Article

Assessing the Impact of Public Rental Housing on the Housing Prices in Proximity: Based on the Regional and Local Level of Price Prediction Models Using Long Short-Term Memory (LSTM)

Hyunsoo Kim, Youngwoo Kwon and Yeol Choi *

Department of Urban Planning & Engineering, Pusan National University, Busan 46241, Korea;
hyunsoo.kim@pusan.ac.kr (H.K.); starbugs7070@pusan.ac.kr (Y.K.)
*** Correspondence: yeolchoi@pusan.ac.kr; Tel.: +82-51-510-2357

check for updates

Abstract: Providing adequate public rental housing (PRH) of a decent quality at a desirable location is a major challenge in many cities. Often, a prominent opponent of PRH development is its host community, driven by a belief that PRH depreciates nearby property values. While this is a persistent issue in many cities around the world, this study proposed a new approach to assessing the impact of PRH on nearby property value. This study utilized a machine learning technique called long short-term memory (LSTM) to construct a set of housing price prediction models based on 547,740 apartment transaction records from the city of Busan, South Korea. A set of apartment characteristics and proximity measures to PRH were included in the modeling process. Four geographic boundaries were analyzed: The entire region of Busan, all neighborhoods of PRH, the neighborhoods of PRH in the "favorable," and the "less favorable" local housing market. The study produced accurate and reliable price predictions, which indicated that the proximity to PRH has a meaningful impact on nearby housing prices both at the city and the neighborhood level. The approach taken by the study can facilitate improved decision making for future PRH policies and programs.

Keywords: public rental housing; housing market analysis; price forecasting; long short-term memory

1. Introduction

Housing is an essential human need that shapes the well-being of all citizens. As most cities around the world are faced with the overwhelming burden of housing costs, establishing housing affordability is a core policy objective in many countries [1]. The rising housing and rental price in urban areas are especially burdensome for the vulnerable sector, as it restricts low-income families from accessing the decent quality of housing, healthcare, education, and other urban services. Affordable rental housing policy is a promising solution to alleviate the housing cost burden in cities, as various types of public rental housing (PRH) policies are encouraged around the world [1,2]. Consequently, a number of innovative policies and strategies are being introduced, and the opportunities of PRH as a building block of sustainable urban development is rapidly growing [3–6]. PRH provisions and policies are connected to ecological, economic, and social aspects of sustainability [7]. While the ecological aspect of sustainability relies on the single-building assessment with a focus on a building's life cycle and performance, the economic and social aspects of sustainability focus on the connection between buildings and people as well as the surrounding environment [8]. The role of community is emphasized in sustainable development, as it is closely related to the life of urban residents, and at the same time, it is the root of some social problems [6,8].

Adequate PRH provisions and policies promote the livelihood and well-being of a community by providing a healthy mix of housing options and allowing all citizens to access a decent quality of housing, health care, education, and other urban services. They also contribute to achieving sustainable urban development by protecting communities from urban issues such as social inequalities. However, providing good PRH of a decent quality at a desirable location is an endemic problem in many cities around the world. In Korea, where this study takes place, the level of housing costs in a household budget of 15% indicates that housing affordability is not a severe problem. However, the quality of housing for low-income groups is poor, and the supply of PRH stock is deficient [9,10]. Along with the overheating housing market in city centers, it is no surprise that the burden of housing costs is in a steep incline, and the need for PRH is growing.

Despite the need for decent PRH stock in desirable locations, a new supply of housing complexes for low-income families often presents conflicts with the surrounding communities. The conflict mainly involves the belief that PRH depreciates the value of nearby properties. This widespread perception, along with the fear of an increase in crime and a concentration of poverty, has created a "not in my back yard" (NIMBY) attitude toward PRH projects, and other forms of affordable housing projects in many cities [11–14].

While this is a continuous phenomenon in many cities, including some of the densest cities in Korea, a vast amount of literature in housing studies have suggested that the effect of affordable housing on surrounding property values is very complex. Notably, the effect can go in either direction, depending on a number of different factors, including the physical and social attributes of affordable housing and their host neighborhoods [15,16]. Traditionally, housing price studies approached the issue with a hedonic pricing method and other forms of multiple regression analyses. However, they often suffered from issues including nonlinearity, heteroscedasticity, spatial dependency, and a selection of independent variables in modeling housing prices [17,18]. It is also critical to reflect on changing local housing markets over time, which often poses difficulties under the provisions of traditional statistical methods such as serial autocorrelation [19].

In the recent wave of utilizing artificial intelligence in various data science applications, some of the machine learning algorithms were presented as feasible alternatives in creating housing price models [20,21]. As they are explained further in the rest of this paper, these algorithms are capable of mimicking the patterns of data without any constraints of functional form, assumption, or data structure, thus providing greater flexibility in understanding the complex nature of a regional housing market. A body of housing price literature has demonstrated the robustness of these models [17,22–24]. For this paper, the advantages of utilizing a machine learning technique can provide a refined understanding of the effect of PRH based on the improved predictive power of housing price models.

The purpose of the study is to investigate the dynamic relationship between PRH and surrounding property values in consideration of various determinants of property prices. The study uses 547,740 apartment sales data over 167 months from January 2006 to November 2019 that were recorded in Busan, South Korea. While there are different types of PRH in Korea, the build-to-rent type of PRH was chosen as the subject of this study as they are the most widely provided type of residence. By employing a machine learning technique called long short-term memory (LSTM), which is a branch of the recurrent neural network (RNN), the study was able to construct a housing price prediction model with improved performance. The neural network not only considers the sequence of housing prices over time to make predictions but also accounts for six apartment characteristics that affect the price. This includes the structural attributes such as the lot size and building age, as well as the attributes that represent the price variations such as the monthly transaction volume and the brand of apartment complexes. In order to assess the impact of PRH in proximity, the model also includes three proximity measures regarding PRH: Two accounting for the nearest PRH residence; and one for all PRH units in proximity.

In the model building process, the study expanded the model in steps by introducing additional sets of variables into the learning process. The performance at each step was evaluated to find

the "construct" of variables with the most predictive power. The process was applied to produce a model for the entire region of Busan and the neighborhoods of PRH to observe the differences in patterns in housing prices. In addition, we split our sample in proximity to PRH into two, based on the changes in their local housing markets determined by the average increase in apartment prices. Accordingly, two models were produced: The neighborhoods of PRH located in the "favorable" local housing market, and those in the "less favorable" local housing market.

Ultimately, the study was designed to focus on answering the following questions:

1. Do proximity measures direct toward PRH increase the predictive power of regional housing prices?
2. Are there any differences in trends between the regional housing prices and the ones in proximity to PRH?
3. For the apartments that are located in proximity to PRH, are there any differences in trends if they are located in the neighborhood with a more "favorable" local housing market?
4. Does LSTM produce improved predictive power than the traditional multivariate time series model?

The findings from the study can provide a comprehensive understanding of the impact of PRH based on the regional and local trends of housing prices. This study can contribute to the literature in two ways: By providing another set of empirical evidence on the machine learning algorithms' applicability in housing price prediction; and by addressing a social concern related to housing policy—in our case, the effect of PRH, based on the use of advanced prediction method. Finally, the information drawn from the model can also be used to foresee the community benefits and assist in making policy decisions concerning PRH.

2. Background

2.1. Public Rental Housing (PRH) Provisions in Korea

PRH is a type of subsidized housing provision that can provide low-income families without homeownership access to affordable rental housing options. Thanks to advancements in construction technology and the management of public housing, PRH has become an effective method for supplying a large number of rental units in good condition. Consequently, it is one of the major modes through which affordable housing is supplied in many countries, including Korea. Statistics indicate that the PRH stock has been gradually increasing over the past decade, and the mode of supply is diversifying. As of 2017, PRH accounted for approximately 7% of the entire housing stock [25]. While it can be developed and managed by both the public and private sectors, government-subsidized PRH is much more prevalent, taking up nearly 87% of all types [26].

There are different types of PRH provisions designed with different tenant eligibility, and they are provided in three channels: Build-to-rent, buy-to-rent, and rent-to-rent. The procedure used to select qualified tenants differs by program. For build-to-rent PRH, the government and private developers acquire and develop the land, build rental housing, manage them, and rent them to eligible tenants. In buy-to-rent PRH, the government or certified contractors purchase housing units, manage them, and rent them to qualified tenants. For the rent-to-rent type of PRH, the government provides a portion of rent or deposit to eligible tenants. Monthly income is a general criterion, with additional point-based criteria such as the number of family members and children, the time during which they have not been homeowners, and the time during which they have held a subscription deposit account. While most programs prioritize vulnerable groups, including extremely-low income families, homeless families, and social security recipients, some programs are designed to prioritize certain groups such as newlyweds, college students, senior citizens with low incomes, and young adults with short work experience. Table 1 presents information regarding different types of PRH, including the mode of supply and tenant eligibility.

Table 1. Type of public rental housing (PRH) in Korea.

Type *		Tenant	Income Quintile **	Rent ***	Unit Size	Total Supply
Build-to-rent	Permanent rental (permanent or 50-year contract)	Extremely low-income household, social/medical security recipients, disabled	1st (less than 50%)	~30%	less than 40 m²	307,115
	Citizens' rental (30-year contract)	Low-income household	2nd to 4th (70–100%)	60%–80%	less than 85m² (mostly less than 60 m²)	509,248
	Rent-to-own (5/10-year contract)	Low-income household, newlywed, multiple children, senior dependent	3rd to 5th (100–120%)	~90%	less than 85 m²	223,412
	'Haengbok' rental	Low-income household, college student, newlywed, young adults	2nd to 5th (80–100%)	60%–80%	less than 45 m²	2041
Buy-to-rent		Low-income household	1st to 2nd (below 50%)	~30%	less than 85 m²	92,004
Rent-to-rent		Low-income household	1st to 2nd (below 50%)	~30%	less than 85 m²	165,764

Data source: Ministry of Land, Infrastructure, and Transport, 2018. * Region-specific programs and discontinued programs are excluded. The maximum allowed rental period is presented in the parentheses. ** Respective income criteria are provided in the parentheses. The percentage indicates the percent average monthly income of urban employers of a region. *** The rent indicates the approximate percentage of average monthly rent in the regional housing market.

The majority of PRH in Korea is the build-to-rent type, which accounts for more than 80% of the total supply. As of 2017, the percentage of apartment housing in all housing types was nearly 51%, making it the dominant housing type. While the number is continuously growing, build-to-rent units are almost entirely apartment or multi-family residential units. In contrast, very few single-family detached housing units remain in the buy-to-rent and rent-to-rent types. The units have four size categories: Less than 40 m², 40 m² to 60 m², 60 m² to 85 m², and 85 m² and larger. Across all PRH programs, units with an area less than 60 m² are the most common (87%), prioritizing individuals or families with smaller households and lower incomes [27]. Most of the build-to-rent type PRH are planned with a 5-, 10-, 30-, and 50-year rental contract. After the contract, the units can be sold in the private market. The rent-to-own type, which has a rental period of 5 to 10 years, gives priority to the tenant for purchase.

Although it has been reported that some older housing, especially the permanent rental housing built in the 90s, requires repairs for essential elements such as water supply, electricity, and heating, the majority of the PRH residents are satisfied with the quality of their housing and neighborhood [28]. A typical build-to-rent PRH is a large-scale apartment complex, which often faces difficulties in land acquisition, especially near urban centers where prices are high and developable land is scarce. Such challenges have forced the construction of PRH to move to the outskirts of cities into areas with less employment and amenity options. Recent efforts toward downscaling the projects and the use of publicly-owned land have allowed PRH to be located closer to urban centers. However, this often faces the opposition of nearby residents who fear the depreciation of local housing prices and negative impacts on neighborhood quality [29].

2.2. The Effect of PRH on Real Estate

The relationship between affordable housing and surrounding property values is very complicated, and a number of previous studies have presented multifaceted approaches using a large amount of housing sales data with rigorous methodologies in many cities worldwide. While there is a variety of affordable housing types and programs, our literature review focuses on the studies that used the government-manufactured or subsidized apartments, which are similar to the PRHs used in this study.

Some of the earliest affordable housing studies have suffered from some limitations, including the proper delineating of neighborhood boundaries, controlling a proper determinant of housing values, acquiring enough sample sizes, and accounting for different types/programs of affordable housing. Although the studies lacked methodological rigor, they appeared to agree that affordable housing does not have a detrimental effect on nearby housing values [30]. A wave of studies that

followed those earlier studies were advanced ones with the capability to handle a large amount of real estate sales data and more sophisticated methodologies such as hedonic price estimations. While some studies still did not find any significant relationship between affordable housing and nearby property values [31–33], others began to find significant effects under different scenarios with more detailed characterizations of affordable housing, programs, and neighborhoods.

Case Studies: The Effect of PRH on Real Estate

A study that was conducted in the American cities of Portland, Seattle, and Cleveland found that affordable housing under the Low Income Housing Tax Credit (LIHTC) program had a positive effect on nearby property values in the immediate neighborhood. However, in the case of Cleveland, the effect gradually decreased as the size of the LIHTC housing complexes increased and became negative after a critical point [34]. Similar results were also observed in a Korean study, as the increase in the size of rental housing complexes had a negative effect on surrounding housing prices [35].

Another study conducted in Seattle presented somewhat similar results for LIHTC residences; the effects were spatially heterogeneous, and the positive spillover effect was associated with the land use and socioeconomic characteristics of the sub-regions examined in the study [19]. This result aligned with studies of other LIHTC cases, which found that developments located in low-income neighborhoods consistently had a positive effect on the property values of their host neighborhoods [36–38]. In contrast, Koschinsky [19] also found that Section 8 housing units, especially in larger concentrations, consistently had negative effects in wealthier regions. Overall, mixed results are often presented regarding the effect of the income status of hosting neighborhoods, though the effects tend to be relatively small [39,40].

Proximity is another variable of interest, as previous studies have applied different proximity measures and found significant effects. The proximity measures in housing price studies often refer to neighborhood boundaries or walkable distances typically ranging from 400 m to 1 km. Generally, the proximity effect is most successfully captured in a half-mile radius—or approximately 800 m, up to 1 km [36,41]. This is also applicable in Busan, the location of this study, as the proximity effects of housing prices can extend as far as 1 km [42]. In terms of the impact of affordable housing, the effect was observed in more immediate neighborhoods as close as 300 m to 400 m [34,39].

According to our review of affordable housing literature, the effect of PRH is expected to be minute—if it exists—and likely to associate with some of the physical factors of PRH and the socioeconomic factors related to host neighborhoods. Based on the studies, one can expect that the concentration of subsidized housing and larger complexes are likely to assert a negative impact on nearby property values. It is also important to note that the lack of compatibility in the quality of PRH with their host neighborhoods is also likely to bring about a detrimental effect. These factors are commonly stated as being important by scholars in affordable housing studies as well as by advocates of affordable housing policies [30,32,38,43–45].

2.3. Application of Advanced Valuation Methods

The hedonic price method is the most widely used property valuation method based on the premise that the price is determined by internal characteristics, such as the structural attributes of housing, and external factors, such as the locational attributes of housing. The advantage of the method is that it isolates the impact of a good's constituent characteristics on its market price [46]. The method is especially useful for the valuation of external factors associated with housing, including environmental goods, urban amenities, and neighborhood characteristics [42,47,48]. Despite its usefulness, the approach has been criticized due to its stringent assumptions in regards to the selection of functional forms and potential factors explaining price relationships, market disequilibrium, and segmentation [49,50]. In response, hedonic pricing has evolved to reflect on the complex and dynamic nature of the housing market. Various measures have been taken—in terms of data mining and modeling techniques—to overcome the limitations and to expand the capability of hedonic pricing. Consequently, several modified forms of regression-based methods and hybrid methods

were introduced in the valuation procedures, including the spatial analysis methods, autoregressive integrated moving average models (ARIMA), and artificial neural networks (ANN).

The recent advancement of spatial datasets and tools, such as the geographic information system (GIS), has allowed for the implementation of GIS-based spatial analytic methods in housing price studies. One of the prominent applications of spatial analytics in housing price studies is the geographically weighted regression (GWR), which is a nonparametric weighted local regression technique that recognizes the spatial correlation of the observations [51]. When integrated with spatiotemporal data, the method is advantageous in characterizing the spatial heterogeneity and nonstationarity of housing prices, concerning potential price determinants at various spatial scales [52,53].

With respect to the analysis of time-series data, the autoregressive integrated moving average (ARIMA) model is one of the most popular methods in housing price studies. Mainly used in price forecasting, ARIMA models are quite flexible in that they can be used to model different types of time series such as the autoregressive (AR), moving average (MA), or the combination of AR and MA (ARMA) series. Despite the simplicity and flexibility, the major drawback of ARIMA is the presumed linear form of the model, which makes it inadequate for representing real-world problems. It is also unsuitable for making long term predictions, especially with a series with one or more turning points [54]. A vector autoregressive (VAR) model has advantages over ARIMA as it can incorporate multiple time series in a single model. VAR models generalize the univariate AR model by allowing for more than one evolving variable [55]. The efficiency and predictive power of VAR forecasting are well established in economic and financial studies.

One of the most recent applications in the advanced valuation method is the evolutionary polynomial regression (EPR) model, which combines the power of the genetic algorithm with numerical regression to develop symbolic models. It is based on the evolutionary computation algorithms that aim to search for the polynomial structure and a set of explanatory vectors that best represents a system through continuous iteration. The methodology is advantageous in that it is capable of returning an explicit model expression for the relationship without requiring a pre-defined functional form and variables [51,56].

2.3.1. Application of Machine Learning

The artificial neural network (ANN) is suited to provide the flexibility needed in modeling housing prices [17,57]. The input and output parameters are identical to those in the hedonic pricing method; however, ANN is both nonlinear and nonparametric in that it does not require a specific functional form or assumptions in characterizing the price relationship. Accordingly, there are three primary components within ANN, which can be referred to as layers: The input layer, hidden layer, and output layer. Each layer consists of nodes, which are connected with the nodes in the adjacent layers, as shown in Figure 1. The hidden layer serves two functions: Applying weights to the inputs and transforming to the outputs through activation functions. This framework allows for the "learning" of data patterns through continuous iterations of feeding inputs to match outputs to compose the best fitting functional form of the relationship, which will develop the final predictions. The ANN's applicability in real estate valuation has significantly improved with the recent progress in computing power as well as the availability of large housing sales data. Compared with the early application of the ANN in real estate valuation [58,59], the extensions of the model were developed with the capability to handle more complex logic by introducing additional layers and advanced algorithms.
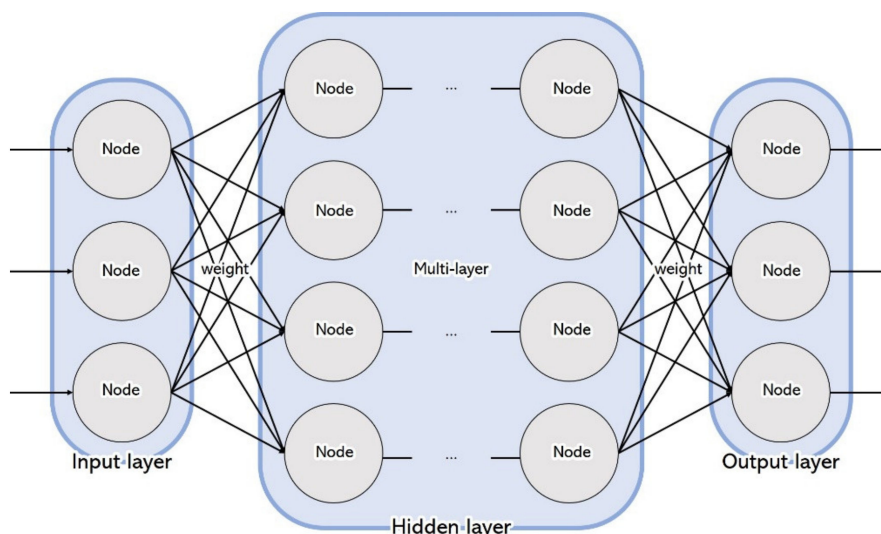
Figure 1. Operational map of artificial neural network (ANN).

The RNN is one of the extensions of ANN that is better suited for handling time series data as it assigns the processes in time step [60,61]. Because of the construction of ANN, which involves treating each item of data as an independent entity, it is not capable of accommodating the hierarchy and scale of time-dependent variables [62]. While the network of the RNN is seemingly similar to that of the ANN, the nodes in the hidden layer(s) of the RNN have the capability of memorizing the outputs from the previous time step in a space called the "hidden state," enabling more suitable learning processes for time series data [63,64]. Another advancement is the backpropagation algorithm, which allows for the efficient training of a multilayer network such as the RNN. In short, backpropagation is the process of computing the gradient of the loss function and the "backward passing" of errors to update the weights used in learning processes. However, the gradient can vanish by becoming too small to matter; moreover, it can also explode by becoming too big for our viewpoint over long time windows caused by the continuous matrix multiplication in backpropagation [65]. The issue of exploding gradients can be reduced by the rescaling of the gradient within given boundaries. However, the issue of vanishing gradients is more complicated as it could become and remain as zero, which would diminish the learning processes.

Following the limitation of the RNN, the LSTM was introduced to mitigate the issue of the vanishing gradient [66]. LSTM is a cell that modifies the "hidden state" in the RNN by adding a "cell state" that can selectively remember the past. The process starts with three inputs into the LSTM cell: The new input value at time t, x_t; the "previously memorized" cell state from the previous time step, C_{t-1}; and the output from the previous time step, h_{t-1}. The values follow the arrows and go through the operations as shown in Figure 2. The operations occur at three "gates" in order to regulate the flow of information through the LSTM: The forget gate (f_t), input gate (i_t), and output gate (o_t). There are two activation functions at the gates, the first of which is the sigmoid function, denoted as 'σ' that outputs a value between 1 and 0 to govern a value to flow or not to flow through the gates. Then there is the hyperbolic tangent function denoted as "tanh," which outputs a value between -1 and 1, which gives weightage to the presentation of their relative importance. "W" and "b" represent the weightage and bias for the respective gate.

At the forget gate, the sigmoid function decides which value should be discarded—or "forgotten"—by examining the value from the previous hidden state and the new input as calculated in Equation (1). At the input gate, the sigmoid function determines how to update the cell state by transforming values to between 0 and 1, as presented in Equation (2). The sigmoid output of the input gate is multiplied with the potential cell state value, C_t', which is calculated with Equation (4). The product is then added to the previous cell state value factored by the output of the forget gate

to produce the new cell state value, C_t, calculated as Equation (5). At the output gate, calculated as Equation (3), the sigmoid function determines how the hidden state value should be updated. The sigmoid output of the output gate is multiplied with the tanh output of the new cell state value to decide which hidden state value should be carried, as presented in Equation (6). Finally, the new cell state and the new hidden state are carried onto the next time step. In series, the neural network is capable of selectively remembering patterns for long durations of time until it is needed, making it effective in handling long and short temporal dependence in sequence prediction modeling [67].

$$f_t = \sigma\big(W_f \times [h_{t-1}, \, x_t] + b_f\big) \tag{1}$$

$$i_t = \sigma(W_i \times [h_{t-1}, \, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o \times [h_{t-1}, \, x_t] + b_o) \tag{3}$$

$$C'_t = tanh(W_C \times [h_{t-1}, \, x_t] + b_i) \tag{4}$$

$$C_t = f_t \times C_{t-1} + i_t \times C'_t \tag{5}$$
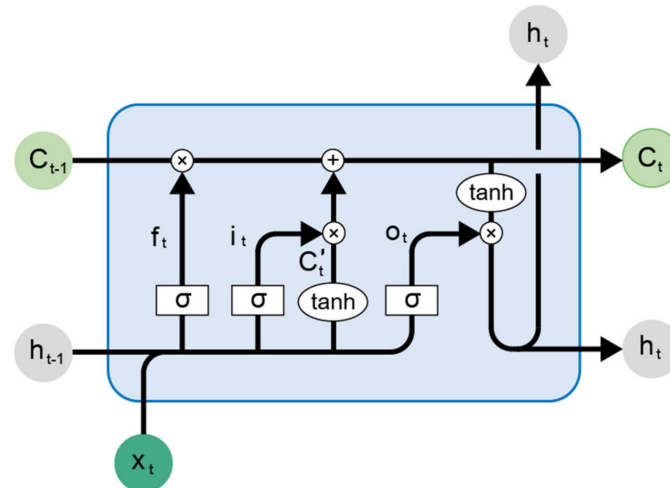
$$h_t = o_t \times \tanh(C_t) \tag{6}$$

Figure 2. Operational map of long short-term memory (LSTM) cell.

2.3.2. Case Studies: Application of Advanced Valuation Method

The ability of machine learning algorithms has been extensively utilized to predict future events in various disciplines, including finance, business, economics, and engineering. The application in analyzing housing price is relatively new compared to those disciplines, and it has recently been gaining much attention in analyzing housing markets in many cities worldwide. This section refers to those studies that specifically utilized the LSTM.

Chen (2017) compared the predictive power between time series models using a single variable in the housing prices from January 2004 to September 2016 in the Chinese cities of Beijing, Shanghai, Guangzhou, and Shenzhen. The study compared the performance of various housing price prediction models, including the autoregressive integrated moving average (ARIMA), simple RNN, and a set of LSTMs with different structures. Although it was a brief study and had limitations, including data shortage and inability to consider regional variations of housing prices, it revealed that the simple LSTM outperforms the others.

In a study conducted in Turkey [23], a comparative analysis was performed between ARIMA, LSTM, and the hybrid model that combines the two methods. The results showed that the hybrid model is the best—based on the mean absolute percentage error and the mean squared error—claiming

that the consideration of both linear and nonlinear components in housing price predictions is essential. Although the results indicate that the hybrid model performs better, the authors agreed that the predictive power of the models is very similar. Similar to the study conducted by Chen (2017), neural network algorithms are superior over traditional algorithms such as ARIMA. The study also was not able to control for the regional variation of housing prices due to the use of countrywide sales data. Furthermore, it is suspected that adjusting hyper-parameters in the analysis could improve the results, such as increasing the number of training epochs, cells, and hidden layers.

A study conducted in Seoul, South Korea [68] investigated the performance of simple RNN and LSTM in predicting apartment price indexes. All sales data for medium and large-sized apartments between January 2006 and October 2017 were used in the analysis with six macroeconomic indicators to reflect the temporal variations of the housing market. The results found conflicting results, showing higher predictive power of a vector autoregressive (VAR) model over LSTM. However, the validity of comparisons to a VAR is limited because it is difficult to identify the price relationship with each variable.

The practicality and efficiency of using LSTM in housing price predictions are evidently appealing, but further improvement in the quality of LSTM applications should be investigated. The necessary improvements include better specification of hyper-parameters, acquiring sufficient data, the inclusion of factors associated with housing price variations, and the employment of efficient comparative models. While only a handful of studies have taken this approach, the application of a neural network such as the LSTM can go beyond the role of housing price prediction, as it can be utilized to address the issues related to housing policies [22].

3. Experiment

3.1. Research Design and Data

The primary purpose of this study is to investigate the effect of PRH on the nearby housing prices using the LSTM method. Consequently, this research is designed to construct the housing price prediction model for the city of Busan, South Korea, where it can examine the changes in housing prices due to the proximity effect of the PRH. The model builds upon all of the apartment sales data for 167 months that occurred between January 2006 and November 2019, which includes 547,740 transactions at 2024 apartment complexes.

Among the different types of PRH, this study examines build-to-rent housing because it is much more common, more affordable in general, and larger in scale. Buy-to-rent and rent-to-rent units are geographically dispersed over different housing types, making them less likely to influence nearby housing prices. It is also important to note that build-to-rent housing is generally developed with amenities, neighborhood commercial units, and some infrastructure improvements in the vicinity, making it more influential in its host neighborhood. Hence, a total of 70 PRH complexes have been selected for the analysis in this study. Figure 3 shows the density map of apartment sales data, which is overlaid with the locations of the PRH units subjected to the analysis.

For better performance of the model, the model also considers two sets of independent variables: Six apartment characteristics that are known to affect housing price, and three proximity measures to PRH.

Six apartment characteristics were selected based on the findings from previous studies. They are total area, floor level, building age, the total number of households in the complex, monthly transaction volume per complex, and whether or not it is a top 10 apartment brand. According to the studies conducted in Busan, the ranking of constructors by performance is a significant factor in apartment prices [42,69].
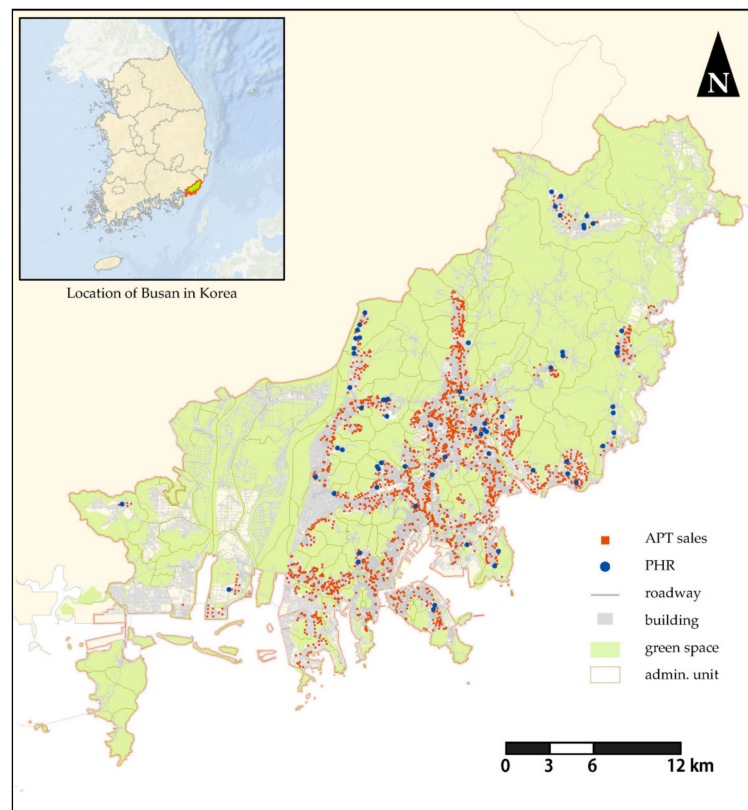
Figure 3. Study area and sample locations.

Three proximity measures to PRH consist of two measures that correspond to the nearest PRH complex, and one measure that corresponds to the comprehensive exposure to PRH units in proximity. This measure combines the extent of the size and distance to all PRH units located within 1 km of each apartment complex. The cutoff distance in defining proximity is set at 1 km based on the evidence that the proximity effect of housing price extends up to 1 km in Busan [42]. The variables used in the study are presented in Table 2, along with the description and data source.

Table 2. Variable description.

Variable	Description	Source
Apartment Price	Monthly average transaction price per m^2 for each APT complex	MOLIT (Ministry of Land, Infrastructure, and Transport) Apartment real transaction price open system
Transaction volume	Total number of transactions made for each APT complex per month	MOLIT (Ministry of Land, Infrastructure and Transport) Apartment real transaction price open system
Total area	Total lot area for each APT complex	MOLIT (Ministry of Land, Infrastructure and Transport) Building Administration System
Floor	Maximum number of floors	MOLIT (Ministry of Land, Infrastructure and Transport) Building Administration System
Age	Building age of APT complex at the time of transaction	MOLIT (Ministry of Land, Infrastructure and Transport) Apartment real transaction price open system
Household	Total number of households	MOLIT (Ministry of Land, Infrastructure and Transport) Building Administration System
Brand	1: Top 10 APT brands determined by the performance of constructor	MOLIT (Ministry of Land, Infrastructure and Transport) Building Administration System
Distance to PRH	Distance to the nearest PRH	Produced in GIS software (ArcGIS 10.3.1)
Household in PRH	Total number of households in the nearest PRH	Produced in GIS software (ArcGIS 10.3.1)
Exposure to PRH	comprehensive exposure to PRH(s) within 1 km $\sum \frac{\text{total number of households}}{\text{distance}}$	Produced in GIS software with Python (ArcGIS 10.3.1)

In an ANN, such as LSTM, the weight of each independent variable that affects the dependent variable constantly varies within the hidden layer, so it does not produce coefficients as found

in regression analysis. On the other hand, it is possible to identify the relative importance of variables by comparing the change in the predictive power of a model when variables are added. Through this approach, it is possible to select a set of independent variables that produces the best prediction model, which is optimized for the housing market in Busan. Following the additive scheme of model building, four prediction models, which are referred to as "constructs" to avoid confusion with the final models of the study, are presented in Table 3.

Construct 1: The most basic price prediction model constructed with the sequence of housing prices.

Construct 2: The apartment characteristics are added to Construct 1. Considering that apartment characteristics are known to be strong determinants of the price, it is likely that the predictive power will increase as they are added in the model. Thus, the change in predictive power will be able to determine the relative importance of apartment characteristics on housing prices.

Construct 3: Two of the proximity measures to the nearest PRH are added to Construct 2. If the addition of those variables increases the predictive power from Construct 2, then it suggests that the distance and size of the nearest PRH have a meaningful effect on housing prices.

Construct 4: The comprehensive measure of exposure to PRH within 1 km is added to Construct 3. If the addition of the variable increases the predictive power from the best performing model in the previous stage, then it suggests that the exposure to closer and larger PRH has a meaningful effect on housing prices.

Table 3. Model building scheme.

	Construct 1	Construct 2	Construct 3	Construct 4
Apartment Price	O	O	O	O
Apartment Characteristics				
Transaction volume		O	O	O
Total area		O	O	O
Floor		O	O	O
Age		O	O	O
Household		O	O	O
Brand		O	O	O
Proximity to PRH(s)				
Distance to PRH			O	O
Household in PRH			O	O
Exposure to PRH				O

The best "construct" was applied to produce the final models using the neural network, as presented in the following descriptions.

Model 1. This model is subjected to the entire region of Busan. Herein, the housing price prediction model at the city level can be referred to as the regional model, which represents the regional housing market.

Model 2. This model is subjected to the neighborhoods (or areas within 1 km) of PRH. The housing price prediction model at the neighborhood level can be referred to as the local model, which represents the local housing market.

Model 3. This model is subjected to the neighborhoods of PRH units that are located in the "favorable" local housing markets, where the average change in housing price is above that of Busan, which is based on the transaction data used in this study. The observed price trends of the neighborhoods in Busan that have "favorable" local housing markets are also included for comparison.

Model 4. This model is subjected to the neighborhoods of PRH units that are located in the less favorable local housing markets, where the average change in housing price is below that

of Busan, based on the transaction data used in this study. The observed price trend of all neighborhoods in Busan that have "less favorable" local housing markets is also included for comparison.

3.2. Specification of Long Short-Term Memory (LSTM)

In the modeling of housing prices using LSTM, this study used Tensorflow, a Python-based open-source software that is widely used—especially in the application of multilayered neural networks. The overall mechanism of the LSTM cell in an RNN is discussed in the previous section (see Section 2.3); the details regarding the implementation of LSTM are presented here.

In order to maximize the performance of the model, the neural network needs to be optimized for the research questions and data. This can be done by selecting suitable hyper-parameters, or ways to set up for the learning processes to make valid predictions. However, there is not a rule of thumb for optimizing neural networks, and it is very difficult to put relative importance on one over another [70]. Thus, the optimization process relied on fine-tuning our network by testing various trials of different combinations of hyper-parameters; moreover, the overall performance determined by the root mean square error (RMSE) was evaluated. The starting point of the optimization was determined based on various sources ranging from the previous housing price literature to the related information archived from the Internet.

Prior to the trials, 70% of the data for training and 30% for validation were assigned. The exponential linear unit (ELU) in the rectified linear unit (ReLU) family was chosen as the activation function between LSTM layers. It is one of the most commonly used activation functions in neural networks—including RNN. Furthermore, it is known to outperform both sigmoid and hyperbolic tangent functions in terms of computational efficiency, and it also can handle a vanishing gradient issue [71]. We employed the He initialization method, which is similar to the Xavier method, yet it is known to provide a controlled initialization—hence, the faster and more efficient gradient descent [72]. We confirmed the suitability of the activation function and the initialization method by observing the weight variation of each layer along with ensuring that the distribution does not lean toward 0 and 1. A dropout value of 0.5, the batch size of 1, and the learning rate of 0.001 were used to allow for more rigorous learning processes as it can obtain larger data for model validation with a higher number of batches used for learning. The Adam optimizer was used in the study because it is computationally efficient, does not suffer any major decrease in accuracy, and has a relatively wide range of successful learning rates [73]. The number of epochs was determined to be 5000, which was done by identifying the "elbow" on the learning curve, where the slope of the RMSE changes. In all of our cases, it occurred near the epoch of 5000.

While the abovementioned hyper-parameters stay fixed for the model-building processes, we adjusted the number of cell units, hidden layers, and sequence lengths to acquire the best performing network. It was given that those parameters are considered to impact the performance of networks more than the others. The number of cell units and hidden layers tested were determined based on the previous cases from various sources with fairly similar data sizes. Considering the commonly discussed cycle of housing prices, we tested the four sequence lengths of 12, 6, 3, and 1 to reflect the annual, semiannual, quarterly, and monthly patterns of housing price variations. Accordingly, we considered the combinations of the hyper-parameters for Models 1 through 4 and evaluated the performances by observing the minimum RMSE, as presented in Table 4.

Table 4. Hyper-parameter optimization trials.

Hyper-Parameters		
#Cell Unit	#Hidden Layer	Sequence Length
3/5/7	3/5	1/3/6/12

Despite the robustness of neural networks presented in previous studies, the performance of the networks is greatly influenced by a set of hyper-parameters, or the specifications of a network's structure, function, learning process, etc. A practical way to ensure the level of performance is to test the performance of neural network against the outcome of a traditional time series analysis. Some of the applicable methodologies include the autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), vector autoregression (VAR) and its extensions. In this study, the VAR model was selected as a comparative model because it allowed for more than one evolving variables and showed to have the best forecasting results [68,74]. The VAR model was applied to Model 1 and Model 2.

VAR model is a stochastic process model that identifies linear interdependencies among multiple time series. It is an extension of the univariate autoregression model to multivariate time series data that consists of a system of relationships in which each housing price is explained by its own lags as well as the lags of other variables [55]. Prior to the modeling of time series using VAR, it is important to ensure the stationarity of the times series data to prevent the risk of spurious regression. In the case of nonstationarity, necessary data transformation, or differencing, is needed prior to the analysis [75]. The stationarity of time series can be tested with the presence of unit root. The augmented Dickey–Fuller (ADF) test was employed to determine the presence of unit root in our data. The results of the tests revealed that each housing price series is integrated of order one, I (1), and thus the first difference of housing price is stationary. Upon transformation, the ADF test confirmed that all variables were significant at the 5% level. For the selection of optimal lag length in VAR, several lag lengths were tested so that the residuals do not have a serial correlation. Relying on the Akaike's information criterion (AIC) and Bayesian information criterion (BIC), both Model 1 and Model 2 were analyzed with the lag length of one (the results for ADF test and VAR specification can be provided upon request). Similar to the approach taken with LSTM, four constructs were tested and evaluated using the RMSE.

4. Results and Discussion

4.1. Analysis of the Data

In this study, data from a total of 157,740 apartment sales were gathered over a period of 167 months from January 2006 to November 2019 in Busan, Korea. These data were used to investigate the impact of PRH on nearby housing prices. In this study, a neural network—the RNN-LSTM—was used to construct a set of housing prediction models that represent the regional and local housing markets with a focus on identifying differences in trends of housing prices by their proximity to PRH. A set of apartment characteristics and proximity measures were used in the sequence prediction of housing prices, which can act as a set of control variables in explaining price trends. The descriptive statistics of the variables used in this study are presented in Table 5.

Although it is not a significant difference, the average transaction price of apartments in the proximity of PRH is slightly higher than that of the entire city. The average monthly transaction volume of PRH neighborhoods is slightly less than that of the entire city. Among PRH neighborhoods, the transaction volume was lower in the "favorable" local housing market and higher in the "less favorable" market. Observing the floor and age variables, the apartment complexes applied in the models are quite similar in height and building age. Compared to Models 2 and 4, the total number of households in Model 3 is smaller, while the total lot area is larger. This implies that the apartments in the neighborhoods of PRH in the "favorable" local housing market have larger unit sizes or larger on-site facilities and amenities. This is also supported by the higher value of the apartment brand variable in Model 3, as the apartments built by prominent apartment builders are known to provide higher-quality on-site facilities and amenities such as park space, neighborhood commercial sites, and parking lots [42].

Table 5. Descriptive statistics.

		Model 1		Model 2		Model 3		Model 4	
Variable	Unit	Mean (Std. dev.)		Mean (Std. dev.)		Mean (Std. dev.)		Mean (Std. dev.)	
Apt. Price	10,000 KRW/m^2	263.27	(73.18)	266.87	(73.76)	292.47	(89.29)	259.31	(69.25)
Trans. vol.	sales/mo.	76.05	(7.97)	68.16	(7.18)	61.58	(19.74)	69.76	(6.07)
Total area	1,000m^2	118.41	(11.89)	98.81	(10.51)	114.99	(35.86)	94.27	(6.84)
Floor	#of floors	21.00	(0.79)	21.40	(0.91)	21.25	(2.67)	21.38	(0.75)
Age	years	14.74	(2.77)	14.38	(2.66)	14.13	(2.32)	14.50	(2.87)
Household	# of households	950.99	(67.51)	885.16	(82.25)	783.75	(162.86)	911.71	(89.91)
Brand	dummy	0.10	(0.03)	0.11	(0.03)	0.18	(0.07)	0.09	(0.02)
Dist. to PRH	m	1571.36	(78.1)	571.30	(17.75)	595.05	(40.78)	565.21	(17.77)
Household in PRH	#of households	633.87	(37.59)	789.52	(58.35)	454.16	(96.02)	878.20	(59.90)
Exposure to PRH	number	0.29	(0.22)	0.64	(0.44)	0.25	(0.07)	0.74	(0.56)

For the apartments located in Busan, the average distance to the nearest PRH is approximately 1.5 km. For the apartments located in the neighborhoods of PRH, the distance to the nearest PRH complex is similar, the size of the complex and the exposure variable revealed some differences for Models 3 and 4. The statistics indicate that the average number of households in the PRH complexes in Model 3 is significantly less than that of the other models, meaning that the PRH complexes in the "favorable" neighborhood are smaller than in the other geographic boundaries used in the study. Furthermore, the average of exposure variable in Model 4 is significantly larger than that of Model 3, meaning that the PRH complexes are more concentrated in "less-favorable" neighborhoods.

Overall, the statistics showed that the models representing different geographic boundaries correspond to different spatial characteristics in terms of prices, apartment characteristics, and PRH characteristics. The differences between Models 3 and 4 are notable, showing that apartment complexes located in neighborhoods with "less favorable" local housing markets are smaller in lot size, have more households, and are less likely to be built by prominent apartment constructors. The PRH complexes located in "favorable" neighborhoods are larger in size and are likely to be in clusters. This aligns with the limitations of the build-to-rent type of PRH discussed in Section 2.1, that large-scale PRH complexes are likely to be located in areas with lower land value. The following analyses examine the effect of PRH variables more comprehensively by testing the effect of PRH variables in the price prediction model in each geographic boundary and comparing the predicted trend to the observations.

Next, the Moran index (Moran's I) for each variable was calculated to assess the spatial pattern of the variables, as presented in Table 6.

Table 6. Moran's I for Busan, Korea.

Variable	Apt. Price	Trans. Vol.	Total Area	Floor	Age	Household	Brand	Distance to PRH	Scale of PRH
Moran's I *	0.892	0.117	0.115	0.162	0.246	0.139	0.589	0.761	0.704

* All variables were significant at the 1% level ($p < 0.01$).

According to the result, the apartment price, brand, and the two proximity variables showed spatial autocorrelation, while other variables showed no correlation. The average transaction price for the apartments was revealed to have high spatial autocorrelation, which was not surprising, given that it is commonly found in real estate prices [42,76]. Considering that the apartments built by prominent constructors are more expensive, the spatial clustering of top-brand apartments aligned with our expectations. The average transaction volume is randomly dispersed, meaning that the number of sales does not have any spatial pattern in Busan. Similarly, the attributes of apartment complexes, including the total lot size, number of floors, building age, and the total number of households, are randomly dispersed throughout the city. Based on the moderate to the high level of spatial autocorrelation found in the proximity variables, the clustering pattern of PRH complexes exists, both in their location and size.

4.2. Model Production

As the performance of machine learning algorithms is greatly influenced by hyper-parameters that specify the network's structure, function, learning process, etc., the selection of parameters optimized for data and research objectives is important. We relied on the trial-and-evaluations of different combinations of hyper-parameters and variables in our model-building process, which featured three sets of hyper-parameters and four "constructs" of variables. As a result, the optimized construct was obtained in terms of the hyper-parameters and independent variables that produced a price prediction model with the most predictive power for the entire region of Busan (Model 1) and for the neighborhoods of PRHs (Model 2). A set of hyper-parameters selected for Model 1 and Model 2 are presented in Tables 7 and 8, respectively.

Table 7. Optimization results for Model 1.

		Construct 1	Construct 2	Construct 3	Construct 4
Hyper-parameters	Fixed	Learning / Validation = 70/30 Activation function = ELU Initialization = He Optimizer = Adam Batch size = 1 Learning rate = 0.001 Dropout rate = 0.5 # of epochs = 5000			
	# of cell unit	3	7	7	7
	# of hidden layer	3	3	5	3
	Sequence length	1	1	1	1
RMSE	LSTM	0.0563	0.0328	0.0317	0.0338
	VAR	0.0849	0.0703	0.0548	0.0658

Table 8. Optimization results for Model 2.

		Construct 1	Construct 2	Construct 3	Construct 4
Hyper-parameters	Fixed	Learning / Validation = 70/30 Activation function = ELU Initialization = He Optimizer = Adam Batch size = 1 Learning rate = 0.001 Dropout rate = 0.5 # of epochs = 5000			
	# of cell unit	3	5	7	5
	# of hidden layer	3	3	5	5
	Sequence length	1	1	1	1
RMSE	LSTM	0.0670	0.0594	0.0573	0.0407
	VAR	0.0796	0.0815	0.0626	0.0494

For Model 1, Construct 3 showed the most predictive power with an RMSE value of 0.0317. Notably, this construct included two variables regarding the nearest PRH, which suggests that those variables have a meaningful effect on predicting housing prices in Busan.

Model 2, which represents the local housing market of the neighborhoods of PRH, showed that Construct 4 produced the most predictive power, with an RMSE value of 0.0407. Given that Construct 4

included all three proximity variables of PRH, this result indicates that the variables are more relevant in the neighborhoods of PRH than in the greater Busan region.

Regarding the use of VAR, the RMSE value of 0.0548 revealed that Construct 3 produced the most predictive power for Model 1. Similarly, the RMSE value of 0.0494 revealed that Construct 4 produced the most predictive power for Model 2. While the selection of a model construct was the same as the LSTM, the LSTM results showed better performance in all cases. Nonetheless, VAR results were also shown to have good predictability.

4.3. Analysis Results of the Proximity Effect of PRH

A total of four price prediction models were produced using the LSTM optimized for the city as well as the PRH neighborhoods. As specified, 30% of the data were used for prediction, which equates to 50 months. The graphical representation of each prediction is presented. The RMSE values indicate the performance of each final model, as presented in Table 9.

Table 9. Performance of final prediction models.

	Subject	RMSE
Model 1 (construct 3)	APT prices for the entire city	0.0317 (0.0548) *
Model 2 (construct 4)	APT prices for the neighborhoods of PRH	0.0407 (0.0494) *
Model 3 (construct 4)	APT prices for the neighborhoods of PRH located in "favorable" local housing market	0.0490
Model 4 (construct 4)	APT prices for the neighborhoods of PRH located in "less favorable" local housing market	0.0489

* The values in parenthesis indicate the RMSE values of the predictions according to the VAR model.

Figure 4 presents the results of Model 1. Comparing the RMSE value of 0.0317 to 0.0548, the LSTM model achieved a lower error rate than the VAR model. The graphical representations show that the trend breaks and the segment slopes in the predictions resemble those of the observations very well. In the case of the VAR model, it can be observed that the predicted slopes often conflict with the observed slopes. Further, the VAR predictions often deviate at the high and low points, meaning that the model tends to overestimate the prices following the inclining price trend and underestimate following the declining price trend.

Figure 5 presents the results of Model 2. Again, the LSTM model (RMSE = 0.0407) achieved better predictive power than the VAR model (RMSE = 0.0494). This is also supported by the graphical representation, as the predicted trend using LSTM is more stable around the observed trend than that of the VAR. Compared to the LSTM model, the predicted trend using VAR showed sharper drops at the low points, meaning that the model tends to underestimate the prices following the declining price trend.

The results showed that price trends in PRH neighborhoods are more turbulent than those of the city overall. This was expected, given that the neighborhood-level models have smaller sample sizes. Regardless, our price prediction models showed good performance when both the city and PRH neighborhoods were concerned. On the other hand, it was not expected to find that the predictive power of the regional model was higher than that of the local model, even though the proximity variables were determined to have meaningful effects in our modeling processes.
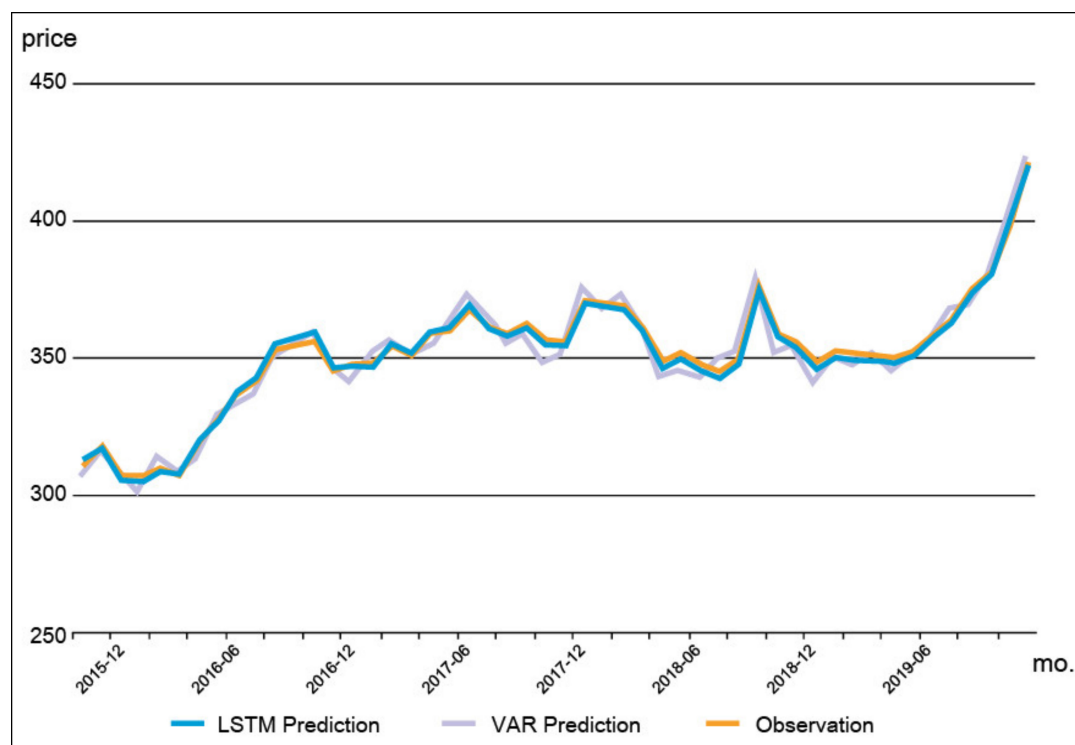
Figure 4. Model 1—Apartment price predictions for the entire city.
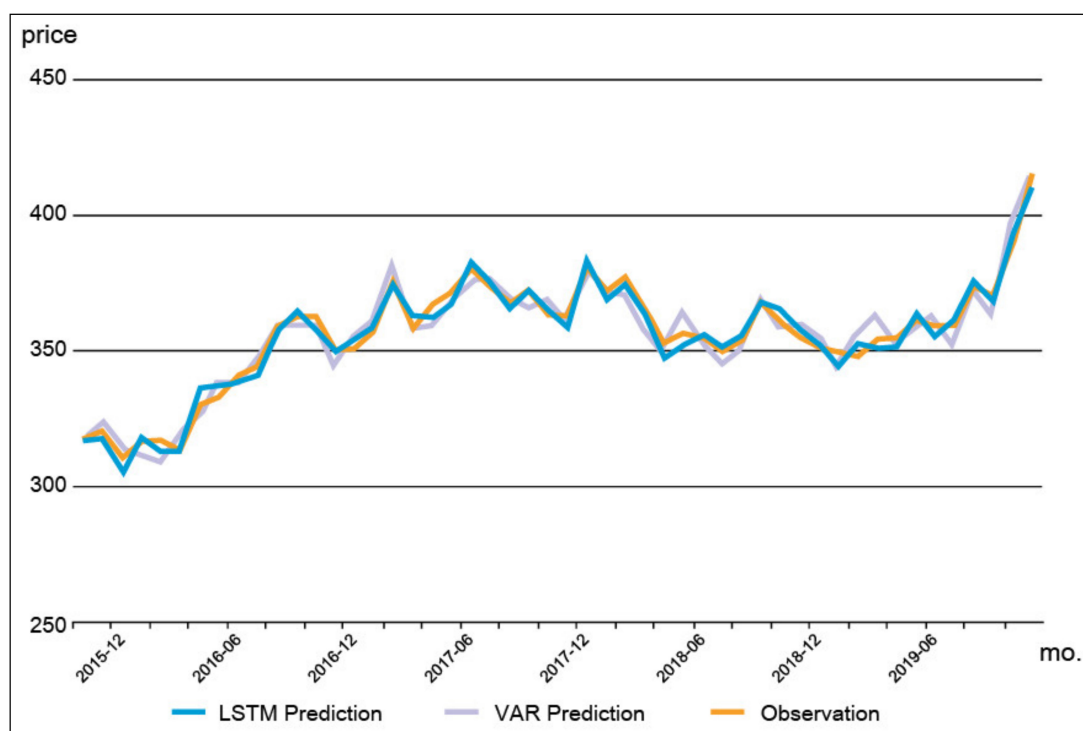

Figure 5. Model 2—Apartment price predictions for PRH neighborhoods.

While Model 1 and Model 2 provided some general information on housing price trends in Busan and PRH neighborhoods, the study identified the need for closer observations at the neighborhood level. Further examination of PRH neighborhoods showed that some of the characteristics, including the average housing price and monthly transaction volume, are quite different by neighborhood.

In particular, it was apparent that most of the newer PRH complexes are located along newly constructed apartment complexes or in recently developed new town projects that often featured a simultaneous construction of multiple apartment complexes. Consequently, the local housing market in those neighborhoods was suspected to be unstable, with a rising trend. It was also found that many of the larger PRHs in our sample were older and were located in areas in which older and cheaper apartments are concentrated. As the PRH neighborhoods were split into two groups based on the average price changes in the housing market over the study period, it was possible to confirm that the price trends, both in prediction and observation, were distinctly different.

Figure 6 presents the results of Model 3. This model was subjected to the neighborhoods of PRH located in the "favorable" local housing market, where the average change in the apartment prices was above the average of the city. As expected, the price variation was much more turbulent than the other cases considered.
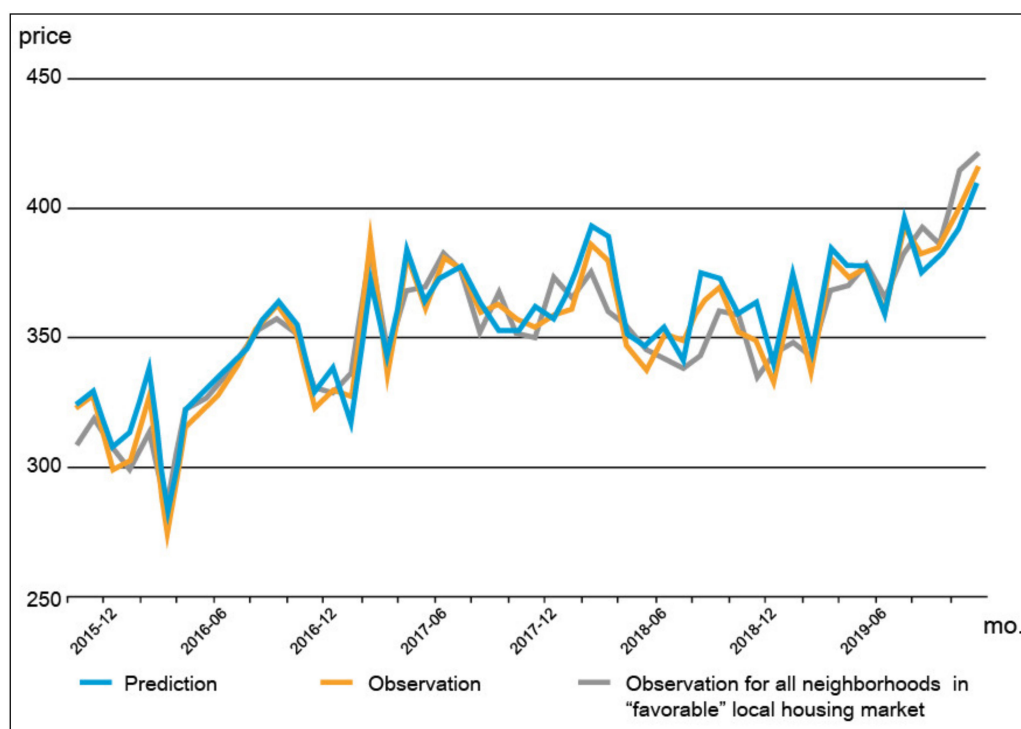
Figure 6. Model 3—Apartment price predictions for PRH neighborhoods in the "favorable" local housing market.

Lastly, Figure 7 presents the results of Model 4, which represents PRH neighborhoods located in the "less-favorable" local housing market in which the average change in apartment price was below the average of the city. With an RMSE of 0.0489, the predictive power of the model was found to be practically the same as Model 3, though the trend is distinctively different. The price trend of Model 4 is much smoother and has fewer jumps or drops compared to the other cases.

Another noteworthy result is that the predicted values are slightly above the observed values throughout the period, except for the last month. The overall trends of predicted and observed values match well, although the prediction started to offset upwards at the beginning of the prediction period and partially recovered toward the end of the prediction period. Additionally, the prices are mostly higher throughout the prediction compared to the rest of the neighborhoods in the "less-favorable" local housing market in Busan. This trend is unique in that it presents a unidirectional error of predictions that were consistent over time. This is likely caused by some unknown factors that were not considered in the study and which may be unique to those neighborhoods. It is difficult to connect this trend to the effect of PRH because the duration of the error is unprecedentedly long compared to the other

models. Similar magnitudes of error were observed in the other models considered in the study, though they were soon caught up by new predictions. In other words, external factors that were not accounted for in the learning process of the model must have depreciated the prices for the duration of the gap.
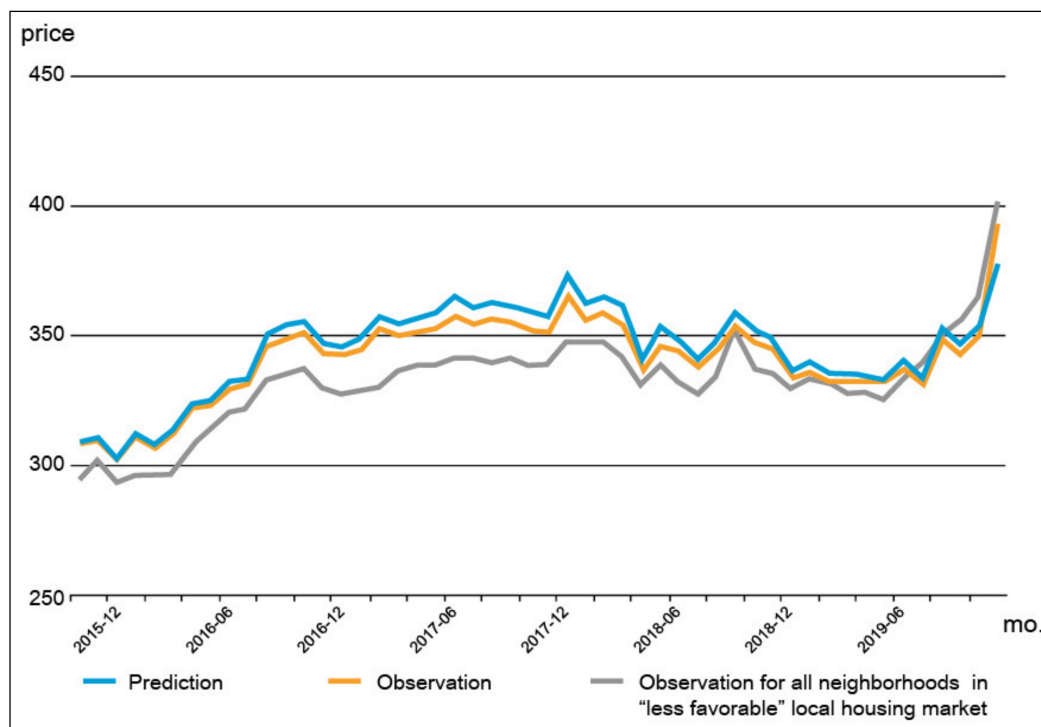

Figure 7. Model 4—Apartment price predictions for the neighborhoods of PRH in the "less favorable" local housing market.

Overall, the impact of PRH on nearby housing prices is meaningful as the proximity variables employed in the models were determined to be the factors that improve the performance of predictions. The variables were determined meaningful in both Model 1 and Model 2, which represented the price prediction for the entire city and for the PRH neighborhoods, respectively. While both LSTM and VAR forecasting were effective, the predictions made with the LSTM was more accurate in all models considered. Figure 8 presents the price predictions constructed in each model.

Given the outstanding accuracy of predictions, the models are capable of making reliable price predictions for the future. Comparing the predictions in Model 1 and 2, the trend of housing prices in the PRH neighborhoods is more turbulent than that of the city. In contrast, the overall trends of the models throughout the prediction period are very similar. The turbulence intensifies in the PHR neighborhoods located in the "favorable" local housing market. While it is unable to address the individual effects of the variables on the predictions, the results indicate that forecasting housing prices in the PHR neighborhoods requires a careful selection of prediction models, especially when making short-term predictions in the neighborhoods with more turbulence in price trend. The results from Model 4 indicated that the effect of external factors on the housing prices is more evident in the PHR neighborhoods located in the "less favorable" local housing market. By comparing with other models, the overall price prediction in Model 4 remained similar until the end of 2018, when it started to fall below. Once the drop occurred at the end of 2018, the price trend remained below the city's trend by a more considerable margin than ever recorded throughout the study period.
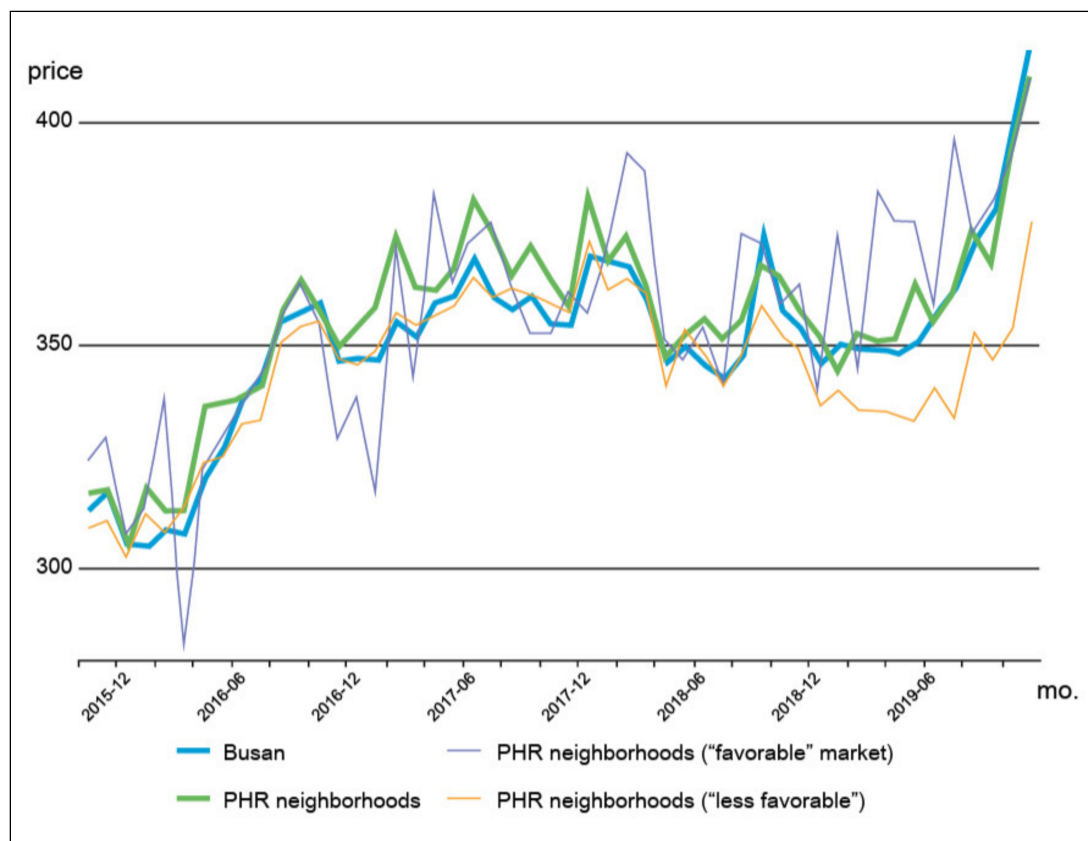
Figure 8. Comparison of the predictions.

5. Conclusions

In this study, the proximity effect of PRH on housing prices was investigated with four price prediction models using long short-term memory (LSTM). The models utilized apartment transaction data over 167 months, along with the potential factors of housing prices. The results produced a set of price-prediction models for the city and neighborhoods of PRH. Despite the limited interpretation of the predictions over the pre-existing price data, the merit of the approach taken in this study lies in its ability to predict housing prices, both in the long and short term, with improved accuracy. This is meaningful, as it can be used to evaluate and simulate changes in the regional and local housing markets.

The model building process showed that the proximity measures have a meaningful impact on housing prices, both for the entire region of the city and PRH neighborhoods. The price trends in PRH neighborhoods showed some turbulence but did not consistently fall below the city's trends. Turbulence intensified for PRH neighborhoods located in the "favorable" housing market, although the general trend remained similar to that of the city. An exceptional case was presented when PRH neighborhoods located in the "less favorable" local housing market were considered, as an offsetting effect was observed between the trends. The results show that the predicted pattern resembled the observed pattern very well, but with a slight upwards offset for almost the entire period. This result is meaningful not only because it is unique to the spatial boundary selected, but also because it presented a unidirectional error of the prediction model that lasted over time. This suggests that the housing prices in those neighborhoods were likely suppressed by other external factors that were not accounted for in this study, such as the effect of housing policy, macroeconomic factors, and local housing market factors. Investigating the drivers of this trend is important because it may hold the key to identifying the determinants of the local price depreciation, which may be more concerning than the effect of PRH in proximity. This can be done using the technique employed in the study, by considering

additional variables such as macroeconomic factors, housing policy factors, and local housing market factors, and identifying the set of variables that closes the gap. It is also recommended that future studies take a combined approach by employing additional methodologies such as regression-based models to identify the individual effects of the variables concerned. This can be useful for detecting the cross-sectional or longitudinal relationship at certain parts of the price trend in question, such as the gap observed in Model 4 in this study. The approach can also determine whether or not the proximity effect is detrimental to nearby housing prices.

The application of machine learning techniques for constructing a housing price model was outstanding. In this study, the LSTM managed to produce a set of price-prediction models efficiently, with less margin of error than a traditional time series model. Although the interpretation of the results was limited in assessing the effects of individual factors, the approach taken produced a set of reliable forecast models for the city and the PRH neighborhoods.

With respect to the implication of the research, the modeling approach employed in this study can assist in the decision-making process in the planning, development, and management of PRH provisions. For the planners, forecasting housing prices in the neighborhoods of PRH can help with the planning processes of PRH as an informative medium for the public and residents who fear the construction of PRH. Considering that a significant number of PRH units in Korea are likely to flow into the private market in the near future, the modeling approach taken in this study can assist in search of investment opportunities by forecasting the potential profit of the units in the private market. Additionally, city planners can supplement the evaluation and simulation processes of the PHR plans to ensure the economic stability and viability of PRH provisions and their host neighborhoods. With further expansion of the modeling approach, various socioeconomic aspects can be tested to address urban issues, including the infrastructure/resource deprivation and spatial polarization due to the deteriorated PRH provisions. The developers of PRH can expand the approach with a more detailed characterization of PRH complexes and neighborhoods in search of better design and locations for PRH.

Lastly, the superiority of the machine learning technique in analyzing large time-series data can be valuable in assessing the impact of various built environment factors, market factors, and the impact of housing policies over time. This can be useful in understanding the effect of certain housing policy and market dynamics over time based on big-data, and simulating the response of the housing market. Its efficiency can also assist in the monitoring of various aspects of PRH to help improve the quality of living for residents and neighborhoods. Ultimately, this effort can create opportunities for PRH and its host neighborhood to benefit from each other and construct a healthy and sustainable community.

References

1. Monroy, A.M.; Gars, J.; Matsumoto, T.; Crook, J.; Ahrend, R.; Schumann, A. *Housing Policies for Sustainable and Inclusive Cities*; OECD Regional Development Working Papers 2020; OECD Publishing: Paris, France, 2020. [CrossRef]
2. World Bank Systems of Cities: Harnessing Urbanization for Growth & Poverty Alleviation. The World Bank Urban & Local Government Strategy. 2010. Available online: http://documents1.worldbank.org (accessed on 1 September 2020).

3. Tusell, M.S. Affordable Housing in Europe: Innovative Public Policies that Can Effectively Address the Housing Crisis. Barcelona Centre for International Affairs. 2017. Available online: https://cidob.org (accessed on 1 September 2020).

4. Dawkins, C. Realizing housing justice through comprehensive housing policy reform. *Int. J. Urban. Sci.* **2020**. [CrossRef]

5. Galster, G.; Lee, K.O. Housing affordability: A framing, synthesis of research and policy, and future directions. *Int. J. Urban Sci.* **2020**. [CrossRef]

6. Galdini, R.; Lucciarini, S. Social Innovation and Environmental Sustainability in Social Housing Policies: Learning from Two Experimental Case Studies in Italy. In *Different Strategies of Housing Design*; IntechOpen Limited: London, UK, 2019. [CrossRef]

7. Li, D.; Chen, Y.; Chen, H.; Guo, K.; Hui, C.M.; Yang, J. Assessing the integrated sustainability of a public rental housing project from the perspective of complex eco-system. *Habitat Int.* **2016**, *53*, 546–555. [CrossRef]

8. Wu, G.; Duan, K.; Zuo, J.; Zhao, X.; Tang, D. Integrated sustainability assessment of public rental housing community based on a hybrid method of AHP-Entropy weigh and cloud model. *Sustainability* **2017**, *9*, 603. [CrossRef]

9. OECD Better Life Index. Available online: http://www.oecdbetterlifeindex.org/topics/housing/ (accessed on 3 January 2020).

10. Park, J.H.; Yu, J.; Geem, Z.W. Optimal project planning for public rental housing in South Korea. *Sustainability* **2020**, *12*, 600. [CrossRef]

11. Currie, J.; Yelowitz, A. Are public housing projects good for kids? *J. Public Econ.* **2000**, *75*, 99–124. [CrossRef]

12. Finkel, M.; Climaco, C.; Elwood, P.; Feins, J.; Locke, G.; Popkin, S. *Learning from Each Other: New Ideas for Managing the Section 8 Certificate and Voucher Programs*; Department of Housing and Urban Development: Washington, DC, USA, 1996.

13. Kean, T.; Ashley, T. *"Not in My Back Yard" Removing Barriers to Affordable Housing*; Department of Housing and Urban Development: Washington, DC, USA, 1991.

14. Scally, C.P.; Koenig, R. Beyond NIMBY and poverty deconcentration: Reframing the outcomes of affordable rental housing development. *Hous. Policy Debate* **2012**, *22*, 435–461. [CrossRef]

15. Read, D.; Tsvetkova, A. Housing and social issues: A cross disciplinary review of the existing literature. *J. Real Estate Lit.* **2012**, *20*, 3–35. [CrossRef]

16. Woo, A. Analysis research trends examining the impacts of subsidized housing developments on nearby property values in the U.S.: A critical review on empirical methodologies. *J. Korea Real Estate Anal. Assoc.* **2015**, *26*, 71–82.

17. Selim, H. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Syst. Appl.* **2009**, *36*, 2843–2852. [CrossRef]

18. Park, S.; Bae, Y. Predicting the real estate price index using machine learning methods and time series analysis model. *Hous. Stud. Rev.* **2015**, *26*, 107–133. [CrossRef]

19. Koschinsky, J. Spatial heterogeneity in spillover effects of assisted and unassisted rental housing. *J. Urban Aff.* **2009**, *31*, 319–347. [CrossRef]

20. Son, H.; Kim, C. A deep learning approach to forecasting monthly demand for residential–sector electricity. *Sustainability* **2020**, *12*, 3103. [CrossRef]

21. Hung, C.; Hung, C.; Lin, S. Predicting time series using integration of moving average and support vector regression. *Int. J. Mach. Learn. Comput.* **2014**, *4*, 491–495. [CrossRef]

22. Hu, L.; He, S.; Han, Z.; Xiao, H.; Su, S.; Weng, M.; Cai, Z. Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy* **2019**, *82*, 657–673. [CrossRef]

23. Temür, A.S.; Melek, A.; Günay, T. Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. *J. Bus. Econ. Manag.* **2019**, *20*, 920–938. [CrossRef]

24. Chen, Y.; Liu, X.; Li, X.; Liu, Y.; Xu, X. Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning. *Appl. Geogr.* **2016**, *75*, 200–212. [CrossRef]

25. Ministry of Land. Infrastructure and Transport. Rental Housing Statistics 2018. Available online: https://stat.molit.go.kr/ (accessed on 20 August 2020).

26. Cheon, H.S. Assessment of public rental housing policy and assignment. *Real Estate Focus* **2017**, *105*, 22–31.

27. Statistics Korea. Population and Housing Census. 2018. Available online: https://kostat.go.kr/ (accessed on 20 August 2020).

28. Lee, Y.M.; Seong, J.W. A Comprehensive Report on the Characteristics and Quality of Life of Public Rental Housing Tenants. Seoul Housing and Communities Corporation. 2017. Available online: https://i-sh.co.kr/ (accessed on 20 August 2020).

29. Hwang, Y.W. *Evaluation of Public Rental Housing Policy and Improvement Plan in Busan Metropolitan Area*; Policy Report 2019-04-860; Busan Development Institute: Busan, Korea, 2019.

30. Nguyen, M.T. Does affordable housing detrimentally affect property values? A review of the literature. *J. Plan. Lit.* **2005**, *20*, 15–26. [CrossRef]

31. Cummings, M.; Landis, J. *Relationships between Affordable Housing Developments and neighboring Property Values: An Analysis of BRIDGE Housing Corporation Developments in the San Francisco Bay Area*; IURD Working paper 599; Institute of Urban and Regional Development, University of California: Berkeley, CA, USA, 1993.

32. Briggs, S.; Darden, X.; Joe, T.; Aidala, A. In the wake of desegregation: Early impacts of scattered-site public housing on neighborhoods in Yonkers, New York. *J. Am. Plan. Assoc.* **1999**, *65*, 27–49. [CrossRef]

33. Kim, J. Analysis on the effect of public housing complex's location conditions on the neighborhood apartment prices in Daegu. *Korea Spat. Plan. Rev.* **2013**, *79*, 23–32.

34. Johnson, J.; Bednarz, B. *Neighborhood Effects of the Low Income Housing Tax Credit Program: Final Report*; Department of Housing and Urban Development: Washington, DC, USA, 2002.

35. Woo, A.; Joh, K.; Van Zandt, S. Unpacking the impacts of the Low-Income Housing Tax Credit program on nearby property values. *Urban Stud.* **2016**, *53*, 2488–2510. [CrossRef]

36. Baum-Snow, N.; Marion, J. The Effects of Low Income Housing Tax Credit Developments on Neighborhoods. *J. Public Econ.* **2009**, *93*, 654–666. [CrossRef] [PubMed]

37. Schwartz, A.; Ellen, I.; Voicu, I.; Schill, M. The external effects of place-based subsidized housing. *Reg. Sci. Urban Econ.* **2006**, *36*, 679–707. [CrossRef]

38. Diamond, R.; McQuade, T. Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development. *J. Political Econ.* **2019**, *127*, 1063–1117. [CrossRef]

39. Lyons, R.F.; Loveridge, S. *An Hedonic Estimation of the Effect of Federally Subsidized Housing on Nearby Residential Property Values*; Department of Applied Economics, University fo Minnesota: St Paul, MN, USA, 1993. [CrossRef]

40. Davison, G.; Han, H.; Liu, E. The impacts of affordable housing development on host neighbourhoods: Two Australian case studies. *J. Hous. Built Environ.* **2017**, *32*, 733–753. [CrossRef]

41. Zahirovich-Herbert, V.; Gibler, K.M. The effect of new residential construction on housing prices. *J. Hous. Econ.* **2014**, *26*, 1–18. [CrossRef]

42. Kim, H.; Lee, G.; Lee, J.; Choi, Y. Understanding the local impact of urban park plans and park typology on housing price: A case study of the Busan metropolitan region, Korea. *Landsc. Urban Plan.* **2019**, *184*, 1–11. [CrossRef]

43. Rossi-Hansberg, E.; Pierre-Daniel, S.; Raymond, O., III. Housing externalities. *J. Political Econ.* **2010**, *118*, 485–535. [CrossRef]

44. Santiago, A.; Galster, G.C.; Tatian, P. Assessing the property value impacts of the dispersed subsidy housing program in Denver. *J. Assoc. Public Policy Anal. Manag.* **2001**, *20*, 65–88. [CrossRef]

45. Galster, G.C.; Tatian, P.; Smith, R. The impact of neighbors who use Section 8 certificates on property values. *Hous. Policy Debate* **1999**, *10*, 879–917. [CrossRef]

46. Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]

47. Melichar, J.; Kaprova, K. Revealing preferences of Prague's homebuyers toward greenery amenities: The empirical evidence of distance-size effect. *Landsc. Urban Plan.* **2013**, *109*, 56–66. [CrossRef]

48. Hickman, E.P.; Gaines, J.P.; Ingram, F.J. The influence of neighbourhood quality on residential values. *Real Estate Apprais. Anal.* **1984**, *50*, 36–42.

49. Herath, S.; Maier, G. *The Hedonic Price Method in Real Estate and Housing Market research: A Review of the Literature*; SRE-Discussion Papers 2010; WU Vienna University of Economics and Business: Vienna, Austria, 2010.

50. Fan, G.; Ong, S.E.; Koh, H.C. Determinants of house price: A decision tree approach. *Urban Stud.* **2006**, *43*, 2301–2315. [CrossRef]

51. Locurcio, M.; Morano, P.; Tajani, F.; Liddo, F.D. An innovative GIS-based territorial information tool for the evaluation of corporate properties: An application to the Italian context. *Sustainability* **2020**, *12*, 5836. [CrossRef]

52. Chen, S.; Zhuang, D.; Zhang, H. GIS-based spatial autocorrelation analysis of housing prices oriented towards a view of spatiotemporal homogeneity and nonstationarity: A case study of Guangzhou, China. *Complexity* **2020**, *2020*, 1079024. [CrossRef]

53. Chica-Olmo, J.; Canp-Guervos, R.; Chica-Revas, M. Estimation of housing price variations using spatio-temporal data. *Sustainability* **2019**, *11*, 1551. [CrossRef]

54. Zhang, G.P. Times series forecasting using a hybrid ARIMA and neural network model. *Nuerocomputing* **2003**, *50*, 159–175. [CrossRef]

55. Sims, C. Macroeconomics and reality. *Econometrica* **1980**, *48*, 1–48. [CrossRef]

56. Morano, P.; Rosato, P.; Tajani, F.; Manganelli, B. Contextualized property market models vs. generalized mass appraisals: An innovative approach. *Sustainability* **2019**, *11*, 4896. [CrossRef]

57. Wong, Y.; Solomon, P.L. Community integration of persons with psychiatric disabilities in supportive independent housing: A conceptual model and methodological considerations. *Ment. Health Serv. Res.* **2002**, *4*, 13–28. [CrossRef] [PubMed]

58. Allan, D.; Hoesli, M.; Bender, A. Environmental variables and real estate prices. *Urban Stud.* **2001**, *38*, 1989–2000. [CrossRef]

59. Worzala, E.; Lenk, M.; Silva, A. An exploration of neural networks and its application to real estate valuation. *J. Real Estate Res.* **1995**, *10*, 185–201.

60. Werbos, P.J. Generalization of backpropagation with application to a recurrent gas market model. *Neural Netw.* **1988**, *1*, 339–356. [CrossRef]

61. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. *Learning Internal Representations by Error Propagation*; California University San Diego La Jolla Inst for Cognitive Science, Institute for Cognitive Science, University of California: San Diego, CA, USA, 1985.

62. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]

63. Hermans, M.; Schrauwen, B. Training and analysing deep recurrent neural networks. In Advances in Neural Information Processing Systems 26, Proceedings of the Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 190–198.

64. Hoiem, D.; Chodpathumwan, Y.; Dai, Q. Diagnosing error in object detectors. *Comput. Vis.* **2012**, *7574*, 340–353. [CrossRef]

65. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]

66. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, 1735–1780. [CrossRef]

67. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef] [PubMed]

68. Lee, T.H.; Jun, M.-J. Prediction of Seoul House Price Index using deep learning algorithms withmultivariate time series data. *SH Urban. Res. Insight* **2018**, *8*, 39–56. [CrossRef]

69. Choi, Y.; Kim, H.J.; Kim, S.J. An Analysis on the Determinants of Mountainous and Coastal Area's Housing Value Caused by the Characteristics of the Natural Environment. *J. Korean Soc. Civ. Eng.* **2013**, *33*, 811–819. [CrossRef]

70. Reimers, N.; Gurevych, I. Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks. *arXiv* **2017**, arXiv:1707.06799.

71. Qiumei, Z.; Dan, T.; Fenghua, W. Improved convolutional neural network based on fast exponentially linear unit activation function. *IEEE Access* **2019**, *7*, 151359–151367. [CrossRef]

72. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034. [CrossRef]

73. Zhang, Z. Improved Adam Optimizer for Deep Neural Networks. In Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2. [CrossRef]

74. Li, Y.; Jiao, C.; Xin, H.; Wang, Y.; Wang, K. Prediction on housing price based on deep learning. *Int. J. Comput. Inf. Eng.* **2018**, *12*, 90–99.

75. Enders, W. *Applied Econometric Time Series*, 3rd ed.; Wiley Series in Probability and Statistics: Hoboken, NJ, USA, 2010.
76. Li, C.W.; Wong, S.K.; Chau, K.W. An analysis of spatial autocorrelation in Hong Kong's housing market. *Pac. Rim Prop. Res. J.* **2011**, *17*, 443–462.