Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.20932

Comparative Study of House Valuation Models Based on Traditional and Modern Methods

Yunxin Guo[1,a,*]

[1]*College of Economic and Management, Fuzhou University, Fuzhou, Fujian, 350102, China*
a. 2535663709@qq.com
**corresponding author*

Abstract: In the context of the swift advancement of the economy and the hastening pace of urbanization, the real estate market has assumed an increasingly pivotal role within the national economy. This study reviews the current status and trends of house valuation methods, analyzes the advantages and disadvantages of traditional and modern methods, and looks ahead to future research directions. The study found that traditional methods are simple to operate but have significant limitations, while modern methods have advantages in handling complex data and capturing market dynamic changes but also have limitations in explanatory power. Hybrid models combine the advantages of traditional and modern methods, improving prediction accuracy and model interpretability.

Keywords: House Appraisal, Valuation Methods, Influencing Factors, Real Estate Valuation

1. Introduction

Real estate, as an important asset form, its price fluctuations have significant impacts on economic operation, residents' lives, and social stability. Therefore, precisely gauging the value of real estate holdings is of utmost significance to all stakeholders in the market. Traditional appraisal techniques, such as the comparative approach, cost method, and profit method, although widely used in practice, often rely on the appraiser's subjective judgment and thus have certain subjectivity and limitations. In tandem with the burgeoning development of information technology and data science, model-based appraisal methods have emerged as a vibrant area of research focus. These methods build mathematical models and use large data sets to analyze the factors that affect real estate value, with the aim of improving the accuracy and objectivity of appraisal. This particular study employs a literature review and comparative analysis approach. First, through a literature review, the existing research findings on real estate appraisal methods are systematically summarized to provide a theoretical foundation for subsequent analysis and discussion. Second, through comparative analysis, the principles, applicable scope, advantages, and limitations of different appraisal methods are deeply explored. Therefore, conducting in-depth research on real estate appraisal methods, discussing the advantages and disadvantages, applicable scope, and improvement directions of various methods, is of great significance for improving appraisal quality and promoting the healthy development of the real estate market.

DOI: 10.54254/2755-2721/2025.20932

2. Theoretical Foundation of House Valuation

2.1. Concept and Connotation of House Valuation

House valuation refers to the scientific and reasonable evaluation of the market value of real estate, which is an important link in the real estate market. Its core goal is to provide reliable value judgments for transaction, financing, and insurance decisions. House valuation not only involves the physical characteristics of the house itself but also includes the surrounding environment, market supply and demand, and legal and policy factors.

2.2. Basic Principles of House Valuation

When undertaking the valuation of a residential property, several cardinal principles must be meticulously adhered to. Firstly, the principle of fairness mandates that the valuation exercise be conducted in an objective and unbiased manner. As stated by the Appraisal Institute, throughout the valuation, one must always maintain independence and objectivity to prevent any conflict of interest from biasing the judgment, ensuring that the result reflects the actual market situation rather than personal opinions [1]. Secondly, the principle of suitability accentuates the importance of judiciously selecting the most appropriate valuation methodology, contingent upon the specific idiosyncrasies of the market conditions and the property in question. Crone T M and Voith R P noted that different types of real estate may demand different methods to ascertain their market value, and the method can be adjusted according to the property's use, location, etc [2]. The principle further stipulates that the basic data and market intelligence harnessed during the house valuation process must be current and up-to-date. Brown and Matysiak emphasized that in rapidly changing market conditions, outdated data can lead to significant estimation errors, so the valuation model needs to be regularly updated with market data and related information [3].

3. Main Methods of House Valuation

3.1. Traditional Econometric Methods

3.1.1. Sales Comparison Approach

The traditional method uses the Sales Comparison Approach as the most widely used approach. The value of the assessed property is assumed to be closely related to the sales price of similar properties in the same market area [4]. The appraiser first selects several similar properties from all recently sold properties and compares them. Since no two properties are exactly the same, the appraiser must adjust the sales price of each comparable property to take into account the differences between the target property and the comparable property in terms of size, age, construction quality, sale date, and surrounding community. The appraiser calculates the distance by weighting the differences in the comparable property's characteristics. The distance D is calculated using the following formula [4]:

$$D = \sqrt{\sum_{i=1}^{n}[A_i(X_i - X_{si}]^{\lambda} + \sum_{j=1}^{m}\left[A_j\delta\left(X_j, X_sj\right)\right]^{\lambda}} \, , \tag{3.1.1.1}$$

3.1.2. Profit Approach

The market value of the property is determined by analyzing the potential cash flow generated by the property. This method does not consider the initial purchase motivation of the buyer but rather the

DOI: 10.54254/2755-2721/2025.20932

potential economic benefits of the property as a production unit. The appraiser evaluates the expected annual income, deducts operating costs, including direct expenses and the remuneration of the operator, interest, and capital return, to calculate the economic rent. Ultimately, the capital value of the property is estimated by multiplying the annual rent by an appropriate capitalization rate, especially suitable for properties closely related to businesses such as hotels and restaurants.

3.1.3. Hedonic Price Model

The Hedonic Price Model postulates that goods, with houses being a prime example, can be considered as a collection of separate components or attributes. It is based on the premise that consumers engage in the purchase of goods with the objective of maximizing their utility function. This model emphasizes that price changes are due to the differences in the goods' characteristics, which are reflected in the market as implicit prices. Typically, the model usually includes some attributes that determine the price of the house, such as land area, age of the house, number of bedrooms, number of bathrooms, and surrounding facilities. The combination of these features determines the final price. As a result, the model can offer a quantitative analysis of the precise impact that each specific feature has on the house price. For instance, it can accurately gauge the effect of adding an extra bedroom or an additional bathroom. A simple hedonic model can be expressed as follows Table 1:

$$\mathrm{PRICE} = \mathrm{f}(\mathrm{LAND}, \mathrm{AGE}, \mathrm{TYPE}, \mathrm{BEDROOMS}, \mathrm{BATHROOMS}, \mathrm{GARAGES}, \mathrm{AMENITIES}, \mathrm{LOCATION}, \epsilon)$$

Table 1: Variable definition

Variable	Definition
PRICE	Price of the house
LAND (+)	Land size (in square meters)
AGE (-)	Age of the house (in years)
TYPE (+)	Type of the house; 1 if the house has a garden, 0 otherwise
BEDROOMS (+)	Number of bedrooms
BATHROOMS (+)	Number of bathrooms
GARAGES (+)	Number of garages
AMENITIES (+)	Amenities around the house; 1 if the house is close to two or more public facilities (e.g., bus stop, school, public park), 0 otherwise
LOCATION	Geographical location; Representation of different geographical locations by dummy variables (e.g. inner city, north, south, etc.)
ε	Error term

3.2. Machine Learning Algorithms

3.2.1. Random Forest (RF)

Random Forest represents an ensemble learning methodology based on decision trees. Its central tenet revolves around amalgamating the predictions of multiple weak learners, specifically decision trees, to enhance both the accuracy and robustness of forecasts. Given that a solitary decision tree is highly susceptible to data perturbations, which can precipitate instability and high variance issues, Random Forest circumvents these limitations by randomly selecting features and training samples to fabricate multiple autonomous decision trees.

DOI: 10.54254/2755-2721/2025.20932

The training process of the Random Forest model is as follows:

Firstly, a subset of features is randomly culled from the original feature assemblage, and concurrently, a subset of samples is randomly cherry-picked for the purpose of training a decision tree. Secondly, the aforementioned steps are reiterated to erect multiple decision trees, thereby assembling what can be metaphorically described as a “forest”. Finally, each decision tree proffers a prediction for the identical sample. Subsequently, the predictions proffered by all decision trees are either averaged or subjected to a voting mechanism, ultimately yielding the prediction outcome of the Random Forest model. This ensemble learning stratagem can proficiently seize the nonlinear relationships and intricate patterns latent within the data, thus augmenting the model’s generalization prowess. In the case of house price evaluation, the advantage of the Random Forest model lies in its ability to integrate the opinions of multiple "experts", i.e., multiple decision trees. Each decision tree makes an independent prediction of the house price based on different features such as house quality, location, and supporting facilities. Since each decision tree uses only a subset of features, it can avoid overfitting problems and reduce the bias of a single decision tree. Ultimately, by combining the predictions of all decision trees, the Random Forest model can provide a more accurate and robust house price evaluation. For example, Wang and Wu used the Random Forest model to evaluate house prices in Arlington County, Virginia, and compared it with the linear regression model [5]. The results show that the Random Forest model can better capture the nonlinear relationship between house prices and features, and provide more accurate estimation results.

3.2.2. Gradient Boosting Machines (GBM)

Gradient boosting constitutes a sophisticated machine learning technique that amalgamates multiple weak prediction models, typically decision trees, to construct a potent prediction model [6]. The core idea is to transform a feature or weak learner, which has a weak prediction on the target variable (Y) into a strong learner. In gradient boosting, the weak learners in each regression tree are first identified, and then equal weights are assigned to each observation while focusing on the fitting error between the weak learner's prediction and the actual value. The gradient boosting model uses the error rate to calculate the gradient of the error function and uses this gradient to adjust the model parameters to reduce the error at each iteration. This process is repeated m times as specified. The gradient boosting method assumes that Y is a real-valued variable and tries to approximate Y as the weighted sum of weak learner functions. The model update follows a specific mathematical formula, including the base learner function, the gradient of the loss function, and the step size, among other concepts. Although the optimal function h for each step cannot be specified in advance, researchers can solve this minimization problem using the steepest descent method. During the model training process, the model performance can be optimized by adjusting the learning rate, tree depth, and iteration number.

3.2.3. Support Vector Machines (SVMs)

SVM is a powerful machine learning algorithm that especially excels in classification problems. Its core idea is to maximize the distance between different categories by finding the optimal hyperplane, thus achieving effective data segmentation. The advantage of this method is that no complex preprocessing is required, and it has strong generalization ability for small sample cases. In the context of house price evaluation, SVM operates by first subjecting the housing feature data to preprocessing and standardization. Subsequently, the support vector regression (SVR) technique is employed to train the model [7]. Optimization of the model is used to predict new house data, and the estimated house price is obtained. Eventually, the trained model is deployed to predict new housing data, yielding the estimated house price. Nevertheless, SVM also has some limitations. Firstly, selecting the appropriate kernel function, regularization, margin, etc., and hyperparameters requires

DOI: 10.54254/2755-2721/2025.20932

certain experience and skills, which often depend on trial-and-error methods and cross-validation to determine the optimal combination. Secondly, the computational complexity of SVM is higher, especially when dealing with large data sets, its efficiency may be affected. And when facing highly correlated variables, SVM's performance may not be as good as other methods, because it assumes that features are independent [8]. However, in contemporary times, numerous house pricing models have embraced hybrid algorithms. Take, for instance, the dynamic pricing procedure put forward by Ozoğur Akyuz et al. in 2023 [9]. This innovative hybrid algorithm amalgamates linear regression, clustering analysis, nearest neighbor classification, and the Support Vector Regression (SVR) method, achieving outcomes that outperform those of single machine learning algorithms [8]. Such hybrid approaches hold great promise in further enhancing the accuracy and reliability of house price predictions.

4. Comparisons of Advantages and Disadvantages of Estimation Methods

4.1. Comparison of Traditional Methods and Machine Learning Methods

Traditional house pricing methods, such as the Hedonic Price Model (HPM) and sales comparison approach, usually have a slower response to market dynamics. This is primarily because they hinge on historical data and static characteristics, rendering them incapable of promptly seizing the fluctuations and transformations within the market. Additionally, traditional methods require manual intervention to adjust model parameters or redefine features when facing structural changes in the market, thus limiting their adaptability to market changes and making them less flexible than machine learning methods. Machine learning methods, such as Support Vector Machines (SVMs), k-Nearest Neighbors (KNN), Random Forest (RF), and the XGBoost(an optimized distributed GBM) algorithm, show higher flexibility and accuracy in handling large and complex data [10]. These methods can capture non-linear relationships in the data and provide more accurate predictions. Choy L H T et al. studied using 24,936 house transaction records and found that machine learning methods had a significantly better explanation power and error minimization than traditional statistical techniques. In Sakri S et al.'s study, the prediction accuracy of the XGBoost algorithm combined with feature regression pricing reached 84.1%, while the accuracy of the traditional feature regression algorithm was only 42% [11].Research shows that machine learning methods may be more advantageous in handling complex house pricing problems.

4.2. Comparison of Modern Methods

Machine learning methods can handle large, multi-dimensional, and heterogeneous data without strict statistical assumptions and provide accurate predictions by learning complex relationships between input data. They can adapt to market changes by updating model parameters to reflect the latest market conditions. Random forest models show lower RMSE values than linear regression models, indicating their insufficiency in modeling complex factors affecting house prices. GBM models achieve good model fitting, slightly better than RF based on MSE, RMSE, and MAPE criteria [12]. R^2 values for deep learning and random forest models are often above 0.8, providing good explanations for house price fluctuations. In contrast, R^2 values for linear regression are often below 0.5, especially when dealing with complex multidimensional features [13]. Xu and Li showed that a model built using multiple machine learning algorithms (including LightGBM and XGBoost) performed significantly better in house price evaluation than traditional linear regression models or single machine learning models [14]. GBM models have been proven superior to traditional single models or simple ensemble models in house price prediction. An intelligent evaluation method by integrating LightGBM and Bayesian optimization strategy achieved a 3.15% improvement in evaluation accuracy compared to traditional single models [15].

DOI: 10.54254/2755-2721/2025.20932

5. Conclusion

This study conducts an in-depth exploration into the current state and future trends of house valuation methods, analyzing the advantages and disadvantages of traditional and modern methods. Traditional methods, such as the sales comparison method, profit method, and hedonic model, still occupy an important position in practical applications, as they are easy to operate and understand, but also have certain limitations, such as strong data dependence and difficulty in capturing market dynamic changes. Machine learning algorithms and fuzzy logic, etc. modern methods have advantages in handling complex data and capturing market dynamic changes, and can provide more accurate prediction results. However, modern methods also have some limitations, such as poor model interpretability and complex parameter selection. Current hybrid methods that combine traditional and modern methods can leverage the strengths of both methods to improve prediction accuracy and model interpretability. In this study, there are limitations in data source and sample range when comparing and analyzing various modern housing valuation models, which may lead to bias in model evaluation results. Specifically, the data sets used by different models are different in terms of source, range and sample size, which may affect the accuracy of model performance comparison and lead to errors in the evaluation of model advantages and disadvantages. Future research endeavors will focus on using identical or closely similar datasets when conducting comparative analyses of different models. Additionally, ensuring an adequately large sample size will be crucial in mitigating the impact of data discrepancies on model evaluation outcomes. On the data front, improving data quality and expanding data sources, such as introducing GIS data and social media data, can better reflect market changes and house characteristics. Furthermore, researching how to improve the model's generalization ability to adapt to different regions and types of real estate valuation needs, and developing more comprehensive model evaluation methods to accurately evaluate the model's predictive performance.

References

[1] *American Institute of Real Estate Appraisers. The Appraisal of Real Estate[M]. The Institute, 1952.*

[2] *Crone T M, Voith R P. Estimating house price appreciation: a comparison of methods[J]. Journal of Housing Economics, 1992, 2(4): 324-338.*

[3] *Brown G R, Matysiak G A. Real estate investment: A capital market approach[J]. 2000.*

[4] *McCluskey, William and Borst, Richard A. (1997)"An evaluation of MRA, comparable sales analysis, and artificial neural networks (ANNs) for the mass appraisal of residential properties in Northern Ireland,"Assessment Journal: Vol. 4: No. 1, Article 5.*

[5] *Wang, C., & Wu, H. (2018). A new machine learning approach to house price estimation. New Trends in Mathematical Sciences, 6(4), 165-171. https://doi.org/10.20852/ntmsci.2018.327*

[6] *Limsombunchai, V., Gan, C., & Lee, M. (2004). House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. American Journal of Applied Sciences, 1(3), 193-201. ISSN 1546-9239. © Science Publications, 2004.*

[7] *Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X*

[8] *Mukhlishin M F, Saputra R, Wibowo A. Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor[C]//2017 1st International Conference on Informatics and Computational Sciences (ICICoS). IEEE, 2017: 171-176.*

[9] *Özöğür Akyüz, S., Eygi Erdogan, B., Yıldız, Ö. et al. A Novel Hybrid House Price Prediction Model. Comput Econ 62, 1215–1232 (2023). https://doi.org/10.1007/s10614-022-10298-8*

[10] *Choy L H T, Ho W K O. The use of machine learning in real estate research[J]. Land, 2023, 12(4): 740.*

[11] *Sakri S, Ali Z, Ismail N H A. Assessment of the Hybrid Deep Learning Models and Hedonic Pricing Model for House Price Prediction[C]//2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU). IEEE, 2024: 127-133.*

[12] *Ho, W. K. O., Tang, B.-S., & Wong, S. W. (2020). Predicting property prices with machine learning algorithms. Journal of Property Research, 37(3), 243-257. https://doi.org/10.1080/09599916.2020.1832558*

DOI: 10.54254/2755-2721/2025.20932

[13] Mim, A., & Li, Z. (2020). A comparative study on the performance of different machine learning algorithms for housing price prediction. Journal of Property Research.

[14] Xu, L., & Li, Z. (2020). A New Appraisal Model of Second-Hand Housing Prices in China's First-Tier Cities Based on Machine Learning Algorithms. Computational Economics. https://doi.org/10.1007/s10614-020-09973-5

[15] GU T, XU G L, LI W L, et al. Intelligent house price evaluation model based on ensemble LightGBM and Bayesian optimization strategy[J]. Journal of Computer Applications, 2020, 40(9): 2762-2767(in Chinese). https://www.cnki.com.cn/Article/CJFDTOTAL-JSJY202009043