Modelling Housing Feature Impacts on Sale Price in Newly Developed Suburbs Relative to a Standard House

Christina Yin-Chieh Lin, Andrew Mason,
Charles Ma, and Andreas W. Kempa-Liehr
Department of Engineering Science
University of Auckland
New Zealand
clin363@aucklanduni.ac.nz

Abstract

There is a recent trend of entire new suburbs that support a local community being designed and built to solve the shortage of affordable housing all around the world. The aim of this study is to anticipate the value of housing features in new suburbs that are still in the planning stage. For this purpose, we are separating price movements over time from the impact of individual housing features in recently developed suburbs. To generate insights on housing features that can be directly interpreted by developers, we propose modelling house prices relative to a standard house representative of local preferences. The proposed model is successfully evaluated on newly developed suburbs in Auckland, New Zealand. The case study on the newly developed suburbs of Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, and Silverdale demonstrates that the proposed modelling approach effectively captured the complex relationship between housing features and sale price relative to a standard house (R-squared 91.5%). The proposed model generalizes to a reasonable extent to house prices in the new suburb of Hobsonville (R-squared 75.0%) without using any historical sale records in this suburb. This indicates that the insights on housing features relative to the standard house are applicable to other new suburbs still in the planning stage and, therefore, have the potential to support future suburb developments.

1 Introduction

Shortage of affordable housing is a growing crisis all around the world. According to statistics published in 2019 by the Organisation for Economic Co-operation and Development (OECD), people in OECD countries now spend, on average, approximately 22% of their disposable income on housing costs [1]. This means that

housing is becoming increasingly unaffordable to middle-class families. Surveys conducted by Gallup's World Poll from 2015 to 2017 across 140 different countries indicate that on average 27% of a country's population cannot afford adequate housing [2]. In addition to the supply and demand issue, new housing developments are often priced for upper-class buyers and overseas investors, and therefore do not resolve the crisis of affordable housing shortage for middle-class buyers [3]. Countries are looking for new housing solutions to create sustainable and affordable housing conditions for middle-class families.

One new solution is the coordinated development of entirely new suburbs that support a local community. Developing an entirely new suburb is time-consuming because of challenges such as planning, land allocation, and funding [4], and it is difficult to predict the expected price and housing demands of a new suburb far into the future. For example, the Old Oak development in west London aims to create a new London suburb that provides 25,500 homes and 65,000 jobs, but the entire project will take an estimated 20 to 30 years to complete [5]. Demands for different housing features are quickly changing as a new generation of middle-class families emerges. It is becoming increasingly difficult to accurately identify specific housing requirements that are needed to create sustainable and affordable suburbs for the middle-class families of a local community [4].

Table 1: Overview of modelling approaches from a selection of past studies on house price prediction and housing feature impacts.

Reference	Modelling Approach	Case Study Location
Wen et al. [6]	Linear Hedonic Model	Mainland China
Bao and Wan [7]	Semi-parametric Regression	Hong Kong
Pace et al. [8]	Autoregressive Model	USA
Bourassa et al. [9]	Spatial Regression	New Zealand
Limsombunchai [10]	Artificial Neural Network	New Zealand
Yoo et al. [11]	Random Forest	USA

House price prediction using historical sales data from the same suburb has been thoroughly explored by past research, but there is a distinct lack of studies that focus on modelling entirely new suburbs with no historical sales. Table 1 provides an overview of the modelling approaches from a selection of past studies on house price prediction with housing feature impacts. Common methods for modelling house prices include linear hedonic models, spatial autoregressive models, random forest models, and artificial neural networks. The main advantage of traditional linear hedonic models is that the impacts of individual housing features on sale price are easy to interpret, but these models are less adaptive to non-linear relationships [6], [7]. Semi-parametric regressions and spatial autoregressive models are also not suitable for analyzing entirely new suburbs because historical sales from the same neighbourhood are unavailable [12]. Non-parametric algorithms such as random forests and artificial neural networks often produce good forecasting results, but the impacts of individual housing features are more difficult to interpret [13], [14]. The 'black box' nature of neural networks means that the contribution of each housing feature to the final sale price cannot be easily derived from the model [14].

This is a major issue for developers who wish to understand buyer preferences to housing features in a new suburb.

Analysis Criteria	Wen et al. [7]	Bao and Wan [8]	Pace et al. [12]	Bourassa et al. [11]	Limsombunchai [17]	Yoo et al. [16]	Our Contribution
Explicitly Models Price Change Over Time	●	○	●	○	○	○	●
Incorporates Non-linear Housing Feature Impact	○	○	○	○	●	●	●
Targets Newly Developed Suburbs	○	○	○	○	○	○	●
Uses a Point of Reference to Produce Interpretable Results	○	○	○	○	○	○	●

Table 2: Analysis criteria for a selection of past studies on house price prediction and housing feature impacts. None of the past studies has specifically targeted newly developed suburbs.

None of the past studies specifically targets new suburbs before or shortly after being built (see Tab. 2 for analysis criteria), and very few papers analyzed the housing feature impacts on price in Auckland suburbs. This paper focuses on analyzing newly developed suburbs in Auckland to understand housing feature impacts on sale price in these new suburbs. A special focus is the formulation of the model, such that developers can easily interpret the results and integrate them into their planning. Defining a reliable point of reference for our model is critical for conveying the results to non-experts in machine learning, such as housing developers [15, 16]. Our chosen point of reference is the price of a standard house in the new suburbs over time, similar to Rehm and Filippova's (2008) method of using a reference standard house to compare price premiums of individual suburbs. Rehm and Filippova's (2008) model includes housing features such as floor area and site area, but the effects are not analyzed relative to the standard house.

We expand on previous research to include features such as number of bedrooms, bathrooms, and garages, which has not been a focus in past research on Auckland housing [18]. The literature review conducted by Fernandez [18] on the application of hedonic price models to the New Zealand housing market also identifies the scope for matching housing features to sale prices and the application of Bayesian approaches. Our model also expands on Rehm and Filippova's (2008) method of estimating reference standard house price for every two-year period by modelling monthly standard house price using a moving average, and thereby recognizing that price change over time is continuous. Modelling housing feature impacts relative to a standard house will create a solution to our regression problem that is easy to interpret and will be beneficial to developers.

The main objectives of this paper are to:

1. Separate the effects of price change over time due to market dynamics from the effects of individual housing features.

2. Understand the non-linear impact of individual housing features on sale price relative to the configuration of a reference standard house.

3. Validate the robustness of our proposed model by making out-of-sample price predictions for a new suburb.

The remainder of this paper is structured as follows. Sec. 2 describes in detail the proposed method of modelling housing feature impacts relative to the configuration of a standard house. Sec. 3 shows the results from applying the proposed modelling method to property sales in seven newly developed suburbs in Auckland, New Zealand-Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, Silverdale, and Hobsonville. Conclusions about the house sales modelling method proposed in this study and the directions of future applications are given in Sec. 4.

2 Modelling Sales in New Suburbs

This section describes the proposed approach to model house sale values in new suburbs using housing examples from Auckland, New Zealand. Our modelling approach is designed with specific emphasis on understanding the impacts of individual housing features that can be controlled by developers (e.g. number of bedrooms, number of bathrooms etc.).

Due to economic uncertainties, it is very difficult to forecast future house price trends over a long time horizon. This study separates the effects of price change over time from the effects of individual housing features by decomposing the problem into two components:

1. Price of a standard house over time

2. Impacts of individual housing, features on sale price relative to the standard house.

The first component estimates the price of a standard house over time, where the features of a standard house are defined by average statistics such as the median of houses in the studied suburbs. The second component models how individual housing features impact sale price relative to the estimated price of a standard house from the first component. Estimating the price of a standard house over time means that the non-linear trend of price increase over time can be incorporated into the same model as other housing features. The estimated housing feature impacts will also be easier to interpret for developers when they are presented relative to a standard house.

Features of a standard house incorporated into our model are floor area $\bar{A}$, land area $\bar{L}$, number of bedrooms $\bar{B}$, number of bathrooms $\bar{C}$, number of garages $\bar{G}$, and a Boolean value $\bar{F}$ that indicates whether a house has at least one free-standing garage. Our standard house is described by a tuple of constants defining its configuration:

$$(\bar{A}, \bar{L}, \bar{B}, \bar{C}, \bar{G}, \bar{F}).$$

For every sale record with index i, we record the sale price P_i, floor area A_i, land area L_i, number of bedrooms B_i, number of bathrooms C_i, number of garages G_i, a Boolean value F_i that indicates whether a house has at least one free-standing garage, suburb S_i, and sale month M_i. A sale record with index i is described by a tuple of values:

$$(P_i, A_i, L_i, B_i, C_i, G_i, F_i, S_i, M_i).$$

We call our model the Standard House Configuration Model (SHCM). We model the sale price P_i relative to the configuration $(\bar{A}, \bar{L}, \bar{B}, \bar{C}, \bar{G}, \bar{F})$ of a standard house with the following formula:

$$\begin{aligned}\log(P_i) =& w_0 + w_A(A_i - \bar{A}) + w_L(L_i - \bar{L}) + \sum_{b \in \mathcal{B} \setminus \bar{B}} w_b^{\mathrm{B}} \mathbb{1}_{B_i = b} \\ &+ \sum_{c \in \mathcal{C} \setminus \bar{C}} w_c^{\mathrm{C}} \mathbb{1}_{C_i = c} + \sum_{g \in \mathcal{G} \setminus \bar{G}} w_g^{\mathrm{G}} \mathbb{1}_{G_i = g} \\ &+ \sum_{f \in \mathcal{F} \setminus \bar{F}} w_f^{\mathrm{F}} \mathbb{1}_{F_i = f} + \sum_{s \in \mathcal{S}} w_s^{\mathrm{S}} \mathbb{1}_{S_i = s} \\ &+ \frac{1}{N(M_i)} \sum_{m \in \mathcal{M}} w_m^{\mathrm{M}} \mathbb{1}_{(M_i - 5 \le m \le M_i + 6)} + \varepsilon_i\end{aligned} \tag{1a}$$

where

$$N(M_i) = \sum_{m \in \mathcal{M}} \mathbb{1}_{M_i - 5 \le m \le M_i + 6} \tag{1b}$$

and

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{1c}$$

All variables and sets of the proposed model are described in Table 3, and ε_i is the residual. We take the natural log of sale price P_i as the target variable, and the rest are explanatory variables. Since we applied the natural log transformation on sale price, estimated impacts of housing features are calculated as multiplicative percentage changes after back transforming fitted coefficients $w_A, w_L, w_1^{\mathrm{B}}, \ldots, w_{5+}^{\mathrm{B}}, w_1^{\mathrm{C}}, \ldots, w_{5+}^{\mathrm{C}}, w_1^{\mathrm{G}}, \ldots, w_{3+}^{\mathrm{G}}, w_0^{\mathrm{F}}$ and w_1^{F} by taking the exponential function. The impacts of floor area A_i and land area L_i are modelled in terms of the difference to their respective configurations $\bar{A}$ and $\bar{L}$ of the standard house. Each level of discrete housing features such as number of bedrooms B_i, number of bathrooms C_i, and number of garages G_i are all encoded into binary variables (represented by indicator function $\mathbb{1}$) so that our model is able to capture the non-linear relationship between housing features and sale value relative to the configuration of the standard house. For example, $\mathbb{1}_{B_i=1} = 1$ if the property with sale record index i has one bedroom, and $\mathbb{1}_{B_i=1} = 0$ otherwise. Most garages are under the main roof of the house, but some houses also have free-standing garages. The impact on sale price of having at least one free-standing garage is captured by the term $\mathbb{1}_F$. By excluding the encoded terms of $\bar{B}, \bar{C}, \bar{G}$ and $\bar{F}$ from our model, the coefficients of these configurations are essentially fixed to zero so that all other feature impacts are estimated relative to the standard house. For example, $w_b^{\mathrm{B}} = 0$ when $b = \bar{B}$, and $w_c^{\mathrm{C}} = 0$ when $c = \bar{C}$. Based on a previous study on Auckland housing by Bourassa et al. [9], using homogeneous geographic submarkets as explanatory

Table 3: Description of variables where i is the sale record index.

Variable	Description
P_i	Full amount paid for a property in NZ dollars
$A_i - \bar{A}$	Difference from standard floor area above ground in 10 square metres
$L_i - \bar{L}$	Difference from standard land area of the property in 100 square metres
$\mathbb{1}_{B_i=b}$	1: if $B_i = b$ where B_i indicates the number of bedrooms e.g. $b \in \mathcal{B}$ where $\mathcal{B}$={1,2,3,4,5+} 0: otherwise
$\mathbb{1}_{C_i=c}$	1: if $C_i = c$ where C_i indicates the number of bathrooms e.g. $c \in \mathcal{C}$ where $\mathcal{C}$={1,2,3,4,5+} 0: otherwise
$\mathbb{1}_{G_i=g}$	1: if $G_i = g$ where G_i indicates the number of garages e.g. $g \in \mathcal{G}$ where $\mathcal{G}$={0,1,2,3+} 0: otherwise
$\mathbb{1}_{F_i=f}$	1: if $F_i = f$ where F_i is a Boolean value indicating whether the house has at least one free-standing garage $f \in \mathcal{F}$ where $\mathcal{F}$={0,1} 0: otherwise
$\mathbb{1}_{S_i=s}$	1: if $S_i = s$ where S_i indicates the suburb where the house is located e.g. $s \in \mathcal{S}$ where $\mathcal{S}$={Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, Silverdale} 0: otherwise
$\mathbb{1}_{M_i-5\leq m\leq M_i+6}$	1: if $M_i - 5 \leq m \leq M_i + 6$ where M_i indicates the sale month index e.g. $m \in \mathcal{M}$ where $\mathcal{M}$={1,2,3,...,274} 0: otherwise
ε_i	Residual

variables is sufficient for capturing spatial variation without more complex statistical methods. This study encodes suburbs S_i into binary variables to represent geographic submarkets. The back-transformed coefficients of each encoded suburb w_s^{S} provide a percentage price change for house values in each suburb.

Sale months extracted from the provided sale dates are indexed as integers. For example, in our Auckland housing case study, March 1996 to December 2018 are indexed from 1 to 216 chronologically. Each unique sale month index m is then encoded into a binary variable similar to the discrete housing features. All model coefficients of encoded sale month indices $w_1^{\mathrm{M}}, \ldots, w_{274}^{\mathrm{M}}$ are averaged across a moving time window so that the estimated price of a standard house to be used as a reference value changes gradually. Our moving time window is defined by the five months before and six months after the time of sale M_i, where $N(M_i)$ gives the number of months in our case study that occurs in this time window (see Eq. 1b). This method is not suitable for forecasting future house prices because we are using information from the future six months after the time of sale. This restriction is suitable for the presented study because the focus is on analyzing

housing feature impacts in new suburbs, and not on future price predictions. The benefit of separating housing feature impacts from the volatility of market dynamics is that our analysis has the potential to be combined with future price forecasts made by domain experts.

When the features of a standard house $(\bar{A}, \bar{L}, \bar{B}, \bar{C}, \bar{G}, \bar{F})$ are substituted into Eq. 1, all the terms corresponding to housing features are eliminated. Our standard house does not have a fixed suburb location, so the term on suburb impact is removed. The estimated price of a standard house $\bar{P}_M$ sold in month $M \in \mathcal{M}$ can therefore be calculated by the formula:

$$\bar{P}_M = \exp\left(w_0 + \frac{1}{N(M)} \sum_{m \in \mathcal{M}} w_m^{\mathrm{M}} \mathbb{1}_{M-5 \leq m \leq M+6}\right). \tag{2}$$

Ridge regression is used to fit the model coefficients, where the regularization parameters are selected from cross-validation. Standard ordinary least squares model is used to calculate p-values and confidence intervals to provide unbiased insights on statistical significance. This proposed modelling approach is used to analyze house sale values in new Auckland suburbs in the following section.

3 Case Study: New Auckland Suburbs

Auckland has an estimated shortage of 34,000 homes that accumulated from 2013 to 2018 as the local population continued to increase [19]. To meet the rising housing demands, Auckland Council has presented a plan to develop outer suburbs such as Fairview Heights and Hobsonville and expand the current city into 'Greater Auckland' [20]. Variations of housing features and house sale prices are very specific to each housing market, but very little research has targeted Auckland suburbs. The housing crisis in Auckland and the plan for rapid expansion mean that modelling sales in new Auckland suburbs would be increasingly relevant to future developers.

We conduct a case study on modelling house sales in six new Auckland suburbs- Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, and Silverdale. Properties with land area above 2 hectares or sale price above two million dollars are not selected for analysis because they are outliers in these newly developed suburbs. We extracted a total of 8937 sale records in these six suburbs from the New Zealand housing data provided by CoreLogic [21]. The locations of the six suburbs, along with Auckland Central, are all shown on the map in Fig. 1.

We will validate the robustness of our model by making price predictions for 893 sales in the suburb of Hobsonville in Sec. 3.2. Hobsonville is a newly developed suburb northwest of Central Auckland built around 2010. The location of Hobsonville is also shown on the map in Fig. 1.

3.1 Impact of Housing Features on Sale Price

This section analyzes the impacts of individual housing features on sale price based on the coefficients from our fitted model. Our ridge regression model has an optimized regularization parameter of 0.08, and an R-squared value of 91.5% (see Eq. 1

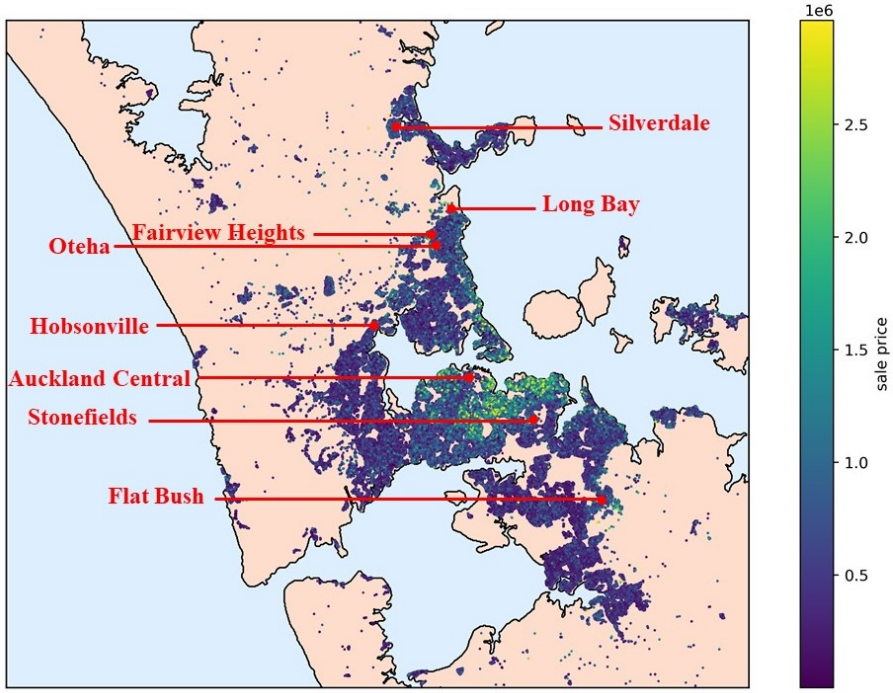

Figure 1: Map of Greater Auckland, showing locations of Auckland Central, Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, Silverdale, and Hobsonville. Fairview Heights, Oteha, Long Bay, and Silverdale are all on the outer edges of the city.

for the full model formula). The estimated percentage change for the standard house price in each suburb, calculated by back-transforming coefficients w_s^{S}, are listed in Table 4. The features of a standard house in the six suburbs are defined in Tab. 5, and all house price impacts are analyzed relative to the configuration of this standard house. Sale months from the six suburbs are indexed from 1 to 274, where $M = 1$ correspond to March 1996, and $M = 274$ corresponds to December 2018. The estimated price of our standard house $\bar{P}_M$, calculated using Eq. 2 for sale month indices $M \in \{1, 2, 3, ..., 274\}$, is shown in Fig. 2. Even though the estimated price of our standard house $\bar{P}_M$ is calculated using a moving average, we still observe discrepancies in periods with rapid changes in sale volume and outlier prices. This trend of increase in the estimated price of a standard house should ideally separate the effects of price change over time from the effects of individual housing features.

To verify that our estimated standard house price $\bar{P}_M$ captures the trend of price change over time in the six suburbs, we convert each property into a standard house by adjusting for housing feature impacts. For example, if a property has four bedrooms, then we divide the true sale price of the property by the estimated price scaling factor caused by having four bedrooms instead of a standard three-bedroom house. We call the resulting value the equivalent standard house price because it represents the hypothetical price of a property if it had been designed as a standard house. The equivalent standard house price also allows us to study the residuals ε_i over time since the effects of all other housing features are eliminated. The

equivalent standard house prices are plotted in Fig. 2, and they follow the trend of our estimated standard house price relatively closely.

Table 4: The estimated percentage change of standard house price for each suburb.

Suburb	Fairview Heights	Oteha	Stonefields	Long Bay	Flat Bush	Silverdale
Percentage Change	-2.51%	-3.76%	23.03%	7.13%	-10.78%	-9.37%

Table 5: Housing features of a standard house in Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, and Silverdale.

Feature	$\bar{A}$	$\bar{L}$	$\bar{B}$	$\bar{C}$	$\bar{G}$	$\bar{F}$
Value	205m^2	399m^2	3	2	1	0

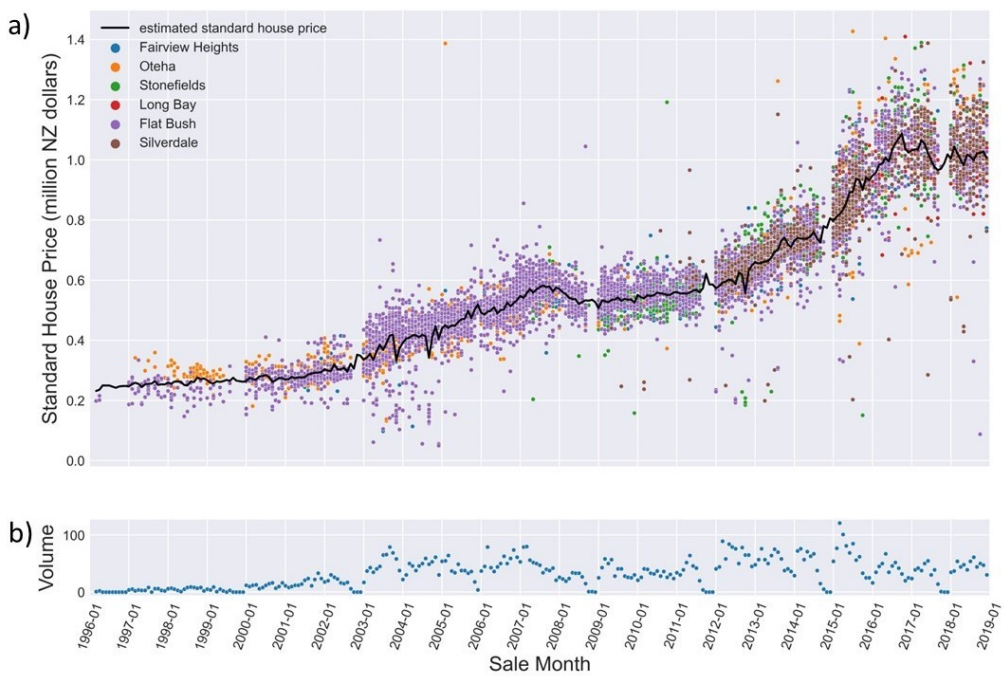

Figure 2: Panel a) shows the equivalent standard house prices after converting each property into a standard house by adjusting for housing feature impacts. The black line represents the estimated standard house price $\bar{P}_M$. Equivalent standard house prices in all six suburbs follow the trend of our estimated standard house price relatively closely. Panel b) shows the monthly sale volume from March 1996 to December 2018.

The estimated percentage change in price for individual housing features, p-values, and 95% confidence intervals (CI) are all listed in Fig. 3. The calculated percentage changes show that the price changes across increasing numbers of bedrooms, bathrooms, and garages are not linear. For every 10m^2 increase in floor area relative to the standard house, price is estimated to increase by approximately 2.79% to 3.03%. A house with only one bedroom is estimated to have a price

approximately 19.97% to 28.53% lower than the price of a standard house with three bedrooms, while a house with two bedrooms is only 1.40% to 4.98% lower in price. A house with four bedrooms is estimated to be 7.02% to 9.43% higher in price than the standard house, but the increase in price for each added bedroom plateaus above four bedrooms. This indicates that most buyers are not satisfied with single-bedroom houses, but five bedrooms or above can become excessive.

A house with only one bathroom reduces sale price by 3.75% to 5.68% compared to a standard house with two bathrooms, while having three bathrooms does not lead to a statistically significant change in price. According to our model, four bathrooms and above become excessive under the same floor area and do not lead to an increase in price. Having no garages only reduces sale price by 0.36% to 4.02% compared to a standard house with a single garage. The highest price change from the number of garages is a house with three garages or above, which leads to a 10.77% to 16.37% price increase compared to the standard house. Lastly, houses with at least one free-standing garage are estimated to have prices 3.20% to 6.83% higher than those that do not. Our fitted model provides insights on the complex relationship between house price and housing features in a new suburb. The next section validates that our insights on housing feature impacts are also applicable to an entirely new suburb that is not part of the case study.

3.2 Validate the Model on the Hobsonville Suburb

We use our ridge regression model fitted on Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, and Silverdale to make house price predictions for sales in the Hobsonville suburb. This is to validate that our fitted model can be used to provide insights on house prices in other new Auckland suburbs before or shortly after they are built. Since historical sales in Hobsonville are not in our training set, we cannot fit a coefficient to the Hobsonville suburb. Instead, we calculate a weighted average from the coefficients of the six suburbs in the training set, where the weighting of each suburb is inversely proportional to the distance from Hobsonville. The weighted average produces a price increase of 0.29% for house prices in Hobsonville. Prediction on Hobsonville using ridge regression produces an R-squared value of 75.0%, while in-sample training on Hobsonville data produces an R-squared value of 82%. Root mean square error (RMSE) is $116,416 after back transformation. Prediction residual is calculated for the ith observation by the following formula:

$$\varepsilon_i = P_i^{\text{actual}} - P_i^{\text{predicted}}. \tag{3}$$

For prediction on Hobsonville using ridge regression, the actual sale prices are plotted against the predicted prices in Fig. 4. Approximately 89% of the predicted values are within plus or minus 20% of the actual sale price. The box plot of residuals against sale years does not show any clear trend of residual changes across time (see Fig. 5). Residuals are relatively uniform across sale years. This implies that the non-linear house price trend over time is effectively captured by the coefficients of the encoded sale month indices. The plot of predicted price against actual price in Fig. 4 shows that there is one clear outlier with a residual just above one million in magnitude. This outlier property is a townhouse sold in 2018 with a floor area

Area	Living Area	Land Area
Price Change	2.92% per $10m^2$	0.64% per $100m^2$
P-value	0.0	1.3×10^{-42}
95% CI	2.79, 3.03	0.55, 0.74

Bedrooms	1	2	3	4	5+
Price Change	-23.93%	-2.76%	0.00%	8.19%	7.21%
P-value	5.2×10^{-22}	5.7×10^{-4}	---	1.4×10^{-43}	2.6×10^{-19}
95% CI	-28.53, -19.97	-4.98, -1.40	---	7.02, 9.43	5.62, 8.89

Bathrooms	1	2	3	4	5+
Price Change	-4.89%	0.00%	-0.12%	-2.24%	-3.68%
P-value	8.0×10^{-21}	---	0.66	0.0037	0.0011
95% CI	-5.68, -3.75	---	-1.16, 0.74	-4.07, -0.80	-6.29, -1.61

Garages	0	1	2	3+
Price Change	-2.77%	0.00%	8.06%	13.55%
P-value	0.02	---	1.4×10^{-28}	6.9×10^{-24}
95% CI	-4.02, -0.36	---	6.52, 9.43	10.77, 16.37

Garage Type	At Least 1 Garage is Free-Standing
Price Change	4.99%
P-value	3.2×10^{-8}
95% CI	3.20, 6.83

Figure 3: Resulting price change from individual housing features compared to the estimated price of a standard house. All percentage changes are calculated from the coefficients of Eq. 1.

of $400m^2$. This is more than twice the median floor area of other townhouses, but the actual sale price is unexpectedly low for unknown reasons.

4 Conclusion

This study aims to analyze the impacts of individual housing features on sale price for houses in new suburbs. Our modelling approach estimates the price of a standard house over time to separate the effects of price change over time from the effects of individual housing features. All impacts on sale price are analyzed relative to the estimated price of the standard house to produce results that are easy to interpret for non-experts. The case study on Fairview Heights, Oteha, Stonefields, Long Bay, Flat Bush, and Silverdale demonstrates that our proposed method effectively captures both the non-linear effects of individual housing features and price change over time. Single-bedroom houses are estimated to be 19.97% to 28.53% lower in price than a standard house with three bedrooms, while houses with four

Figure 4: Scatter plot of actual sale price against predicted price of Hobsonville using ridge regression. Approximately 89% of the predicted values are within plus or minus 20% of the actual sale price.

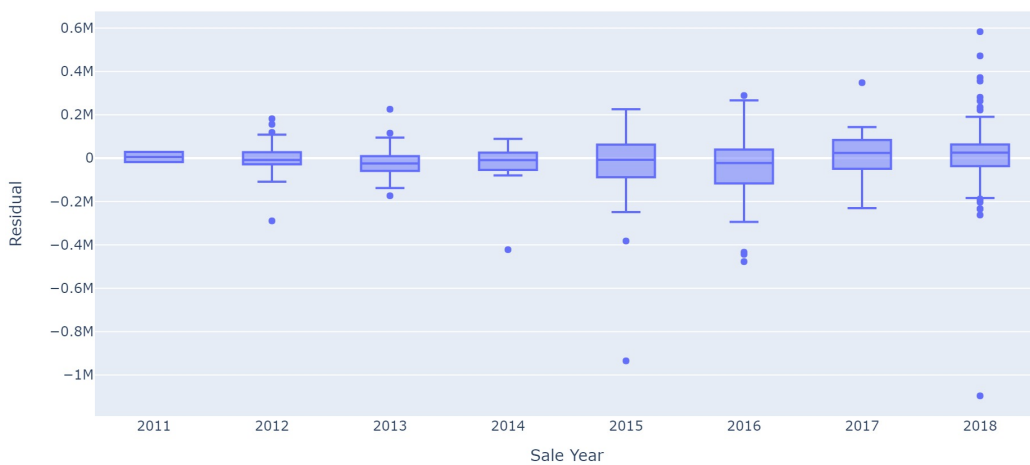

Figure 5: Box plot of prediction residual against sale year of Hobsonville using ridge regression. Residuals across sale years are relatively uniform.

bedrooms lead to around 7.02% to 9.43% increase in price. Our model is also able to predict house prices in the new suburb of Hobsonville with reasonable accuracy, and the prediction residuals are relatively uniform across sale years. This validates that our model is applicable to entirely new suburbs not in the case study, and has the potential to assist with the planning phase of suburb development by identifying buyer preferences for new suburbs.

References

[1] Affordable housing: A growing concern for people and governments. Organisation for Economic Co-operation and Development. Dec 2019. [Online]. Available: https://www.youtube.com/watch?v=tnjN2JYW6vQ

[2] T. Coupe, "How global is the affordable housing crisis?" *International Journal of Housing Markets and Analysis*, 2020.

[3] L. Kusisto and P. Grant, "Affordable housing crisis spreads throughout world; shortages persist despite millions of dollars invested and hundreds of thousands of units built," *The Wall Street Journal*, 2019.

[4] J. Ambrose, "The ongoing housing shortage," *Property Journal*, vol. Jul/Aug, pp. 23–23, 2019.

[5] E. Braidwood, "Liz Peace: 'creating a new London suburb'," *Architects' Journal*, Aug 2017. [Online]. Available: https://www.architectsjournal.co.uk/news/old-oak-commons-liz-peace-were-creating-a-new-london-suburb

[6] H. Z. Wen, J. F. Lu, and L. Lin, "An improved method of real estate evaluation based on Hedonic price model," in *2004 IEEE International Engineering Management Conference (IEEE Cat. No.04CH37574)*, vol. 3, Oct 2004, pp. 1329–1332 Vol.3.

[7] H. X. Bao and A. T. Wan, "On the use of spline smoothing in estimating hedonic housing price models: empirical evidence using Hong Kong data," *Real Estate Economics*, vol. 32, no. 3, pp. 487–507, Sep 2004, doi: 10.1111/j.1080-8620.2004.00100.x.

[8] R. Pace, R. Barry, O. W. Gilley, and C. F. Sirmans, "A method for spatial–temporal forecasting with an application to real estate prices," *International Journal of Forecasting*, vol. 16, no. 2, pp. 229–246, 2000, doi: 10.1016/S0169-2070(99)00047-3.

[9] S. C. Bourassa, E. Cantoni, and M. Hoesli, "Spatial dependence, housing submarkets, and house price prediction," *The Journal of Real Estate Finance and Economics*, vol. 35, pp. 143–160, 2007, doi: 10.1007/s11146-007-9036-8.

[10] V. Limsombunchai, "House price prediction: hedonic price model vs. artificial neural network," in *New Zealand Agricultural and Resource Economics Society Conference*, 2004, pp. 25–26.

[11] S. Yoo, J. Im., and J. E. Wagner, "Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY," *Landscape and Urban Planning*, vol. 107, no. 3, pp. 293–306, 2012, doi: 10.1016/j.landurbplan.2012.06.009.

[12] W.-C. Liao and X. Wang, "Hedonic house prices and spatial quantile regression," *Journal of Housing Economics*, vol. 21, no. 1, pp. 16–27, 2012, doi: 10.1016/j.jhe.2011.11.001.

[13] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016, doi: 10.1007/s11749-016-0481-7.

[14] Z. Liu, S. Yan, J. Cao, T. Jin, J. Tang, J. Yang, and Q. Wang, "A Bayesian approach to residential property valuation based on built environment and house characteristics," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec 2018, pp. 1455–1464, doi: 10.1109/BigData.2018.8622422.

[15] K. A. Sargent-Cox, K. J. Anstey, and M. A. Luszcz, "Patterns of longitudinal change in older adults' self-rated health: The effect of the point of reference." *Health Psychology*, vol. 29, no. 2, p. 143, 2010, doi: 10.1037/a0017652.

[16] K. Manderbacka, I. Kåreholt, P. Martikainen, and O. Lundberg, "The effect of point of reference on the association between self-rated health and mortality," *Social science & medicine*, vol. 56, no. 7, pp. 1447–1452, 2003, doi: 10.1016/s0277-9536(02)00141-7.

[17] M. Rehm and O. Filippova, "The impact of geographically defined school zones on house prices in New Zealand," *International Journal of Housing Markets and Analysis*, vol. 1, no. 4, pp. 313–336, nov 2008.

[18] M. A. Fernandez, "A review of applications of hedonic pricing models in the new zealand housing market," 2019, available: https://knowledgeauckland.org.nz/publications/a-review-of-applications-of-hedonic-pricing-models-in-the-new-zealand-housing-market/.

[19] G. Ninness, "New figures show Auckland's housing shortage is still getting worse but should start to decline in the next one to two years," *interest.co.nz*, Nov 2018. [Online]. Available: https://www.interest.co.nz/property/97023/new-figures-show-aucklands-housing-shortage-still-getting-worse-should-start-decline

[20] Auckland Council, "What will Auckland look like in the future?" 2020. [Online]. Available: https://www.aucklandcouncil.govt.nz/plans-projects-policies-reports-bylaws/our-plans-strategies/auckland-plan/development-strategy/future-auckland/Pages/what-auckland-look-like-future.aspx

[21] CoreLogic, "Residential property sales statistics. data files for New Zealand house sales," Auckland, Updated 29 June 2020, 2020, accessed: 06.02.2020.