Residential Price Forecasting at National and Regional Levels

Dr. Ian WILSON, Stuart PARIS, Dr. Andrew WARE and David JENKINS, United Kingdom

Key words: SFR's; Sustainable Value; Forecasting; Neural Networks; Gamma Test.

ABSTRACT

In the UK professional valuers currently assess residential values (Single Family Residences) for sale and loan based on current bid prices. This approach has some merits. The comparative valuation methodology on which it is based is straightforward and there is often sufficient data to form the basis of a judgement. Possibly too, the measure of value, commonly known as 'open market value', despite its transient nature, is understood by consumers.

Such an approach makes no attempt to predict periodic market crises or to produce estimates of longer-term sustainable value. As a result, consumers entertain substantial (hidden) risk. Prices can fall with undesirable personal consequences. The lack of labor mobility engendered leads to disruptive social and economic consequences. In worst case scenarios, generalized synchronous downturns, lenders are also exposed to risk.

European financial institutions and governments have expressed a desire for the development of theory and technique in this area: ideally a measure of value would be evinced capable of expressing an 'underlying' price. The Engineering and Physical Sciences Research Council, supported by the Office of Science & Technology, UK, have funded the research described in this paper. The ultimate aim is to develop a means of providing a more informed objective opinion about how the property market might move and what the consequences would be for a particular consumer.

The research starts from the basic proposition that predicting trends within housing markets may help in the development of a new measure of value that reflects 'economic sustainability'. The inputs for analysis are national and regional economic, social and residential property transaction time-series data. The most direct output is the trend of open market value. Given that the period of greatest risk occurs during the first 3 years of the mortgage,[1] predictions of open market value 12 quarters ahead were thought to be a useful target. Such a series might inform notions of sustainability.

Work has also been undertaken to identify a measure of house price/ affordability that showed no overall trend. Such a representation has two advantages. First is that the measure represents a time independent indicator of the state of the housing market in relation to its underlying average. Second is that it is easier to isolate those factors causing the measure to vary over time, while allowing the removal of the date as an explicit variable in models.

The tools chosen to undertake the analysis are neural networks for two reasons. First is that neural networks are recognised as powerful and appropriate forecasting tools. Second is that the research team have prior experience of successfully building neural network architectures for the estimation of residential open market values.[2]

Neural networks were trained using as inputs economic data and the previous quarter's measure of house prices. If successful the neural networks should be able to identify subtle and consistent relationships between these factors, enabling them to predict potential future patterns of change. The outputs were measures of the current quarter's house price and a surrogate affordability ratio.

The *Gamma* test, a non-linear analysis and modelling tool, was used to optimise the networks.

At the half-way stage in the research programme (January 2002) results described give confidence that artificial neural networks can be used to forecast national and regional price trends within the housing market under various conditions. The research has identified a number of issues currently the subject of empirical work:

- identifying the limits of Takens' theorem within the domain
- introducing Economic Trend Data
- measuring the level of 'noise' within data and determining an appropriate response
- acquiring appropriate data to migrate from national to regional aggregates.

CONTACT

Dr Ian Wilson, Reasearch Fellow and David H Jenkins, BSc MPhil FRICS, Senior Lecturer
School of Technology
University of Glamorgan
CF37 1DL
UNITED KINGDOM
Tel + 44 1443 482 268 (Wilson)
+ 44 1443 482 336 (Jenkins)
Fax + 44 1443 482 169
E-mail: idwilson@glam.ac.uk and dhjenkin@glam.ac.uk

Residential Price Forecasting at National and Regional Levels

Dr. Ian WILSON, Stuart PARIS, Dr. Andrew WARE and David JENKINS, United Kingdom

1. INTRODUCTION

Property professionals are well aware of the importance of residential owner occupation to the macro economy. A few headlines will illustrate how well founded is that awareness:

- In the UK and USA, investment in residential property since W.W.II has typically accounted for between 3% and 4% of GDP per annum[3]. In developing countries the contribution may be much higher (the World Bank suggests 10% per annum in China for example).
- In Sweden housing wealth as a fraction of household net wealth has fluctuated between 50% and 75% during the post-war period[4]
- The value of owner occupied domestic stock in the UK was estimated to be £1450 billion in 2000 [5]

Major changes in the housing market can have a significant impact on the economy as a whole. The large fall in house prices in the UK between 1989 and 1992 (when real house prices fell by about 25%) was a period during which the savings rate almost doubled (from just over 6% to around 12%), GDP stagnated and business confidence slumped[6]. Many individuals were left in a situation of negative equity, when they had paid (and were usually still paying) more for their homes than they could realise by selling. This had social cost in increased repossessions and reduced labour mobility.

While valuation methodology cannot be blamed for these consequences, in the UK at least it is likely that professional estimates of open market value tend to be driven by the highest bid price in a rising market and that professional valuers (real estate appraisers) tend to be driven by lenders anxious to meet market share targets.[7] Ultimately, this state of affairs is of no benefit to individual consumers nor to macroeconomic performance.

Objectively consumers need a longer run measure of value that answers questions about the sustainability of the investment to which they are committing. At the very least, consumers want to know not only the bid price but the likely trend of values.

This paper describes the attempt of the authors to model house price trends using artificial neural networks (ANNs) applied to published economic, social and house price data series. Section 2 examines preliminary issues: the identification of data and the selection of appropriate modelling approaches. Section 3 presents a summary of results while section 4 describes issues for further work.

2. DATA AND MODELLING

2.1 Data

Theoretical market models indicate that the main variables expected to influence house prices at both the national and regional levels are[8]: incomes; interest rates (real or nominal); the general level of prices; household wealth; demographic variables; the tax structure; financial liberalisation; the housing stock. Some of these data are available (generally at the national level) from UK Government Departments and bodies such as the Bank of England.

The economic data sets used in this analysis (Income, Retail Prices, Interest Rate and Unemployment Percentage Rate) and the UK time-series data on national and regional house prices are described and critically analysed on our website (www.glam.ac.uk/aitech) and in earlier publications.[9] There are 3 all UK house price indices (DTLR Survey of Mortgage Lenders, Halifax plc and the Nationwide Building Society) which exhibit some variation between them, though they all portray an overall upward trend in prices with similar magnitudes. Much of this movement can be ascribed to inflation, as recorded in a general increase in RPI. However, even when inflation is taken into account, there still remains an upward trend in UK house prices. The Nationwide Building Society has produced a graph (Figure 1) showing RPI deflated house prices and the underlying real trend in prices.

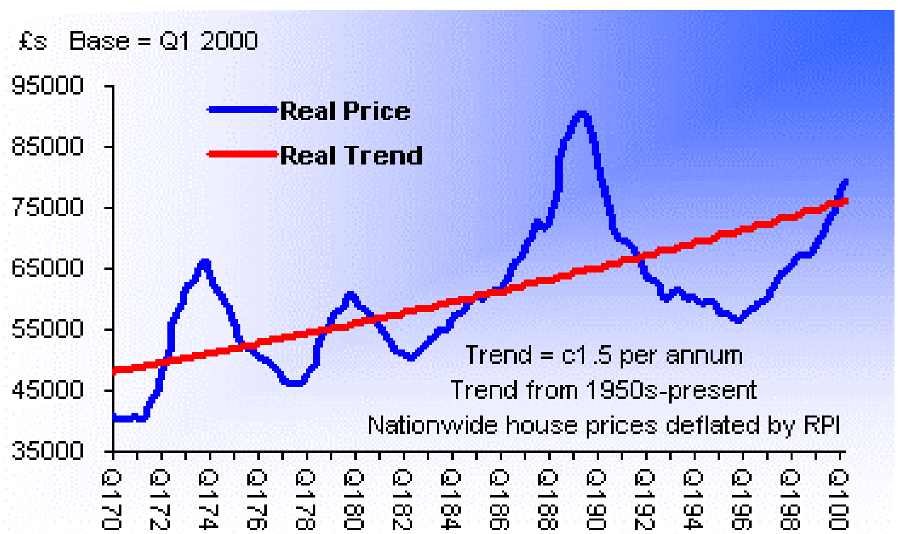

Figure 1 Nationwide House Prices deflated by RPI and Real Trend

At an early stage in the project it was decided that benefits could be gained from the identification of a measure of prices which showed no overall trend. The conclusion from the Merrill Lynch forecasting model[10] that real house prices are strongly linked to movements in personal disposable income suggested that house prices and incomes might show similar trends.

Figure 2 is a graph of the Nationwide House Price Index and the seasonally adjusted Average Earnings Index (suitably scaled) adjusted for RPI together with the addition of the DTLR /

SML quarterly house price index adjusted for inflation. This shows that the trend of average earnings appears to be very similar to the real trend in prices shown in the Nationwide graph. And the trend of the DTLR index in Figure 2 appears to have a similar slope to the trend of the average earnings index, suggesting that the ratio of house price to average earnings could be used as a measure of prices showing no overall trend.

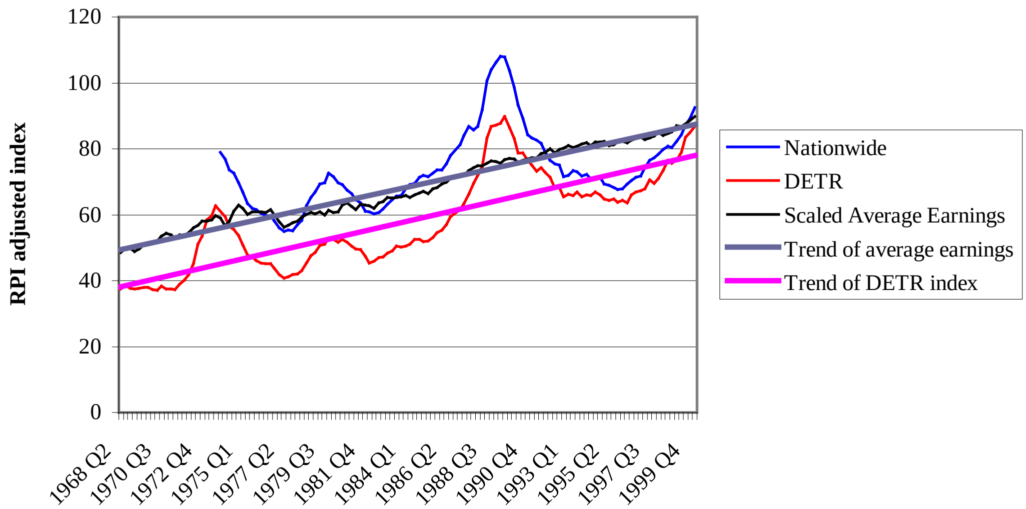

Figure 2 Comparison of the RPI adjusted DTLR, Nationwide and Average Earnings Indices

The results of plotting the ratio of Nationwide house price index to average earnings index (HPI/ AEI) is shown at Figure 3 (alongside forecasts of its value). It can be seen that the trend is approximately horizontal; the value of the ratio represents a surrogate measure of affordability and a time independent indicator of the state of the housing market in relation to its underlying average. Major departures from the trend are visible suggesting moments in the series where other causal factors may be identified.

2.2 Modelling

According to general economic theory, if the price or open market value of dwellings is determined by supply and demand, the value is discovered when simultaneous equations representing a function of demand and a function of supply are solved. Given the multitude and complexity of factors that contribute to both the supply and demand functions, house-price determination studies invariably limit the number of variables with the intention of reducing the problem to manageable proportions. The outcome of this is that the models generated can only be applied within narrow time and space zones because of the restrictions imposed by the reduction.

Using a stratified ANN model, researchers at Glamorgan demonstrated how these restrictions can be overcome in relation to space within a relatively static time frame [[11]]. To achieve the same result for time frames (that is, to forecast time series of prices), it was proposed that an analogous approach be implemented using AI techniques. Forecasting is an area where ANNs are considered to provide a great deal of promise [[12]]. A stratified ANN approach allows different data sets to be introduced at say national and sub-regional levels or, by identifying clusters in the time series, permits different networks to be deployed at different stages in the house price cycle.

Three types of AI techniques that are being examined within this project are:

- supervised learning networks like the multi-layered perceptron (MLP)
- Kohonen Self-Organizing Maps
- genetic algorithms

A strength of ANNs is their ability to cope with non-linear functions. They have been used successfully in a number of experiments to derive house prices in snapshots of limited housing markets. It was therefore decided to apply MLP networks to time series as an initial technique for forecasting house price movements.

Using a MLP the time series is partitioned into three data sets:

- training set – used to train the network and identify the underlying function
- validation set - used to test the network during training. If the quality of the training set is good, then the error in predicting the outputs from the independent test set will fall during training.
- testing set - used to test the accuracy of the network once trained.

Initial results reported in Section 3.1 reflect this approach. However, given documented limitations of MLP networks it was additionally decided to deploy the Gamma test to optimise the performance of the MLPs. The Gamma (near neighbour) test is a data analysis algorithm that estimates the Mean Squared Error (MSE) that can be achieved by a model constructed using this data. This test can be used to simplify the process of constructing an ANN and optimise its performance. A description of the Gamma Test appears on the website (www.glam.ac.uk/aitech) and in Stefánsson[13].

3. SUMMARY OF RESULTS

3.1 MLP

Initial empirical work introduced the national data series to MLP networks in order to predict the values of the affordability ratio, HPI/AEI, referred to earlier and Nationwide average house price.

The inputs consisted of just the current values of the economic data for the relevant quarter (Basic networks) or of the current values plus the change in each indicator from the previous quarter (Basic + Change network). The value of HPI/AEI was used as the output of each of the networks. A further network (2-year Window network) was trained using a two-year moving window of the values of HPI/AEI as the only inputs - that is, for any quarter, the inputs were the values of HPI/AEI for the previous eight quarters.
In each case except the network shown as Basic A, a three-year set of quarterly data was held

back to use as a forecasting set. In the case of Basic A, two three-year periods were held back – these were chosen so that the first value of HPI/AEI for each forecasting period was equal to the underlying trend value, but the directions of movement of the ratio were opposite. The remaining data were used in the training and testing sets. Each fifth set of data was used for testing – this interval was chosen so that each quarter was equally represented in the set to ensure that any seasonal effects would average out.

Each set of data was used as the input for two networks – one with the date included as an input and one with it excluded.

The trained networks were used to make successive one period (quarter) ahead forecasts. This was done by rolling the sample forward one period, using the predicted measure of house prices as an input, and making another one-step ahead forecast and so on. The results of the forecasts are graphed at Figure 3.

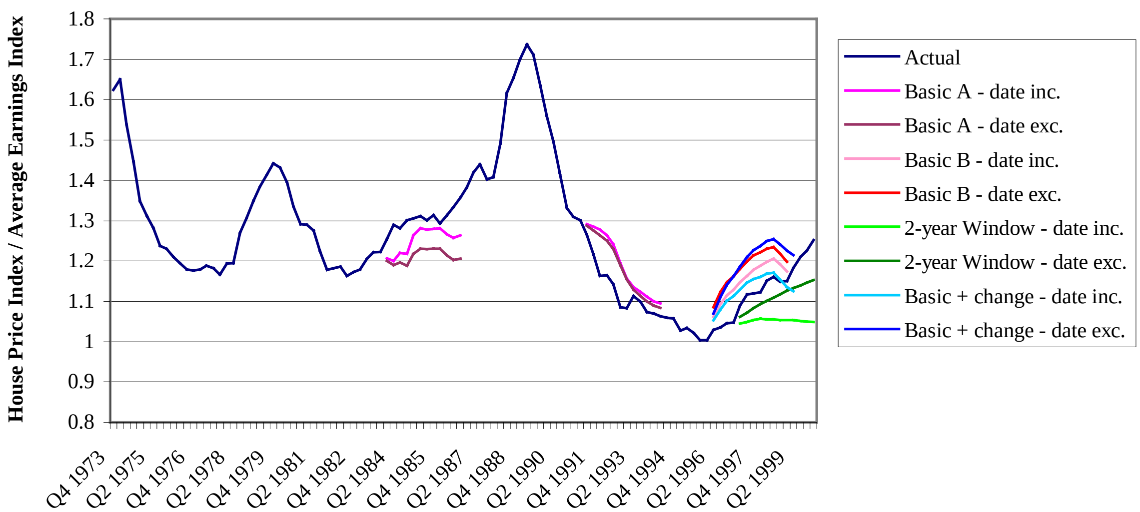

Figure 3 Actual and Predicted Values of the Nationwide House Price Index / Average Earnings Index

The forecast ratios of HPI/AEI, together with the value of the average earnings index, may be used to predict the average house price. The actual and predicted values of the Nationwide average house prices are graphed at Figure 4.

Although the networks have predicted the changing values of HPI/AEI, and average house price, with varying errors, the overall shapes of the curves have been similar to the actual result observed. This is particularly significant in the case of network Basic A, since a single network has successfully forecast two opposite market movements starting from approximately the same value of HPI/AEI.

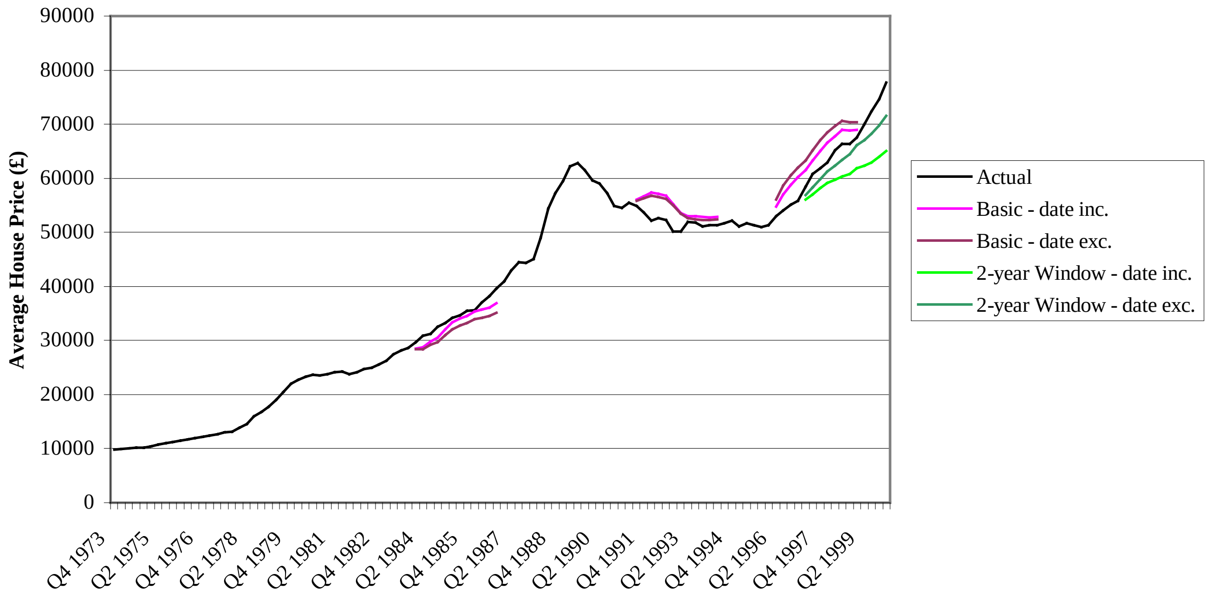

Figure 4 Actual and Predicted Values of Nationwide Average House Prices

The forecasts of HPI/AEI show similar shapes for the forecast curves whether date is included or excluded. This suggests that at least some of the market driving forces have been identified in defining the economic data as the inputs. The lower errors of some of the forecasts when date is included as input suggests that other information may be contained within the time-series.

3.2 Gamma Test

The *Gamma* test estimates the least Mean Squared Error (MSError) that any smooth data model (e.g. a trained feed forward neural network) can achieve on the given data without over training. It can be used with multiple column Input/ Output data files and single or multiple time series. As before, the investigation used the time series HPI/AEI, annual increase in the DTLR House Price Index, annual increase in average earnings, annual increase in RPI, bank interest rate and the percentage unemployment rate.

A number of fully-connected-feed-forward back-propagation ANNs were trained with window lengths from one to 16 lagged observations. The final 12 quarters were removed from the dataset to compare with subsequent forecasts, and the whole of the rest of the dataset was used to train the ANN. In each case, the networks were trained until the MSError predicted by the Gamma test for the new, reduced dataset was reached. The trained networks were used to produce forecasts for 12 quarters by one-step-ahead forecasts iterated for 12 steps. The results for the annual percentage change in the DTLR house price index using window length 15 are shown at Figure 5.

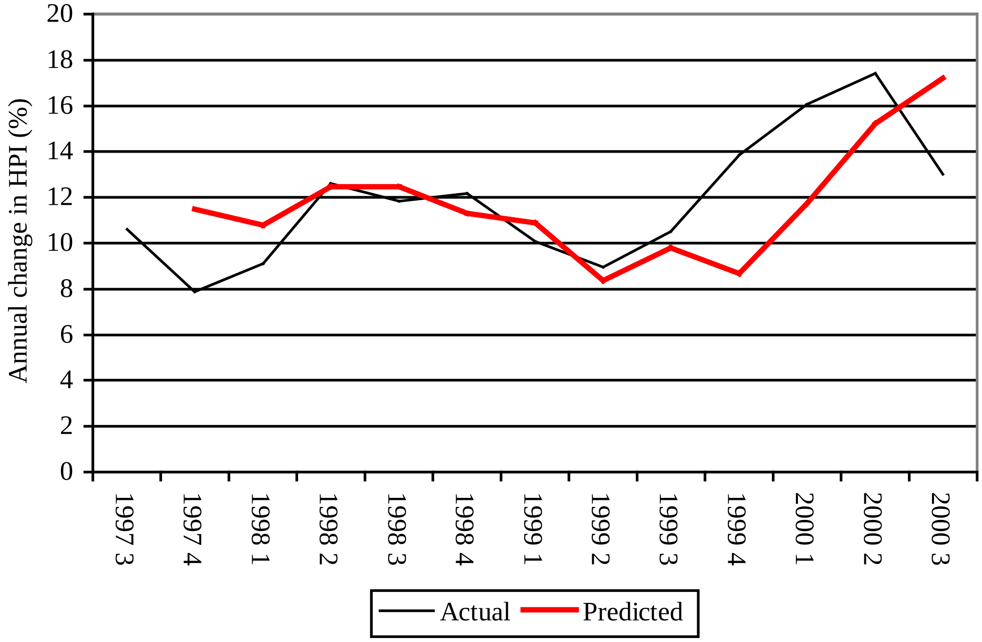

Figure 5 Annual % Change in DTLR House Price Index (detail)

The improved performance attained using the Gamma Test required experimentation to establish optimum window length. There is no room here to describe the nature and extent of this empirical work[14]; suffice it to say that the Gamma Test is of practical utility in optimising window length in the area of time series modelling.

4. UNFINISHED BUSINESS

The low average percentage error of the forecast over 12 successive quarters portrayed in Figure 5 is encouraging. Such an interval represents the period during which the parties to a mortgage of residential property are at greatest risk. However, there are several aspects of the development of methodology that remain to be investigated at the national level.

Even then useful forecasts of average UK house price, though they may be of interest to governments and corporations, are of only passing interest to consumers who are more concerned about the future value of their particular homes. A key objective for the research is to produce forecasts at the regional level.

4.1 Refining national forecasts

4.1.1 Embedding: identifying the limits of Takens' theorem

The *Gamma* test requires the selection of those inputs which can best be used to predict a selected output (some inputs may be noisy or irrelevant). Such options are all designed to produce an *embedding* – a selection of inputs chosen from all the inputs, and designated by a string of '1's and '0's called a *mask*.

Tsui [15] has suggested that, rather than the standard approach of taking regular past time lags as suggested by Takens, a good model can be constructed using an irregular embedding. That is, a good (possibly optimal) model may exist that requires only a subset of the inputs. This will be investigated by undertaking *Full Embedding* for single time series. (Full Embedding tries every combination of inputs to determine which combination yields the smallest absolute Gamma value.)

4.1.2 Economic Trend Data

We have demonstrated that reasonable levels of forecasting accuracy can be obtained using the values of basic economic data as input. However, it is likely that the housing market is affected by more than just the absolute level of a particular indicator. The individuals who drive the market are more likely to be influenced by their impressions of how different parts of the economy are moving (and on their own prospects within it) than by a snapshot of individual economic indicators.

This suggests that a model should include information on trends within the economy. However, this must be balanced against the need to keep the number of inputs to the ANNs relatively low in relation to the number of data sets available for training. As a means of achieving this, the use of the Gamma Test to establish irregular embedding for multiple time series will also be investigated.

4.1.3 Measuring the level of noise and determining an appropriate response

ANNs are thought to be adept at identifying the underlying trend within a time series and empirical work to date has shown this to be true for the housing data. However, as ANNs learn from past data, it is not possible for them to predict significant changes from the trend that are caused by catastrophic events not previously encountered. This should not be seen as a fatal flaw in ANN technology as no predictive technique would be able to cope with such occurrences. The yardstick for measuring the success of an ANN must be its ability to offer a better prediction than other techniques in reasonably foreseeable circumstances. It is envisaged that in our final system means of using heuristic reasoning based on the analysis of past crises will be used to assist in prediction following any future crisis.

4.2 Regional forecasts

Alongside the national average house price, Nationwide Building Society also publishes average price data for the UK regions. As a Welsh University collaborating with the largest Welsh firm of estate agents/ surveyors our aim is to forecast house prices for the region of Wales and ultimately for the conurbation of the Welsh capital, Cardiff and its hinterland, the South Wales Valleys.

At this stage we have produced a preliminary forecast for Wales simply by substituting the Welsh regional time series for the UK series. Table 1 shows forecasts of the average Welsh house price for the 12 quarters from Q3 2001. Currently we are substituting for other UK data series and intend to present further results at Washington.

X Series	Actual Series	Predicted Series	Error Series
1999_Q4	£57,697.35	£57,901.68	£204.33
2000_Q1	£59,437.05	£58,737.75	-£699.30
2000_Q2	£60,523.65	£60,416.42	-£107.22
2000_Q3	£60,778.50	£61,476.70	£698.20
2000_Q4	£60,916.34	£61,132.22	£215.88
2001_Q1	£63,431.75	£63,496.92	£65.17
2001_Q2	£65,337.70	£65,855.86	£518.16
2001_Q3	£66,454.98	£66,550.31	£95.33
2001_Q4	£68,191.00	£67,444.60	-£746.40
2002_Q1	-	£68,227.35	-
2002_Q2	-	£68,554.30	-
2002_Q3	-	£71,253.83	-
2002_Q4	-	£71,363.76	-
2003_Q1	-	£71,297.50	-
2003_Q2	-	£72,741.70	-
2003_Q3	-	£72,256.44	-
2003_Q4	-	£74,321.41	-
2004_Q1	-	£74,161.82	-
2004_Q2	-	£73,070.85	-
2004_Q3	-	£72,719.19	-
2004_Q4	-	£69,863.36	-

Table 1. 12 quarters forecast of the average house price for Wales based on the Nationwide Building Society Index

REFERENCES

[1] Jenkins, D.H. (2000) "Residential Valuation Theory and Practice", (ISBN 1 902375 50 5), Chandos Publishing (Oxford) Ltd.

[2] Lewis, O., Ware, J. A. & Jenkins, D. H., 1997, "A Novel Neural Network Technique for the Valuation of Residential Property." Journal of Neural Computing and Applications, Springer Verlag.

[3] Miles, D. (1994) "Housing, Financial Markets and the Wider Economy", ISBN 0 471 95210 9, Wiley.

[4] Englund, P. & Ioannides, Y. M. Berg (1988), cited in "House Price Dynamics: An International Empirical Perspective." Journal of Housing Economics, Volume 6, pp 119-136, 1997

[5] See 1

[6] See 3

[7] Gronow SA, Ware JA, Jenkins DH, Lewis OM, Almond NI 1996, "A Comparative Study of Residential Valuation Techniques and the Development of a House Value Model and Estimation System", Economic & Social Research Council.

[8] Meen, G. & Andrew, M. (1998) "Modelling Regional House Prices: A Review of the Literature." (ISBN 0 7049 1305 4), University of Reading for the Department of the Environment, Transport and the Regions.

[9] See for example, Wilson I D, Paris S D, Ware J A, Jenkins D H. Residential Property Price Time Series Forecasting with Neural Networks. The Twenty-first SGES International

Conference on Knowledge Based Systems and Applied Artificial Intelligence, Cambridge, UK, December 2001.
[10] Miles, D. and Andrew, S. (1997) "Merrill Lynch Model of the UK housing market", Merrill Lynch, Pierce, Fenner and Smith Limited.
[11] Lewis, O. M., Ware, J. A. and Jenkins, D. H. (2001) "Identification of Residential Property Sub-Markets Using Evolutionary and Neural Computing Techniques." Journal of Neural Computing and Applications, Springer-Verlag.
[12] Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998) "Forecasting With Artificial Neural Networks: The State of the Art." International Journal of Forecasting, Volume 14, pp. 35-62.
[13] Stefánsson A, Koncar N, Jones A J. A Note of the Gamma Test. Neural Computing and Applications 1997; 5:131-133
[14] See www.glam.ac.uk/aitech
[15] Tsui A P M. Smooth Data Modelling and Stimulus-Response via Stabilisation of Neural Chaos. PhD thesis, Imperial College of Science, Technology and Medicine, University of London, UK, 1999.

BIOGRAPHICAL NOTES

Ian D. Wilson, Principal Investigator on this project, is a Research Fellow at the University of Glamorgan who has 18 years background in intelligent computing applications.
http://web.glam.ac.uk/schools/sot/doms/Research/ai.php

Stuart Paris worked as a Civil Servant for over 20 years, before joining the Artificial Intelligence Technologies Modelling Group at the University of Glamorgan in 1999.

Andrew Ware obtained his PhD in Robotic Vision in 1992 and has since worked on a variety of Artificial Intelligence related projects. He is currently Head of the Artificial Intelligence Research Group at the University of Glamorgan.

David Jenkins is a Chartered Surveyor who spent over 20 years in professional practice before becoming a Senior Lecturer at Glamorgan University in 1993, where he is the Award leader for Property Appraisal.

Fuller biographical details available at http://web.glam.ac.uk/schools/sot/Staff/