Second International Conference on Construction in the 21st Century (CITC-II)
"Sustainability and Innovation in Management and Technology"
10-12 December, 2003, Hong Kong

The Impact of Construction Type on Single-Family Home Values Using Hedonic Estimation and Artificial Neural Network

Mohammed E. Haque, Ph.D., P.E.
Associate Professor, Department of Construction Science
Texas A&M University, College Station, TX 774843-3137, USA

Amro Taibah, Ph.D.
Assistant Professor, Department of Landscape Architecture
King Abdulaziz University, Jeddah, 21452, Saudi Arabia

Abstract

Estimating the value of a property concerns builders, developers, homebuyers, appraisers, economists, and policy makers among many others. Nonetheless, real estate valuation is a complex process considering the range of variables that are known to play a role in determining such a value. This paper investigated the impact of construction quality, resembled by various types of siding materials, on residential property value of single-family homes (SFH) within the City of College Station's urban setting by employing a comparative hedonic estimation and artificial neural network (ANN) model. The study relied on a large sample of SFHs that were sold during the period from 1997 to 2000 in College Station, Texas. The study used the homes that are highly homogeneous in their structural attributes in order to eliminate their impact on the home values. The main aim of this study was to find out the impact of four main types of siding materials that included brick veneer, frame, stucco, and mixed on home values. The results indicated that stucco siding had the most significant impact on the property value. In comparing the hedonic results and ANN results in this study found that both analytical methods support one another and have assigned similar weights to the various construction types that have been studied. In addition, ANN showed to have a higher predictive accuracy level than did the hedonic estimation. The estimated implicit values of different siding materials were a measure of the importance of such a material to the homebuyer and was resembled in the form of a paid premium. The findings extend the body of literature concerned with real estate value analysis and have significant implications in the realm of fund allocation decision making for a real estate developer.

Keywords

Residential Construction, Siding Materials, Real Estate Value Analysis, Hedonic Estimation, Artificial Neural Network.

1. Introduction

Estimating the value of a property concerns builders, developers, homebuyers, appraisers, economists, and policy makers among many others. Nonetheless, real estate valuation is a complex process considering the range of variables that are known to play a role in determining such a value. For example, in the case of single-family home (SFH) valuation, Can (1992) indicates that the price of a home is determined as a function of various structural, neighborhood, and location variables used to determine property value when using a hedonic estimation procedure.

In addition, it is well established in the literature that the larger portion of a property's value is mainly estimated by its structural attributes that include, but are not limited to, heated area, number of stories, number of bedrooms, type of construction, etc. (Sirpal, 1991). Hence, knowing to what degree a particular component impacts the overall value can provide valuable information pertaining to the significance of that particular component.

To a real estate developer, such information would provide valuable insight pertaining to allocation of funds decision-making. Generating higher proceeds from financing and/or from eventual sale of a property is the main goal of the real estate developer. He or She accomplishes this by creating a real estate market value far in excess of the real estate development cost. Creating real estate value is the most effective and powerful tool in real estate development used to maximize returns on real estate investments. It is through the reduction of needed equity, increasing returns from financing, and increasing returns from sales proceeds that higher returns are generated (Sharkawy, 1994).

2. Methodology

This study is designed to explore the potentials of combining conventional econometric methods for property value analysis, namely the hedonic price function, with ANN capabilities in assessing change, due to construction type factors, in residential property values. The methodology employed intends to compare results from both approaches to validate one another and to better understand the nature of the relationship between construction type and market value.

2.1 Study population

The study population consists of a sample of 145 highly homogeneous confirmed sales of SFHs in the city of College Station, Texas that sold during the period from 1997-2000.

SFHs of the same structural characteristics were used in this analysis to eliminate the effects of such variables. This will improve the model performance and accuracy in capturing the effect of construction-type on home values. In addition, the study used price per square foot (PPSQFT) for this analysis to minimize variation within the sample. An initial regression was conducted including a variable accounting for the year-sold for each observation. All PPSQFT values were then transformed to 2000 value.

The use of estimated appreciation rates for value transformation was not possible because all market reports found considered the Cities of Bryan and College Station as one continuous market. Taibah (2002), who studied the impact of amenities on SFH values in the College Station market, indicates that both cities can be treated as separate markets with different characteristics and real estate dynamics and therefore we chose not to use such published rates in this study. This above standardization is important for the ANN model training and specification. The ANN model architecture and learning method is based on error backward propagation. Building the learning curve of the model on values of the same year increases the accuracy of the ANN prediction.

2.2 Data Collection

The data, obtained from the Multiple Listing Services (MLS), includes price data, structural attributes data, and address information. A local real estate broker provided data strictly for research purposes.

The MLS data was validated and integrated with data from the Brazos County Appraisal District (BCAD). After cleaning, validating, and integration, only data that have complete records has been deemed suitable for this analysis.

2.3 Study variables

The variables included in the analysis are defined in the following Table 1, and include sales price per square foot and building age. In addition, the data includes the type of construction variables as designated by the MLS data. These include construction-veneer, construction-frame, construction-mix, and construction-stucco. These categories represent the type of construction system used for each home and particularly the exterior siding material.

Table 1: Variable definitions.

VARIABLES	DEFINITIONS
	DEPENDENT VARIABLE
PPSQFT00	Market price/Heated Area adjusted to 2000 value
	STRUCTURAL ATTRIBUTES
AGE_YR	Dummy--Age of the building at the time it is sold in years
	CONNSTRUCTION ATTRIBUTES
CONST_VENEER	At least 85% of exterior is brick veneer
CONST_FRAME	Asbestos/T111 plywood siding
CONST_MIX	Combination of two materials
CONST_STUCO	Stucco-paint/sheet sidings.

According to Taibah (2002) who studied the College Station market dynamics at length, the inclusion of location variables in the model, such as Texas A&M University as the main employer in the area that resembles a central business district (CBD), was found to be insignificant. The author indicates that this might be explained be the relatively shorter commuting distances and absence of traffic congestions.

3. Results of Hedonic Procedure

An analysis of the data was performed using the hedonic estimation procedure in order to ascertain the relationship between property market value and the type of siding material used. The hedonic price function describes the relationship between the observed prices of commodities and the characteristics associated with the commodities, based on the hypothesis that "goods are valued for their utility-bearing attributes or characteristics," (Rosen, 1974). Hence, the coefficient estimates represent the premium paid by the homeowner for the aesthetic utility that any siding material might bare. The following model was used for the analysis:

$$PPSQFT = \beta_0 + \beta_1 AGE\_YR + \beta_2 CONST\_VENEER + \beta_3 CONST\_FRAME + \beta_4 CONST\_MIX + \beta_5 CONST\_STUCO + e \quad (1)$$

Where β_0 = Intercept

β_1, β_2, etc. = coefficients of the independent variables, and

e = error term.

Results of the analysis are shown in Table 2.

Based on the results of the analysis, the hedonic equation can be written as follows:

$$PPSQFT00 = 52.419 - 0.515*AGE\_YR + 16.595*CONST\_VENEER + 15.734*CONST\_FRAME + 18.323*CONST\_MIX + 37.642*CONST\_STUCO + e \quad (2)$$

The F-value of the model was found to be statistically significant at the 0.0001 levels. The F statistic tests how well the model, as a whole, accounts for the dependent variable's behavior. The estimation efficacy of the model was found to be moderately high with a coefficient of determination (R^2) of 0.622. This is an indication that the model is able to explain approximately 62% of the total variation in the property's market value.

The results, shown in Table 2, indicate that "construction type" is an important attribute that has a significant effect on SFH market value. Moreover, the findings indicate that certain types of construction have a more considerable impact than others. For example, based on the study sample, it was found that using various types of stucco siding in exterior finishing could have a noticeably higher impact on future market price than other sidings. Compared to brick veneer, which is the most common exterior finishing material in the College Station market, Stucco finishing was found to reflect an increase difference of approximately $21 per square foot in market price. In addition, it was found that plywood sidings had the least premium estimated at approximately $15 per square foot. Homes with mixed siding materials in the sample were mainly brick veneer mixed with either wood or stucco sidings. This category also showed an estimated hedonic price of approximately $18 per square foot.

Table 2: Hedonic coefficient estimates.

Variable	Coefficients	Std. Error	T-Stat.	Sig.
(Constant)	52.419	7.784	6.734	0.000
AGE_YR	-0.515	0.073	-7.068	0.000
CONST_VE	16.595	7.745	2.143	0.034
CONST_FR	15.734	7.868	2.000	0.047
CONST_MX	18.323	7.725	2.372	0.019
CONST_ST	37.642	8.062	4.669	0.000

4. Evaluation of the Results

4.1 Artificial Neural Network Approach

The results of the analysis provided by the hedonic procedure were verified using Artificial Neural Network approach. Neural computing is a relatively new field of artificial intelligence (AI), which tries to mimic the structure and operation of biological neural systems, such as the human brain, by creating an Artificial Neural Network (ANN) on a computer. An ANN is a modeling technique that is useful to address problems where solutions are not clearly formulated or to validate the results obtained through other modeling techniques (Chester, 1993). The network has the ability to learn by example. Patterns in a series of input and output values of example cases are recognized. This acquired "knowledge" can then be used by the network to predict unknown output values for a given set of input values (Obermeier, 1989). An ANN is composed of simple interconnected elements called processing elements (PEs) or artificial neurons that act as microprocessors. Figure 1 illustrates a simple processing element of an ANN with the analogy of the human brain (Haque and Mund, 2002).

Each PE has an input and an output side. The connections on the input side correspond to the dendrites of the biological original and provide the input from other PEs while the connections on the output side correspond to the axon and transmit the output. Synapses are mimicked by providing connection weights between the various PEs and transfer functions or thresholds within the PEs. Currently, back-propagation is the most popular, effective, and easy to learn model for complex networks (Haque and Sudhakar, 2001a,b). For the last few years, the second author has been using various ANN back-propagation Multi-layer Perceptron (MLP) modeling techniques in materials science and engineering (Haque and Sudhakar, 2001a,b), and construction management (Choudhury and Haque, 2001). To develop a back-propagation neural network, a developer inputs known information, assigns weight to the connections within the network architecture, and runs in the networks repeatedly until the output is satisfactorily accurate. The weighted matrix of interconnections allows the neural networks to learn and remember (Obermeier and Barron, 1989).

4.2 Development of ANN model

Back-propagation networks are most useful for problems involving forecasting and pattern recognition. "NeuroShell 2" was used in the present analysis to implement back-propagation training. In essence, back-propagation training adapts a gradient-descent approach of adjusting the ANN weights. During training, an ANN is presented with the data of thousands of times (called cycles). After each cycle, the error between the ANN outputs and the actual outputs are propagated backward to adjust the weights in a manner that is mathematically guaranteed to converge (Rumelhart et al., 1986).

4.3 Training the model

The superiority of an ANN to algorithmic and other network based systems is its ability to be trained on historical information as well as real-time data. Training is the act of continuously adjusting connection weights of the inputs until they reach unique values that allow the network to produce outputs that are close enough to the desired outputs. The accuracy of the developed model, therefore, depends on these weights. Once optimum weights are reached, the weights and biased values encode the network's state of knowledge. Thereafter, using the network on new cases is merely a matter of simple mathematical manipulation of these values.

4.4 Results of ANN Approach

The parameter inputs considered in the proposed ANN model were the ones that were found to be statistically significant using the hedonic procedure (AGE_YR, CONST_VENEER, CONST_FRAME, CONST_MIX, CONST_STUCO). The output parameter was price per square foot adjusted to 2000 value (PPSQFT00). During training the model, 130 observations were used out of a total of 145 observations. The remaining 15 observations, separated arbitrarily from the total number of observations, were used during testing the trained model. Several different ANN models with different layers/slabs connections, weights, and activation functions (including logistic, linear, Symmetric Logistic, and Gaussian) were trained. In addition, pattern selections including "Rotation" and "Random" were applied with weight updates using Vanilla, Momentum and TurboProp. The presented ANN- Jordan-Elman recurrent network with linear and logistic mixed activation function ("Rotation" for pattern selection, and "Momentum" for weight updates) was the best one among all other trials. Jordan-Elman recurrent network is one of the several different variations of backpropagation networks that NeuroShell 2 offers (NeuroShell 2: User's Manual, 1996). In this research, Jordan-Elman recurrent network converged very rapidly to produce a very high predictive efficacy as indicated in system performance bellow. Figure 2 shows the actual and ANN predicted price ($ per Sq. Ft.) during the Network Training Phase, which indicates very low errors between the actual price and the ANN predicted price. The trained model was evaluated using the data that was not used during the training. Figure 3 depicts the actual and ANN Predicted Price during the network evaluation phase, which indicates very small difference between the actual and ANN predicted price.

5. System Performance

The neural network used for the presented model demonstrated a good statistical performance as indicated by the R^2 and r-values as shown in Table 5. During network training, R^2 was obtained as 0.7788 and 0.8627 during network evaluation, which were close to +1.0 indicating a very good fit between the actual and the network prediction. R^2 is a statistical indicator usually applied to multiple regression analysis, and can be calculated using the following formulae:

$$R^2 = 1 - (SSE/SS_{yy}) \qquad (3)$$

Where $SSE = \Sigma(y -)^2$, $SS_{yy} = \Sigma(y -)^2$, y is the actual value, is the predicted value of y, and is the mean of the y values.

The correlation coefficient, r is a statistical measure of the strength of the relationship between the actual vs. predicted outputs. During network training, *r-values* were obtained as 0.9109, and 0.9786 during network evaluation, which were very close to +1.0 indicating a very good fit between the actual and the network prediction. The formula for r:

$$r = SS_{xy} / \sqrt{(SS_{xx}\, SS_{yy})} \qquad (4)$$

Where

$$SS_{xy} = \Sigma xy - (1/n)\{(\Sigma x)(\Sigma y)\}$$
$$SSxx = \Sigma x^2 - (1/n)(\Sigma x)^2$$
$$SS_{yy} = \Sigma y^2 - (1/n)(\Sigma y)^2$$

Where n equals the number of patterns, x refers to the set of actual outputs, and y refers to the predicted outputs. Table 3 shows the overall statistical performance of the trained ANN model.

Table 3: Statistical performance of the ANN trained model

Items	**ANN – training phase**	**ANN – evaluation phase**
Pattern processed	130	15
R Squared	0.7788	0.8627
Correlation Coefficient, r	0.9109	0.9786
Mean Absolute Error	4.336	3.551
Minimum Absolute Error	0.191	0.177
Maximum Absolute Error	18.211	15.373
Percent within 5%	48.462	53.333
Percent within 5% to 10%	27.692	40.000
Percent within 10% to 20%	17.692	6.667
Percent within 20% to 30%	5.385	0

Percent > 30%	0.769	0

5.1 Price prediction

Hypothetical data was developed in order to compare the performance of the hedonic estimation model and the ANN model. The purpose is to further the validation and evaluate the performance similarities or differences between the two methods.

Ten imaginary data points were used with building ages of 5 and 10 years old and siding material characteristics that were chosen arbitrarily. The two models produced price values that were within acceptable variation of each other. The ANN R-square of 0.75 indicates that the variation between predicted values of the ANN model and the hedonic model did not exceed 25%. The following Figure 4 shows a comparison of the performance by both models.

6. Conclusion

Estimating the value of a property concerns builders, developers, homebuyers, appraisers, economists, and policy makers among many others. By employing a comparative hedonic-artificial neural network approach, this paper studied the impact of four main types of siding materials that included brick veneer, frame, stucco, and mixed on home values. In comparing the hedonic results and ANN results this study found that both analytical methods support one another and have assigned similar weights to the various construction types that have been studied. In addition ANN showed to have a higher predictive accuracy level than did the hedonic estimation. The study showed that various construction types have various impacts on market values of SFHs. The estimated implicit values of different siding materials are a measure of the importance of such a material to the homebuyer and are resembled in the form of a paid premium. The findings extend the body of literature concerned with real estate value analysis and have significant implications in the realm of fund allocation decision making for a real estate developer.

7. References

Can, A. (1992). Residential Quality Assessment: An Alternative Approach Using GIS, *Annals of Regional Science*, 26, 97-110.

Chester, M. (1993). *Neural Networks: A Tutorial*, Englewood Cliffs, NJ: Prentice Hall, 1993.

Choudhury, I. and Haque, M.E. (2001). A Study of Cross-cultural Training in International Construction Using General Linear Model Procedure and Artificial Neural Network Approach, proceedings of the 3rd International Conference on Construction Project Management (3ICCPM) Conference, pp. 444-453, 29 - 30, Singapore.

Haque, M.E. and Mund, A. (2002). Shoring Loads in Multistory Structure: An Artificial Neural Network Model," Proceedings of the First International Conference on Construction in the 21st Century - Challenges and Opportunities in Management and Technology, Miami, Florida.

Haque, M.E., and Sudhakar, K.V. (2001a). Prediction of Corrosion-Fatigue behavior of DP Steel through Artificial Neural Network, International Journal of Fatigue, Vol. 23 (1), pp. 1-4.

Haque, M.E., and Sudhakar, K.V. (2001b). ANN-based Prediction Model for Fatigue Crack Growth in DP Steel. International Journal of Fatigue & Fracture of Engineering Materials and Structures, Vol. 24 (1), pp. 63-68.

Obermeier, K. and Barron, J. (1989). Time to Get Fried Up, *BYTE*, 14(8), 227-233.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy*, 82:1, 34-55.

Rumelhart, D., Hinton, G., and Williams, R. (1986). Parallel distributed processing, Cambridge, MA: MIT Press.

Sharkawy, A. (1994). A Physical-Financial Model for Design Economy Trade-Offs, *Real Estate Research Issues: Essays in Honor of James A. Graaskamp*, Delisle, J., Ed., American Real Estate Society, Madison, Wisconsin.

Sirpal, R. (1994). Empirical Modeling of the Relative Impacts of Various Sizes of Shopping Centers on the Values of Surrounding Residential Properties, *Journal of Real Estate Research*, 9:4, 487-505.

Taibah, A. (2002). The Impact of Amenities on Residential Property Value: A Hybrid Geoinformatic-Hedonic Approach, Dissertation, Department of Landscape Architecture and Urban Planning, Texas A&M University.

Ward Systems Group, NeuroShell 2 User's Manual. Ward Systems Group, Inc., Maryland, USA, 1996.