Malaysian Management Journal 3 (1), 61-69 (1999)

Multi Layer Perceptron Modelling in the Housing Market

KU RUHANA KU MAHAMUD, AZURALIZA ABU BAKAR
& NORITA MD. NORWAWI
School of Information Technology
Universiti Utara Malaysia

ABSTRACT

The study examines the use of multi layer perceptron network (MLP) in predicting the price of terrace houses in Kuala Lumpur (KL). Nine factors that significantly influence the price were used in this attempt. Housing data from 1994 to 1996 were presented to the network for training. Tested results from the model obtained for various years were compared using regression analysis. The study provides the predictive ability of the trained MLP model that can be used as an alternative predictor in real estate analysis.

ABSTRAK

Kajian ini menguji penggunaan rangkaian perceptron multi aras (MLP) untuk meramal harga rumah teres di Kuala Lumpur (KL). Sembilan faktor yang mempengaruhi harga secara signifikan digunakan. Data perumahan dari tahun 1994 hingga 1996 dipersembahkan kepada rangkaian untuk sesi latihan. Hasil ujian model yang diperoleh untuk beberapa tahun dibandingkan dengan kaedah regresi. Kajian ini menyediakan kebolehan peramalan bagi model rangkaian multi aras yang terlatih dan boleh digunakan sebagai alat peramal alternatif dalam analisis hartanah.

INTRODUCTION

The housing industry has been a very important sector in Malaysia's economic growth. House pricing consistently fluctuates in accordance with many factors. Housing developers classify the factors involved in determining prices into internal factors and external factors. Internal factors are marketing objectives, marketing mix strategy and cost, while external factors include market and demand, consumer perceptions of price and value, competitors' cost prices and offers, economic conditions which are non-linear, such as recession, inflation and interest rates, government policies and social concerns.

The unpredictable movement of prices in the housing industry has shown the need for a systematic model to overcome the problem of forecasting the price movement of houses in Malaysia. The ability to deal with non-linear factors which are found in the housing industry is a major advantage of artificial neural networks (ANNs) modelling. Other techniques such as regression and heuristics are unable to give a precise prediction in price movements as price movements forecasting is an example of the real world system in that they change over time.

ANNs have been shown to be successful as a predictive tool. One of the most commonly used ANNs tool for prediction, classification and forecasting is the multi layer perceptron (MLP). Some organizations have developed sophisticated systems that require the training of hundreds or even thousands of ANNs on a weekly basis to predict stock market index movements as well as individual stock price behaviour (Patterson, 1996). It has also been shown that neural network modelling is a more superior method compared with other techniques because of its ability to handle incomplete and noisy data (Hardgrave,Wilson & Walstrom, 1994). Neural network is also preeminently suited to detect non-linear patterns without ex-ante knowledge of the underlying model (Baestaens, Van Den Bagh & Wood, 1994).

Chapman (1994) discusses the development of a neural network model for use as part of a stock market trading system and some of the important considerations and decisions in the development of such systems. Some research examples using neural networks in stock market applications include forecasting the value of a stock or index, recognition of patterns in trading charts, prediction of ratings of corporate bonds, estimation of the market price of options, and direct generation of buy, sell and hold trading signals.

Richeson, Zimmerman and Barnett (1994) developed a neural network model to predict the payment performance on consumer loans by identifying loans as either performing or non-performing. The models help the loan officer function more efficiently by prescreening and eliminating unqualified applicants. The models can also assist the bank manager or auditor in evaluating how well the bank is adhering to its own credit standards.

Hardgrave et al. (1994) explored the biological inspired, non-parametric statistical approach of neural networks, in terms of their ability to predict academic success in an MBA program. They found that neural networks perform at least as well as traditional methods such as regression and discriminant analysis, and are worthy of further investigation. Saiful Hafizah et al. (1996) discussed the neural networks model for predicting the profitability of selected firms on the Kuala Lumpur Stock Exchange (KLSE). Their research employed MLP to achieve better results in predicting the profitability of listed companies in the KLSE.

In this paper, an original application of MLP modelling in the housing market is presented. To our knowledge, no studies have been done on the application of ANNs in the Malaysian housing industry. Furthermore MLP can be applied as a systematic and efficient forecasting method in non-linear problems for business applications.

THE MODEL

Neural networks are built of many nodes, called neurons or processing elements. Each node is a very simple unit which takes many inputs simultaneously and sums them. Each node produces a response dependent on the level of inputs received. If the sum of the inputs is high, the node reacts strongly; if the sum of the inputs is low, the response is low. This triggers an activation function that might add weight to high value patterns and ignore low value patterns.

MLP, which is one of the neural network's architecture, is a feed forward network with one or more layers of nodes between the input and output nodes. The additional layers contain hidden units or nodes that are not directly connected to both the input and output nodes. It is a supervised learning process with back propagation that feeds the errors between predicted and actual outputs backwards through the network to adjust the weights on the network connections.

In this study, a structured design approach recommended by Turban (1992) for developing neural network applications was adopted. The housing data are of terrace houses in KL, spanning the period from 1991 to 1996. Nine variables which were used in determining the house price are the year the data was collected, land area, house type, type of land ownership, built up area, age of the house, distance from the city, environment and building quality. The nine variables are the same as those used by INSPEN (1996) to produce the Malaysian house price index.

Binary patterns have been used to represent the data as MLP processing elements produced output signals that vary in magnitude, usually in the range of 0 and 1. Binary patterns are the easiest patterns to represent a network because they provide a very direct method of representing an input pattern. Several important decisions are required in MLP modelling such as the topology, i.e., number of processing elements and their configurations (input, layer, output), size of training and test data, network tool, learning algorithm and the activation function. The input layer consists of nodes that represent the nine variables mentioned earlier. The output layer has only one node which represents the price of the house.

In principle, the more examples or input data used for training, the better the model obtained. An increase in the number of observations will lead to a proportional decrease in noise effects. In this study 100 inputs from three years (i.e., 1994, 1995, 1996) have been used as exemplars giving a total of 300 inputs. 80% of the inputs were used as training data, 15% as validation data and 5% as test data (SPSS Inc, 1997). The selection of learning algorithm influences how well a neural network is able to solve a problem. The conjugate learning algorithm has been used in building the network in conjunction with the nature of the inputs. It measures the gradient of the error surface after each pass. The algorithm then alters the weights of the node inputs using a compromise between the direction of the steepest gradient and the previous direction of change. Three activation functions were used for trial in developing the model. They are the linear function, sigmoid function and the hyperbolic tangent function (Skapura, 1995).

The goal of building a network is not to memorise the training set but to learn something about the past that will generalise in the future when given a new example of the problem. The training process can be excessive and the best way to represent input data and the choice of architecture is subject to trial and error. Figure 1 shows the model structure of the network.

ANALYSIS OF RESULTS

Various different models were obtained in the training process. The criterion of choosing the optimum/best model using this network is based on the highest percentage of prediction. The best model obtained used the conjugate gradient as the learning algorithm and the linear activation function with one hidden layer and five hidden nodes, which have resulted in a 100% correct prediction. The associated root mean square (RMS) is 0.061717 which gives the mean square error (MSE) of 0.0038089. The small error statistic indicates that the trained MLP network has a very good simulation fit. Table 1 depicts part of the modelling phase performed while obtaining the best model. However, the model had to be tested for its validity with different sets of test data.

Test sets were presented to the trained network for testing the predictive performance of the best model obtained. Table 2 shows the results for four different years when data were presented for testing using the MLP and multiple regression (MR) approaches. Multiple regression is a classical statistical method that fits the examples to a straight line or plane. The algorithm estimates the equation of the line or plane by finding the lowest error measure for the examples (SPSS Inc, 1997). The regression model used was,

$$y = 0.004613\,x_1 + 0.36482\,x_2 + 0.05108\,x_3 + 0.03805\,x_4 + 0.24442\,x_5 + 0.02651\,x_6 - 0.20767\,x_7 + 0.09402\,x_8 + 0.05726\,x_9 - 92.04159$$

where the dependent variable is the price, represented by y. The independent variables are as mentioned earlier, represented by x_1, x_2 ... x_9 respectively. This is the same model used by INSPEN (1996). All MSE values obtained from the two approaches fulfilled the required error of less than 0.01, except for 1997. However the values obtained for 1997 still indicated a small error that can be accepted. The MSE values obtained from the MLP approach provided a better result than the one from regression. Figures 2 and 3 depict the graphs for the targeted and the predicted output for 100 data from 1996 and 1997. The 1997 housing data used for the testing and the predicted price obtained are shown in the Appendix.

Figure 1
The Network Model Structure

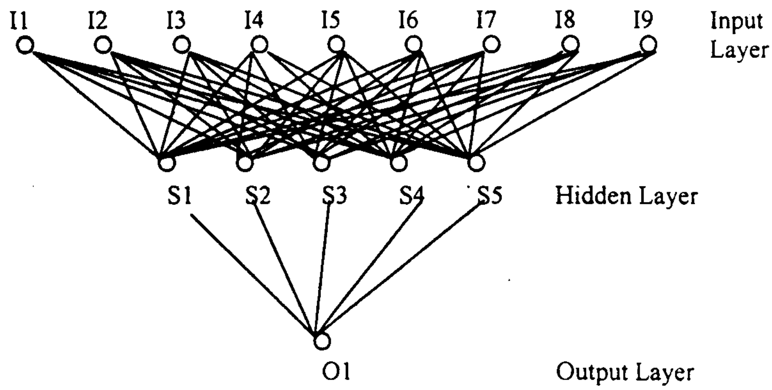


Table 1
MLP Modelling Process

Model	Act.funct	Learn.algo	Distribution	RMS	Mean Abs	% Prediction
1 (9-3-1)	sigmoid	conj.grad	standard	0.071315	0.054999	80.00
2 (9-4-1)	sigmoid	conj.grad	standard	0.061517	0.048291	86.67
3 (9-5-1)	sigmoid	conj.grad	standard	0.081211	0.057658	80.00
4 (9-6-1)	sigmoid	conj.grad	standard	0.076261	0.060379	86.67
5 (9-7-1)	sigmoid	conj.grad	standard	0.082582	0.054931	80.00
6 (9-3-1)	tanh	conj.grad	standard	0.073867	0.053065	86.67
7 (9-4-1)	tanh	conj.grad	standard	0.059695	0.045983	86.67
8 (9-5-1)	tanh	conj.grad	standard	0.111897	0.087805	66.67
9 (9-6-1)	tanh	conj.grad	standard	0.093520	0.063304	80.00
10(9-7-1)	tanh	conj.grad	standard	0.128567	0.079643	73.33
11(9-3-1)	linear	conj.grad	standard	0.061668	0.049229	93.33
12(9-4-1)	linear	conj.grad	standard	0.062770	0.045233	86.67
13(9-5-1)	linear	conj.grad	standard	0.061717	0.049156	100.00
14(9-6-1)	linear	conj.grad	standard	0.061476	0.048923	93.33
15(9-7-1)	linear	conj.grad	standard	0.061406	0.048699	93.33

Multi Layer Perceptron Modelling Housing Market
Type of NN : MLP - multilayer perceptron
Learning Algorithm: i) steepest descent ii) conjugate gradient
Activation function: i) linear ii) hyperbolic tangent iii) sigmoid
Num. of data : 100 from each year (94,95,96) = 300 training data
Data Allocation: 80% - training, 15% - validation, 5% - test

Table 2
The MSE Result of Test Data

Year	Data Size	MSE	
		MLP	MR
1994	600	0.00393500	0.0048104
1995	600	0.00447900	0.0076862
1996	100	0.00139988	0.0026157
1997	100	0.01496900	0.0174810

Figure 2
Predicted vs Actual House Prices for 1996

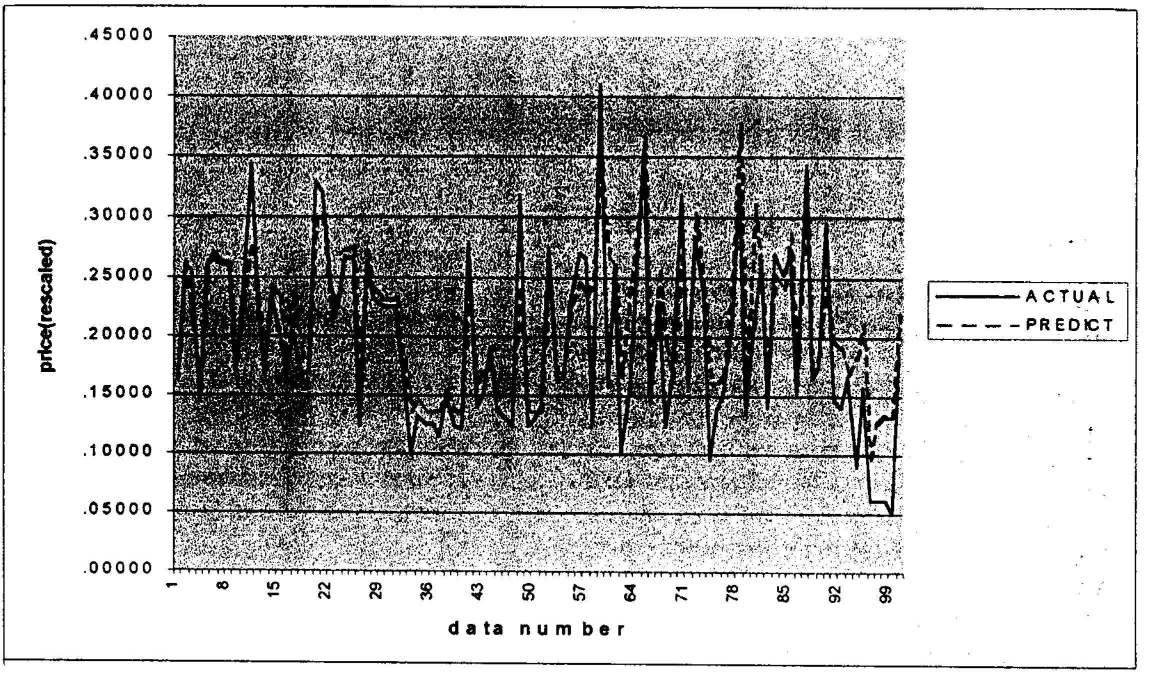
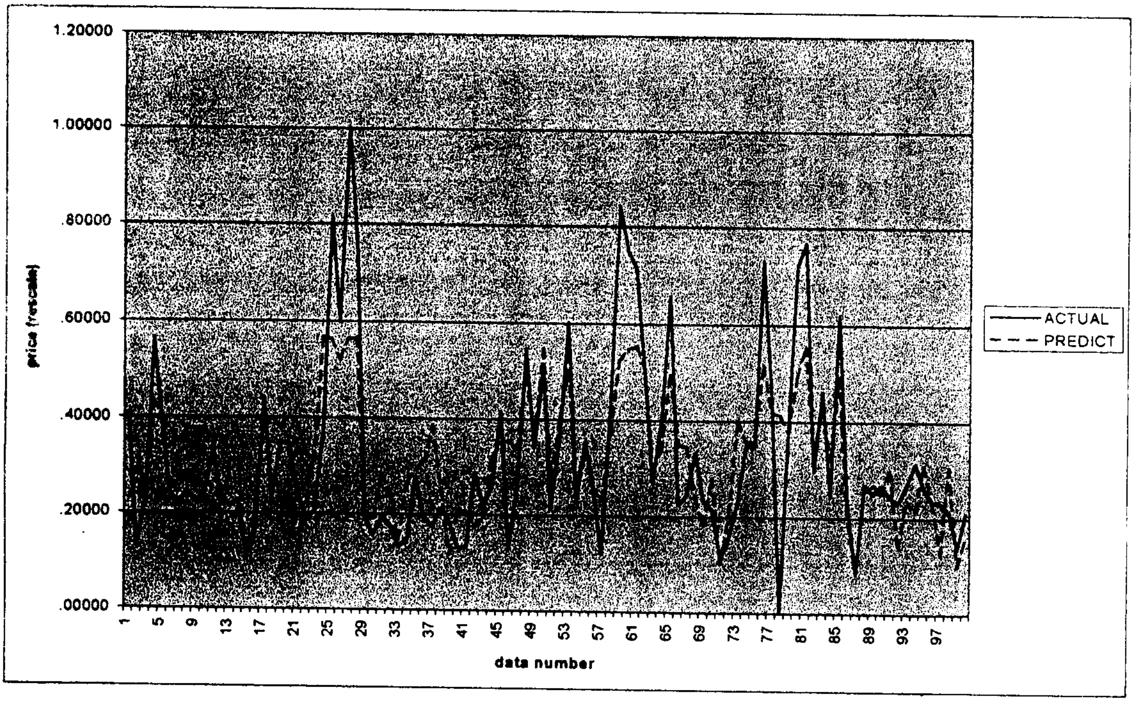


Figure 3
Predicted vs Actual House Prices for 1997



CONCLUSION AND RECOMMENDATION

This study shows that the MLP network can be used in modelling the Malaysian housing market and can be a useful and efficient way of modelling and analysing real estate markets to optimise allocation of resources. The results obtained from MLP are comparable to the one obtained by regression. This shows that MLP possesses considerable potential as an alternative to regression models. Further work on the MLP can be done to prove or test the generality of the model obtained. This can be done by providing test data sets of terrace houses from various states and years.

REFERENCES

Baestaens, D.E., Van Den Bagh, W.M. & Wood, D. (1994). *Neural Network Solutions for Trading in Financial Market*, New Jersey: Pitman.

Chapman, A.J. (1994). Stock Market Trading Systems through Neural Networks: Developing a Model, *International Journal of Applied Expert Systems*, **2(2)**: 90 -99.

Hardgrave, B.L.C., Wilson, R.L. & Walstrom, K.A. (1994). Predicting Graduate Student Success: A Comparison of Neural Networks and Traditional Techniques, *Computers Operational Research*, **21(3)**: 249 - 263.

INSPEN (1996). *Malaysian House Price Index: A Technical Summary*, Kuala Lumpur: Ministry of Finance Malaysia.

Neural Connection 2.0 User's Guide (1997). Chicago: SPSS Inc/ Recognition Systems Inc.

Patterson, W. D. (1996). *Artificial Neural Networks, Theory and Applications*, Prentice Hall: Singapore.

Richeson, L., Zimmerman, A.R. & Barnett, K.G. (1994). Predicting Consumer Credit Performance: Can Neural Networks Outperform Traditional Statistical Methods?, *International Journal of Applied Expert Systems*, **2(2)**: 115 - 129.

Saiful Hafizah, Siti Mariyam Shamsudin, Noriza Majid & Noriszura Ismail (1996). *Analisis Pelaburan Saham Dengan Rangkaian Neural, Prosiding Simposium Kebangsaan Sains Matematik ke- VII*, organised by Institut Teknologi MARA & Persatuan Matematik Malaysia, 369 - 375.

Skapura, D.M. (1995). *Building Neural Networks*, New York: Addison Wesley.

Turban, E. (1992). *Expert Systems and Applied Artificial Intelligence*, New York: MacMillan Publication Co.

APPENDIX

	Year	L/area	H/type	O/ship	B/area	Age	Dist	A/qlty	B/qlty	A/price	P/price
1	1997	121	0.67	0	122	5	12	1.00	1.00	245500	217402
2	1997	184	0.00	1	74	14	10	1.00	1.00	95000	216592
3	1997	55	0.67	1	85	7	17	1.00	1.00	150000	138773
4	1997	262	0.67	0	129	7	17	1.00	1.00	330000	259652
5	1997	174	0.67	0	136	12	16	1.00	1.00	230000	236496
6	1997	134	0.67	1	135	0	17	1.00	1.00	147750	224503
7	1997	130	0.67	1	127	0	17	1.00	1.00	146000	214065
8	1997	130	0.67	1	160	0	18	0.50	1.00	156000	227959
9	1997	130	0.67	1	160	0	18	0.50	1.00	146000	227959
10	1997	130	0.67	1	160	0	18	0.50	1.00	156000	227959
11	1997	227	0.00	0	98	17	16	1.00	1.00	205000	210900
12	1997	55	0.67	0	93	6	18	1.00	1.00	151000	123291
13	1997	134	0.67	1	126	0	18	0.50	1.00	140363	192848
14	1997	134	0.67	1	126	0	18	0.50	1.00	147750	192848
15	1997	82	0.67	1	55	11	18	0.50	0.50	86000	80879
16	1997	88	0.67	1	88	11	15	0.50	0.50	110000	136332
17	1997	292	0.00	0	94	5	16	1.00	1.00	264000	229563
18	1997	223	0.67	0	120	0	15	1.00	1.00	150000	240789
19	1997	161	0.67	0	130	2	14	1.00	1.00	213500	230918
20	1997	189	0.67	0	130	2	14	1.00	1.00	212000	243257
21	1997	130	0.67	1	160	0	17	0.50	1.00	156000	233640
22	1997	133	0.67	0	130	3	17	1.00	1.00	130000	202044
23	1997	164	0.67	0	109	0	17	1.00	1.00	133888	191520
24	1997	317	0.67	0	160	2	14	1.00	1.00	224440	332179
25	1997	167	0.67	1	176	12	10	1.00	1.00	465000	328853
26	1997	179	0.00	1	178	18	14	1.00	1.00	350000	306318
27	1997	174	0.67	1	160	20	8	1.00	1.00	565000	329690
28	1997	178	0.67	1	157	24	8	1.00	1.00	440000	330068
29	1997	133	0.00	0	75	26	14	0.50	0.50	128000	126666
30	1997	52	0.67	1	80	10	12	0.50	0.50	110000	128361
31	1997	176	0.00	0	88	17	16	1.00	1.00	130000	177582
32	1997	149	0.00	1	83	27	15	0.50	0.50	120000	155205
33	1997	55	0.67	1	57	8	12	0.50	0.50	97000	103802
34	1997	137	0.00	1	76	23	10	0.50	0.50	105000	168824
35	1997	143	0.00	1	95	28	12	0.50	0.50	175000	183069
36	1997	149	0.00	1	101	28	12	0.50	0.50	130000	192211
37	1997	148	0.67	1	127	34	12	0.50	0.50	120000	232873
38	1997	127	0.00	1	99	24	15	0.50	0.50	140000	161447
39	1997	82	0.67	0	88	11	20	0.50	0.50	125000	87300
40	1997	66	0.67	1	62	8	8	0.50	0.50	96000	136775
41	1997	133	0.00	0	83	27	14	0.50	0.50	97000	135819
42	1997	82	0.67	0	89	16	15	0.50	0.50	180000	119111
43	1997	55	0.67	1	80	10	12	0.50	0.50	143000	129679

44	1997	100	0.67	1	94	8	12	1.00	1.00	165000	197195
45	1997	155	0.67	0	147	4	14	0.50	0.50	250000	214200
46	1997	184	0.00	1	74	14	10	1.00	1.00	95000	216592
47	1997	108	0.67	1	101	7	12	1.00	1.00	162888	207849
48	1997	196	0.67	0	155	13	6	0.50	0.50	320000	290556
49	1997	106	0.67	1	131	3	15	1.00	1.00	210000	220583
50	1997	305	0.67	1	129	10	13	1.00	1.00	300000	320715
51	1997	82	0.67	1	97	17	10	0.50	0.50	140000	174639
52	1997	116	0.67	1	130	25	8	0.50	0.50	220000	240503
53	1997	254	0.67	1	131	13	14	1.00	1.00	350000	296150
54	1997	55	0.67	0	93	7	7	0.50	0.50	150000	152721
55	1997	108	0.67	1	102	7	12	1.00	1.00	218000	208924
56	1997	82	0.67	1	97	17	12	0.50	0.50	165000	163289
57	1997	55	0.67	0	83	14	6	0.50	0.50	90000	150847
58	1997	250	0.00	1	92	28	12	0.50	0.50	260000	226944
59	1997	199	0.67	1	176	10	16	1.00	1.00	480000	307955
60	1997	178	0.67	1	127	25	4	1.00	1.00	430000	320704
61	1997	223	0.67	1	157	4	11	1.00	1.00	410000	323485
62	1997	149	0.67	1	166	15	12	1.00	1.00	300000	300141
63	1997	153	0.67	0	145	19	20	0.50	0.50	170000	184149
64	1997	174	0.67	0	168	18	22	0.50	0.50	240000	206516
65	1997	175	0.67	1	146	20	12	1.00	1.00	380000	292251
66	1997	130	0.67	1	127	1	17	1.00	1.00	146000	214529
67	1997	130	0.67	1	160	1	18	0.50	0.50	156000	211187
68	1997	98	0.00	0	98	17	16	1.00	1.00	205000	154065
69	1997	55	0.67	0	93	6	18	1.00	1.00	151000	123291
70	1997	134	0.67	1	126	1	18	0.50	0.50	140363	176076
71	1997	82	0.67	1	55	11	18	0.50	0.50	86000	80879
72	1997	88	0.67	1	88	11	15	0.50	0.50	110000	136332
73	1997	223	0.67	0	120	1	15	1.00	1.00	150000	241259
74	1997	130	0.67	0	130	2	14	1.00	1.00	213500	217267
75	1997	130	0.67	0	130	2	14	1.00	1.00	212000	217267
76	1997	167	0.67	1	165	19	13	1.00	1.00	420000	303181
77	1997	143	0.67	1	151	23	12	0.50	0.50	275000	251535
78	1997	153	0.67	0	161	20	12	0.50	0.50	25000	247361
79	1997	153	0.67	1	132	19	13	0.50	0.50	230000	227781
80	1997	149	0.67	1	177	19	10	0.50	0.50	415000	291825
81	1997	184	0.67	1	184	6	13	1.00	1.00	440000	325170
82	1997	122	0.00	1	122	27	12	0.50	0.50	185000	202617
83	1997	143	0.67	1	151	11	10	0.50	0.50	275000	257254
84	1997	149	0.00	1	96	29	10	0.50	0.50	156000	198599
85	1997	147	0.67	1	177	19	10	0.50	0.50	360000	290945
86	1997	82	0.67	1	80	4	14	0.50	0.50	150000	127395
87	1997	51	0.67	0	56	19	15	0.50	0.50	70000	71089
88	1997	92	0.67	0	103	4	10	0.50	0.50	170000	161436
89	1997	139	0.00	0	117	4	16	1.00	0.50	160000	169385
90	1997	143	0.00	1	90	25	15	0.50	0.50	169000	159212
91	1997	98	0.67	0	98	2	8	0.50	1.00	157000	186309

92	1997	82	0.00	0	88	11	16	0.50	0.50	150000	99898
93	1997	133	0.00	1	75	19	12	0.50	0.50	170000	152748
94	1997	82	0.67	1	100	11	16	0.50	0.50	195000	141014
95	1997	145	0.00	1	89	34	10	0.50	0.50	168000	191612
96	1997	109	0.67	0	83	13	10	0.50	0.50	152000	151473
97	1997	82	0.67	1	55	11	16	0.50	0.50	150000	92230
98	1997	137	0.67	1	85	27	10	0.50	0.50	140000	190559
99	1997	55	0.67	1	57	7	16	0.50	0.50	100000	80625
100	1997	55	0.67	1	80	8	15	0.50	0.50	138000	111713

L/area - Land area
H/type - House type
O/ship - Type of ownership
A/qlty - Area quality
B/qlty - Building quality
A/price - Actual price
P/price - Predicted price