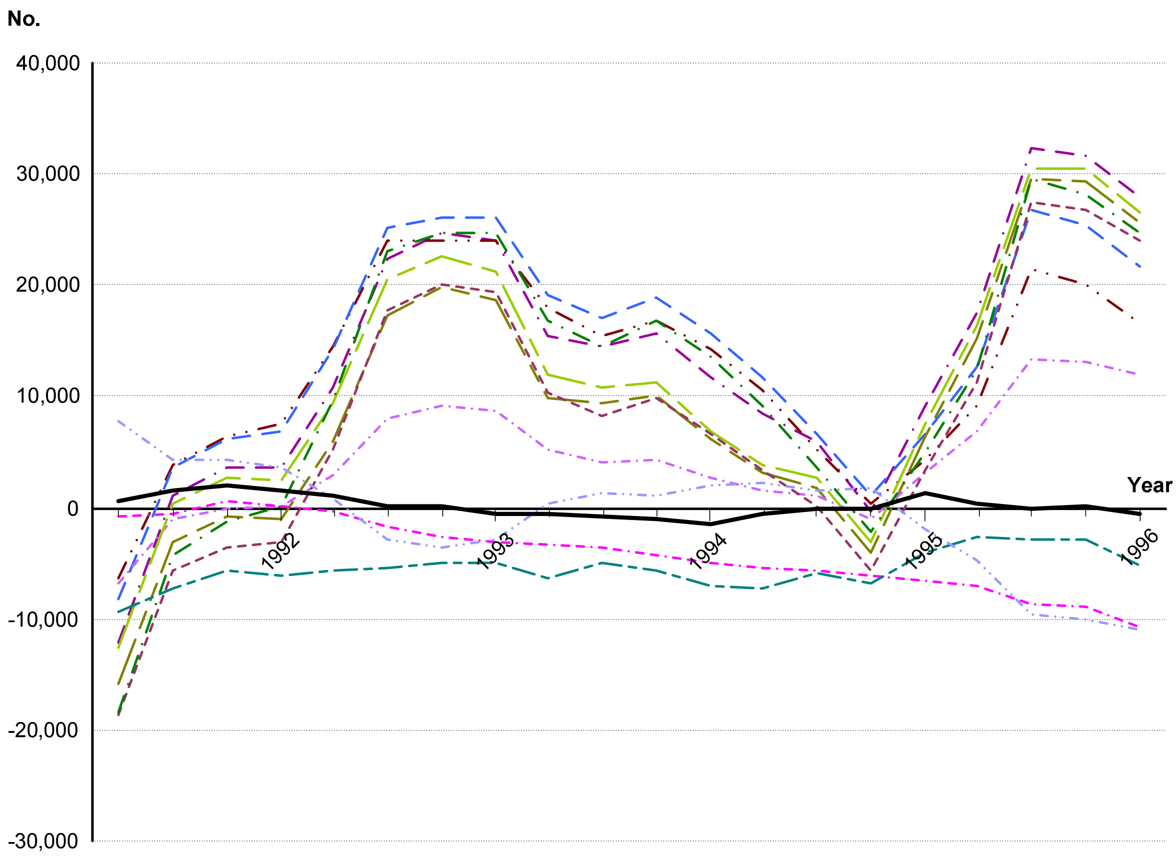
(a) equal weights

Figure 3: Results of 5-year GA model



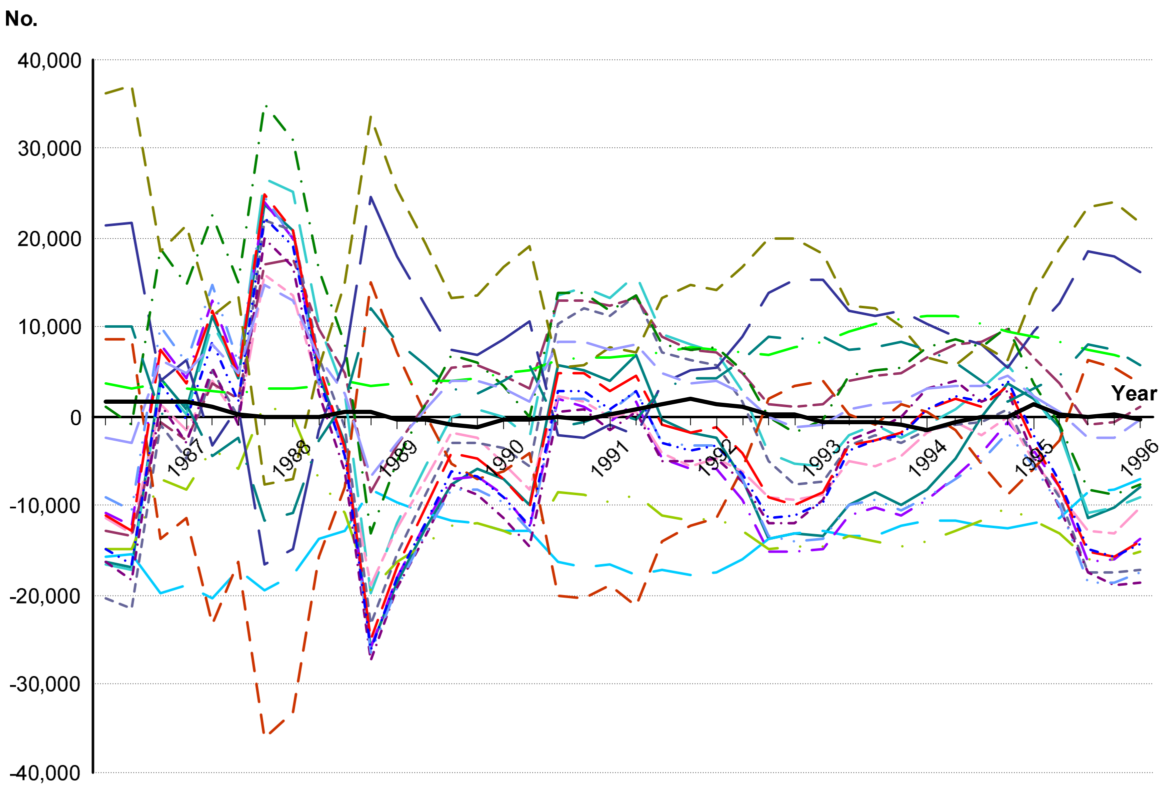
(b) special weights

Figure 3 (cont'd): Results of 5-year GA model



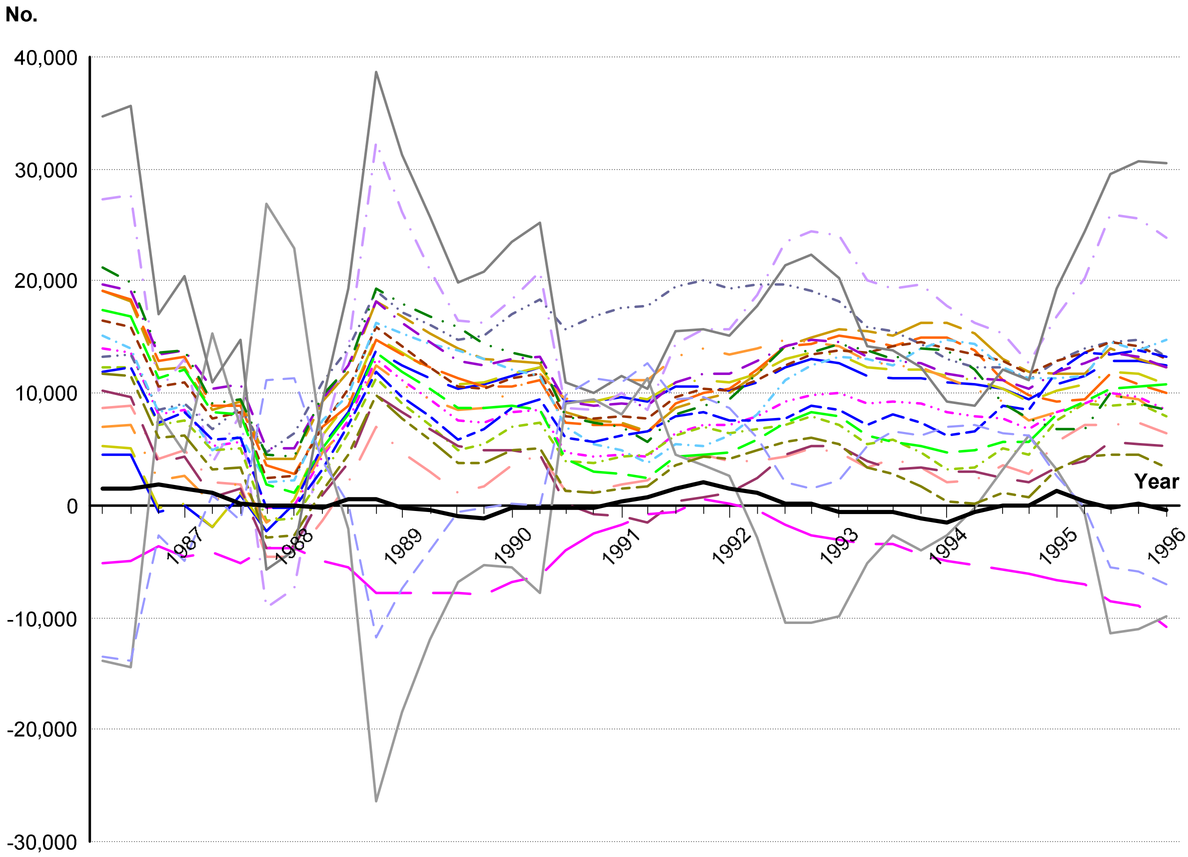
(c) calculated weights

Figure 3 (cont'd): Results of 5-year GA model



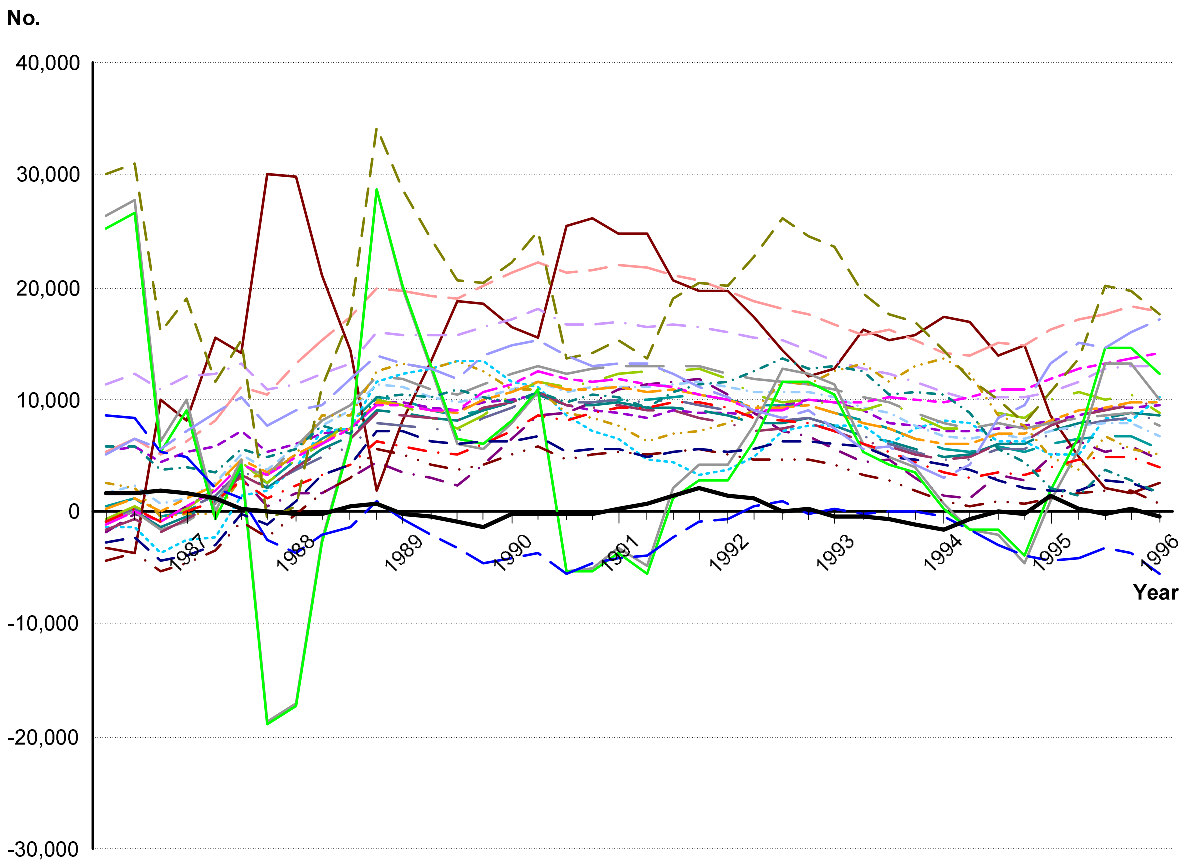
(a) equal weights

Figure 4: Results of 10-year GA model



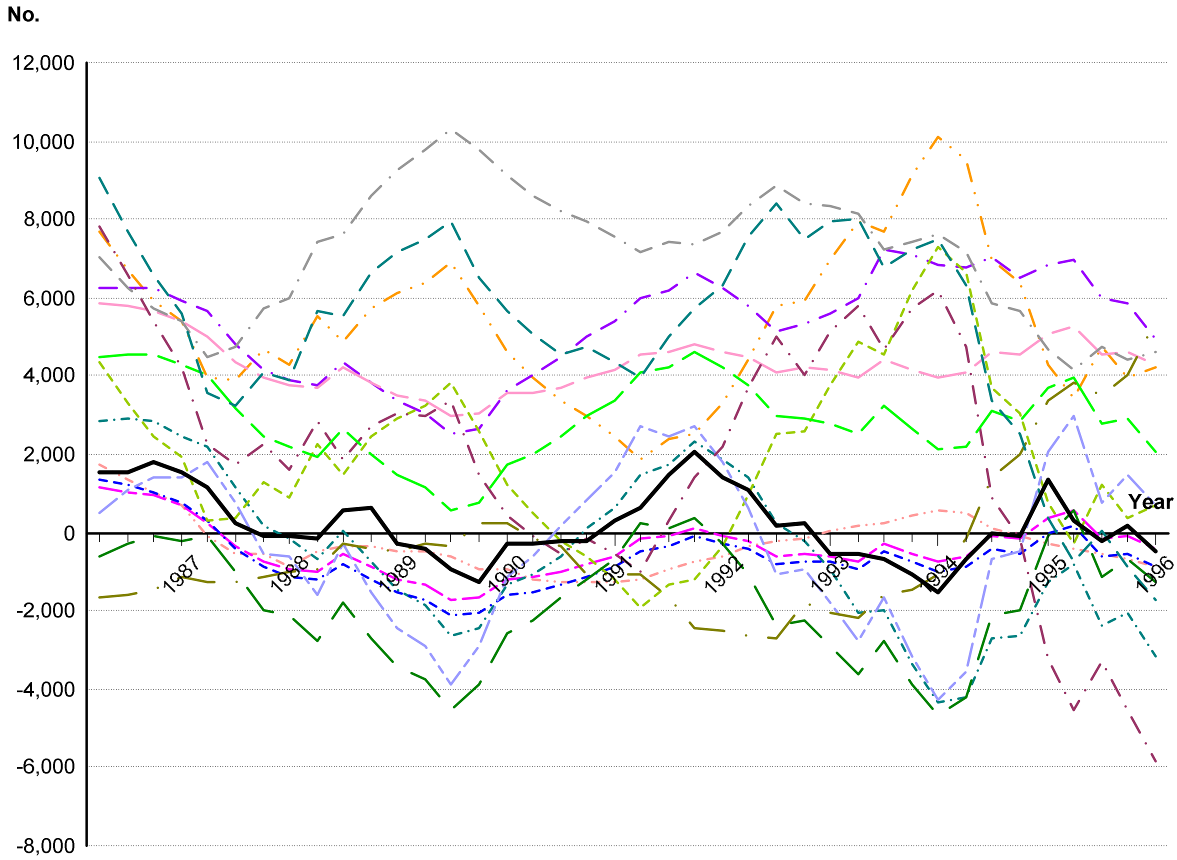
(b) special weights

Figure 4 (cont'd): Results of 10-year GA model



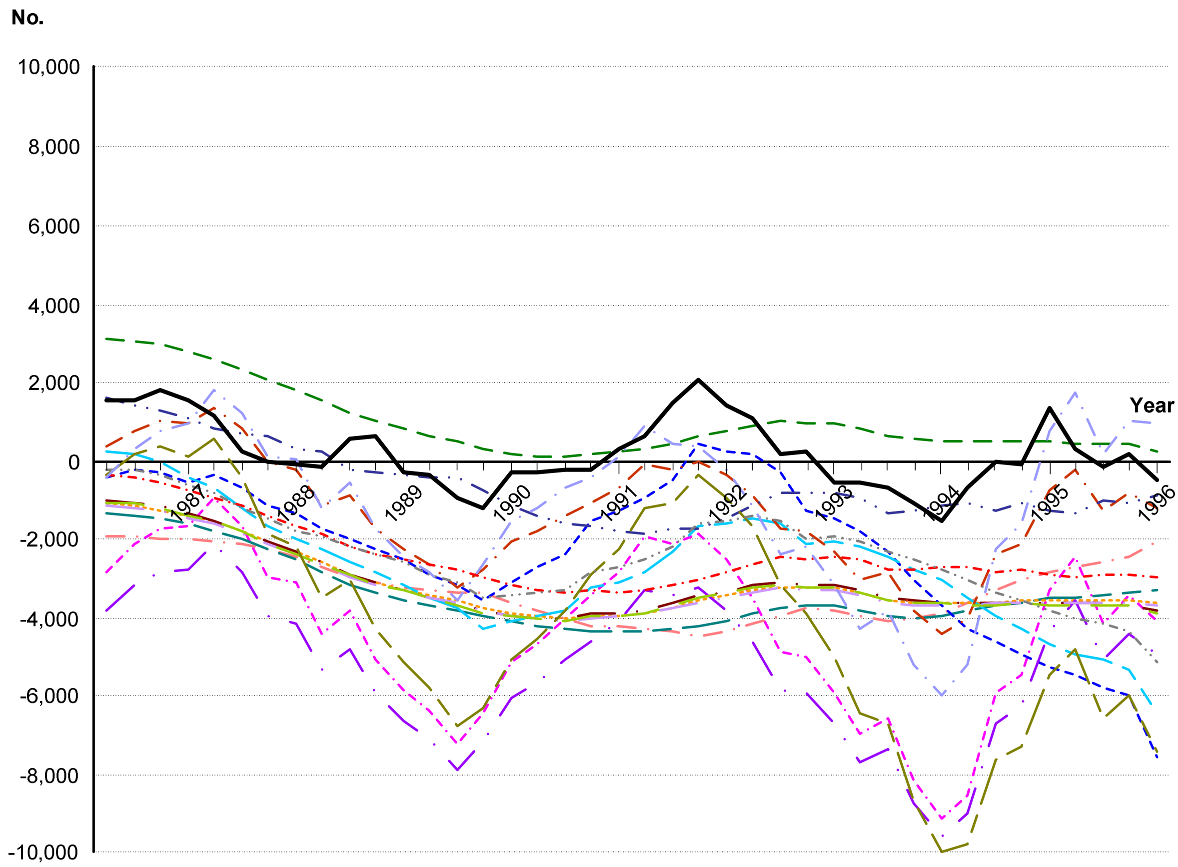
(c) calculated weights

Figure 4 (cont'd): Results of 10-year GA model



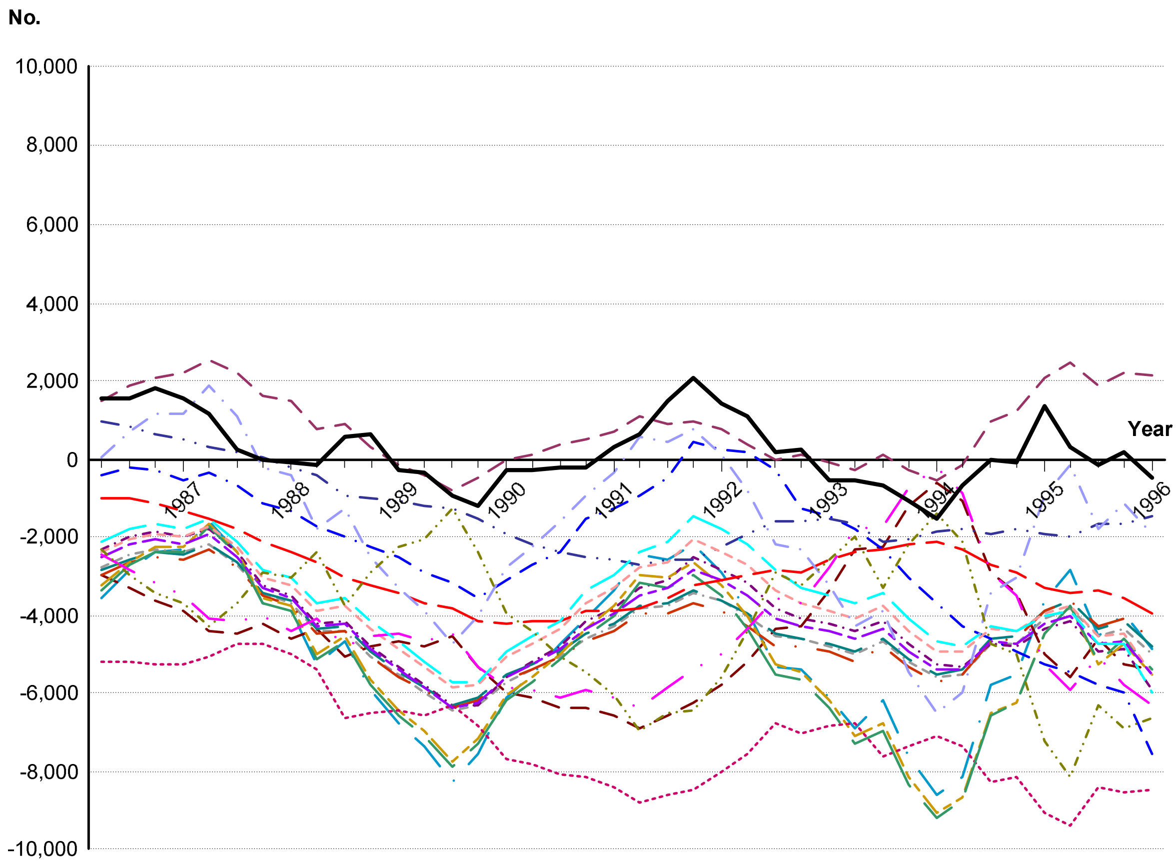
(a) equal weights

Figure 5: Results of 10-year GA-LRA model



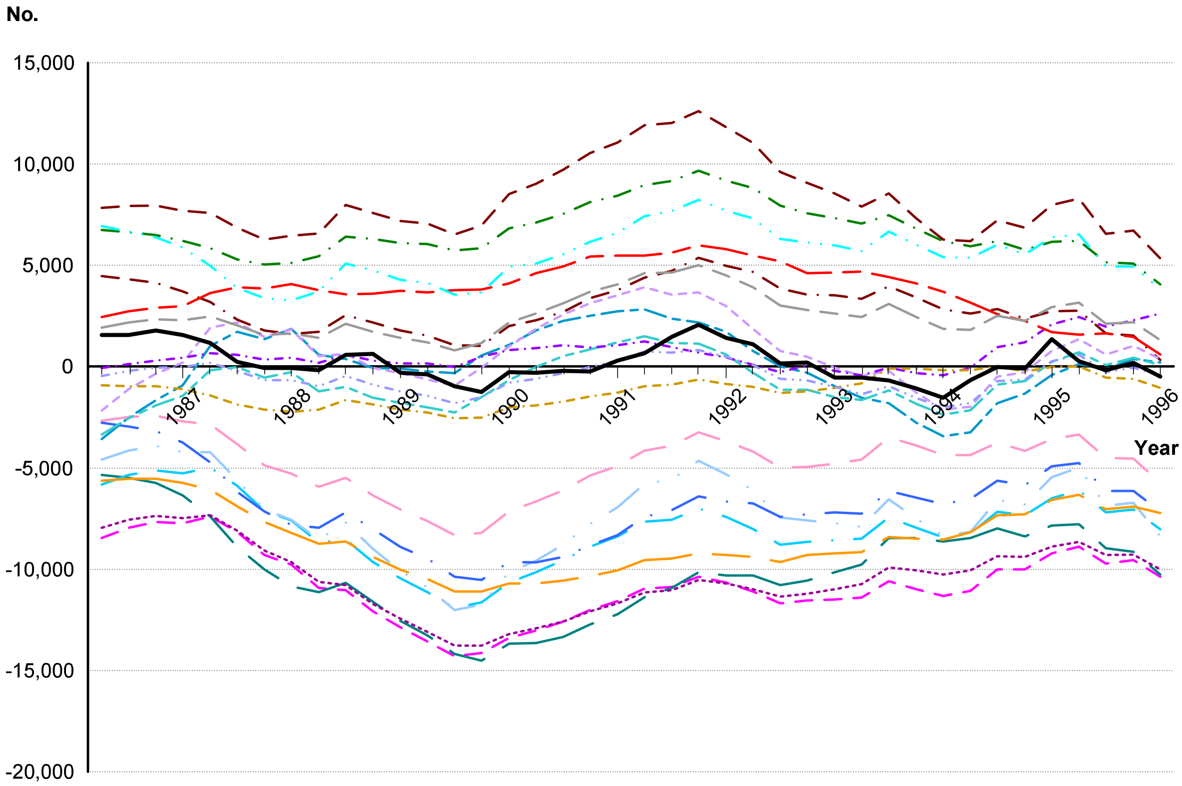
(b) special weights

Figure 5 (cont'd): Results of 10-year GA-LRA model



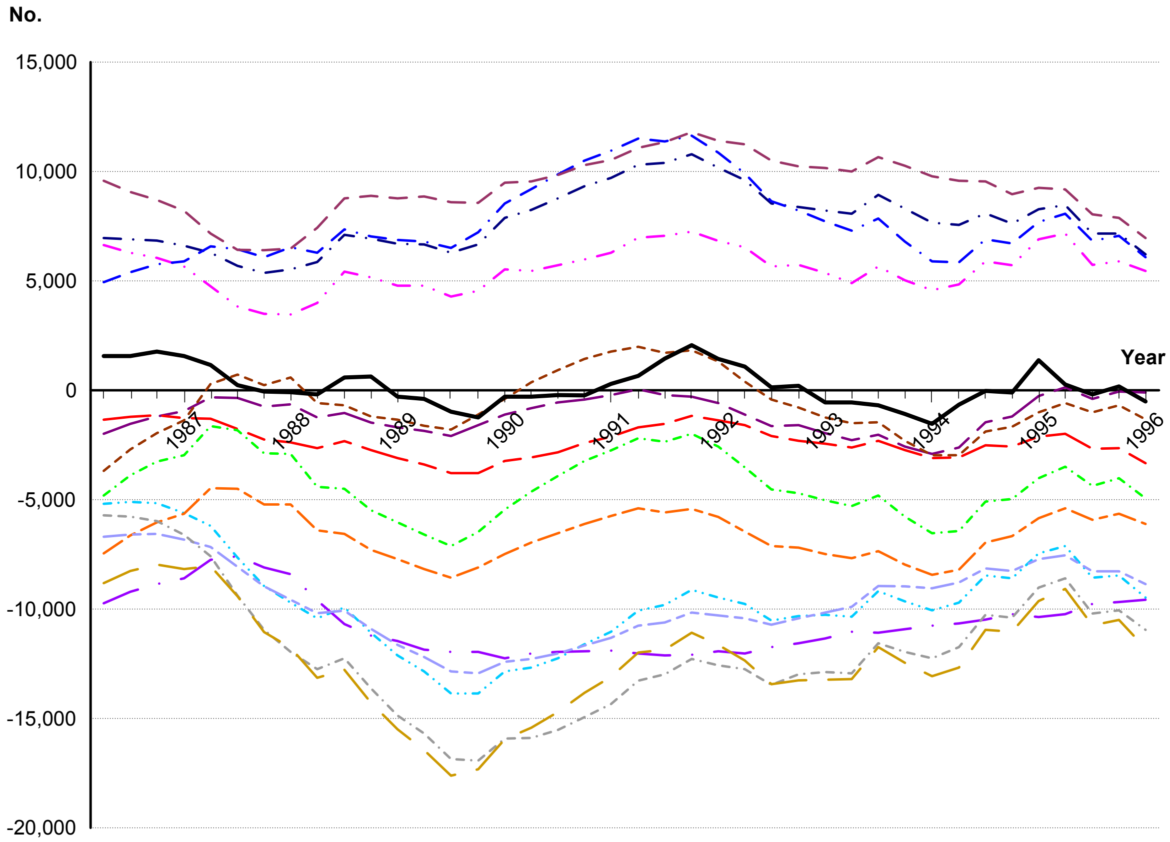
(c) calculated weights

Figure 5 (cont'd): Results of 10-year GA-LRA model



(a) 300 generations

Figure 6: Results of 10-year GA-LRA(AMR) model



(b) 1,000 generations

Figure 6 (cont'd): Results of 10-year GA-LRA(AMR) model

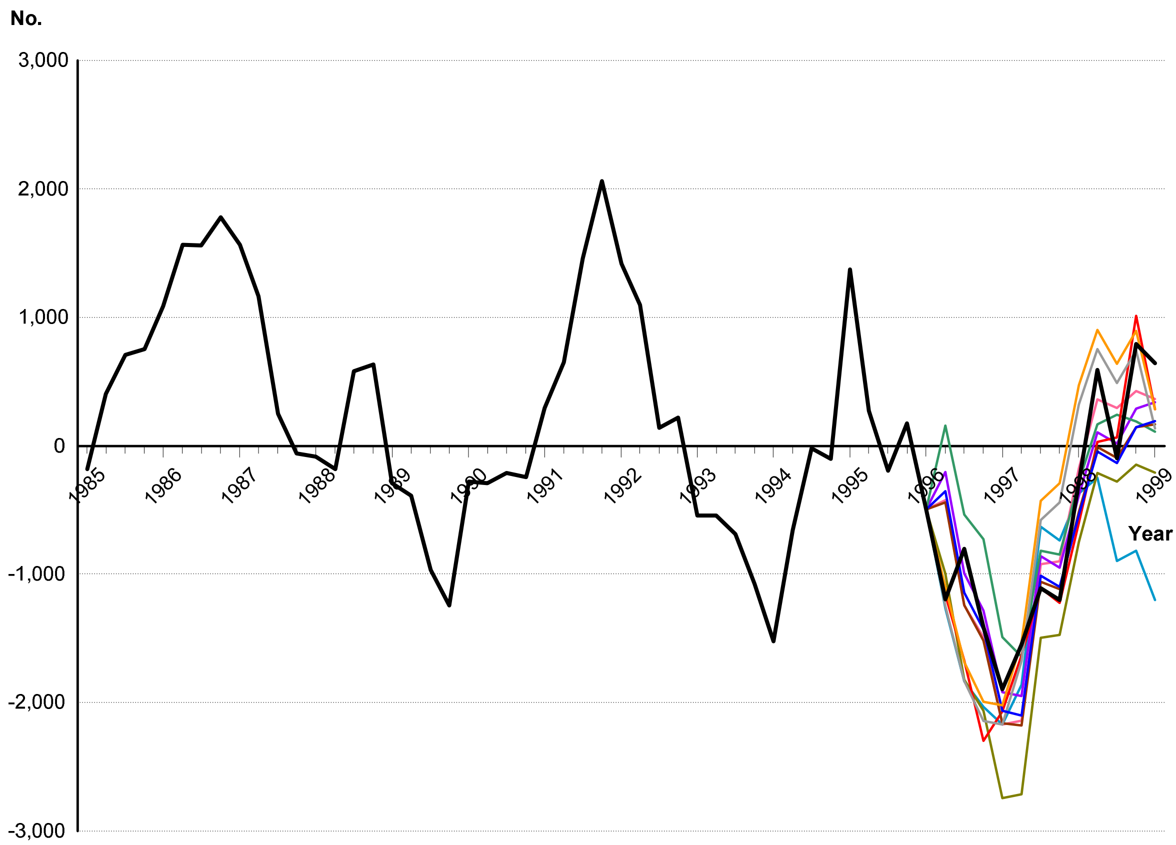


Figure 7: Forecast made by the 10-year GA-LRA(AMR) model

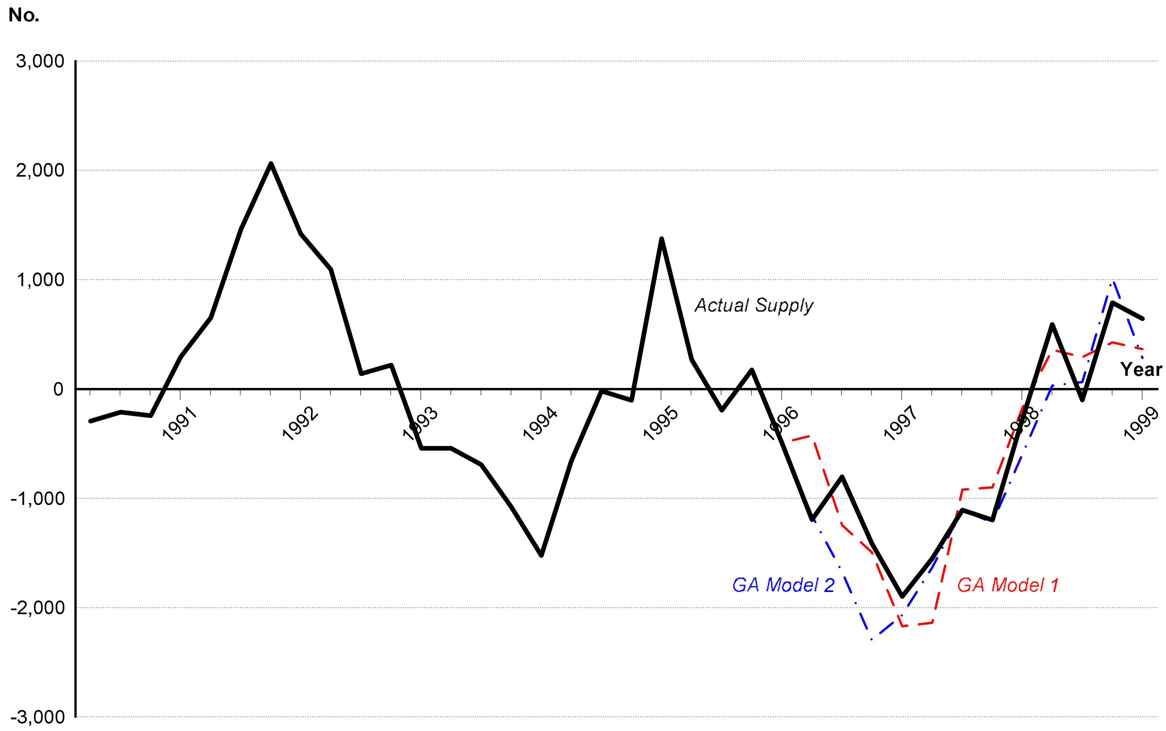


Figure 8: Forecast made by the 2 selected GA-LRA(AMR) models

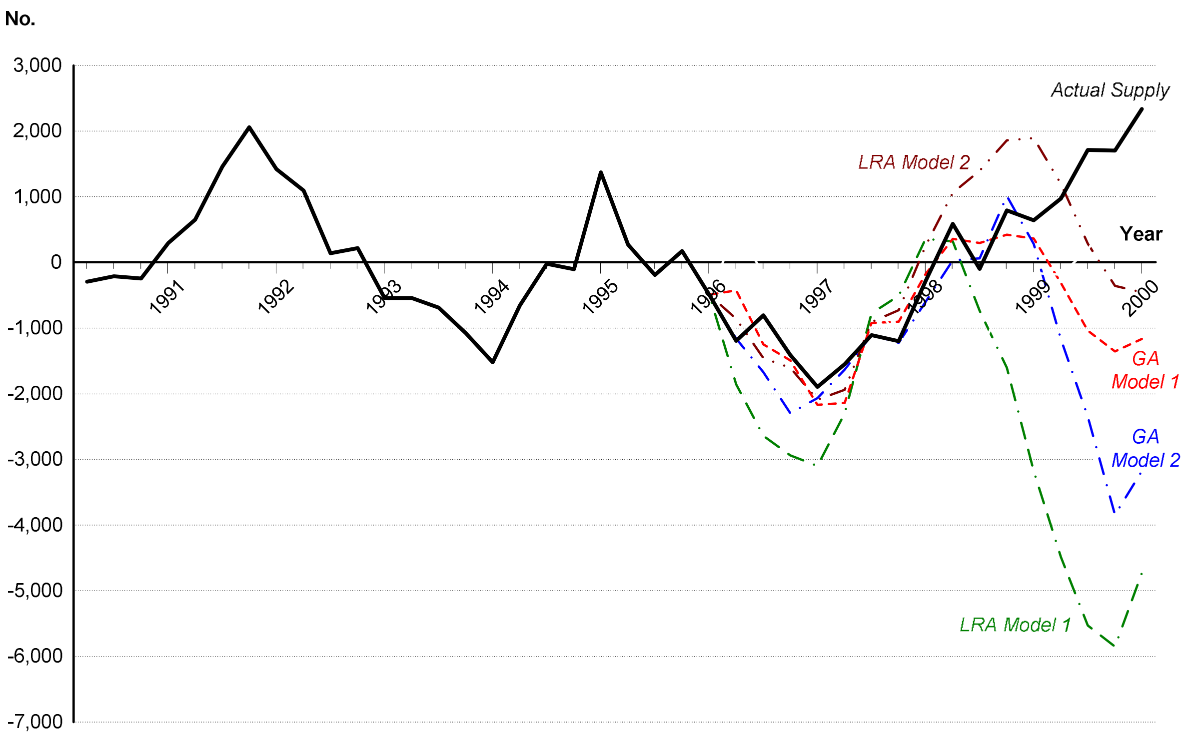


Figure 9: Forecast between 1996 and1999

Table 1: List of leading indicator variables

Economic Indicators	***Abbreviation***
Newly completed public housing	PUBH
Disposal of government land	LAND
Unemployment rate	UER
Property index	PROIN
Heng Sang index	HIS
Gross domestic product	GDP
Gross domestic product – construction	GCON
Composite consumer price index on housing item	HCPI
Total housing stock	HSTOCK
Government consumption expenditure	GCE

Table 2: Univariate statistics

	Candidate indicators	***Max. value***	***Min. value***	***Adopted value (+/-)***
0	Intercept term	5,332.012	-3,477.671	6000
1	PUBH	0.436	0.098	1
2	PROIN	52.541	-85.288	100
3	HIS	0.568	-1.594	5
4	GDP	89.316	-188.444	200
5	GCE	87,805.205	-130,529.099	150000
6	HSTOCK	8.070	-17.479	50
7	LAND	0.013	-0.017	1
8	GCON	11,639.434	-57,425.481	70000
9	HCPI	122,666.405	-72,976.654	150000
10	UER	3,636.312	-770.603	5000

Table 3: Pearson correlation value with different quarter lead (significant value with P value < 0.05)

	0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
PUBH	-0.1798	-0.0557	0.1291	0.3036	0.5048	**0.6161**	0.5428	0.4756	0.3247
PROIN	-0.1680	-0.2632	-0.3237	-0.3690	**-0.3920**	-0.3801	-0.3535	-0.2923	-0.2311
HIS	-0.0420	-0.0436	-0.0800	-0.1177	-0.1929	-0.2968	-0.3726	-0.4272	**-0.4423**
GDP	-0.3085	-0.3697	**-0.3902**	-0.3892	-0.3853	-0.3629	-0.3079	-0.2142	-0.0938
GCE	-0.3330	**-0.3345**	-0.3015	-0.2632	-0.2061	-0.1375	-0.0815	-0.0240	0.0377
HSTOCK	-0.5165	**-0.5076**	-0.4482	-0.3669	-0.2845	-0.2040	-0.1284	-0.0524	0.0260
LAND	0.2684	**0.2714**	0.2151	0.1119	0.0161	-0.0547	-0.0829	-0.1390	-0.1632
GCON	-0.2070	-0.1198	-0.0348	0.0738	0.1857	0.2945	0.3999	0.4550	**0.4694**
HCPI	-0.0127	0.0642	0.1353	0.2022	0.2592	0.2965	**0.3163**	0.3048	0.2693
UER	-0.2292	-0.2233	-0.2113	-0.1685	-0.0795	0.0467	0.1775	0.2801	**0.3421**

Note: Bold figure in the table are those with the largest magnitude after analysis

Table 4: Candidate indicators with their corresponding quarter lead over housing supply

Candidate indicator	***Quarter lead***
PUBH	5
PROIN	4
HIS	8
GDP	2
GCE	1
HSTOCK	1
LAND	1
GCON	8
HCPI	6
UER	8

Table 5: Step-wise selection

The REG Procedure
Model: MODEL1
Dependent Variable: PRI
Stepwise Selection: Step 1
Variable PUBH Entered: R-Square = 0.3993 and C(p) = 164.8468
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	13133325	13133325	28.58	<.0001
Error	43	19758480	459500		
Corrected Total	44	32891805			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	381.26263	104.35716	6133227	13.35	0.0007
PUBH	0.26543	0.04965	13133325	28.58	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2
Variable HSTOCK Entered: R-Square = 0.6470 and C(p) = 81.9669

Stepwise Selection: Step 3
Variable HIS Entered: R-Square = 0.7310 and C(p) = 55.1808

Stepwise Selection: Step 4
Variable LAND Entered: R-Square = 0.7719 and C(p) = 43.1633

Stepwise Selection: Step 5
Variable PROIN Entered: R-Square = 0.7924 and C(p) = 38.1312

Stepwise Selection: Step 6
Variable GDP Entered: R-Square = 0.8116 and C(p) = 33.5435

Stepwise Selection: Step 7
Variable GCON Entered: R-Square = 0.8225 and C(p) = 31.8319

Stepwise Selection: Step 8
Variable UER Entered: R-Square = 0.8942 and C(p) = 9.2684

Stepwise Selection: Step 9
Variable GDP Removed: R-Square = 0.8939 and C(p) = 7.3626

Stepwise Selection: Step 10
Variable HSTOCK Removed: R-Square = 0.7219 and C(p) = 5.5663

All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.
The REG Procedure
Model: MODEL1
Dependent Variable: PRI
Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	PUBH		1	0.3993	0.3993	164.8470	28.58	<.0001
2	HSTOCK		2	0.2477	0.6470	81.9669	29.47	<.0001
3	HIS		3	0.0840	0.7310	55.1808	12.80	0.0009
4	LAND		4	0.0409	0.7719	43.1633	7.17	0.0107
5	PROIN		5	0.0205	0.7924	38.1312	3.86	0.0567
6	GDP		6	0.0192	0.8116	33.5435	3.88	0.0562
7	GCON		7	0.0108	0.8225	31.8319	2.26	0.1415
8	UER		8	0.0717	0.8942	9.2684	24.38	<.0001
9		GDP	7	0.0003	0.8939	7.3626	0.09	0.7615
10		HSTOCK	2	0.0004	0.7219	5.5663	0.05	0.7215

Table 6: The actual and predicted values and prediction index

Year & Quarter	***Actual***	***Linear Regression Analysis***						***Genetic Algorithm - Linear Regression Analysis (with Adaptive Mutation Rate)***					
		LRA Model 1			*LRA Model 2*			*GA-LRA(AMR) Model 1*			*GA-LRA(AMR) Model 2*		
		Predict	*Diff.*	*Index*	*Predict*	*Diff.*	*Index*	*Predict*	*Diff.*	*Index*	*Predict*	*Diff.*	*Index*
1996Q1	-1,198	-1,861	-663	5.73	-861	337	2.91	-426	772	6.67	-1,161	37	0.32
Q2	-803	-2,640	-1,837	15.87	-1,458	-655	5.66	-1,247	-444	3.84	-1,673	-870	7.52
Q3	-1,411	-2,938	-1,527	13.19	-1,607	-196	1.69	-1,493	-82	0.71	-2,297	-886	7.66
Q4	-1,899	-3,102	-1,203	10.39	-2,090	-191	1.65	-2,171	-272	2.35	-2,068	-169	1.46
1997Q1	-1,550	-2,302	-752	6.50	-1,943	-393	3.40	-2,139	-589	5.09	-1,632	-82	0.71
Q2	-1,108	-776	332	2.87	-910	198	1.71	-923	185	1.60	-1,108	0	0.00
Q3	-1,200	-508	692	5.98	-731	469	4.05	-901	299	2.58	-1,223	-23	0.20
Q4	-283	356	639	5.52	251	534	4.61	-178	105	0.91	-600	-317	2.74
1998Q1	590	321	-269	2.32	1,058	468	4.04	360	-230	1.99	31	-559	4.83
Q2	-100	-736	-636	5.50	1,393	1,493	12.90	293	393	3.40	63	163	1.41
Q3	789	-1,606	-2,395	20.69	1,859	1,070	9.24	425	-364	3.14	1,012	223	1.93
Q4	643	-3,176	-3,819	33.00	1,895	1,252	10.82	364	-279	2.41	282	-361	3.12
Index Sum				127.56			62.68			34.69			31.90