Article

Modelling residential building costs in New Zealand: Time series transfer function approach

Linlin Zhao [1], Jasper Mbachu [2], Zhansheng Liu [1,*], Huirong Zhang [1]

[1] Beijing University of Technology, Beijing, China; llzhao@bjut.edu.cn; huirongzhang@bjut.edu.cn
[2] Bond University, Gold Coast, Australia; jambachu@bond.edu.au
* Correspondence: liuzhansheng83@sohu.com;

Abstract: Cost estimating based on building cost index plays an important role in project planning and management by providing accurate cost information. Recently, tremendous advances in cost estimating has been made but serious inaccuracies in it are still too frequently witnessed. This study aims to improve estimating accuracy for residential building costs in New Zealand. In this study, the New Zealand house prices index is involved in the transfer function models to produce forecasts of building costs for one-storey house, two-storey house, and town house in New Zealand. To demonstrate the effectiveness of the proposed models, this study compares the estimate results of the transfer function models with the univariate ARIMA models. The results indicate that the proposed transfer function models can achieve better outcomes than ARIMA models by considering the causality between building costs and New Zealand house prices. During the modelling process, the better cost estimation approach can be identified, and the movements of building costs are shown.

Keywords: Transfer function models; ARIMA models; model selection; building cost index; New Zealand

---

1. Introduction

In today's highly competitive market, companies are seeking solutions for predicting project cost, considered important for the commercial success of a project [1]. Project cost estimates are needed by clients, consultants, contractors, or organizations for purposes such as a project feasibility study, financial evaluation of alternatives, and the formulation of the initial budget [2]. Accurate forecasting the building project's cost is crucial to adequately manage resources [3]. Besides, an accurate building cost estimates can properly allocate resources so that the total project cost can be reduced. Poor estimating of the cost may disrupt the project or may even lead to bankruptcy for the company. Furthermore, in [4], the study also pointed out that inaccurate cost estimating is one of the main reasons causing project cost overruns.

To some extent, building cost in New Zealand increased considerably during 2002-2012 [5]. Cost fluctuations usually result in inaccurate estimations. As accuracy plays an important role in preparing budgets and bids in most building projects, industry professionals and organizations show great interest in more accurate cost estimating. An effective cost estimating tool is important. However, traditional approaches are usually blamed for lacking accuracy [6].

The study aims to provide an accurate estimating method for building cost. Time series analysis is a classic and powerful approach used in predicting the future of stochastic processes. ARIMA and transfer function models' development follow the procedures outlined by [7]. The transfer function model is another method of time series forecasting, incorporating more variables (including more predicting information) into the model for producing more accurate forecasts. This is one of the main

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

reasons that transfer function models can have comparative advantages over ARIMA models. This study takes into account the influence of house prices in an ARIMA model to develop the transfer function model for building costs forecasting.

The rest of the study is organized as follows. Section 2 presents the previous related works, Section 3 illustrates the proposes time series model. Section 4 presents the application of the model on the cost series and the model effectiveness assessment. The results are discussed in the Section 6. The final section presents the conclusion of the study.

2. Literature Review

There is limited literature on methods that estimate the $-per-square metre cost rate of building projects, also referred to as unit cost. The project cost is then predicted by multiplying the cost rate by the total square metres of the project. In the last few decades, several methods for deriving a building cost forecasting model have been proposed. The most common used techniques include regression methods [8, 9], artificial neutral network algorithms [2, 10], and case-based reasoning [11, 12].

Future changes in some variables are usually foreshadowed by changes in other variables; the latter variables are called leading indicators [13]. Leading indicators have been successfully used in forecasting. For example, in [14], the study used financial and economic indicators in forecasting US recessions. The application of leading indicators is based on the view that repetitive sequences occur in the business cycle. The cycles include booms and busts in various activities, but these booms and busts do not happen at exactly the same time for different activities; some activities are leading and some lagging. Although a leading indicators model is usually referred to as a method without theory, existing literature and empirical evidence give clues as to the selection of appropriate indicators.

In [15, 16], the studies pointed out that housing supply usually exhibits a lag. Therefore, house prices should influence future housing supply due to developers making decisions about whether to increase the housing supply frequently depend on present house prices. Based on the findings of [17], house prices affect changes in building construction costs since they depend on the raw material market for housing and are subject to the effect of the derived demand for housing. Moreover, in [18], the study also provided evidence that building construction costs are sensitive to housing prices.

When a time series is examined, questions usually arise about the relationship and impact of other series on it over time. If the relationship or impact is important, a dynamic model incorporating the relationship or impact is necessary. A transfer function model was introduced to relate the endogenous response series to the exogenous series. Examples of transfer function applications abound in business, economics, and engineering. In business, it is used widely in modelling sales and advertising [19]. In economics, the transfer function was used for predicting business cycles. It also models the effect of personal disposal income on real nondurable consumption in the UK [20]. According to [21], statistical and engineering process controls are associated with modelling the transfer functions between inputs and outputs. Transfer function models are fundamental components in many complicated systems.

The objectives of this study are as follows: first, to expand the very limited existing literature on leading indicator models in building and construction; second, to provide a comprehensive study in terms of comparing in a building context the forecasting performance of ARIMA models and transfer function (TF) models; and third, to examine the application of the proposed model on three building cost series in order to draw general conclusions from the empirical results with greater confidence.

3. Methodology

3.1. Data

To develop cost estimating models for one-storey house, two-storey house and town house in New Zealand, New Zealand house price (NZHP) is involved as an explanatory variable. The series data were obtained from the Reserve Bank of New Zealand and QV cost builder. Changes in house prices can be measured in many ways. The house price index produced by the Reserve Bank of New

doi:10.20944/preprints201903.0132.v1

Zealand has become a favoured benchmark in recent years. Building cost index, which is published quarterly in New Zealand by QV cost builder, is defined as the average cost per square metre of the building. The BCI includes raw material costs, labour costs, and equipment costs.

The 72 observations used were quarterly observations starting with the first quarter of 2001, through until the end of 2018. The training sample is from 2001:Q1 to 2014:Q4 (total 56 observations) while the validation sample is from 2015:Q1 to 2018:Q4. Those four time series are plotted in Figure 1. It is evident from the graph that the four series are autocorrelated and highly unlikely stationary. Note that the study also provided data for the house price stretching three periods back before the first quarter of 2001 to accommodate the lagged relationships.

Figure 1. Time series plots

3.2. Seasonal ARIMA Model

The seasonal ARIMA model in its seasonal form is usually given as in Equation (1), denoted as ARIMA(p,d,q)(P,D,Q)$_s$. p, d, and q refer to autoregressive order, integrated, and moving average term of the non-seasonal part of the model, respectively, while P, D, and Q have the same role for the seasonal part of the model, and s indicates the length of seasonality.

$$\varphi_p(B)\emptyset_P(B)(1-B)^d(1-B^s)^D z_t = \theta_q(B)\vartheta_Q(B)a_t, \quad (1)$$

Where

$$\varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p,$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,$$

$$\emptyset_P(B) = 1 - \emptyset_1 B^s - \emptyset_2 B^{2s} - \cdots - \emptyset_P B^{Ps},$$

$$\vartheta_Q(B) = 1 - \vartheta_1 B^s - \vartheta_2 B^{2s} - \cdots - \vartheta_Q B^{Qs},$$

$$B z_t = z_{t-1},$$

$$B^s z_t = z_{t-s},$$

and $\varphi_1, \varphi_2, \ldots, \varphi_p$ are the parameters of the non-seasonal autoregressive term of the model; $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the non-seasonal moving average terms of the model; $\emptyset_1, \emptyset_2, \ldots, \emptyset_P$ are the parameters of the seasonal autoregressive term of the model; $\vartheta_1, \vartheta_2, \ldots, \vartheta_Q$ are the parameters

doi:10.20944/preprints201903.0132.v1

of the seasonal moving average terms of the model; B is the backshift operator; d and D indicate regular and seasonal differencing, respectively; a_t is a white noise process; z_t is the data series.

Sample autocorrelation function (SAC) and sample partial autocorrelation function (SPAC) are usually used to identify an ARIMA model, providing systematic guidance about the underlying correlated behaviour of the series. In fact, the procedure follows five steps: (i) stationarity examination, (ii) model identification, (iii) parameter estimation, (ix) model verification, (x) forecasting.

3.3. Transfer Function

The general transfer function is shown in Equation (2)

$$z_t = \mu + \frac{C\omega(B)}{\delta(B)} B^b z_t^{(x)} + a_t, \quad (2)$$

Where

$$\omega(B) = 1 - w_1 B - w_2 B^2 - \cdots - w_s B^s,$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r,$$

and z_t represents the stationary Y_t values, $z_t^{(x)}$ represents the stationary X_t values, μ is a constant term, C is a scale parameter, b is the order of delay that is the time delay between changes in X_t and the impact on Y_t, s is the order of regression, r represents the order of decay.

3.3.1. Stationary Transformation

If the time series is stochastic, the variables are usually centered or differenced to attain a condition of stationarity. In general, it is necessary to apply differencing to either the input series $z_t^{(x)}$ or the output series z_t or both, in order to achieve stationarity. Moreover, the input series $z_t^{(x)}$ or the output series z_t do not need to be differenced in the same way. The ACFs and PACFs of the cost series are shown in Figure 2, which indicate they are non-stationary variables. Therefore, they need to be differenced to be stationary variable for further modelling.

3.3.2. Prewhiten *Time Series $z_t^{(x)}$ and z_t*

It is recognised that autocorrelation in the input series may contaminate the cross-correlation between the input and output series [22]. Autocorrelation is a major reason for spurious relationships. For example, two unrelated time series that are internally autocorrelated, sometimes by chance, can produce significant cross correlations. Thus, a prewhitening filter is suggested to neutralize this autocorrelation [23]. This filter can transform the input series into white noise. It is formulated from the ARIMA models. The first step in the prewhitening process is to select an ARIMA model describing the $z_t^{(x)}$ series. The ARIMA model to describe $z_t^{(x)}$ is expressed in Equation (3)

$$\varphi_p(B)\emptyset_P(B^L)z_t^{(x)} = \theta_q(B)\vartheta_Q(B^L)\alpha_t, \quad (3)$$

So

$$\alpha_t = \frac{\varphi_p(B)\emptyset_P(B^L)z_t^{(x)}}{\theta_q(B)\vartheta_Q(B^L)},$$

This inverse filter developed from X_t is then applied to Y_t series substituting z_t for $z_t^{(x)}$ in the above equation. Then β_t is obtained as shown in Equation (4).

$$\beta_t = \frac{\varphi_p(B)\emptyset_P(B^L)z_t}{\theta_q(B)\vartheta_Q(B^L)}, \quad (4)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

3.3.3. Determine Model Orders

The order of b, r, and s determine the structure of the transfer function. The cross-correlation function between the α_t and β_t can be used to tentatively determine the model orders b, s, and r. The cross-correlation subjected to the identical transformation remains the same. After both X_t and Y_t series are prewhitened, direct estimation of the orders is made possible from the examination of the SCC. The shape of the cross-correlation between the two series explores the orders (b, r, and s) of the transfer function. The cross-correlation function at lag k can be described in Equation (5).

$$r_k(\alpha_t, \beta_t) = \frac{\sum_{t=b}^{n-k}(\alpha_t - \bar{\alpha})(\beta_{t+k} - \bar{\beta})}{\left[\sum_{t=b}^{n}(\alpha_t - \bar{\alpha})^2\right]^{1/2}\left[\sum_{t=b}^{n}(\beta_{t+k} - \bar{\beta})^2\right]^{1/2}}, \quad (5)$$

where $\bar{\alpha}$ is the mean of α_t values, $\bar{\beta}$ is the mean of β_t values.

To interpret the cross-correlation, it is supposed that there are no spikes at negative lags. If there were, this indicates that the z_t have an effect on $z_t^{(x)}$. In that case, the transfer function cannot be used. There can be no feedback from Y_t to X_t. In other words, the X_t in the transfer function model must be exogenous. One of the assumptions of the transfer function is that the relationship proceeds from X_t to Y_t. A spike at negative lags can possibly occur, indicating a feedback, simultaneity or a reverse effect. Apparent spike at negative lags may result from the failure to prewhiten that fails to trim out contaminating autocorrelation within the input series or may be due to the reverse effect or feedback in the relationship. No such spikes exist, the next is to identify the lag at which the first spike occurs in the cross-correlation plot. This lag is b, the number of periods before the X_t begin to influence Y_t.

Furthermore, the practice has suggested that after the first spike a clear dying-down pattern (exponential or sinusoidal) may exist in the SCC. The value of s is the number of lags that lie between the first spike and the beginning of the dying-down pattern. Sometimes, the s is not obvious due to the beginning of the dying-down pattern being questionable. In some cases, the value of s is somewhat arbitrarily determined. In addition, the value of r is determined by examining the dying-down pattern after lag b+s. Specifically, if the sample cross-correlation is dying down in an oscillatory or compound exponential fashion, it is reasonable to assign r=2. If it is dying down in a damped exponential fashion, it is reasonable to set r=1. The value of r is zero if there is no decay. Fine-tuning the identification process may require some trial and error with a view toward examining the parameters for significance and minimising the errors.

3.3.4. Estimation and Diagnosis Checking

After determining the values of b, r, and s, the model parameters can be estimated by least squares. To minimize sums of squared residuals the iterations continue until they do not improve significantly. The next step is to evaluate the model adequacy by examining the model residuals. The residuals can be examined by their ACF and PACF as well as the Ljung-Box Q test. The autocorrelation function and partial autocorrelation function of the residuals are employed. If no spikes exist in either RSAC or RSPAC, it is reasonable to conclude that the residuals are independent which meets the model assumption. Otherwise, if there are spikes in either RSAC or RSPAC, this indicates that residuals are dependent. New parameters should be incorporated into the model to account for those spikes.

A comparative evaluation of alternative models is necessary by examining the residuals or their error measures such as sums squared residuals, mean absolute errors, and mean absolute percentage errors. The comparison also includes the evaluation of the forecasts produced by those models. The MAPE is usually employed for investigation of the forecasts against the validation sample.

4. Data Analysis

The ARIMA models and transfer function models for the building costs of these three building types as well as the results of building cost forecasting are discussed in this section. To compare the

 doi:10.20944/preprints201903.0132.v1

forecasting performance of proposed ARIMA models and transfer function models, the MAPE is introduced. It can be expressed in Equation (6).

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right|}{n} \times 100\%, \tag{6}$$

where y_i is the actual observed value, $\hat{y}$ is the forecasting values, n is number of forecasting values.

4.1. Seasonal ARIMA Model

An important characteristic of seasonal time series is that the observations that are s time intervals apart are similar if the seasonal period is s [24]. In this study, s=4 quarters. Thus, for example, the observed value in the second quarter of one year will be alike to, or correlated with, that in the second quarter of the following year. However, it should be noted that the value in the second quarter is also correlated with that in the immediately preceding quarter, the first quarter. Therefore, there are two relationships going on simultaneously: (i) between observations for successive quarters within the same year, and (ii) between observations for the same quarter in successive years. It is necessary to develop two time series models – one for modelling the correlation between successive quarters within the years and one that can describe the relationship between same quarters in successive years and then combine the two. The model development process for seasonal models is similar to that used for regular time series models. First, the difference operations can also be used to make time series data stationary. However, for seasonal data, both regular difference and seasonal difference ($\nabla_s z_t = z_t - z_{t-s}$) can be used.

4.1.1. Univariate ARIMA Model for Building Cost of One-Storey House in New Zealand (HBC1)

For the building cost data for one-story house in New Zealand, denoted as HBC1, with quarterly seasonality, a regular first difference and a seasonal difference were applied $\nabla\nabla_4 z_t = \nabla(z_t - z_{t-4}) = z_t - z_{t-1} - z_{t-4} + z_{t-5}$. Finally, when the data was differenced twice the autocorrelation function (ACF) indicates that the data is stationary as the spikes fade out at both seasonal and non-seasonal lags, as shown in Figure 3 (a1). Having transformed the data to stationary, it is ready to identify the model. The patterns of the ACF and PACF shown in Figure 3 (a1) (a2) provide guidance to identify the stationary seasonal model. From Figure 3 (a1) showing the ACF for the stationary time series HBC1, it can be seen that there is a significant negative spike at lag 4, after which the seasonal autocorrelation pattern cuts off. It indicates that a moving average model applied to the four-quarter seasonal pattern. Moreover, it can also be seen that there is a significant autocorrelation at lag 1 and lag 3, after which the ACF cuts off. This indicates a first order and third order moving average terms in the regular time series model.

The ARIMA (0,1,3)(0,1,1)$_4$ was obtained to model the time series data HBC1. Moreover, the Ljung-Box chi-square statistics indicate that for all lags there is no significant autocorrelation remaining in the residuals, as shown in Figure 4 (a). The residual tests confirm that the model provides a fairly useful description of the dynamics in HBC1 data. Then the model was selected as a yardstick for the corresponding transfer function model.

 doi:10.20944/preprints201903.0132.v1

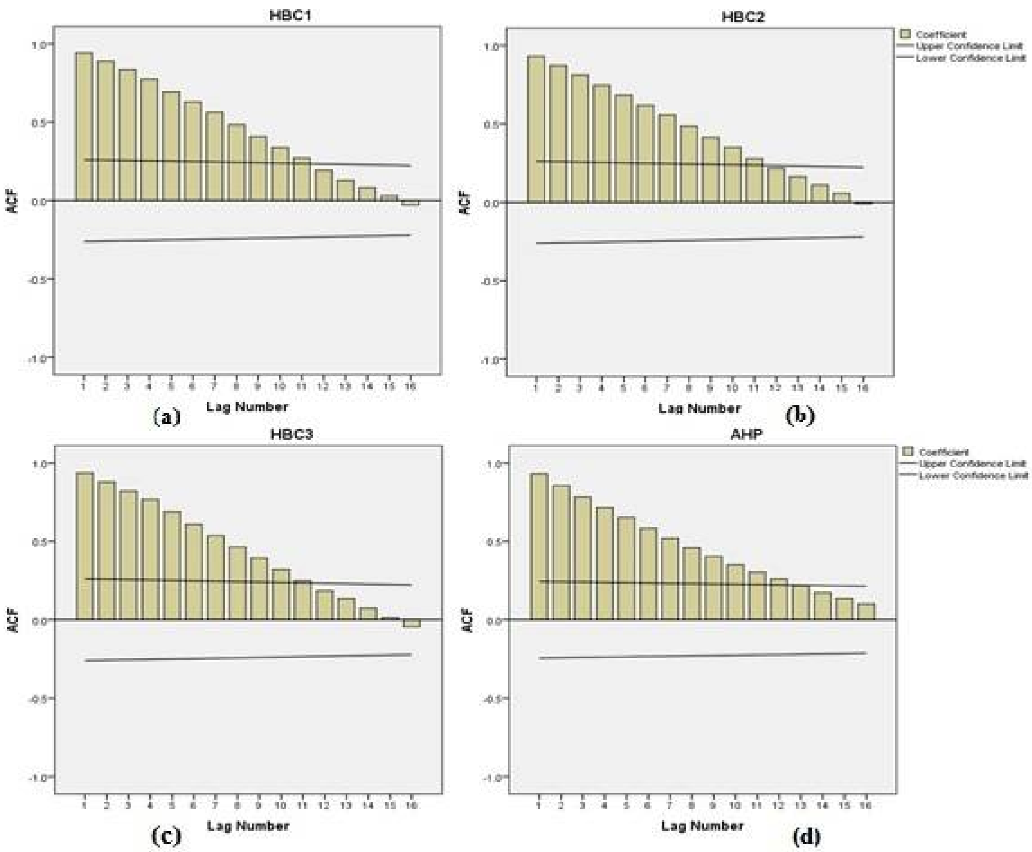

Figure 2. (a) The ACF of HBC1; **(b)** The ACF of HBC2; **(c)** The ACF of HBC3; **(d)** The ACF of AHP

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

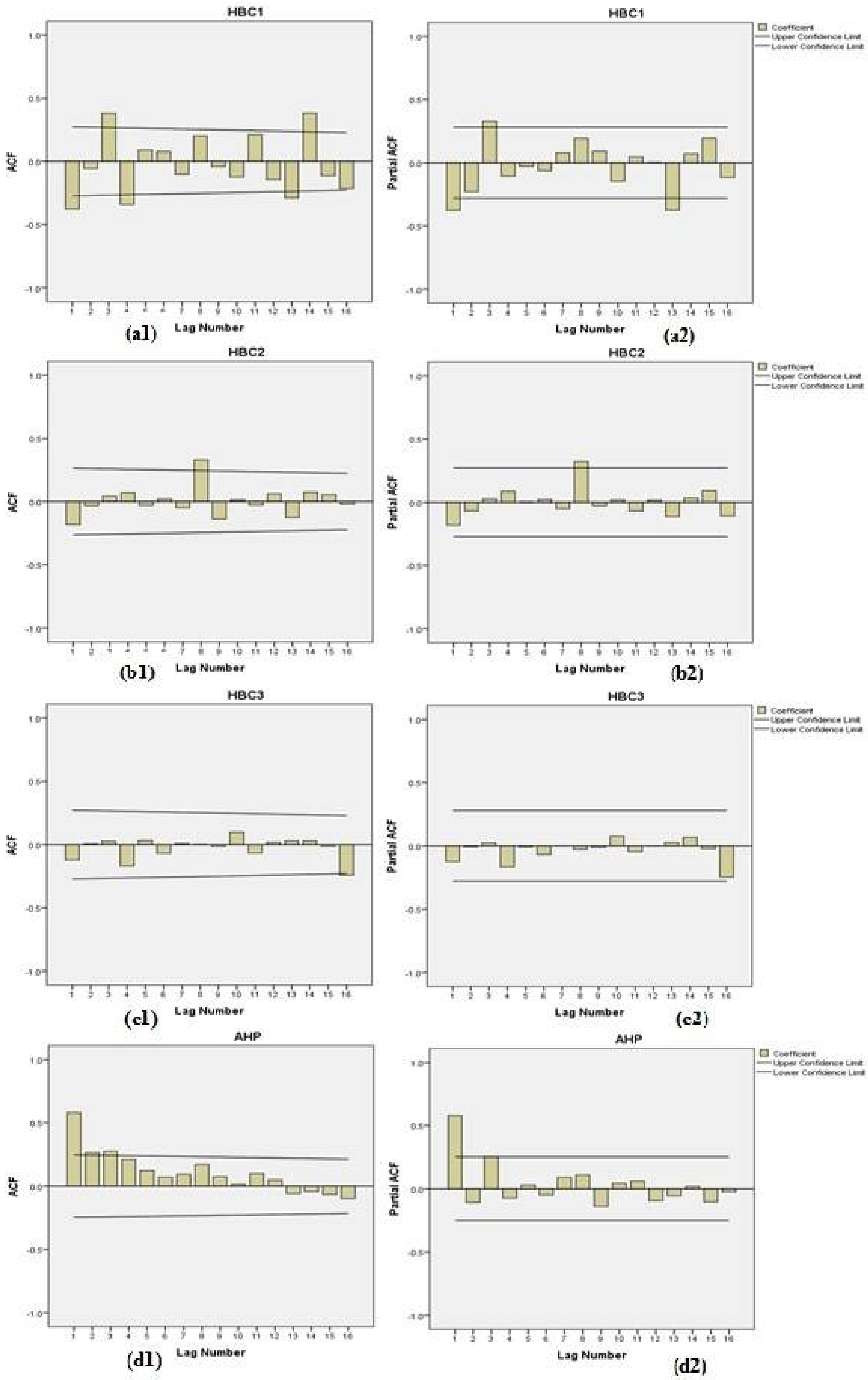

Figure 3. (a1) The ACF of stationary HBC1; **(a2)** The PACF of stationary HBC1; **(b1)** The ACF of stationary HBC2; **(b2)** The PACF of stationary HBC2; **(c1)** The ACF of stationary HBC3; **(c2)** The PACF of stationary HBC3; **(d1)** The ACF of stationary AHP; **(d2)** The PACF of stationary AHP

 doi:10.20944/preprints201903.0132.v1

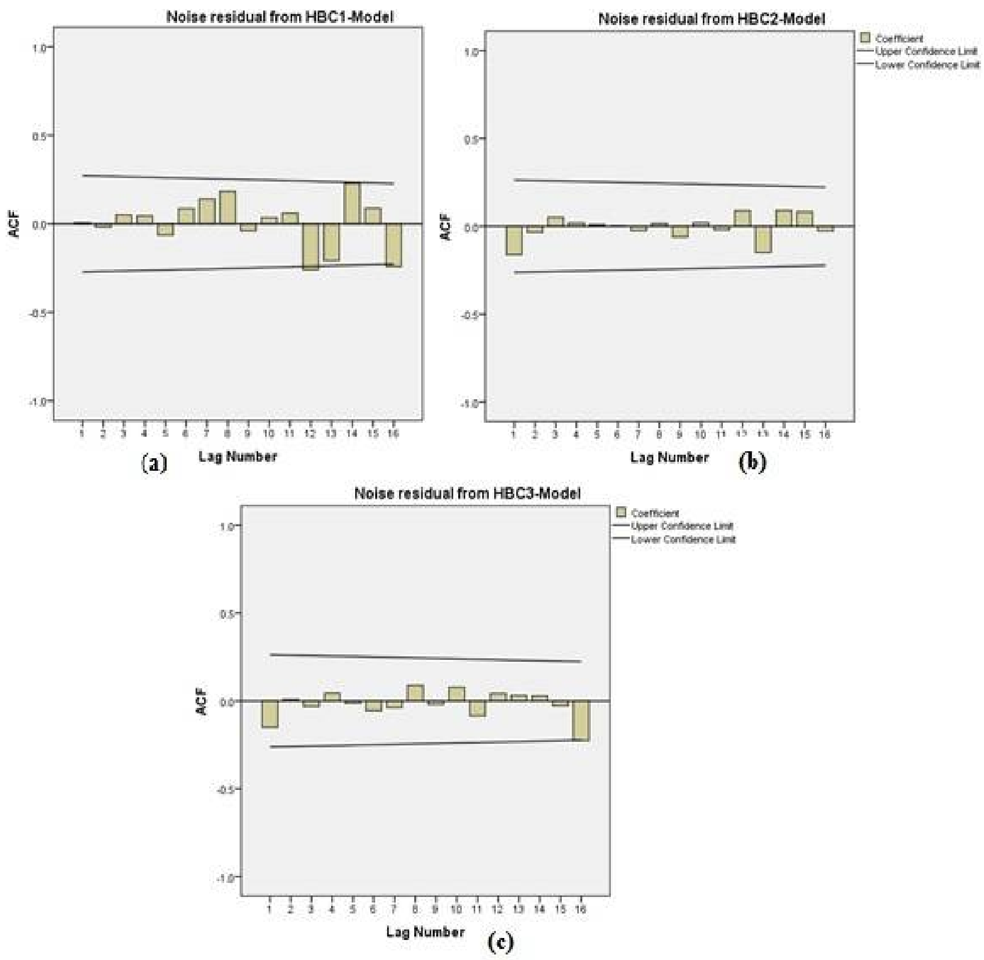

Figure 4. **(a)** The ACF of the residuals of ARIMA model for HBC1; **(b)** The ACF of the residuals of ARIMA model for HBC2; **(c)** The ACF of the residuals of ARIMA model for HBC3

4.1.2. Univariate ARIMA Model for Building Cost of Two-Storey House in New Zealand (HBC2)

The building cost of two storey house in New Zealand is denoted as HBC2. Basse on the ACF and PACF shown in Figure 3 (b1) (b2), the estimated seasonal ARIMA model for forecasting building cost of two-storey house in New Zealand is found to be ARIMA $(0,1,0)(0,0,2)_4$. The residuals shown in Figure 4 (b) test indicate that the model is adequate. The forecast error for two years ahead forecasts measured by MAPE and RMSE are illustrated in Table 1. The values of MAPE and RMSE are 2.177 and 61.76, respectively.

4.1.3. Univariate ARIMA Model for Building Cost of Town House in New Zealand (HBC3)

The building cost of town house in New Zealand is denoted as HBC3. Seasonal ARIMA models are fitted to stationary building cost series of town house, and the cost series require a regular difference and a seasonal difference to achieve stationarity. According to Figure 3 (c1) (c2), the seasonal ARIMA model is ARIMA$(0,1,0)(1,0,0)_4$. The final ARIMA model estimated and selected for forecasting future building cost of town house in New Zealand is ARIMA$(0,1,0)(1,0,0)_4$. The model is adequate and consistent with the underlying theory as the residuals shown in Figure 4 (c).

Table 1. Univariate ARIMA models for cost series

Cost series	Model	Parameters	Estimate	t-statistics	Ljung-Box		
					Statistics	df	p-value
HBC1	ARIMA(0,1,3)(0,1,1)4	θ_1	0.335	2.379	24.77	14	0.375
		θ_3	-0.297	-2.022			
		ϑ_1	0.447	3.177			
	R^2= 0.96, RMSE=35.60, MAPE=1.644, MAE=24.95, BIC=7.453						
HBC2	ARIMA(0,1,0)(0,0,2)4	ϑ_2	-0.394	-2.848	6.035	16	0.988
	R^2= 0.943, RMSE=61.76, MAPE=2.177, MAE=37.41, BIC=8.392						
HBC3	ARIMA(0,1,0)(1,0,0)4	$\emptyset_1$	0.594	5.483	8.168	17	0.963
	R^2= 0.969, RMSE=45.75, MAPE=1.576, MAE=28.92, BIC=7.719						

4.2. Transfer Function Model

The transfer function models will be developed below. In all cases only the first 56 observations were used, which indicate the data set up to 2014:Q4 were used to specify the models. This study used approach in [7] to transfer function identification. After some initial analysis, the New Zealand house price series (AHP) was used as the input variable. The ACF and PACF of AHP are shown in Figure 3 (d1) (d2). Using the proposed identification method, the ARIMA model for New Zealand house price (AHP) can be expressed in Equation (7).

$$(1-0.827B)(X_t - X_{t-1}) = a_t, \tag{7}$$

The model residuals were checked for independence. There is no significant autocorrelation at the 5% level. The results of the transfer function models for the residential building costs in New Zealand are shown in Table 2.

4.2.1. Transfer function model for building cost of one-storey house in New Zealand.

Building cost of one-storey house in New Zealand, denoted as HBC1, could be influenced by the New Zealand house prices. Building costs and house prices are positively correlated so that an increase in house prices leads to an increase in the building costs and vice versa. Thus, finding a mathematical relationship between these two variables in the transfer function form can be particularly valuable. The preliminary identification of series with the SAC and SPAC to test for stationarity and seasonality was performed. With the observation of slow damping of the SAC correlogram, the regular first differencing and seasonal differencing of order four were required to bring about stationarity. After regular and seasonal differencing of HBC1 output series, the series is stationary.

As shown in Figure 5 (a), the first cross-correlation which is placed out of its acceptance region is at lag zero, after which the CCF damping down is in an oscillation pattern. Thus, b is zero, s is zero and r =2 (oscillation pattern) The model parameters were estimated, then $C = 0.414, \delta_1 = -0.959, \delta_2 = -0.894$. Then the model can be expressed in Equation (8).

$$z_t^{HBC1} = \frac{C}{(1-\delta_1 B - \delta_2 B^2)} z_t^{(HP)} + a_t = \frac{0.414}{(1+0.959B+0.894B^2)} z_t^{(HP)} + a_t, \tag{8}$$

The model residuals should be examined to meet the assumption of error independence. According to Figure 6 (a1)(a2), it can be inferred that the residuals are not a white noise process, therefore, to redevelop the model is necessary. After examining the pattern of RACF and RPACF, a moving average term and a seasonal moving average term should be included in the model. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

moving average parameters θ_1=0.298, ϑ_1 =0.251, respectively. Therefore, the model can be expressed in Equation (9).

$$z_t^{HBC1} = \frac{C}{(1-\delta_1 B-\delta_2 B^2)} z_t^{(HP)} + (1-\theta_1 B)(1-\vartheta_1 B^4)\varepsilon_t = \frac{0.414}{(1+0.959B+0.894B^2)} z_t^{(HP)} + (1-0.298B)(1-0.251B^4)\varepsilon_t, \quad (9)$$

In order to access the model expressed in the above Equation, the autocorrelation function of the model residuals was considered. The residuals shown in Figure 7 suggest that the residuals satisfy the condition of a white noise process. Moreover, the results of Ljung-Box test also indicate the absence of the autocorrelation in the residuals. Therefore, the model in Equation (9) is adequate.

4.2.2. Transfer function model for building cost of two-storey house in New Zealand.

Building cost of two-storey house in New Zealand denoted as HBC2. The preliminary identification of series with the SAC and SPAC to test for stationarity and seasonality was performed. With the observation of slow damping of the SAC correlogram, the regular first differencing was required to bring about stationarity. After the first regular differencing of HBC2 output series, the series is stationary.

First, this ARIMA model was used to filter the building cost for two-storey house. Then, the cross-correlation between these two pre-whitened series was computed as shown in Figure 5 (b). The first significant spike appears at lag 3, suggesting a delay of three periods. The orders of the transfer function b=3, s=1, and r=1 can be identified. The ordinary least square method was used to estimate the model parameters. The transfer function is shown in Equation (10).

$$z_t^{HBC2} = \frac{-1.341(1+1.257B)}{1-0.429B} B^3 z_t^{(HP)} + a_t, \quad (10)$$

The model residuals should be examined to meet the assumption of error independence. According to Figure 6 (b1) (b2), it can be inferred that the residuals are a white noise process. The residuals shown in Figure 6 (b1) (b2) suggest that the residuals satisfy the condition of a white noise process. Moreover, the results of Ljung-Box test also indicate the absence of the autocorrelation in the residuals. Therefore, the model in Equation (10) is adequate.

4.2.3. Transfer function model for building cost of town house in New Zealand.

Building cost of town house in New Zealand is denoted as HBC3.With the observation of slow damping of the SAC correlogram, the regular first differencing and seasonal differencing of order four were required to bring about stationarity. After regular and seasonal differencing of HBC3 output series, the series is stationary shown from the SAC. In the identification of the transfer function model, it is necessary to apply the same pre-whiting transformation to both the input and output series. Thus, the ARIMA model for house price (AHP) also applied to the building cost series of town house. However, before the pre-whiten progress, the cost series has been transformed to stationary by a regular difference and a seasonal difference.

Figure 5 (c) shows the first significant spike is at lag 0 on the cross-correlation function, indicating no delay time. The orders of the transfer function b=3, s=0, and r=2 can be identified. The ordinary least square method was used to estimate the model parameters. The transfer function is shown in Equation (11).

$$z_t^{HBC3} = \frac{C}{(1-\delta_1 B-\delta_2 B^2)} B^3 z_t^{(HP)} + a_t = \frac{0.089}{(1-1.614B+0.849B^2)} B^3 z_t^{(HP)} + a_t, \quad (11)$$

The residuals were checked for autocorrelation, the chi-square values for all the lags were found to be not significant at the 5% level, as shown in Figure 6 (c1) (c2). This suggests that the model residuals do not have problems with autocorrelation and, indeed, white noise.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

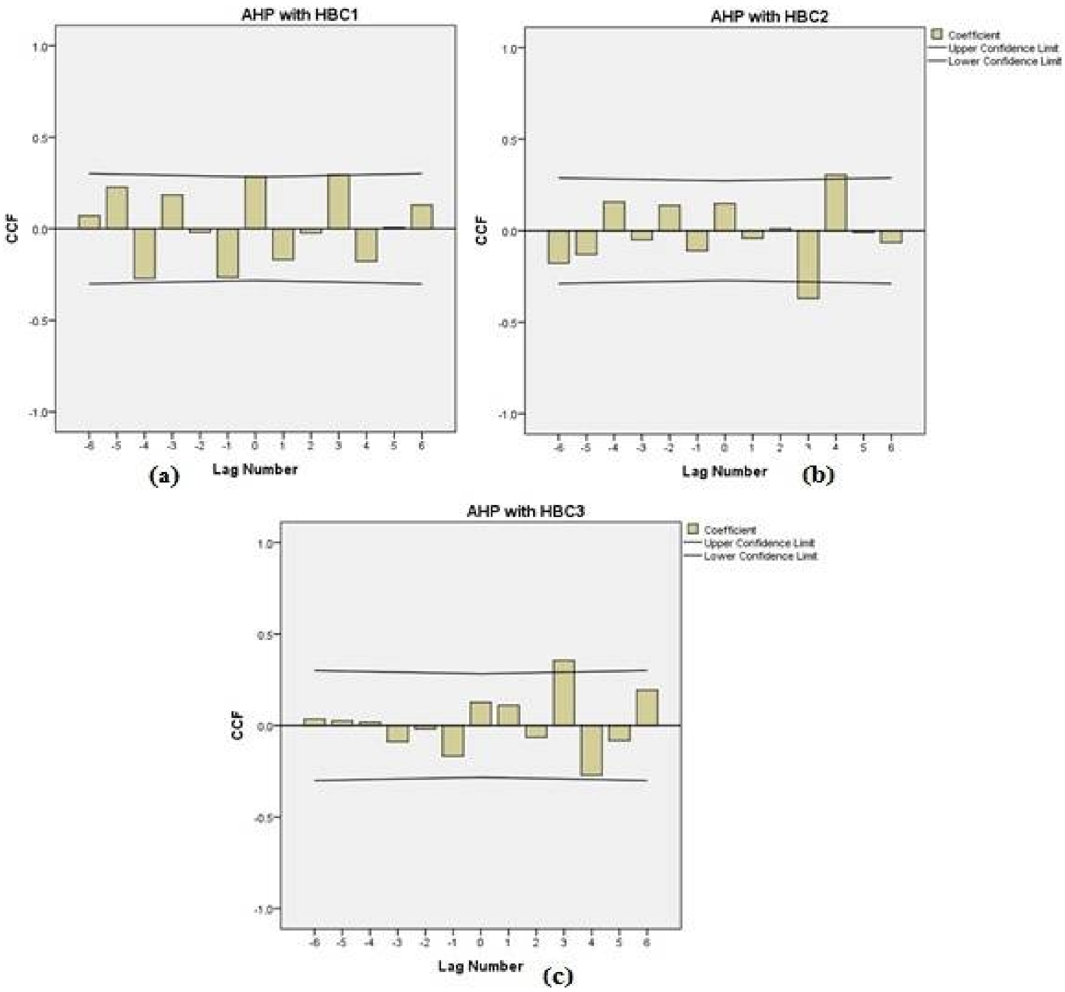

Figure 5. (a) The CCF of AHP and HBC1; **(b)** The CCF of AHP and HBC2; **(c)** The CCF of AHP and HBC3;

 doi:10.20944/preprints201903.0132.v1

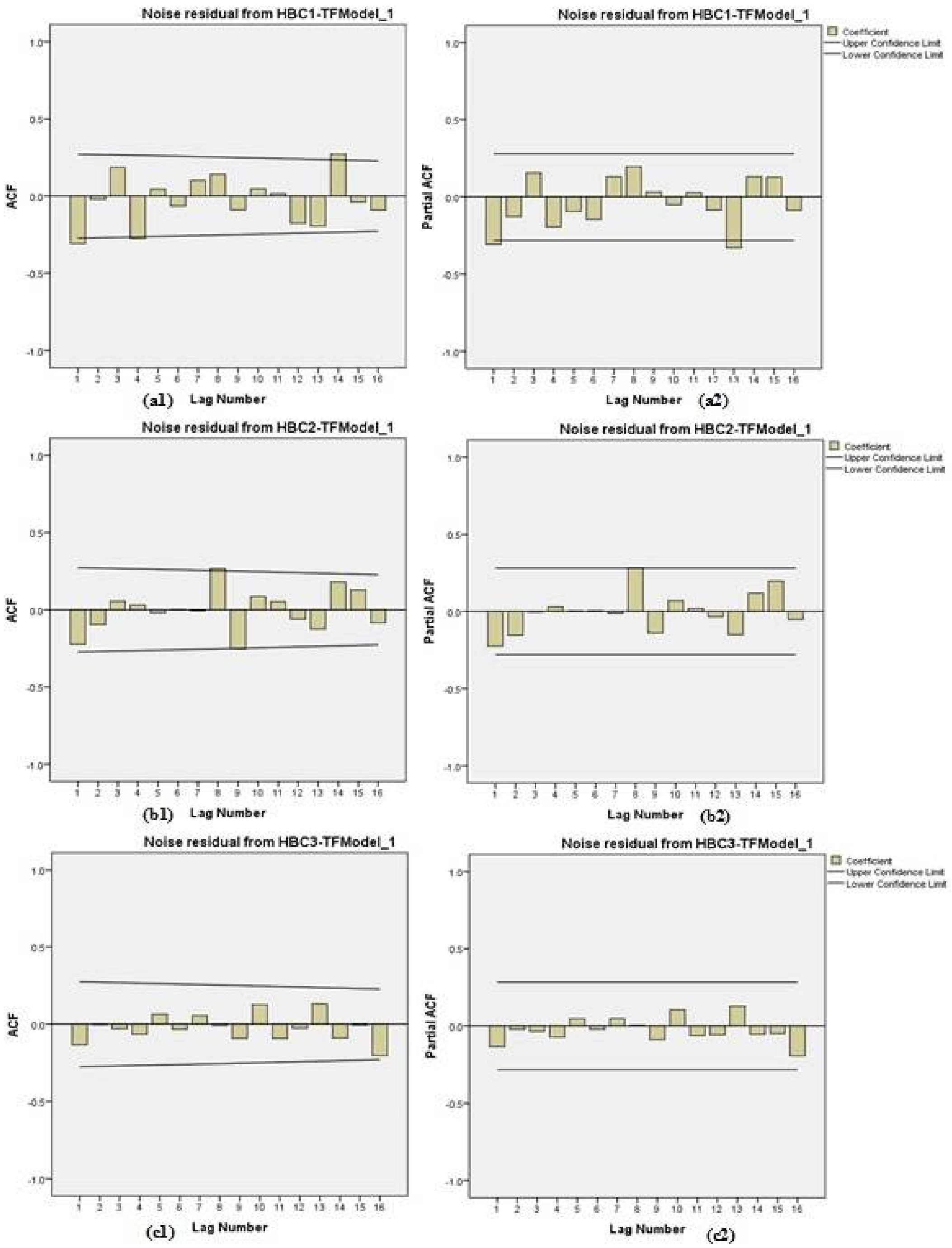

Figure 6. **(a1)**The ACF of the residuals of transfer function model for HBC1; **(a2)**The PACF of the residuals of transfer function model for HBC1; **(b1)**The ACF of the residuals of transfer function model for HBC2; **(b2)**The PACF of the residuals of transfer function model for HBC2; **(c1)**The ACF of the residuals of transfer function model for HBC3; **(c2)**The PACF of the residuals of transfer function model for HBC3

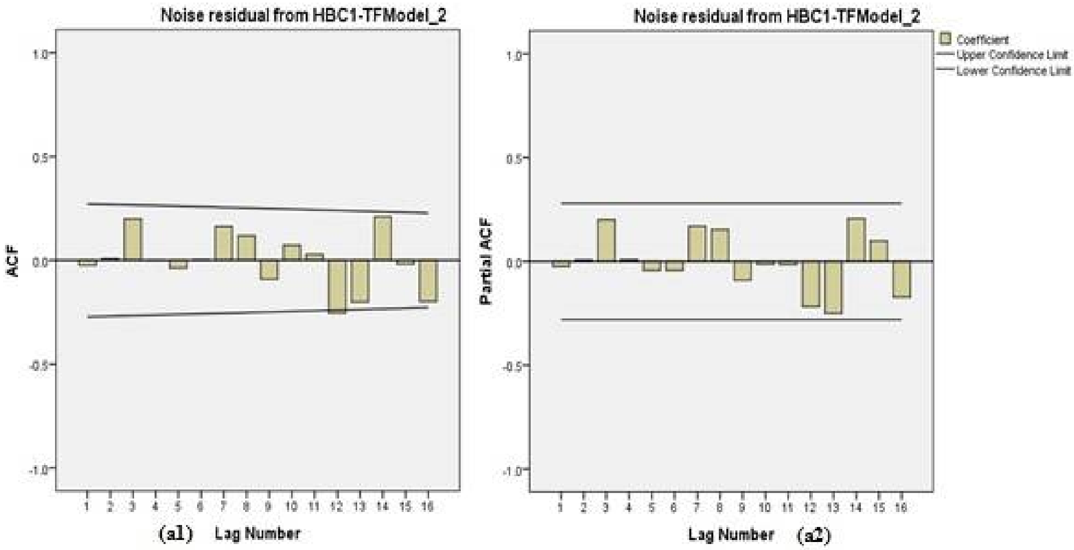

Figure 7. (a) The ACF of the residuals of the re-model for HBC1; **(b)** The PACF of the residuals of the re-model for HBC1

Table 2. Transfer function parameters in the different cost series

Cost series	Model order			TF model parameters	Estimate	t-statistics	p-value
	b	s	r				
HBC1	0	0	2	C	0.414	2.189	0.034
				δ_1	-0.959	-9.110	0
				δ_2	-0.894	-8.434	0
				θ_1	0.298	2.007	0.051
				ϑ_1	0.311	2.630	0.211
	R^2= 0.960, RMSE=35.83, MAPE=1.690, MAE=25.63, BIC=7.543						
HBC2	3	1	1	C	-1.341	-2.969	0.005
				ω_1	1.257	3.596	0.001
				δ_1	0.429	2.103	0.041
	R^2= 0.932, RMSE=59.55, MAPE=2.476, MAE=43.56, BIC=8.405						
HBC3	3	0	2	C	1.088	3.327	0.002
				δ_1	-0.790	-4.529	0
				δ_2	-0.645	-3.698	0.001
	R^2= 0.952, RMSE=47.62, MAPE=1.850, MAE=34.73, BIC=7.961						

4.3. Forecast Evaluation

The above proposed models were evaluated for their out-of-sample forecasting performance. The models were developed using data up to 2014:Q4 and then forecasts were made for the following 16 quarters. As the time index advanced, the sample size increases, the identified model is re-estimated and used for forecasting over the validation sample. However, this model evolution would not affect the forecasting performance as the model orders do not allow changing as more observations are accumulated. Also, the model comparison is carried out in a meaningful and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

systematic way by using formal statistical tests. And the model parameters do not reveal substantial variations and maintain the same sign over the whole validation sample. The MAPE values for the transfer function models and seasonal ARIMA models for the period 2015:Q1-2018:Q4 are presented in Table 3. The results clearly indicate that the transfer function model outperforms the seasonal ARIMA models for each building cost series considered.

The improvement of forecasting efficiency from the use of transfer function models over the ARIMA models is best illustrated by observing the error measurements. For the building cost of one-storey house (HBC1) in New Zealand, there is a reduction of about 39.1% in the MAPE from the seasonal ARIMA model to the transfer function model. For the building cost of two-storey house in New Zealand (HBC2), the reduction from using the transfer function model is even more pronounced: there is a 53.5% reduction in the MAPE from the ARIMA model to the transfer function model. For the building cost of town house in New Zealand (HBC3), the transfer function models provide a 79.9% MAPE reduction over the seasonal ARIMA model. Note that the transfer function models are, for all three cost series, superior to the seasonal ARIMA models.

Table 3. Comparison of model performance (MAPE)

Cost series	ARIMA	TF
HBC1	2.091	1.273
HBC2	2.935	1.365
HBC3	5.766	1.161

5. Results Discussion

The fluctuations in the building cost index are problematic for cost estimating. The ARIMA models completely take into account the dynamic process of cost series, the seasonality, and serial correlation in the residuals to obtain precise cost forecasts. Although the model can dynamically model the time series data, it cannot investigate the causality between independent variables and dependent variables. Based on the knowledge and existing literature, building costs are not only influenced by the construction industry, but also are influenced by other factors, such as house prices, economic conditions, population, incomes, and financial loans. The inclusion of explanatory variables into the models can provide opportunities to forecast building cost more accurately. New Zealand house prices index was selected as an explanatory variable for predicating building cost in this study.

It is supposed that the transfer function model can describe more about the characteristics of the output than a univariate model by modelling the dynamic relationship between the input and output variables. The gain from including New Zealand house price in the transfer function models is significant. This is consistent with the findings of [25] which indicated that model inclusion of an exogenous variable can improve the forecasting performance. House is one kind of building construction products, which connect the building construction sector to the housing sector. So, the building costs subject to the effects of the changes in house prices. New Zealand house prices stimulate changes in the three building costs, although building costs are usually viewed as fundamentals. The findings are also supported by the studies of [26, 27].

The analysis results suggest that the simplest model is not always an appropriate model for predicting, particularly when additional information is available. Univariate models like ARIMA models should primarily be used as a benchmark in comparing forecasting performance. When explanatory variables are available they can be included in the model to gain efficiency. Multivariate models can then be realised to improve forecasting performance.

6. Conclusion

Accurate cost estimating can help in developing accurate budgets and preventing under- or over-estimation. Based on the Box-Jenkins model development process, transfer function models and seasonal ARIMA models for building costs of one-storey house, two-storey house, and town house in New Zealand, were developed. The predictability of the transfer function models was compared

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 March 2019 doi:10.20944/preprints201903.0132.v1

with the univariate ARIMA models. Transfer function models surpassed the seasonal ARIMA models. Due to the inherent complexity in the process, the transfer function models that contain an exogenous variable (more information) are better at forecasting than the simpler ARIMA models. The results also indicate that the use of house price significantly reduces the forecasting error in building costs, relative to seasonal ARIMA models. The study has shown that the inclusion of an explanatory variable input within an ARIMA framework leads to an improvement in forecasting performance. More interestingly, the results also demonstrate that variations in house prices can result in fluctuations in residential building costs.

The contributions of this study are mainly in three areas. First, the main contribution of this study is to enhance cost estimation accuracy and guide the industry practitioners in the preliminary design stage. The developed models reduce the time and cost required to develop a cost list for a given project. Moreover, the forecasts produced by the transfer function models are statistically reliable considering the 95% upper and lower confidence limits. Second, the inclusion of New Zealand house prices in the transfer function model significantly improves the cost forecasting performance. The findings contribute to the present body of knowledge on cost estimating and may serve as a valuable guide for future model development. Third, the results of the study prove that New Zealand house prices are significant influencing factors of residential building cost.

The scope of this study is limited to building cost index. Although building cost index has been extensively employed by industry professionals and organisations, it cannot represent all types of building costs in New Zealand's construction industry. It is expected that this study may assist industry professionals to prepare more accurate cost estimates, budgets, and bids. Future studies can be conducted by using the multivariate regression method to produce more accurate forecasts by including other variables such as employment, the labour cost index, and the producer price index. Moreover, whether a hybrid model that combines different forecasting techniques can produce a substantive improvement in forecasting should be investigated. In addition, an extension of the current study could include: the use of a vector structural model; employ intervention models; and, finally, evaluate whether non-linear models represent an improvement over linear models.

Author Contributions: Conceptualization, L.Z., J.M.; methodology, L.Z., H.Z.; software, Z.L., H.Z.; validation, L.Z., Z.L., H.Z.; writing—original draft preparation, L.Z.; writing—review and editing, J.M.; supervision, J.M.; project administration, J.M.; funding acquisition, L.Z

Funding: This research was funded by China Scholarship Council, grant number 201206130069, Massey University, grant number 09166424, and The APC was funded by the both grants.

Acknowledgments: The authors would like to thank the China Scholarship Council (CSC) for its support through the research project and also the Massey University. The authors would like to thank Reserve Bank of New Zealand and Ministry of Business, Innovation and Employment for providing data to conduct this research. In addition, I would like to thank all practitioners who contribute in this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gido, J.; Clements, J. P., *Successful project management*. Sixth edition ed.; Cengage Learning: Stamford, CT, USA, 2015.
2. Sonmez, R., Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering* **2004,** *31*, 677-683.
3. Takano, Y.; Ishii, N.; Muraki, M., Determining bid markup and resources allocated to cost estimation in competitive bidding. *Automation in Construction* **2018,** *85*, 358-368.
4. Merrow, E. W., *Concepts, Strategies and Practices for Success*. John Wiley and sons: New Jersey, 2011.
5. MBIE *NZ Sector Report 2013 – Construction* Ministry of Building Innovation and Employment: Wellington, New Zealand, 2013.
6. Jrade, A. A conceptual cost estimating computer system for building projects. Concordia University, Canada, 2000.
7. Box, G. E. P.; Jenkins, G. M., *Time Series Analysis: forecasting and control*. Holden-Day: Oakland, California, 1976.
8. Lowe, D. J.; Emsley, M. W.; Harding, A., Predicting construction cost using multiple regression techniques. *ASCE Journal of Construction Engineering and Management* **2006,** *132*, 750-758.
9. Hwang, S., Dynamic regression models for prediction of construction costs. *Journal of Construction Engineerign and Management* **2009,** *135* (5), 360-367.
10. Yip, H.-I.; Fan, H.; Chiang, Y.-H., Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models. *Automation in Construction* **2014,** *38*, 30-38.
11. Morcous, G.; Rivard, H.; Hanna, A. M., Case-based reasoning system for modeling infrastructure deterioration. *Journal of Computing in Civil Engineering* **2002,** *16*, 104-114.
12. Yau, N. J.; Yang, J. B., Case-based reasoning in construction management. *Computer-Aided Civil and Infrastructure Engineering* **1998,** *13*, 143-150.
13. Kulendran, N.; Witt, S. F., Leading indicator tourism forecasts. *Tourism Management* **2003,** *24*, 503-510.
14. Qi, M., Predicting US recessions with leading indicators via neural network models. *International Journal of Forecasting* **2001,** *17*, 383-401.
15. Chen, M.-C.; Tsai, I.-C., A cobweb theory of house price incorporating investor behavior. *Academia Economic Papers* **2007,** *35*, 315-344.

 doi:10.20944/preprints201903.0132.v1

16. Janssen, J.; Kruijt, B.; Needham, B., The honeycomb cycle in real estate. *The Journal of Real Estate Research* **1994,** *9*, 237-251.
17. Tsai, I.-C., Housing supply, demand and price: Construction cost, rental price and house price indices. *Asian Economic Journal* **2012,** *26* (4), 381-396.
18. Sunde, T.; Muzindutsi, P.-F., Determinants of house prices and new construction activity: An empirical investigation of the Namibian housing market. *The Journal of Developing Areas* **2017,** *51* (3), 390-407.
19. Makridakis, S.; Wheelwright, S. C.; McGee, V., *Forecasting: methods and applications*. 2nd ed.; Wiley: New York, 1983.
20. Mills, T. C., *Time series techniques for economists*. Cambridge University Press: New York, 1990.
21. Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C., *Time series analysis: forecasting and control*. Englewood Cliffs: New Jercy, 1994.
22. Yaffee, R.; McGee, M., *Introduction to time series analysis and forecasting with applications of SAS and SPSS*. Academic Press Inc: New York, 2000.
23. Box, G. E.; Jenkins, G.; Reinsel, G., *Time sereis analysis*. 4th ed.; Wiley: New Jersey, 2008.
24. Bisgaard, S.; Kulahci, M., *Time series analysis and forecasting by example*. John Wiley & Sons, Inc New Jersey, 2011.
25. Ashley, R., One the usefulness of macroeconomic froecasts as inputs to forecasting models. *Journal of Forecasting* **1983,** *2*, 211-223.
26. Grimes, A., Housing supply, land costs and price adjustment. *Real Estate Economics* **2010,** *38* (2), 325-353.
27. Tsai, I.-C., Housing supply, demand and price: Construction cost, rental price and house price indices. *Asian Economic Journal* **2012,** *26* (4).