Intelligent Systems with Applications 12 (2021) 200052

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.elsevier.com/locate/iswa

# House price forecasting with neural networks

Xiaojie Xu*, Yun Zhang

*North Carolina State University, Raleigh, NC 27695, USA*

ARTICLE INFO

*Article history:*
Received 5 May 2021
Revised 23 August 2021
Accepted 12 September 2021
Available online 20 September 2021

*Keywords:*
House price
Neural network
Forecasting

ABSTRACT

The house market has been rapidly growing for the past decade in China, making price forecasting an important issue to the people and policy makers. We approach this problem by exploring neural networks for forecasting of house prices from one hundred major cities for the period of June 2010–May 2019, serving as the first study with such wide coverage for the emerging Chinese market through a machine learning technique. We aim at constructing simple and accurate neural networks as a contribution to pure technical forecasting of house prices. To facilitate the analysis, we investigate different model settings over the algorithm (the Levenberg-Marquardt, scaled conjugate gradient, and Bayesian regularization), delay (from two to six), hidden neuron (two, three, five, and eight), and data spitting ratio (70%–15%–15%, 60%–20%–20%, and 80%–10%–10% for trainingvalidationtesting), and arrive at a rather simple neural network with only four delays and three hidden neurons that leads to stable performance of 1% average relative root mean square error across the one hundred cities for the training, validation, and testing phases. We demonstrate the usefulness of the machine learning approach to the house price forecasting problem in the Chinese market. Our results could be used on a standalone basis or combined with fundamental forecasting in forming perspectives of house price trends and conducting policy analysis. Our empirical framework should not be difficult to deploy, which is an important consideration to many decision makers, and has potential to be generalized for house price forecasting of other cities in China.



## 1. Introduction

House markets have been rapidly growing for the past decade in China. Naturally, house prices have become the people and policy makers' key concern. Price trends are of extreme importance because they undoubtedly influence people's decisions on real estate investment as well as cities to reside and work. Therefore, house price forecasting has drawn much attention from different forecasters.

In econometrics, much effort has been devoted to accurate and robust forecasting of economic and financial time series. Basic models, such as the autoregressive, vector autoregressive, and vector error correction approaches, as well as their numerous variations, have been widely employed to produce forecasts for different uses and purposes that are summarized in Fig. 1 (e.g. Cabrera et al., 2011; Hyvärinen et al., 2010; Kawahara et al., 2011; Kim et al., 2007; Kouwenberg and Zwinkels, 2014; Liu and Wu, 2020; Milunovich, 2020; Shimizu et al., 2006; Webb et al., 2016; Wei and Cao, 2017; Xu, 2014a; 2015a; 2017a; 2017b; 2017c; 2018b; 2018e; 2019a; 2019c; 2020; Xu and Thurman, 2015a; Xu and Zhang, 2021b; Yang et al., 2018; Zohrabyan et al., 2008). Recently, machine learning models and algorithms, including the support vector regression, regression tree, random forecast, bagging, boosting, and ensemble learning, are found to be useful and attractive solutions to a wide variety of forecasting problems, including those related to house prices, which are summarized in Fig. 1 (e.g. Chen et al., 2017; Embaye et al., 2021; Fu, 2018; Gu et al., 2011; Ho et al., 2021; Huang, 2019; Li et al., 2009; 2020; Liu and Liu, 2019; Milunovich, 2020; Pai and Wang, 2020; Park and Bae, 2015; Plakandaras et al., 2015; Rafiei and Adeli, 2016; Rico-Juan and de La Paz, 2021; Shahhosseini et al., 2019; Wang et al., 2014; Wu et al., 2009; Xu and Li, 2021; Xu and Zhang, 2021c; Yan and Zong, 2020; Yu et al., 2018). For example, Karasu et al. (2017b) study the solar radiation prediction with the Gaussian process regression and find that the method leads to high accuracy. Karasu et al. (2017c) explore the nonlinear autoregressive neural network for the wind speed prediction and determine that the approach results in good performance. Karasu and Altan (2019) propose a high performance recognition model for solar radiation es-

* Corresponding author.
*E-mail addresses:* xxu6@ncsu.edu (X. Xu), yzhang43@ncsu.edu (Y. Zhang).

https://doi.org/10.1016/j.iswa.2021.200052
2667-3053/

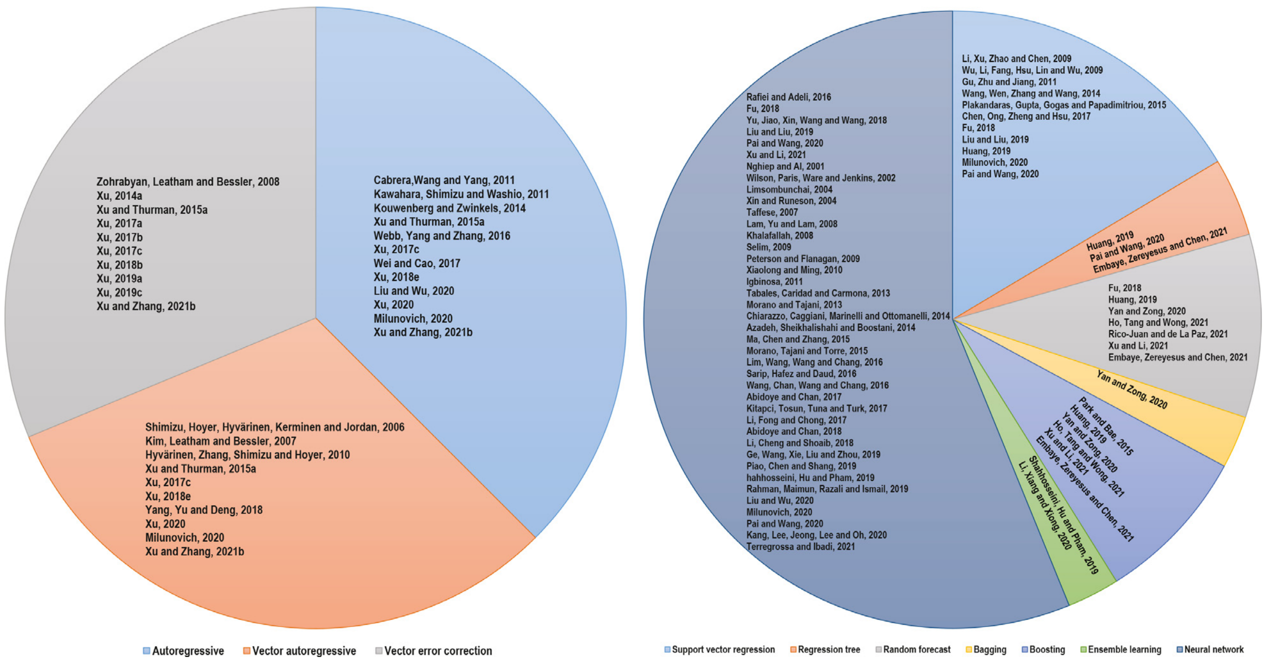


**Fig. 1.** Econometric time series forecasting (left) and machine learning house price time series forecasting (right).

timations based on the random forest with feature selection approach, which copes with nonlinear dynamics in time series. Altan et al. (2019) build a novel hybrid forecasting model based on the long short-term memory neural network and empirical wavelet transform decomposition along with the cuckoo search algorithm for the digital currency price time series prediction, which successfully captures nonlinear properties. Karasu et al. (2020) construct a novel forecasting model based on the support vector regression with the wrapper-based feature selection approach using the multi-objective particle swarm optimization and radial basis function based techniques for the crude oil price time series prediction, which successfully captures nonlinear properties and leads to good performance in terms of the precision and volatility. Altan et al. (2021) investigate a new hybrid model based on the long short-term memory network and decomposition methods with the grey wolf optimizer for the wind speed prediction and find that the model well captures nonlinear characteristics and leads to better performance as compared to single forecasting models. In particular, previous studies show that neural networks could have great potential for forecasting economic and financial time series that might be highly noised and chaotic (e.g. Wang and Yang, 2010; Wegener et al., 2016; Yang et al., 2010; 2008), including house prices summarized in Fig. 1 (e.g. Abidoye and Chan, 2017; 2018; Azadeh et al., 2014; Ćetković et al., 2018; Chiarazzo et al., 2014; Fu, 2018; Ge et al., 2019; Igbinosa, 2011; Kang et al., 2020; Khalafallah, 2008; Kitapci et al., 2017; Lam et al., 2008; Li et al., 2018; 2017; Lim et al., 2016; Limsombunchai, 2004; Liu and Wu, 2020; Ma et al., 2015; Milunovich, 2020; Morano and Tajani, 2013; Morano et al., 2015; Nghiep and Al, 2001; Pai and Wang, 2020; Peterson and Flanagan, 2009; Piao et al., 2019; Rahman et al., 2019; Sarip et al., 2016; Selim, 2009; Shahhosseini et al., 2019; Tabales et al., 2013; Taffese, 2007; Terregrossa and Ibadi, 2021; Wang et al., 2016; Wilson et al., 2002; Xiaolong and Ming, 2010; Xin and Runeson, 2004; Yasnitsky et al., 2021; Yu et al., 2018). The literature has shown that neural networks could lead to high accuracy under many different forecasting settings (Karasu et al., 2017a; 2017c; Wang and Yang, 2010; Wegener et al., 2016; Yang et al., 2010; 2008). This could benefit from neural networks' capability of self-learning for forecasting (Karasu et al., 2020) and capturing nonlinearities (Altan et al., 2021) often inhabiting in economic and financial time series. One of the greatest advantages of neural networks over other nonlinear models for time series is that a class of multilayer neural networks could approximate a large class of functions well (Yang et al., 2008).

From a geographical perspective, machine learning models are used for house price forecasting across many different countries and/or regions, including China (Fu, 2018; Ge et al., 2019; Gu et al., 2011; Ho et al., 2021; Lam et al., 2008; Li et al., 2009; 2018; 2020; Liu and Wu, 2020; Liu and Liu, 2019; Ma et al., 2015; Piao et al., 2019; Wang et al., 2014; Wei and Cao, 2017; Xin and Runeson, 2004; Xu and Li, 2021; Yan and Zong, 2020; Yu et al., 2018), Singapore (Lim et al., 2016; Wang et al., 2016), the United States (Ge et al., 2019; Huang, 2019; Khalafallah, 2008; Kouwenberg and Zwinkels, 2014; Nghiep and Al, 2001; Park and Bae, 2015; Peterson and Flanagan, 2009; Plakandaras et al., 2015; Shahhosseini et al., 2019), the United Kingdom (Wilson et al., 2002), Australia (Milunovich, 2020), Spain (Rico-Juan and de La Paz, 2021), Iran (Azadeh et al., 2014; Rafiei and Adeli, 2016), Tanzania (Embaye et al., 2021), Uganda (Embaye et al., 2021), Malawi (Embaye et al., 2021), New Zealand (Limsombunchai, 2004), Europe (Ćetković et al., 2018), Turkey (Kitapci et al., 2017; Selim, 2009; Terregrossa and Ibadi, 2021), Italy (Chiarazzo et al., 2014; Morano and Tajani, 2013; Morano et al., 2015), Nigeria (Abidoye and Chan, 2017; 2018; Igbinosa, 2011), Malaysia (Rahman et al., 2019; Sarip et al., 2016), Russia (Yasnitsky et al., 2021), and South Korea (Kang et al., 2020), which are summarized in Fig. 2. The wide adoption of different machine learning models reveals their potential as useful tools for house price forecasting. Most of the studies focus on one or a couple of locations. For China, Xin and Runeson (2004), Lam et al. (2008), Li et al. (2018), and Ho et al. (2021) concentrate on Hong Kong, Gu et al. (2011) on Tangshan, Wang et al. (2014) on Chongqing, Ma et al. (2015) and Fu (2018) on Shanghai, Yu et al. (2018), Ge et al. (2019), Yan and Zong (2020), and Li et al. (2020) on Beijing, Liu and Liu (2019) on Shenzhen, Piao et al. (2019) on Dalian, Liu and Wu (2020) on Kunming, Changchun, Xuzhou, and Handan, and Xu and Li (2021) on Beijing, Shanghai, Guangzhou, and Shenzhen. For the United States, Nghiep and Al (2001) concentrate on Rutherford County of Tennessee,

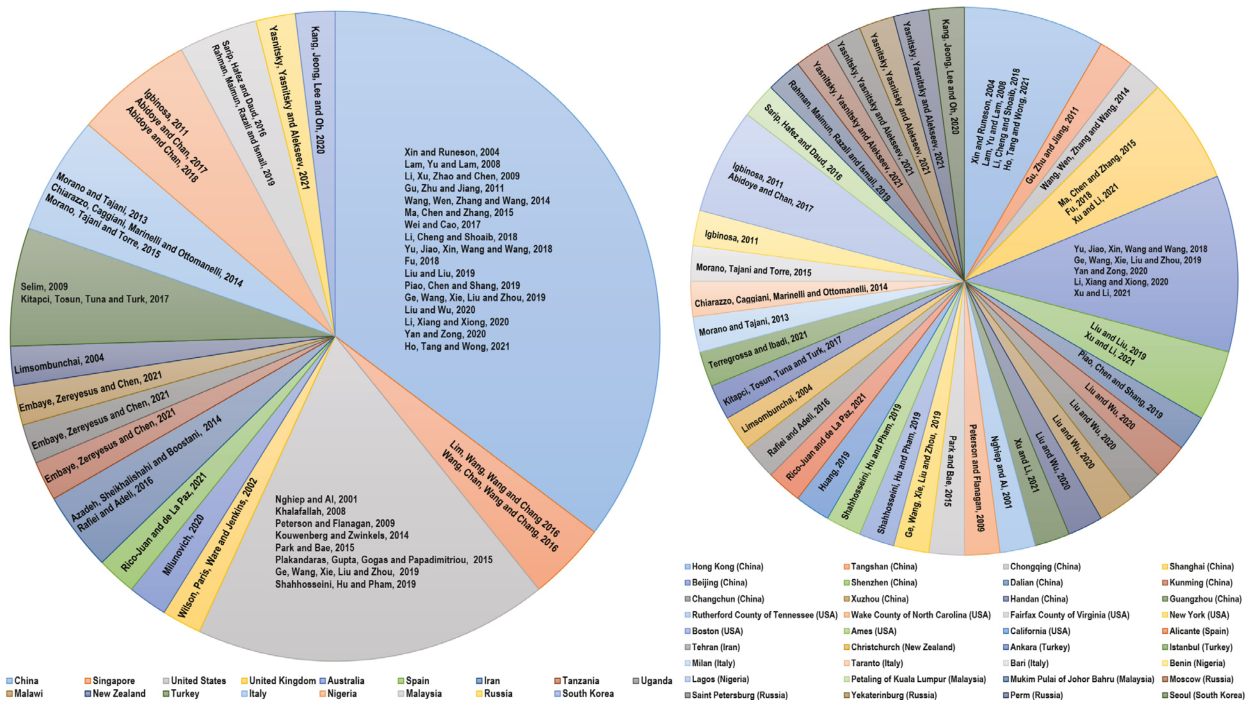


**Fig. 2.** Use of machine learning for house price forecasting across different countries (left) and different cities/counties/states (right).

Peterson and Flanagan (2009) on Wake County of North Carolina, Park and Bae (2015) on Fairfax County of Virginia, Ge et al. (2019) on New York City, Shahhosseini et al. (2019) on Boston and Ames, and Huang (2019) on California. For Spain, Rico-Juan and de La Paz (2021) concentrate on Alicante. For Iran, Rafiei and Adeli (2016) concentrate on Tehran. For New Zealand, Limsombunchai (2004) concentrates on Christchurch. For Turkey, Kitapci et al. (2017) concentrate on Ankara and Terregrossa and Ibadi (2021) on Istanbul. For Italy, Morano and Tajani (2013) concentrate on Milan, Chiarazzo et al. (2014) on Taranto, and Morano et al. (2015) on Bari. For Nigeria, Igbinosa (2011) concentrates on Benin and Lagos and Abidoye and Chan (2017, 2018) on Lagos. For Malaysia, Sarip et al. (2016) concentrate on Petaling of Kuala Lumpur and on Rahman et al. (2019) Mukim Pulai of Johor Bahru. For Russia, Yasnitsky et al. (2021) concentrate on Moscow, Saint Petersburg, Yekaterinburg, and Perm. For South Korea, Kang et al. (2020) concentrate on Seoul. These applications of machine learning for house price forecasting across different cities/counties/states are summarized in Fig. 2.

When researching machine learning models, previous studies differ in model building approaches and forecasting purposes, depending on their focuses. These could be constructing models from house related properties for valuations (Abidoye and Chan, 2017; 2018; Chiarazzo et al., 2014; Ho et al., 2021; Igbinosa, 2011; Kang et al., 2020; Morano and Tajani, 2013; Morano et al., 2015; Peterson and Flanagan, 2009; Rafiei and Adeli, 2016; Selim, 2009; Wu et al., 2009; Xu and Li, 2021), from macroeconomics for valuations (Kang et al., 2020; Rafiei and Adeli, 2016), from house prices themselves for technical forecasts (Ge et al., 2019; Gu et al., 2011; Li et al., 2009; 2020; Liu and Wu, 2020; Ma et al., 2015; Wang et al., 2014; Xiaolong and Ming, 2010; Xin and Runeson, 2004), from house related properties for technical forecasts (Chen et al., 2017; Embaye et al., 2021; Fu, 2018; Ge et al., 2019; Huang, 2019; Igbinosa, 2011; Kang et al., 2020; Khalafallah, 2008; Kitapci et al., 2017; Lam et al., 2008; Li et al., 2018; Limsombunchai, 2004; Liu and Liu, 2019; Morano and Tajani, 2013; Nghiep and Al, 2001; Pai and Wang, 2020; Park and Bae, 2015; Piao et al., 2019; Rahman et al., 2019; Rico-Juan and de La Paz, 2021; Sarip et al., 2016; Shahhosseini et al., 2019; Tabales et al., 2013; Terregrossa and Ibadi, 2021; Yan and Zong, 2020; Yasnitsky et al., 2021; Yu et al., 2018), and from macroeconomics for technical forecasts (Azadeh et al., 2014; Ćetković et al., 2018; Kang et al., 2020; Khalafallah, 2008; Lam et al., 2008; Li et al., 2009; 2018; Lim et al., 2016; Liu and Liu, 2019; Milunovich, 2020; Piao et al., 2019; Plakandaras et al., 2015; Rico-Juan and de La Paz, 2021; Tabales et al., 2013; Wang et al., 2016; Wei and Cao, 2017; Wilson et al., 2002; Xin and Runeson, 2004; Yasnitsky et al., 2021), which are summarized in Fig. 3. The literature reviewed here shows that some common house related properties include the house age, type, number of units, lot size, number of stories, baths, exterior composition, and location, and some common macroeconomic factors include the gross domestic product, gross national product, consumer price index, stock market index, interest rate, default rate, and unemployment.

Some studies focus on using one specific machine learning model for house price forecasting (Abidoye and Chan, 2017; Azadeh et al., 2014; Ćetković et al., 2018; Chen et al., 2017; Chiarazzo et al., 2014; Gu et al., 2011; Igbinosa, 2011; Kang et al., 2020; Khalafallah, 2008; Kitapci et al., 2017; Lam et al., 2008; Li et al., 2018; 2020; Ma et al., 2015; Morano et al., 2015; Piao et al., 2019; Rafiei and Adeli, 2016; Rahman et al., 2019; Shahhosseini et al., 2019; Wang et al., 2016; 2014; Wilson et al., 2002; Xiaolong and Ming, 2010; Xin and Runeson, 2004; Yasnitsky et al., 2021), some on comparing several different machine learning models (Embaye et al., 2021; Fu, 2018; Ge et al., 2019; Ho et al., 2021; Huang, 2019; Li et al., 2009; 2017; Liu and Liu, 2019; Pai and Wang, 2020; Park and Bae, 2015; Rico-Juan and de La Paz, 2021; Sarip et al., 2016; Wu et al., 2009; Xu and Li, 2021; Yan and Zong, 2020; Yu et al., 2018), some on comparing machine learning models with traditional or linear econometric models (Abidoye and Chan, 2018; Ge et al., 2019; Li et al., 2017; Lim et al., 2016; Limsombunchai, 2004; Liu and Wu, 2020; Milunovich, 2020; Morano and Tajani, 2013; Nghiep and Al, 2001; Peterson and Flanagan, 2009; Plakandaras et al., 2015; Selim, 2009; Tabales et al., 2013; Yu et al., 2018), and some on putting forward model combina-

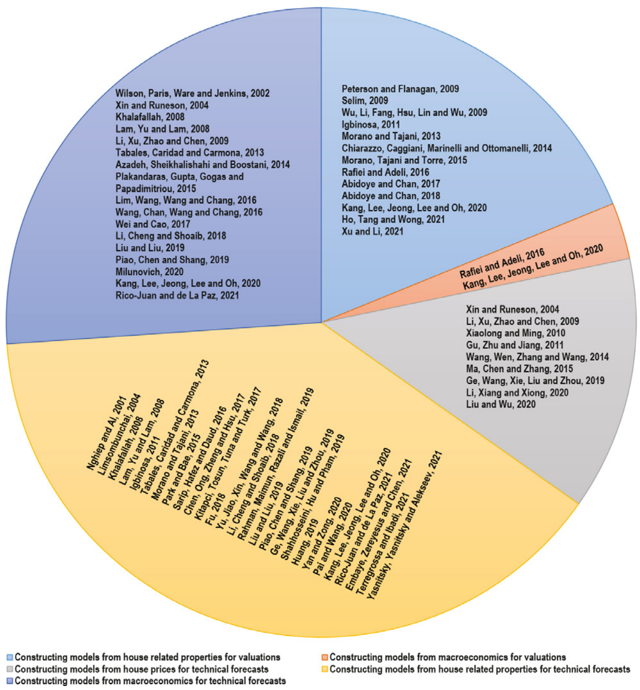


**Fig. 3.** Constructing machine learning models from different variables for different house price forecasting purposes.

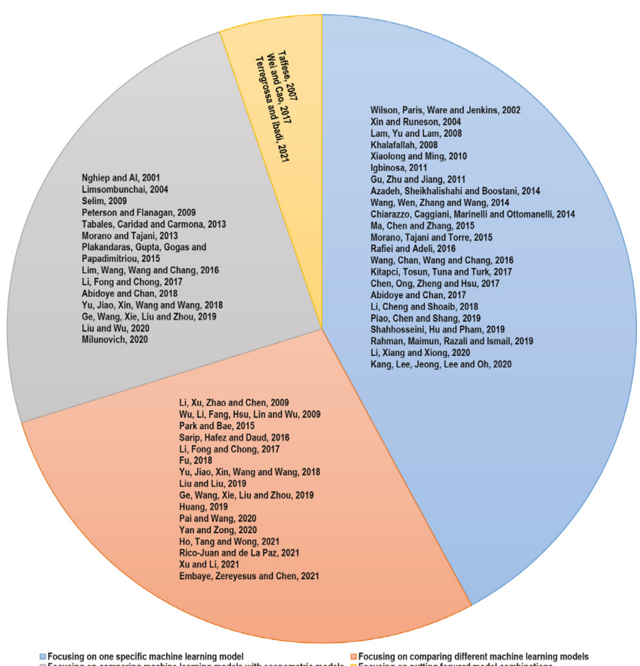


**Fig. 4.** Building machine learning house price forecasting models with different focuses.

tions (Taffese, 2007; Terregrossa and Ibadi, 2021; Wei and Cao, 2017). These studies are also summarized in Fig. 4. Specifically, Wilson et al. (2002), Xin and Runeson (2004), Lam et al. (2008), Khalafallah (2008), Xiaolong and Ming (2010), Azadeh et al. (2014), Chiarazzo et al. (2014), Ma et al. (2015), Morano et al. (2015), Rafiei and Adeli (2016), Wang et al. (2016), Kitapci et al. (2017), Abidoye and Chan (2017), Ćetković et al. (2018), Li et al. (2018), Piao et al. (2019), Rahman et al. (2019), Kang et al. (2020), Igbinosa (2011), and Yasnitsky et al. (2021) focus on the neural network, Gu et al. (2011), Chen et al. (2017), and Wang et al. (2014) on the support vector machine, and Shahhosseini et al. (2019) and Li et al. (2020) on the ensemble learners. Li et al. (2009) compare the support vector machine with back-propagation neural network and find that the former leads to better performance. Wu et al. (2009) compare the integrated hybrid genetic-based support vector regression, back-propagation neural network, and fuzzy neural network and find that the support vector regression leads to the best performance. Park and Bae (2015) compare the C4.5, RIPPER (Repeated Incremental Pruning to Produce Error Reduction), Naïve Bayesian, and AdaBoost (Adaptive Boosting) algorithms and find that RIPPER is optimal. Sarip et al. (2016) compare the neural network, fuzzy inference system, and fuzzy least-squares regression and find that the fuzzy least-squares regression leads to the best performance. Li et al. (2017) compare the GMDH (Group Method of Data Handling) neural network, single exponential smooth, double exponential smooth, autoregressive integrated moving average, and back-propagation neural network and find that the GMDH neural network is optimal. Yu et al. (2018) compare the convolution neural network and LSTM (Long Short-Term Memory) neural network and find that the latter beats the former. Liu and Liu (2019) find that the LSTM neural network is better than the back-propagation neural network, support vector regression, and evolution LTSM. Ge et al. (2019) propose the LSTM dense neural network that beats the support vector regression and vector autoregression. Huang (2019) compares the linear regression, decision tree, boosting, random forest, support vector machine and finds that the support vector machine results in the best performance. Yan and Zong (2020) find that XGBoost (eXtreme Gradient Boosting) beats the linear regression, random forest, Ridge, LASSO (least absolute shrinkage and selection operator), bagging, and boosting. Ho et al. (2021) find that the random forest and gradient boosting machine both beat the support vector machine. Rico-Juan and de La Paz (2021) find that the random forest beats the AdaBoost, CatBoost, decision tree, linear LASSO, linear Ridge, liner regression, nearest neighbors, and XGBRegressor. Embaye et al. (2021) find that the Ridge, LASSO, bagging, random forest, and boosting all beat the linear regression. As one might have expected, empirical evidence is mixed on choices of machine learning models for house price forecasting. For studies focusing on comparing machine learning models with traditional or linear econometric models, Nghiep and Al (2001), Selim (2009), Peterson and Flanagan (2009), Tabales et al. (2013), Morano and Tajani (2013), Abidoye and Chan (2018), and Limsombunchai (2004) find that the neural network beats the multiple linear regression. Plakandaras et al. (2015) propose a hybrid forecasting approach combining the EEMD (Ensemble Empirical Mode Decomposition) with the support vector regression, which beats the random walk, Bayesian autoregression, and Bayesian vector autoregression. Lim et al. (2016) find that the multilayer perceptron neural network beats the autoregressive integrated moving average. Milunovich (2020) explores traditional time series models, machine learning procedures, and deep learning neural networks, and finds mixed evidence. Some other studies propose model combinations to further improve forecasting. For example, Taffese (2007) presents the combination of the case-based reasoning and neural network. Wei and Cao (2017) introduce a dynamic model averaging method. Terregrossa and Ibadi (2021) proposes combining the multiple linear regression and neural network.

In terms of performance, machine learning models for house price forecasting show promising accuracy in the literature. Empirical evidence reveals a mean absolute percentage error of less than 1% (Ho et al., 2021; Liu and Wu, 2020; Liu and Liu, 2019; Ma et al., 2015; Pai and Wang, 2020; Shahhosseini et al., 2019), between 1% and 2% (Khalafallah, 2008; Li et al., 2009; Lim et al., 2016; Liu and Wu, 2020; Ma et al., 2015; Pai and Wang, 2020; Rico-Juan and de La Paz, 2021; Shahhosseini et al., 2019; Wilson et al., 2002; Xiaolong and Ming, 2010), between 2% and 3% (Lim et al., 2016; Liu and Liu, 2019; Pai and Wang, 2020; Plakandaras et al., 2015), between 3% and 4% (Liu and Wu, 2020; Rafiei and Adeli, 2016; Shahhosseini et al., 2019; Wang et al., 2014; Wilson et al., 2002), between 4% and 5% (Kang et al., 2020; Rahman et al., 2019; Wang et al., 2014; Wu et al., 2009), between 5% and 6% (Kang et al., 2020; Li et al., 2020; Liu and Wu, 2020; Plakandaras et al., 2015; Wu et al., 2009), between 6% and 7% (Kang et al., 2020; Liu and Wu, 2020), between 7% and 8% (Kang et al., 2020), and between 8% and 9% (Kang et al., 2020; Pai and Wang, 2020). In particular, the error below 1% is achieved through the neural network (Ma et al., 2015), random forecast (Ho et al., 2021; Shahhosseini et al., 2019), support vector machine (Ho et al., 2021; Pai and Wang, 2020; Shahhosseini et al., 2019), modified Holt's exponential smoothing with the whale optimization algorithm (Liu and Wu, 2020), and gradient boosting (Ho et al., 2021). The error between 1% and 2% is achieved through the neural network (Khalafallah, 2008; Lim et al., 2016; Ma et al., 2015; Wilson et al., 2002; Xiaolong and Ming, 2010), the support vector machine (Li et al., 2009; Pai and Wang, 2020), XGBoost (Shahhosseini et al., 2019), and modified Holt's exponential smoothing with the whale optimization algorithm (Liu and Wu, 2020). The error between 2% and 3% is achieved through the neural network (Liu and Liu, 2019), support vector machine (Liu and Liu, 2019; Plakandaras et al., 2015), and classification and regression tree (Pai and Wang, 2020). The error between 3% and 4% is achieved through the neural network (Shahhosseini et al., 2019), support vector machine (Wang et al., 2014), deep restricted Boltzmann machine (Rafiei and Adeli, 2016), and modified Holt's exponential smoothing with the whale optimization algorithm (Liu and Wu, 2020). The error between 4% and 5% is achieved through the neural network (Rahman et al., 2019), support vector regression (Wang et al., 2014; Wu et al., 2009), and genetic algorithm (Kang et al., 2020). The error between 5% and 6% is achieve through the neural network (Liu and Wu, 2020; Wu et al., 2009), support vector regression (Li et al., 2020; Plakandaras et al., 2015), and genetic algorithm (Kang et al., 2020). The error between 6% and 7% is achieve through the neural network (Liu and Wu, 2020) and genetic algorithm (Kang et al., 2020). The error between 7% and 8% is achieve through the genetic algorithm (Kang et al., 2020). The error between 8% and 9% is achieve through the neural network (Pai and Wang, 2020) and genetic algorithm (Kang et al., 2020). These studies are also summarized in Fig. 5. As one might have expected, accuracy that a machine learning approach could achieve for house price housing would depend on the specific problem being considered.

Continuing in this vein, we focus on neural networks for forecasting of house prices from one hundred major cities in China for June 2010–May 2019, a period during which the house market rapidly grows. Our study is the first one with such wide coverage for the emerging Chinese market through a machine learning technique for house price forecasting. The analysis expands knowledge of machine learning house price forecasting in the Chinese market as previous studies generally focus on one certain city (Fu, 2018; Ge et al., 2019; Ho et al., 2021; Lam et al., 2008; Li et al., 2018; 2020; Liu and Liu, 2019; Ma et al., 2015; Piao et al., 2019; Wang et al., 2014; Xin and Runeson, 2004; Yan and Zong, 2020; Yu et al., 2018). The analysis also sheds light on more recent evidence of usefulness of machine learning models for house price forecasting in China as previous studies examine time periods of 1981Q1–2002Q4 (Xin and Runeson, 2004), 2000Q1–2010Q3 (Wang et al., 2014), July 2013–December 2013 (Ma et al., 2015), January 2010–July 2017 (Yu et al., 2018), December 2010–October 2017 (Liu and Liu, 2019), January 2013–February 2017 (Piao et al., 2019), January 2011–December 2017 (Ge et al., 2019), January 2005–November 2018 (Li et al., 2020), and June 1996–August 2014 (Ho et al., 2021). Our goal is to construct simple and accurate neural networks as

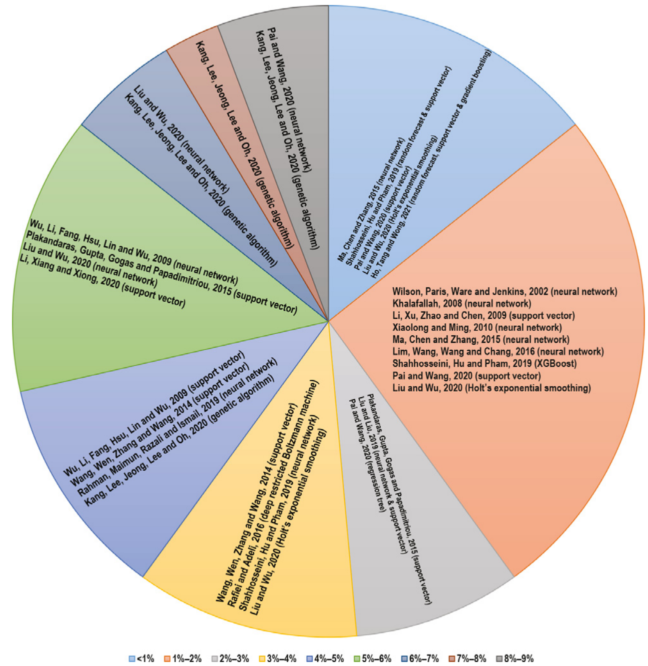


**Fig. 5.** Accuracy of different machine learning house price forecasting models based on the mean absolute percentage error.

a contribution to pure technical forecasting. To facilitate the analysis, we explore different model settings over the algorithm (the Levenberg-Marquardt, scaled conjugate gradient, and Bayesian regularization), delay (from two to six), hidden neuron (two, three, five, and eight), and data spitting ratio (70%–15%–15%, 60%–20%–20%, and 80%–10%–10% for trainingvalidationtesting), and arrive at a rather simple neural network with only four delays and three hidden neurons that leads to stable performance of 1% average relative root mean square error across the one hundred cities for the training, validation, and testing phases. Results here could be used on a standalone basis or combined with fundamental forecasting in forming perspectives of house price trends and conducting policy analysis. Our empirical framework should not be difficult to deploy[1], which is an important consideration to many decision makers, and has potential to be generalized for house price forecasting of other cities in China.

## 2. Data

We use data from the China Real Estate Index System (CREIS), an analytical platform designed to reflect market conditions and development trends of house markets in major Chinese cities. It was originated in 1994, which was initiated by the Development Research Center of the State Council, Real Estate Association, and National Real Estate Development Group Corporation. In 1995 and 2005, CREIS was audited by experts from the Development Research Center of the State Council, Ministry of Construction, Ministry of Land and Resources, Banking Regulatory Commission, Real Estate Association, and different universities. Currently, it periodically publishes different housing price indices that include the one hundred city index, city composite index, residential index, hedonic index, office building index, retail index, villa price index, second-hand housing sales index, and rental price index, and has

[1] We use MATALB 2019b to run our final model. The computer used has an Intel(R) Core(TM) i7-9700 CPU and a RAM size of 32GB( $= 4 \times 8$GB). Each RAM is DDR4 with the frequency of 2666MHz. The runtime of our final model is 21.888389 seconds for all cities.

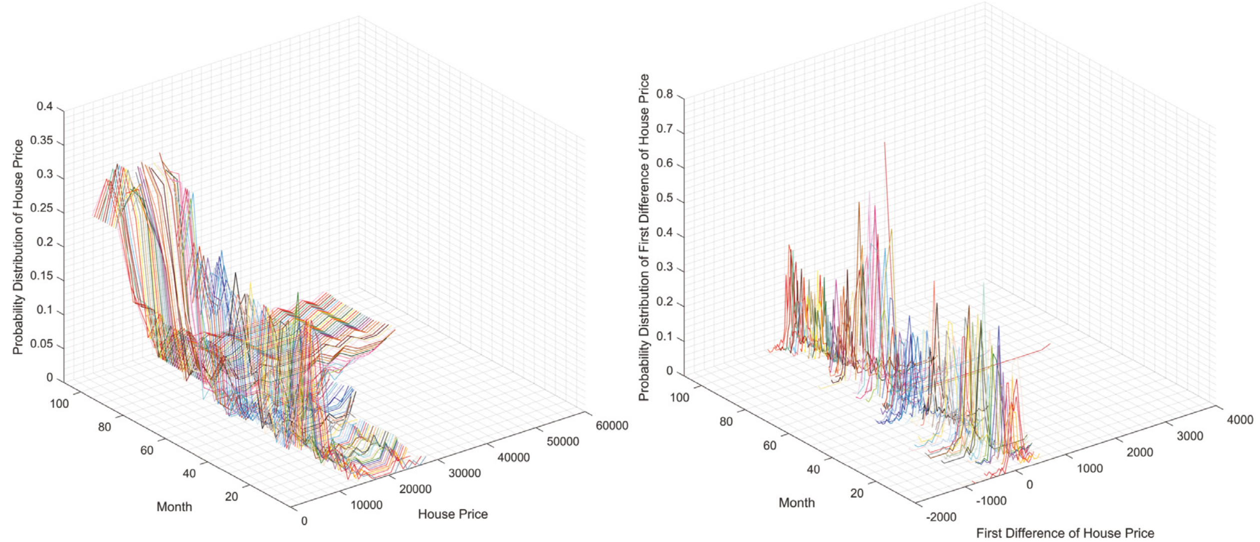


**Fig. 6.** Probability distribution plots of house prices (RMB/m$^2$) of ninety-nine major Chinese cities (left) and their first differences (right). Months are from June 2010 to May 2019.

become the system with the widest coverage in terms of house markets (Xu and Zhang, 2021e). We use the one hundred city index, which became available in CREIS in 2010.

For a given city, its house price index is calculated as: $P_j^t = \frac{\sum P_{ij}^t \times Q_{ij}}{\sum Q_{ij}}$, where $P_j^t$ represents the average house price in the $j$th city at time $t$, $P_{ij}^t$ represents the house price of the $i$th project in the $j$th city at time $t$, and $Q_{ij}$ represents the construction area of the $i$th project in the $j$th city.

The data range from June 2010 to May 2019. Probability distributions of house price series and their first differences, are plotted in Fig. 6. It is worth noting that there are actually ninety-nine, instead of one hundred, cities covered because the data for a city called "Yancheng" are no longer available after December 2016 (Xu and Zhang, 2021d). The ninety-nine cities covered are Sanya, Shanghai, Beijing, Xiamen, Shenzhen, Dongguan, Nanjing, Taizhou of Zhejiang province, Dalian, Tianjing, Ningbo, Guangzhou, Kunshan, Hangzhou, Haikou, Wenzhou, Zhuhai, Fuzhou, Suzhou, Foshan, Nanning, Nanchang, Nantong, Hefei, Jiaxing, Changzhou, Changshu, Langfang, Zhangjiagang, Chengdu, Yangzhou, Wuxi, Kunming, Wuhan, Shantou, Jinan, Huzhou, Shaoxing, Zhengzhou, Ordos, Jinhua, Qingdao, Zhongshan, Urumchi, Lanzhou, Hohhot, Harbin, Taiyuan, Weihai, Huizhou, Liuzhou, Jiangyin, Shenyang, Quanzhou, Taizhou of Jiangsu province, Zhanjiang, Yantai, Shijiangzhuang, Qinhuangdao, Xian, Chongqing, Zhenjiang, Changchun, Changsha, Dongying, Baoding, Baotou, Beihai, Jilin, Tangshan, Yichang, Xuzhou, Rizhao, Guilin, Jiangmen, Luoyang, Zibo, Huaian, Mianyang, Liaocheng, Wuhu, Xining, Guiyang, Ganzhou, Lianyungang, Handan, Yinchuan, Anshan, Maanshan, Baoji, Suqian, Dezhou, Xinxiang, Zhuzhou, Xiangtan, Weifang, Heze, Yingkou, and Hengshui.

## 3. Method

We utilize the nonlinear autoregressive (NAR) model for forecasting, which could be represented as $x_t = f(x_{t-1}, \ldots, x_{t-d})$, where $x$ is the house price of a certain city under consideration, $t$ is used to index time, $d$ is the number of delays, and $f$ is the function. The model is capable of self-learning for forecasting (Karasu et al., 2020) and capturing nonlinearities (Altan et al., 2021) often inhabiting in economic and financial time series. One of the greatest advantages of neural networks over other nonlinear models for time series is that a class of multilayer neural networks could approximate a large class of functions well (Wang and Yang, 2010; Yang et al., 2010; 2008). Note that $f$ is unknown in advance and the estimated model could be expressed as: $x_t = \alpha_0 + \sum_{j=1}^{k} \alpha_j \phi\left(\sum_{i=1}^{d} \beta_{ij} x_{t-i} + \beta_{0j}\right) + \varepsilon_t$, where $k$ is the number of hidden layers with the transfer function $\phi$, $\beta_{ij}$ is the parameter corresponding to the weight of the connection between the input unit $i$ and the hidden unit $j$, $\alpha_j$ is the weight of the connection between the hidden unit $j$ and the output unit, $\beta_{0j}$ and $\alpha_0$ are the constants corresponding, respectively, to the hidden unit $j$ and the output unit, and $\varepsilon$ is the error term. We focus on the short-term forecasting exercise of one-month ahead.

We use the NAR based on a two-layer feedforward network that has a logistic sigmoid transfer function in the form of $\phi(z) = \frac{1}{1+e^{-z}}$ in the hidden layer and a linear transfer function in the output layer. We note that the output $x(t)$ is fed back through delays to the input of the network and model training would be in the form of open loops for efficiency, in which the true output is used rather than feeding back the one estimated. Specifically, employing the open loop would ensure that the input to the feedforward network is more accurate and the resulting network would have an architecture that is pure feedforward. Readers can refer to Karasu et al. (2017a) and Karasu et al. (2017c) for a good representation of the NAR neural network model.

Our final forecasting models are based on three hidden neurons and four delays. We adopt the Levenberg-Marquardt (LM) algorithm (Levenberg, 1944; Marquardt, 1963) to estimate the models, and split the data randomly for training, validation, and testing based on the ratio of 70% vs. 15% vs. 15%.

The LM algorithm approximates the second-order training speed to avoid the expensive Hessian matrix ($H$) computation (Paluszek and Thomas, 2020). The approximation utilized is $H = J^T J$, where $J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}$ for a non-linear function $f(x_1, x_2)$ whose $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$. $g = J^T e$ represents the gradient, where $e$ denotes an error vector. The rule of $x_{k+1} = x_k - \left[J^T J + \mu I\right]^{-1} J^T e$ is employed to update weights and biases, where $I$ is the identity matrix. The algorithm would be similar to Newton's method when $\mu = 0$ and it would become gradient descent with small step sizes when $\mu$ is large. $\mu$ would be decreased because of less need for faster gradient descent after successful steps. The LM algorithm not only has good attributes of steepest-descent algorithms and Gauss-Newton techniques but also avoids many of their limitations. In particularly, it can deal with the slow convergence issue efficiently (Hagan and Menhaj, 1994).

There are many algorithms that could be used for model training. Here, we also take into consideration the scaled conjugate gradient (SCG) (Møller, 1993) and Bayesian regularization (BR) (Foresee and Hagan, 1997; MacKay, 1992) algorithms. The SCG and BR algorithms, as well as the LM algorithm, have been applied in many different varieties of areas (e.g. Doan and Liong, 2004; Kayri, 2016; Khan et al., 2019; Selvamuthu et al., 2019; Xu and Zhang, 2021a). Comparative studies of the three algorithms can be found from, e.g., Baghirli (2015) and Al Bataineh and Kaur (2018).

Backpropagation algorithms perform weight adjustments in the steepest descent as the performance function would decrease rapidly in the direction, which however, does not always represent the fastest convergence. Conjugate gradient algorithms conduct searches along the conjugate direction, which in general, lead to faster convergence as compared to the steepest descent. Most algorithms adopt the learning rate to determine the length of the updated weight step size. For conjugate gradient algorithms, the step size is modified in iterations. Thus, the search is performed along the conjugate gradient direction to decide the step size to reduce the performance function. Besides, to avoid the line search in conjugate gradient algorithms that is time consuming, the SCG algorithm can be taken into consideration, which is fully-automated and supervised and is faster as compared to the LM backpropagation.

The Bayesian regularized NN does not require a lengthy cross validation process. Bayesian regularization would convert a nonlinear regression to a statistical problem in the fashion of a ridge regression and the algorithm would consider the probabilistic nature of weights in a NN related to data. If more hidden layers of neurons are built in a NN, chances of overfitting would increase dramatically. For the BR algorithm, all models that are unreasonable complicated would be penalized through pushing extra linkage weights towards zeros and the NN would concentrate on training and calculating weights that are non-trivial. Certain parameters would converge to constants as the NN grows. The BR NNs tend to be more parsimonious as compared to basic backpropagation NNs and reduce chances of overfitting because of volatilities and noises in data.

Finally, in arriving at our final models aforementioned, we also consider different model settings over the delay, hidden neuron, and data spitting ratio in addition to the algorithm. To be more specific, we explore delays of two, three, four, five, and six, hidden neurons of two, three, five, and eight, and data spitting ratios of 60% vs. 20% vs. 20%, 70% vs. 15% vs. 15%, and 80% vs. 10% vs. 10% for training, validation, and testing. Table 1 lists all model settings investigated, where the setting #22 is used to build our final forecasting models.

## 4. Result

We run each model setting in Table 1 across house prices of the ninety-nine cities and calculate distribution statistics, i.e. the minimum, mean, median, maximum, and standard deviation, of the relative root mean square error (RRMSE) for training, validation, and testing based on each setting over the ninety-nine series. The results are shown in Fig. 7. Balancing model performance and its stability, we arrive at fifteen candidates shown in Table 2 with the mean RRMSE reported for training, validation, and testing. As the BR algorithm does not contain the validation result, house prices of some cities, i.e. those segmented into the validation set, are skipped. Considering also that the BR algorithm is much more time consuming than the LM or SCG algorithm and these fifteen candidates lead to close performance, we drop settings #3, #6, #33, #39, #51, #81, #111, #123, #126, #153, #168, and #171 from further consideration. The setting #31 is more complicated than the setting #22 as the former relies on more hidden neurons than the latter. Considering also that overall performance across the ninety-nine cities is slightly better by adopting the setting #22 than #31, we prefer the setting #22 over #31. The difference between the setting #22 and #142 is that the latter utilizes more data for training and fewer data for validation and testing than the former. In terms of performance, these two settings arrive at close results. Therefore, we prefer the setting #22 that leaves more data for validation and testing.

Having the setting #22 as our final choice, we analyze sensitivities of performance to different settings by changing one setting each time. The results are shown in Fig. 8, where distribution statistics, i.e. the minimum, mean, median, maximum, and standard deviation, of the RRMSE for training, validation, and testing based on each setting over the ninety-nine series are displayed. The comparison between the setting #22 and settings #23 and #24 tests the sensitivity to the algorithm, between the setting #22 and settings #16, #19, #25, and #28 tests the sensitivity to the delay, between the setting #22 and settings #7, #37, and #52 tests the sensitivity to the hidden neuron, and between the setting #22 and settings #82 and #142 tests the sensitivity to the data splitting ratio. These results support the setting #22 as the final choice, which leads to the average RRMSEs across cities of 0.98%, 1.01%, and 1.00% for the training, validation, and testing phases, respectively. We could also observe that other distributional statistics (the minimum, median, maximum, and standard deviation) concerning the RRMSE are also stable based on the setting #22 across the training, validation, and testing.

We show model performance (the RRMSE) across house prices of the ninety-nine cities based on the setting #22 in Fig. 9. We also calculate percentage errors across time for the ninety-nine cities and show probability distribution plots of percentage errors across time in Fig. 10. From Fig. 9, we can discover that the house price of Wenzhou is the most difficult to forecast. This is related to sharp house price increases in Wenzhou from 2010 to 2011 due to huge inflows of private capital into the local real estate market, followed by sharp decreases from 2012 to the third quarter of 2015 due to significantly strengthened regulations and mild increases since the fourth quarter of 2015 when the market rejoins the relatively normal development phase. The good news is that the model could still generate decent performance with the maximum RRMSE slightly below 2% for Wenzhou under the rapidly changing market and regulatory environment. On the other hand, we can observe that the model leads to better performance for Ningbo and Taizhou, Zhejiang, which are closely connected to Wenzhou in the economy. As the issue related to the private capital is less severe in Ningbo and Taizhou, Zhejiang, causing price fluctuations that are not as sharp as those for Wenzhou, one might have expected such better performance from the model. For the four most significant cities, Beijing, Shanghai, Guangzhou, and Shenzhen, the model generates good performance in the sense that the RRMSEs across the training, validation, and testing are all below 1.00% for each city. There are nearly seven hundred cities in China. If more data become available in the future, the neural network might find its potential to be further generalized to wider coverage of house price forecasting in China.

The nonlinearity in higher moments of financial and economic time series is well established in the literature (e.g. Altan et al., 2019; Karasu et al., 2020; Wang and Yang, 2010; Yang et al., 2010; 2008). One of the greatest advantages of neural networks over other nonlinear models for time series is that a class of multilayer

**Table 1**
Explored model settings.

| Setting | Algorithm | Delay | Hidden neuron | Training ratio | Validation ratio | Testing ratio | Setting | Algorithm | Delay | Hidden neuron | Training ratio | Validation ratio | Testing ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #1 | LM | 2 | 2 | 70% | 15% | 15% | #91 | LM | 2 | 5 | 60% | 20% | 20% |
| #2 | SCG | 2 | 2 | 70% | 15% | 15% | #92 | SCG | 2 | 5 | 60% | 20% | 20% |
| #3 | BR | 2 | 2 | 70% | 15% | 15% | #93 | BR | 2 | 5 | 60% | 20% | 20% |
| #4 | LM | 3 | 2 | 70% | 15% | 15% | #94 | LM | 3 | 5 | 60% | 20% | 20% |
| #5 | SCG | 3 | 2 | 70% | 15% | 15% | #95 | SCG | 3 | 5 | 60% | 20% | 20% |
| #6 | BR | 3 | 2 | 70% | 15% | 15% | #96 | BR | 3 | 5 | 60% | 20% | 20% |
| #7 | LM | 4 | 2 | 70% | 15% | 15% | #97 | LM | 4 | 5 | 60% | 20% | 20% |
| #8 | SCG | 4 | 2 | 70% | 15% | 15% | #98 | SCG | 4 | 5 | 60% | 20% | 20% |
| #9 | BR | 4 | 2 | 70% | 15% | 15% | #99 | BR | 4 | 5 | 60% | 20% | 20% |
| #10 | LM | 5 | 2 | 70% | 15% | 15% | #100 | LM | 5 | 5 | 60% | 20% | 20% |
| #11 | SCG | 5 | 2 | 70% | 15% | 15% | #101 | SCG | 5 | 5 | 60% | 20% | 20% |
| #12 | BR | 5 | 2 | 70% | 15% | 15% | #102 | BR | 5 | 5 | 60% | 20% | 20% |
| #13 | LM | 6 | 2 | 70% | 15% | 15% | #103 | LM | 6 | 5 | 60% | 20% | 20% |
| #14 | SCG | 6 | 2 | 70% | 15% | 15% | #104 | SCG | 6 | 5 | 60% | 20% | 20% |
| #15 | BR | 6 | 2 | 70% | 15% | 15% | #105 | BR | 6 | 5 | 60% | 20% | 20% |
| #16 | LM | 2 | 3 | 70% | 15% | 15% | #106 | LM | 2 | 8 | 60% | 20% | 20% |
| #17 | SCG | 2 | 3 | 70% | 15% | 15% | #107 | SCG | 2 | 8 | 60% | 20% | 20% |
| #18 | BR | 2 | 3 | 70% | 15% | 15% | #108 | BR | 2 | 8 | 60% | 20% | 20% |
| #19 | LM | 3 | 3 | 70% | 15% | 15% | #109 | LM | 3 | 8 | 60% | 20% | 20% |
| #20 | SCG | 3 | 3 | 70% | 15% | 15% | #110 | SCG | 3 | 8 | 60% | 20% | 20% |
| #21 | BR | 3 | 3 | 70% | 15% | 15% | #111 | BR | 3 | 8 | 60% | 20% | 20% |
| #22 | LM | 4 | 3 | 70% | 15% | 15% | #112 | LM | 4 | 8 | 60% | 20% | 20% |
| #23 | SCG | 4 | 3 | 70% | 15% | 15% | #113 | SCG | 4 | 8 | 60% | 20% | 20% |
| #24 | BR | 4 | 3 | 70% | 15% | 15% | #114 | BR | 4 | 8 | 60% | 20% | 20% |
| #25 | LM | 5 | 3 | 70% | 15% | 15% | #115 | LM | 5 | 8 | 60% | 20% | 20% |
| #26 | SCG | 5 | 3 | 70% | 15% | 15% | #116 | SCG | 5 | 8 | 60% | 20% | 20% |
| #27 | BR | 5 | 3 | 70% | 15% | 15% | #117 | BR | 5 | 8 | 60% | 20% | 20% |
| #28 | LM | 6 | 3 | 70% | 15% | 15% | #118 | LM | 6 | 8 | 60% | 20% | 20% |
| #29 | SCG | 6 | 3 | 70% | 15% | 15% | #119 | SCG | 6 | 8 | 60% | 20% | 20% |
| #30 | BR | 6 | 3 | 70% | 15% | 15% | #120 | BR | 6 | 8 | 60% | 20% | 20% |
| #31 | LM | 2 | 5 | 70% | 15% | 15% | #121 | LM | 2 | 2 | 80% | 10% | 10% |
| #32 | SCG | 2 | 5 | 70% | 15% | 15% | #122 | SCG | 2 | 2 | 80% | 10% | 10% |
| #33 | BR | 2 | 5 | 70% | 15% | 15% | #123 | BR | 2 | 2 | 80% | 10% | 10% |
| #34 | LM | 3 | 5 | 70% | 15% | 15% | #124 | LM | 3 | 2 | 80% | 10% | 10% |
| #35 | SCG | 3 | 5 | 70% | 15% | 15% | #125 | SCG | 3 | 2 | 80% | 10% | 10% |
| #36 | BR | 3 | 5 | 70% | 15% | 15% | #126 | BR | 3 | 2 | 80% | 10% | 10% |
| #37 | LM | 4 | 5 | 70% | 15% | 15% | #127 | LM | 4 | 2 | 80% | 10% | 10% |
| #38 | SCG | 4 | 5 | 70% | 15% | 15% | #128 | SCG | 4 | 2 | 80% | 10% | 10% |
| #39 | BR | 4 | 5 | 70% | 15% | 15% | #129 | BR | 4 | 2 | 80% | 10% | 10% |
| #40 | LM | 5 | 5 | 70% | 15% | 15% | #130 | LM | 5 | 2 | 80% | 10% | 10% |
| #41 | SCG | 5 | 5 | 70% | 15% | 15% | #131 | SCG | 5 | 2 | 80% | 10% | 10% |
| #42 | BR | 5 | 5 | 70% | 15% | 15% | #132 | BR | 5 | 2 | 80% | 10% | 10% |
| #43 | LM | 6 | 5 | 70% | 15% | 15% | #133 | LM | 6 | 2 | 80% | 10% | 10% |
| #44 | SCG | 6 | 5 | 70% | 15% | 15% | #134 | SCG | 6 | 2 | 80% | 10% | 10% |
| #45 | BR | 6 | 5 | 70% | 15% | 15% | #135 | BR | 6 | 2 | 80% | 10% | 10% |
| #46 | LM | 2 | 8 | 70% | 15% | 15% | #136 | LM | 2 | 3 | 80% | 10% | 10% |
| #47 | SCG | 2 | 8 | 70% | 15% | 15% | #137 | SCG | 2 | 3 | 80% | 10% | 10% |
| #48 | BR | 2 | 8 | 70% | 15% | 15% | #138 | BR | 2 | 3 | 80% | 10% | 10% |
| #49 | LM | 3 | 8 | 70% | 15% | 15% | #139 | LM | 3 | 3 | 80% | 10% | 10% |
| #50 | SCG | 3 | 8 | 70% | 15% | 15% | #140 | SCG | 3 | 3 | 80% | 10% | 10% |
| #51 | BR | 3 | 8 | 70% | 15% | 15% | #141 | BR | 3 | 3 | 80% | 10% | 10% |
| #52 | LM | 4 | 8 | 70% | 15% | 15% | #142 | LM | 4 | 3 | 80% | 10% | 10% |
| #53 | SCG | 4 | 8 | 70% | 15% | 15% | #143 | SCG | 4 | 3 | 80% | 10% | 10% |
| #54 | BR | 4 | 8 | 70% | 15% | 15% | #144 | BR | 4 | 3 | 80% | 10% | 10% |
| #55 | LM | 5 | 8 | 70% | 15% | 15% | #145 | LM | 5 | 3 | 80% | 10% | 10% |
| #56 | SCG | 5 | 8 | 70% | 15% | 15% | #146 | SCG | 5 | 3 | 80% | 10% | 10% |
| #57 | BR | 5 | 8 | 70% | 15% | 15% | #147 | BR | 5 | 3 | 80% | 10% | 10% |
| #58 | LM | 6 | 8 | 70% | 15% | 15% | #148 | LM | 6 | 3 | 80% | 10% | 10% |
| #59 | SCG | 6 | 8 | 70% | 15% | 15% | #149 | SCG | 6 | 3 | 80% | 10% | 10% |
| #60 | BR | 6 | 8 | 70% | 15% | 15% | #150 | BR | 6 | 3 | 80% | 10% | 10% |
| #61 | LM | 2 | 2 | 60% | 20% | 20% | #151 | LM | 2 | 5 | 80% | 10% | 10% |
| #62 | SCG | 2 | 2 | 60% | 20% | 20% | #152 | SCG | 2 | 5 | 80% | 10% | 10% |
| #63 | BR | 2 | 2 | 60% | 20% | 20% | #153 | BR | 2 | 5 | 80% | 10% | 10% |
| #64 | LM | 3 | 2 | 60% | 20% | 20% | #154 | LM | 3 | 5 | 80% | 10% | 10% |
| #65 | SCG | 3 | 2 | 60% | 20% | 20% | #155 | SCG | 3 | 5 | 80% | 10% | 10% |
| #66 | BR | 3 | 2 | 60% | 20% | 20% | #156 | BR | 3 | 5 | 80% | 10% | 10% |
| #67 | LM | 4 | 2 | 60% | 20% | 20% | #157 | LM | 4 | 5 | 80% | 10% | 10% |
| #68 | SCG | 4 | 2 | 60% | 20% | 20% | #158 | SCG | 4 | 5 | 80% | 10% | 10% |
| #69 | BR | 4 | 2 | 60% | 20% | 20% | #159 | BR | 4 | 5 | 80% | 10% | 10% |
| #70 | LM | 5 | 2 | 60% | 20% | 20% | #160 | LM | 5 | 5 | 80% | 10% | 10% |
| #71 | SCG | 5 | 2 | 60% | 20% | 20% | #161 | SCG | 5 | 5 | 80% | 10% | 10% |
| #72 | BR | 5 | 2 | 60% | 20% | 20% | #162 | BR | 5 | 5 | 80% | 10% | 10% |
| #73 | LM | 6 | 2 | 60% | 20% | 20% | #163 | LM | 6 | 5 | 80% | 10% | 10% |

(*continued on next page*)

**Table 1** (*continued*)

| Setting | Algorithm | Delay | Hidden neuron | Training ratio | Validation ratio | Testing ratio | Setting | Algorithm | Delay | Hidden neuron | Training ratio | Validation ratio | Testing ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #74 | SCG | 6 | 2 | 60% | 20% | 20% | #164 | SCG | 6 | 5 | 80% | 10% | 10% |
| #75 | BR | 6 | 2 | 60% | 20% | 20% | #165 | BR | 6 | 5 | 80% | 10% | 10% |
| #76 | LM | 2 | 3 | 60% | 20% | 20% | #166 | LM | 2 | 8 | 80% | 10% | 10% |
| #77 | SCG | 2 | 3 | 60% | 20% | 20% | #167 | SCG | 2 | 8 | 80% | 10% | 10% |
| #78 | BR | 2 | 3 | 60% | 20% | 20% | #168 | BR | 2 | 8 | 80% | 10% | 10% |
| #79 | LM | 3 | 3 | 60% | 20% | 20% | #169 | LM | 3 | 8 | 80% | 10% | 10% |
| #80 | SCG | 3 | 3 | 60% | 20% | 20% | #170 | SCG | 3 | 8 | 80% | 10% | 10% |
| #81 | BR | 3 | 3 | 60% | 20% | 20% | #171 | BR | 3 | 8 | 80% | 10% | 10% |
| #82 | LM | 4 | 3 | 60% | 20% | 20% | #172 | LM | 4 | 8 | 80% | 10% | 10% |
| #83 | SCG | 4 | 3 | 60% | 20% | 20% | #173 | SCG | 4 | 8 | 80% | 10% | 10% |
| #84 | BR | 4 | 3 | 60% | 20% | 20% | #174 | BR | 4 | 8 | 80% | 10% | 10% |
| #85 | LM | 5 | 3 | 60% | 20% | 20% | #175 | LM | 5 | 8 | 80% | 10% | 10% |
| #86 | SCG | 5 | 3 | 60% | 20% | 20% | #176 | SCG | 5 | 8 | 80% | 10% | 10% |
| #87 | BR | 5 | 3 | 60% | 20% | 20% | #177 | BR | 5 | 8 | 80% | 10% | 10% |
| #88 | LM | 6 | 3 | 60% | 20% | 20% | #178 | LM | 6 | 8 | 80% | 10% | 10% |
| #89 | SCG | 6 | 3 | 60% | 20% | 20% | #179 | SCG | 6 | 8 | 80% | 10% | 10% |
| #90 | BR | 6 | 3 | 60% | 20% | 20% | #180 | BR | 6 | 8 | 80% | 10% | 10% |

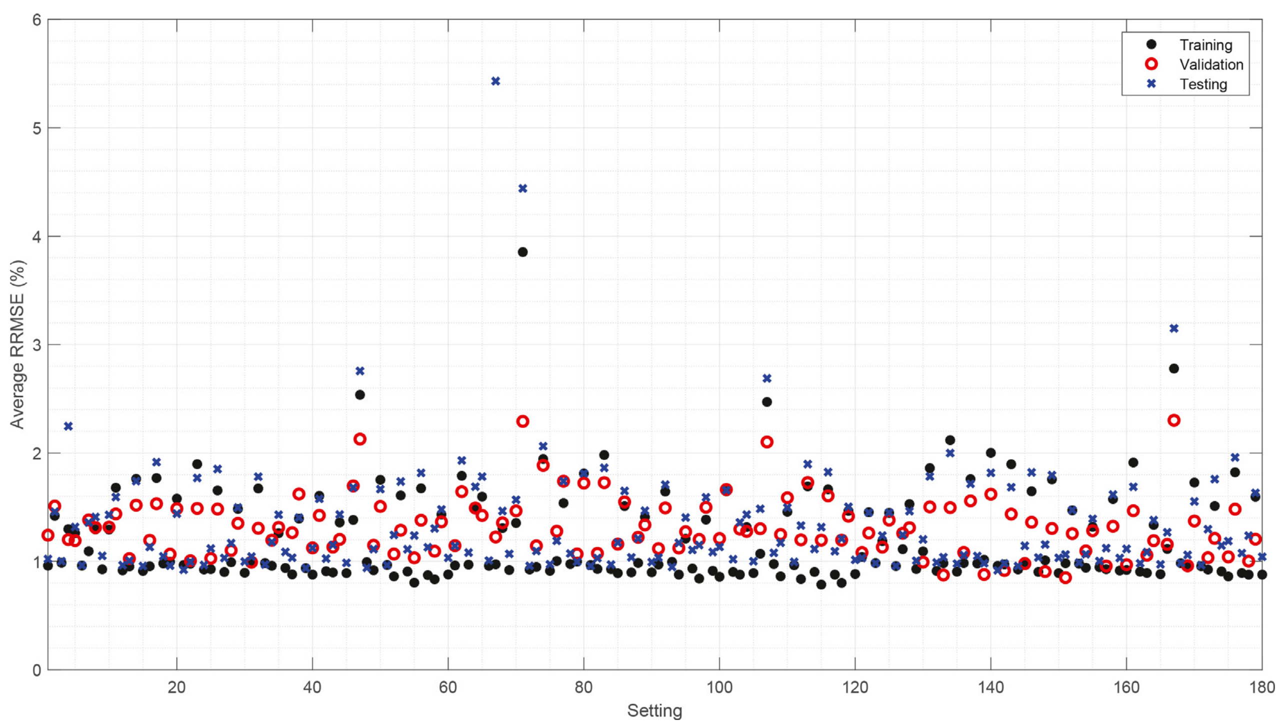


**Fig. 7.** Distribution statistics of the RRMSE across all model settings.

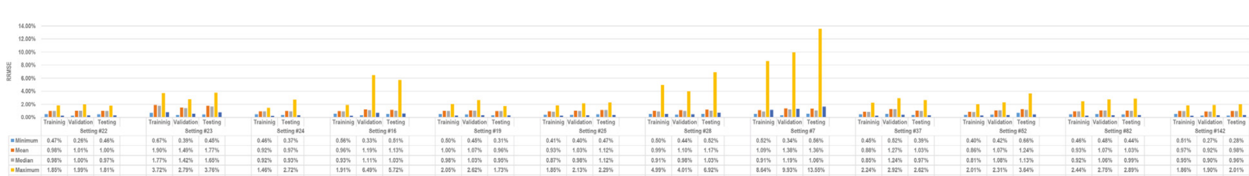

**Fig. 8.** Sensitivities of model performance (the RRMSE) to different settings.

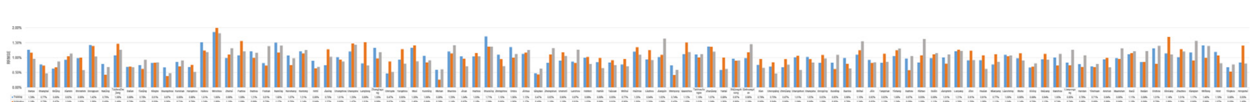

**Fig. 9.** Model performance (the RRMSE) across house prices of the ninety-nine cities based on the setting #22.

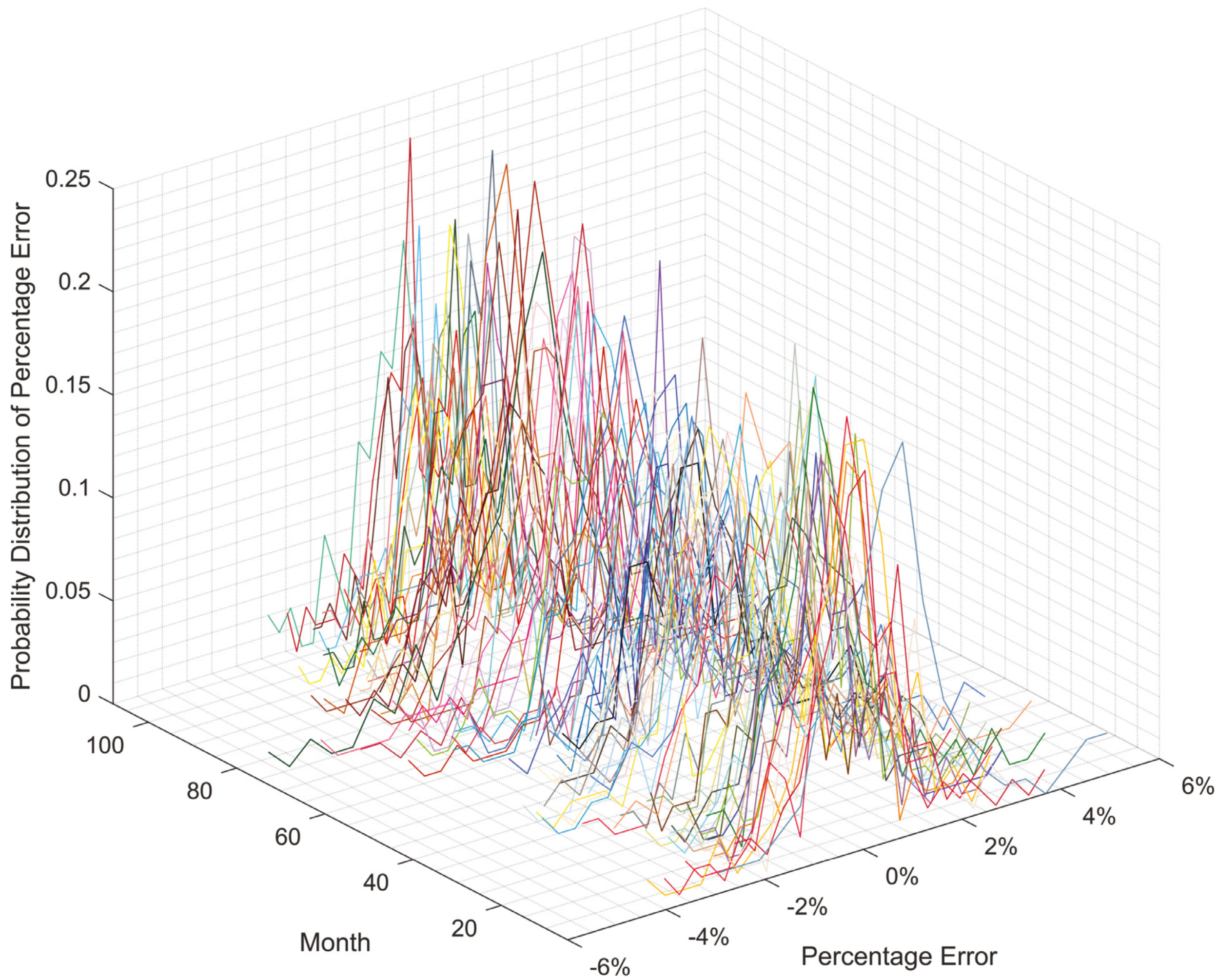


**Fig. 10.** Probability distribution plots of percentage errors across time for the ninety-nine cities based on the setting #22.

**Table 2**
Mean RRMSEs of fifteen candidates.

| Setting | Training | Validation | Testing |
|---|---|---|---|
| #3 | 0.99% | | 1.00% |
| #6 | 0.96% | | 0.96% |
| #22 | 0.98% | 1.01% | 1.00% |
| #31 | 0.98% | 0.99% | 1.04% |
| #33 | 0.98% | | 0.97% |
| #39 | 0.94% | | 0.94% |
| #51 | 0.97% | | 0.97% |
| #81 | 0.97% | | 0.96% |
| #111 | 0.96% | | 1.00% |
| #123 | 0.99% | | 0.98% |
| #126 | 0.96% | | 0.96% |
| #142 | 0.97% | 0.92% | 0.98% |
| #153 | 0.98% | | 0.99% |
| #168 | 0.98% | | 0.98% |
| #171 | 0.96% | | 0.97% |

neural networks could approximate a large class of functions well (Wang and Yang, 2010; Yang et al., 2010; 2008). Rather than utilizing one specific nonlinear function between inputs and the output for common nonlinear models, neural networks' basic structure could combine many 'basic' nonlinear functions through the multilayer structure. These could contribute to neural networks' capability of self-learning for forecasting of chaotic time series (Karasu et al., 2020) and capturing their nonlinearities (Altan et al., 2021). With good performance achieved here, we empirically demonstrate the usefulness of the neural network approach to the house price forecasting problem in the Chinese market.

## 5. Conclusion

House price forecasting is a key concern for the people and policy makers. In the current study, we focus on this problem in a data set of monthly house prices from ninety-nine major cities in China for the period of June 2010–May 2019 during which the house market is rapidly growing. Particularly, we explore the univariate neural network model with great potential for forecasting noised and chaotic economic data. By investigating different model settings over the algorithm, delay, hidden neuron, and data spitting ratio, we arrive at a rather simple neural network with four delays and three hidden neurons that generates stable performance of 1% average relative root mean square error across the ninety-nine cities. The models may serve as a standalone tool or be combined with fundamental forecasting approaches (Wei and Cao, 2017; Xu, 2018a; 2018c) in forming perspectives of house price trends and conducting policy analysis. Our empirical framework should not be difficult to deploy, which is important to many decision makers (Brandt and Bessler, 1983), and has potential to be generalized for house price forecasting of other cities in China or other countries or regions.

Future research of interest might be examining the potential of combining time series models that incorporate fundamental economic factors and pure technical machine learning models for house price forecasting. It should also be a worthwhile avenue for future research to explore economic significance of forecasting using the neural network (e.g. Wang and Yang, 2010; Yang et al., 2010; 2008). Wegener et al. (2016) point out that cointegration degrees and system linearities (Xu, 2018d) could be important to forecasting accuracy under the neural network framework. As including error correction terms could help deal with the information loss issue caused by data differencing in a cointegrated system (Xu, 2015b; 2019b), hybrid forecasting approaches that combine cointegration analysis and neural networks (e.g. Wegener et al., 2016) can be of interest to pursue as well for future studies, maybe for long-term forecasting in particular. This would also need the extension of current univariate modeling here to multivariate through appropriate ways to select relevant price information (Xu, 2014b; Xu and Thurman, 2015b) from other cities for forecasting of a city's house price. As we focus on one specific set of parameters for all cities in the current study, future work on searching for different sets of parameters across cities might help further improve performance. It should also be an interesting and worthwhile avenue for considerations for future research that explores model uncertainty within the neural network framework (e.g. Gal and Ghahramani, 2016; Guo et al., 2017; Lakshminarayanan et al., 2016).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Xiaojie Xu:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yun Zhang:** Writing – original draft, Writing – review & editing.

## References


Abidoye, R. B., & Chan, A. P. (2017). Modelling property values in nigeria using artificial neural network. *Journal of Property Research, 34*(1), 36–53. https://doi.org/10.1080/09599916.2017.1286366.

Abidoye, R. B., & Chan, A. P. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal, 24*(1), 71–83. https://doi.org/10.1080/14445921.2018.1436306.

Al Bataineh, A., & Kaur, D. (2018). A comparative study of different curve fitting algorithms in artificial neural network using housing dataset. In *Naecon 2018-IEEE national aerospace and electronics conference* (pp. 174–178). IEEE. https://doi.org/10.1109/NAECON.2018.8556738.

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons & Fractals, 126*, 325–336. https://doi.org/10.1016/j.chaos.2019.07.011.

Altan, A., Karasu, S., & Zio, E. (2021). A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Applied Soft Computing, 100*, 106996. https://doi.org/10.1016/j.asoc.2020.106996.

Azadeh, A., Sheikhalishahi, M., & Boostani, A. (2014). A flexible neuro-fuzzy approach for improvement of seasonal housing price estimation in uncertain and non-linear environments. *South African Journal of Economics, 82*(4), 567–582. https://doi.org/10.1111/saje.12047.

Baghirli, O. (2015). Comparison of Lavenberg-Marquardt, scaled conjugate gradient and Bayesian regularization backpropagation algorithms for multistep ahead wind speed forecasting using multilayer perceptron feedforward neural network.

Brandt, J. A., & Bessler, D. A. (1983). Price forecasting and evaluation: An application in agriculture. *Journal of Forecasting, 2*(3), 237–248. https://doi.org/10.1002/for.3980020306.

Cabrera, J., Wang, T., & Yang, J. (2011). Linear and nonlinear predictablity of international securitized real estate returns: A reality check. *Journal of Real Estate Research, 33*(4), 565–594. https://doi.org/10.1080/10835547.2011.12091317.

Ćetković, J., Lakić, S., Lazarevska, M., Žarković, M., Vujošević, S., Cvijović, J., & Gogić, M. (2018). Assessment of the real estate market value in the european market by artificial neural networks application. *Complexity, 2018*. https://doi.org/10.1155/2018/1472957.

Chen, J.-H., Ong, C. F., Zheng, L., & Hsu, S.-C. (2017). Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management, 21*(3), 273–283. https://doi.org/10.3846/1648715X.2016.1259190.

Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A neural network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia, 3*, 810–817. https://doi.org/10.1016/j.trpro.2014.10.067.

Doan, C. D., & Liong, S.-y. (2004). Generalization for multilayer neural network bayesian regularization or early stopping. In *Proceedings of asia pacific association of hydrology and water resources 2nd conference* (pp. 5–8).

Embaye, W. T., Zereyesus, Y. A., & Chen, B. (2021). Predicting the rental value of houses in household surveys in tanzania, uganda and malawi: Evaluations of hedonic pricing and machine learning approaches. *Plos One, 16*(2), e0244953. https://doi.org/10.1371/journal.pone.0244953.

Foresee, F. D., & Hagan, M. T. (1997). Gauss-newton approximation to bayesian learning. In *Proceedings of international conference on neural networks (ICNN'97): 3* (pp. 1930–1935). IEEE. https://doi.org/10.1109/ICNN.1997.614194.

Fu, T. (2018). Forecasting second-hand housing price using artificial intelligence and machine learning techniques. In *2018 8th international conference on mechatronics, computer and education informationization (MCEI 2018)* (pp. 269–273). Atlantis Press. https://doi.org/10.2991/mcei-18.2018.54.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050–1059). PMLR.

Ge, C., Wang, Y., Xie, X., Liu, H., & Zhou, Z. (2019). An integrated model for urban subregion house price forecasting: A multi-source data perspective. In *2019 ieee international conference on data mining (ICDM)* (pp. 1054–1059). IEEE. https://doi.org/10.1109/ICDM.2019.00123.

Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications, 38*(4), 3383–3386. https://doi.org/10.1016/j.eswa.2010.08.123.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *International conference on machine learning* (pp. 1321–1330). PMLR.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks, 5*(6), 989–993. https://doi.org/10.1109/72.329697.

Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research, 38*(1), 48–70. https://doi.org/10.1080/09599916.2020.1832558.

Huang, Y. (2019). Predicting home value in california, united states via machine learning modeling. *Statistics, Optimization & Information Computing, 7*(1), 66–74. https://doi.org/10.19139/soic.v7i1.435.

Hyvärinen, A., Zhang, K., Shimizu, S., & Hoyer, P. O. (2010). Estimation of a structural vector autoregression model using non-Gaussianity. *Journal of Machine Learning Research, 11*(5).

Igbinosa, S. O. (2011). Determinants of residential property value in nigeria–a neural network approach. *African Research Review, 5*(5), 152–168. https://doi.org/10.4314/afrrev.v5i5.13.

Kang, J., Lee, H. J., Jeong, S. H., Lee, H. S., & Oh, K. J. (2020). Developing a forecasting model for real estate auction prices using artificial intelligence. *Sustainability, 12*(7), 2899. https://doi.org/10.3390/su12072899.

Karasu, S., & Altan, A. (2019). Recognition model for solar radiation time series based on random forest with feature selection approach. In *2019 11th international conference on electrical and electronics engineering (ELECO)* (pp. 8–11). IEEE. https://doi.org/10.23919/ELECO47770.2019.8990664.

Karasu, S., Altan, A., Bekiros, S., & Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy, 212*, 118750. https://doi.org/10.1016/j.energy.2020.118750.

Karasu, S., Altan, A., Saraç, Z., & Hacioğlu, R. (2017a). Estimation of fast varied wind speed based on narx neural network by using curve fitting. *International Journal of Energy Applications and Technologies, 4*(3), 137–146.

Karasu, S., Altan, A., Saraç, Z., & Hacioğlu, R. (2017b). Prediction of solar radiation based on machine learning methods. *The Journal of Cognitive Systems, 2*(1), 16–20.

Karasu, S., Altan, A., Saraç, Z., & Hacioğlu, R. (2017c). Prediction of wind speed with non-linear autoregressive (NAR) neural networks. In *2017 25th signal processing and communications applications conference (SIU)* (pp. 1–4). IEEE. https://doi.org/10.1109/SIU.2017.7960507.

Kawahara, Y., Shimizu, S., & Washio, T. (2011). Analyzing relationships among arma processes based on non-Gaussianity of external influences. *Neurocomputing, 74*(12–13), 2212–2221. https://doi.org/10.1016/j.neucom.2011.02.008.

Kayri, M. (2016). Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: A comparative empirical study on social data. *Mathematical and Computational Applications, 21*(2), 20. https://doi.org/10.3390/mca21020020.

Khalafallah, A. (2008). Neural network based model for predicting housing market performance. *Tsinghua Science and Technology, 13*(S1), 325–328. https://doi.org/10.1016/S1007-0214(08)70169-X.

Khan, T. A., Alam, M., Shahid, Z., & Mazliham, M. (2019). Comparative performance analysis of Levenberg-Marquardt, Bayesian regularization and scaled conjugate gradient for the prediction of flash floods. *Journal of Information Communication Technologies and Robotic Applications*, 52–58.

Kim, J., Leatham, D. J., & Bessler, D. (2007). Reits' dynamics under structural change with unknown break points. *Journal of Housing Economics, 16*(1), 37–58. https://doi.org/10.1016/j.jhe.2007.04.001.

Kitapci, O., Tosun, Ö., Tuna, M. F., & Turk, T. (2017). The use of artificial neural networks (ann) in forecasting housing prices in ankara, turkey. *Journal of Marketing and Consumer Behaviour in Emerging Markets*, (1 (5)), 4–14.

Kouwenberg, R., & Zwinkels, R. (2014). Forecasting the us housing market. *International Journal of Forecasting, 30*(3), 415–425. https://doi.org/10.1016/j.ijforecast.2013.12.010.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv:1612.01474,.

Lam, K. C., Yu, C., & Lam, K. (2008). An artificial neural network and entropy model for residential property price forecasting in hong kong. *Journal of Property Research, 25*(4), 321–342. https://doi.org/10.1080/09599910902837051.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics, 2*(2), 164–168. https://doi.org/10.1090/qam/10666.

Li, D.-Y., Xu, W., Zhao, H., & Chen, R.-Q. (2009). A SVR based forecasting approach for real estate price prediction. In *2009 international conference on machine learning and cybernetics: 2* (pp. 970–974). IEEE. https://doi.org/10.1109/ICMLC.2009.5212389.

Li, R. Y. M., Cheng, K. Y., & Shoaib, M. (2018). Walled buildings, sustainability, and housing prices: an artificial neural network approach. *Sustainability, 10*(4), 1298. https://doi.org/10.3390/su10041298.

Li, R. Y. M., Fong, S., & Chong, K. W. S. (2017). Forecasting the reits and stock indices: Group method of data handling neural network approach. *Pacific Rim Property Research Journal, 23*(2), 123–160. https://doi.org/10.1080/14445921.2016.1225149.

Li, Y., Xiang, Z., & Xiong, T. (2020). The behavioral mechanism and forecasting of Beijing housing prices from a multiscale perspective. *Discrete Dynamics in Nature and Society, 2020*. https://doi.org/10.1155/2020/5375206.

Lim, W. T., Wang, L., Wang, Y., & Chang, Q. (2016). Housing price prediction using neural networks. In *2016 12th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)* (pp. 518–522). IEEE. https://doi.org/10.1109/FSKD.2016.7603227.

Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network. In *New zealand agricultural and resource economics society conference* (pp. 25–26).

Liu, L., & Wu, L. (2020). Predicting housing prices in china based on modified Holt's exponential smoothing incorporating whale optimization algorithm. *Socio-Economic Planning Sciences, 72*, 100916. https://doi.org/10.1016/j.seps.2020.100916.

Liu, R., & Liu, L. (2019). Predicting housing price in china based on long short-term memory incorporating modified genetic algorithm. *Soft Computing, 23*(22), 11829–11838. https://doi.org/10.1007/s00500-018-03739-w.

Ma, H., Chen, M., & Zhang, J. (2015). Study on the prediction of real estate price index based on HHGA-RBF neural network algorithm. *International Journal of u-and e-Service, Science and Technology, 8*(7), 109–118. https://doi.org/10.14257/ijunesst.2015.8.7.11.

MacKay, D. J. (1992). Bayesian interpolation. *Neural Computation, 4*(3), 415–447. https://doi.org/10.1162/neco.1992.4.3.415.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics, 11*(2), 431–441. https://doi.org/10.1137/0111030.

Milunovich, G. (2020). Forecasting australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting, 39*(7), 1098–1118. https://doi.org/10.1002/for.2678.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*(4), 525–533. https://doi.org/10.1016/S0893-6080(05)80056-5.

Morano, P., & Tajani, F. (2013). Bare ownership evaluation. hedonic price model vs. artificial neural network. *International Journal of Business Intelligence and Data Mining, 8*(4), 340–362. https://doi.org/10.1504/IJBIDM.2013.059263.

Morano, P., Tajani, F., & Torre, C. M. (2015). Artificial intelligence in property valuations: An application of artificial neural networks to housing appraisal. *Advances in Environmental Science and Energy, Planning*, 23–29.

Nghiep, N., & Al, C. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research, 22*(3), 313–336. https://doi.org/10.1080/10835547.2001.12091068.

Pai, P.-F., & Wang, W.-C. (2020). Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences, 10*(17), 5832. https://doi.org/10.3390/app10175832.

Paluszek, M., & Thomas, S. (2020). *Practical MATLAB deep learning: A project-based approach*. Apress.

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications, 42*(6), 2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040.

Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research, 31*(2), 147–164. https://doi.org/10.1080/10835547.2009.12091245.

Piao, Y., Chen, A., & Shang, Z. (2019). Housing price prediction based on CNN. In *2019 9th international conference on information science and technology (ICIST)* (pp. 491–495). IEEE. https://doi.org/10.1109/ICIST.2019.8836731.

Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the us real house price index. *Economic Modelling, 45*, 259–267. https://doi.org/10.1016/j.econmod.2014.10.050.

Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management, 142*(2), 04015066. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001047.

Rahman, S. N. A., Maimun, N. H. A., Razali, M. N. M., & Ismail, S. (2019). The artificial neural network model (ann) for malaysian housing market analysis. *Planning Malaysia, 17*(9). https://doi.org/10.21837/pm.v17i9.581.

Rico-Juan, J. R., & de La Paz, P. T. (2021). Machine learning with explainability or spatial hedonics tools? an analysis of the asking prices in the housing market in alicante, spain. *Expert Systems with Applications, 171*, 114590. https://doi.org/10.1016/j.eswa.2021.114590.

Sarip, A. G., Hafez, M. B., & Daud, M. N. (2016). Application of fuzzy regression model for real estate price prediction. *Malaysian Journal of Computer Science, 29*(1), 15–27. https://doi.org/10.22452/mjcs.vol29no1.2.

Selim, H. (2009). Determinants of house prices in turkey: Hedonic regression versus artificial neural network. *Expert systems with Applications, 36*(2), 2843–2852. https://doi.org/10.1016/j.eswa.2008.01.044.

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation, 5*(1), 16. https://doi.org/10.1186/s40854-019-0131-7.

Shahhosseini, M., Hu, G., & Pham, H. (2019). Optimizing ensemble weights for machine learning models: a case study for housing price prediction. In *Informs international conference on service science* (pp. 87–97). Springer. https://doi.org/10.1007/978-3-030-30967-1_9.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research, 7*(10).

Tabales, J. M. N., Caridad, J. M., & Carmona, F. J. R. (2013). Artificial neural networks for predicting real estate price. *Revista de Métodos Cuantitativos para la Economía y la Empresa, 15*, 29–44.

Taffese, W. Z. (2007). Case-based reasoning and neural networks for real estate valuation. In *Artificial intelligence and applications* (pp. 98–104).

Terregrossa, S. J., & Ibadi, M. H. (2021). Combining housing price forecasts generated separately by hedonic and artificial neural network models. *Asian Journal of Economics, Business and Accounting*, 130–148. https://doi.org/10.9734/ajeba/2021/v21i130345.

Wang, L., Chan, F. F., Wang, Y., & Chang, Q. (2016). Predicting public housing prices using delayed neural networks. In *2016 ieee region 10 conference (tencon)* (pp. 3589–3592). IEEE. https://doi.org/10.1109/TENCON.2016.7848726.

Wang, T., & Yang, J. (2010). Nonlinearity and intraday efficiency tests on energy futures markets. *Energy Economics, 32*(2), 496–503. https://doi.org/10.1016/j.eneco.2009.08.001.

Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by pso. *Optik, 125*(3), 1439–1443. https://doi.org/10.1016/j.ijleo.2013.09.017.

Webb, R. I., Yang, J., & Zhang, J. (2016). Price jump risk in the us housing market. *The Journal of Real Estate Finance and Economics, 53*(1), 29–49. https://doi.org/10.1007/s11146-015-9518-z.

Wegener, C., von Spreckelsen, C., Basse, T., & von Mettenheim, H.-J. (2016). Forecasting government bond yields with neural networks considering cointegration. *Journal of Forecasting, 35*(1), 86–92. https://doi.org/10.1002/for.2385.

Wei, Y., & Cao, Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from china. *Economic Modelling, 61*, 147–155. https://doi.org/10.1016/j.econmod.2016.12.002.

Wilson, I. D., Paris, S. D., Ware, J. A., & Jenkins, D. H. (2002). Residential property price time series forecasting with neural networks. In *Applications and innovations in intelligent systems ix* (pp. 17–28). Springer. https://doi.org/10.1007/978-1-4471-0149-9_2.

Wu, C.-H., Li, C.-H., Fang, I.-C., Hsu, C.-C., Lin, W.-T., & Wu, C.-H. (2009). Hybrid genetic-based support vector regression with feng shui theory for appraising real estate price. In *2009 first asian conference on intelligent information and database systems* (pp. 295–300). IEEE. https://doi.org/10.1109/ACIIDS.2009.41.

Xiaolong, H., & Ming, Z. (2010). Applied research on real estate price prediction by the neural network. In *2010 the 2nd conference on environmental science and information application technology: 2* (pp. 384–386). IEEE. https://doi.org/10.1109/ESIAT.2010.5567321.

Xin, J. G., & Runeson, G. (2004). Modeling property prices using neural network model for hong kong. *International Real Estate Review, 7*(1), 121–138.

Xu, L., & Li, Z. (2021). A new appraisal model of second-hand housing prices in China's first-tier cities based on machine learning algorithms. *Computational Economics, 57*(2), 617–637. https://doi.org/10.1007/s10614-020-09973-5.

Xu, X. (2014a). Causality and price discovery in US corn markets: An application of error correction modeling and directed acyclic graphs. *Technical Report*. https://doi.org/10.22004/ag.econ.169806.

Xu, X. (2014b). Price discovery in us corn cash and futures markets: The role of cash market selection. In *Selected paper prepared for presentation at the agricultural & applied economics associations 2014 AAEA annual meeting, Minneapolis, MN*. https://doi.org/10.22004/ag.econ.169809.

Xu, X. (2015a). Causality, price discovery, and price forecasts: Evidence from us corn cash and futures markets,.

Xu, X. (2015b). Cointegration among regional corn cash prices. *Economics Bulletin, 35*(4), 2581–2594.
Xu, X. (2017a). Contemporaneous causal orderings of us corn cash prices through directed acyclic graphs. *Empirical Economics, 52*(2), 731–758. https://doi.org/10.1007/s00181-016-1094-4.
Xu, X. (2017b). The rolling causal structure between the chinese stock index and futures. *Financial Markets and Portfolio Management, 31*(4), 491–509. https://doi.org/10.1007/s11408-017-0299-7.
Xu, X. (2017c). Short-run price forecast performance of individual and composite models for 496 corn cash markets. *Journal of Applied Statistics, 44*(14), 2593–2620. https://doi.org/10.1080/02664763.2016.1259399.
Xu, X. (2018a). Causal structure among us corn futures and regional cash prices in the time and frequency domain. *Journal of Applied Statistics, 45*(13), 2455–2480. https://doi.org/10.1080/02664763.2017.1423044.
Xu, X. (2018b). Cointegration and price discovery in us corn cash and futures markets. *Empirical Economics, 55*(4), 1889–1923. https://doi.org/10.1007/s00181-017-1322-6.
Xu, X. (2018c). Intraday price information flows between the csi300 and futures market: An application of wavelet analysis. *Empirical Economics, 54*(3), 1267–1295. https://doi.org/10.1007/s00181-017-1245-2.
Xu, X. (2018d). Linear and nonlinear causality between corn cash and futures prices. *Journal of Agricultural & Food Industrial Organization, 16*(2), 20160006. https://doi.org/10.1515/jafio-2016-0006.
Xu, X. (2018e). Using local information to improve short-run corn price forecasts. *Journal of Agricultural & Food Industrial Organization, 16*(1). https://doi.org/10.1515/jafio-2017-0018.
Xu, X. (2019a). Contemporaneous and granger causality among us corn cash and futures prices. *European Review of Agricultural Economics, 46*(4), 663–695. https://doi.org/10.1093/erae/jby036.
Xu, X. (2019b). Contemporaneous causal orderings of csi300 and futures prices through directed acyclic graphs. *Economics Bulletin, 39*(3), 2052–2077.
Xu, X. (2019c). Price dynamics in corn cash and futures markets: cointegration, causality, and forecasting through a rolling window approach. *Financial Markets and Portfolio Management, 33*(2), 155–181. https://doi.org/10.1007/s11408-019-00330-7.
Xu, X. (2020). Corn cash price forecasting. *American Journal of Agricultural Economics, 102*(4), 1297–1320. https://doi.org/10.1002/ajae.12041.
Xu, X., & Thurman, W. (2015a). Forecasting local grain prices: An evaluation of composite models in 500 corn cash markets. *Technical Report*. https://doi.org/10.22004/ag.econ.205332.
Xu, X., & Thurman, W. N. (2015b). Using local information to improve short-run corn cash price forecasts, http://doi.org/10.22004/ag.econ.285845.
Xu, X., & Zhang, Y. (2021a). Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture, 184*, 106120. https://doi.org/10.1016/j.compag.2021.106120.
Xu, X., & Zhang, Y. (2021b). Individual time series and composite forecasting of the chinese stock index. *Machine Learning with Applications, 5*, 100035. https://doi.org/10.1016/j.mlwa.2021.100035.
Xu, X., & Zhang, Y. (2021c). Network analysis of corn cash price comovements. *Machine Learning with Applications, 6* 100140. https://doi.org/10.1016/j.mlwa.2021.100140.
Xu, X., & Zhang, Y. (2021d). Network analysis of housing price comovements of a hundred chinese cities,.
Xu, X., & Zhang, Y. (2021e). Second-hand house price index forecasting with neural networks,.
Yan, Z., & Zong, L. (2020). Spatial prediction of housing prices in Beijing using machine learning algorithms. In *Proceedings of the 2020 4th high performance computing and cluster technologies conference & 2020 3rd international conference on big data and artificial intelligence* (pp. 64–71). https://doi.org/10.1145/3409501.3409543.
Yang, J., Cabrera, J., & Wang, T. (2010). Nonlinearity, data-snooping, and stock index ETFreturn predictability. *European Journal of Operational Research, 200*(2), 498–507. https://doi.org/10.1016/j.ejor.2009.01.009.
Yang, J., Su, X., & Kolari, J. W. (2008). Do euro exchange rates follow a martingale? some out-of-sample evidence. *Journal of Banking & Finance, 32*(5), 729–740. https://doi.org/10.1016/j.jbankfin.2007.05.009.
Yang, J., Yu, Z., & Deng, Y. (2018). Housing price spillovers in china: A high-dimensional generalized var approach. *Regional Science and Urban Economics, 68*, 98–114. https://doi.org/10.1016/j.regsciurbeco.2017.10.016.
Yasnitsky, L. N., Yasnitsky, V. L., & Alekseev, A. O. (2021). The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time. *Complexity, 2021*. https://doi.org/10.1155/2021/5392170.
Yu, L., Jiao, C., Xin, H., Wang, Y., & Wang, K. (2018). Prediction on housing price based on deep learning. *International Journal of Computer and Information Engineering, 12*(2), 90–99. https://doi.org/10.5281/zenodo.1315879.
Zohrabyan, T., Leatham, D. J., & Bessler, D. A. (2008). Cointegration analysis of regional house prices in US. *Technical Report*. https://doi.org/10.22004/ag.econ.48138.