

Article

Transfer Function Analysis: Modelling Residential Building Costs in New Zealand by Including the Influences of House Price and Work Volume

Linlin Zhao ¹, Jasper Mbachu ², Zhansheng Liu ^{1,*} and Huirong Zhang ¹

¹ College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100091, China; llzhao@bjut.edu.cn (L.Z.); zhanghuirong@bjut.edu.cn (H.Z.)

² Faculty of Society and Design, Bond University, Gold Coast 4226, Australia; jmbachu@bond.edu.au

* Correspondence: liuzhansheng@bjut.edu.cn

Received: 23 May 2019; Accepted: 18 June 2019; Published: 25 June 2019



Abstract: An accurate cost estimate not only plays a key role in project feasibility studies but also in achieving a final successful outcome. Conventionally, estimating cost typically relies on the experience of professionals and cost data from previous projects. However, this process is complex and time-consuming, and it is challenging to ensure the accuracy of the estimates. In this study, the bivariate and multivariate transfer function models were adopted to estimate and forecast the building costs of two types of residential buildings in New Zealand: Low-rise buildings and high-rise buildings. The transfer function method takes advantage of the merits of univariate time series analysis and the power of explanatory variables. In the dynamic project conduction environment, simply including building cost data in the cost forecasting models is not valid for making predictions, because the change in demand must be considered. Thus, the time series of house prices and work volume were used to explain exogenous effects in the transfer function model. To demonstrate the effectiveness of transfer function models, this study compared the results generated by the transfer function models with autoregressive integrated moving average models. According to the forecasting performance of the models, the proposed approach achieved better results than autoregressive integrated moving average models. The proposed method can provide accurate cost estimates that can help stakeholders in project budget planning and management strategy making at the early stage of a project.

Keywords: ARIMA; transfer function model; time series analysis; short-term forecasts; residential building costs

1. Introduction

The preliminary cost estimation for building projects are usually the basis of project financial feasibility and cost budgeting in the early stages of planning and for effective and efficient project control, monitoring and execution [1]. Reliable and accurate cost estimation of building projects is very important for project stakeholders. However, it is common that the final project cost greatly exceeds the initial cost estimates [2]. According to the findings of [3], nine out of ten cost overrun projects were caused by inaccurate cost estimates in the early stages. A fast, inexpensive and comparatively accurate early-stage cost estimation is nevertheless essential in project decision-making and project feasibility studies [4]. However, owing to the level of project uncertainty and relative short time for preparing cost estimates with relative limited data and information, cost estimates are highly subject to a wide array of inaccuracies. Conventionally, the cost estimates mainly depend on experience of industry professionals and unit cost rate from previous projects. However, the process is complex, and it is

challenging to ensure accuracy. Hence, this study introduces the transfer function model to estimate building costs.

A review of the published research and studies on the performance of the forecasting models explored that the efficiency of the univariate model approaches is limited. The accuracy of the forecasting models has been improved when other explanatory variables are included. In this study, three forecasting models, namely, autoregressive integrated moving average (ARIMA) model, bivariate transfer function model (one explanatory variable), multivariate transfer function model (two explanatory variables), were applied to residential building costs of New Zealand. The performance of the forecasting models is determined by the error measures of the models. According to [5], the ARIMA model was used as a benchmark model because of its standard modelling process and good forecasting performance.

At present, there is a limited number of empirical research and studies focused on building cost predication, particularly for residential building costs. However, the construction sector is important to the economy due to its strong backward and forward linkage with other sectors [6–8]. The residential sector was selected due to residential construction represents an average of 38% of the total construction output in New Zealand [9]. Moreover, transfer function models including two explanatory variables for forecasting building costs remain unexplored. In the current study, time-series data on building costs of low-rise residential buildings (house and town house) and high-rise buildings (apartments) in New Zealand were used as representative cases to fill this gap in knowledge. Thus, the study illustrates both univariate model and multivariate model approaches to forecasting of two different time-series data on residential building costs.

The remaining of the study is organised as follows: Section 2 summarizes the related literature, and Section 3 presents the data and methodology that are used in the study. The univariate models, bivariate models and multivariate models for residential building costs were developed in Section 4, and the forecasting performance of them was also compared in Section 5. Section 6 discusses the results, and Section 7 summarises and concludes the findings and implications.

2. Literature Review

2.1. The Factors Impact Building Costs

Building cost estimates traditionally include expenditures for labor, material, equipment and construction method, but they are also impacted by many other factors such as industry factors, market conditions, and macro-economic environment. The fluctuation of variables, such as interest rates, the consumer price index, house prices and the building service index may influence the cost [10]. In order to accurately estimate the building cost, it is necessary to identify the main determinant factors. Some studies have been conducted in this area. For example, [11] found that the construction cost can be impacted by factors such as gross domestic product (GDP), crude oil prices, consumer price index (CPI), unemployment rate, producer price index (PPI), number of building permits and money supply.

In [12], a study was conducted in the relationship between the real estate sector and construction industry and highlighted that increasing the price for real estate has created a favourable investment climate and pushed the building cost up. Furthermore, [13] also addressed that building costs will rise and fall, with time-lags, in relation to changes in work volume. According to [14], building costs are determined by short-run resource costs plus a mark-up that is a positive function of general level of construction activity. [15] also stated that the existence of a competitive environment significantly increases building costs due to prices in goods and service sectors that are sensitive to changes in the demand. The increased work volume in the construction industry also imposes increasing pressure on the supply chain. Indeed, unavailability of resources and unreliable distributions are usually considered as the main reasons of project delays leading to additional costs [16]. This study incorporates two variables in the transfer function modelling process as they may have more direct effects on movements of building costs.

2.2. Estimating Methods

According to the study by [17], developing an accurate forecasting model for the cost index would be useful for both decisionmakers and policy analysts. The estimating method also plays a key role in estimating accuracy. As a common practice, cost estimating is usually based on the experience of industry professionals, and the unit cost rate from previous projects. However, the process is complex and time-consuming, and it is challenging to ensure the accuracy. Some studies have proposed methods for forecasting building costs. Many methods have been employed to forecast the construction cost index, including time series methods [18,19], multiple liner regression [20,21], support vector machines (SVM) [22,23] and ANNs [24–26].

Univariate time series methods, including seasonal ARIMA models and the exponential smoothing method, have been compared in a study by [27], in order to select a forecasting model for the construction cost index. [19] developed an interrupted time-series model for forecasting construction cost. The cost forecasts are more accurate and reliable than the forecasts obtained from the conventional ARIMA model, since the 2008 economic regression is considered. However, the univariate time series methods only examine the cost trend, they are unable to model the explanatory power of cost-influencing factors.

Multiple regression analysis has been used to forecast construction cost, since the construction cost index may be influenced by several financial and macro-economic variables. The multiple regression analysis was considered as a reliable and powerful parametric estimating approach and has a great potential for cost estimating [4]. However, this approach requires clear understanding of the modelling technique and collinearity problem. Moreover, in study [26], the authors addressed the advantages of the multivariate models as they can explain the impact of the independent variables on the dependent variable and have relative robust theoretical framework compared with univariate models. However, identifying the appropriate independent variable in multivariate models is difficult. [21] used multiple regression analysis (MRA) and ANNs for estimating the construction cost at the early stage of a project, the better to budget for it. The forecasting performance of the models indicates that the multiple regression model can generate better cost estimates than the ANN model.

Furthermore, some studies have been performed to forecast the construction cost by using other variables that are assumed to have an explanatory relationship with the construction cost. For example, [28] built a vector error correction (VEC) model for forecasting the tender price index by incorporating financial and macroeconomic variables, such as gross domestic product (GDP), bank interest rate, consumer price index, money supply, and unemployment rate. The support vector machine (SVM) technique has been used to forecast cost index in some studies [22,23]. The core of SVM is the kernel function. The performance of the kernel function is greatly associated with the values of important parameters. However, an appropriate technique for selecting the parameters is lacking.

Study [29] indicated that one of the main disadvantages of the forecasting models is the assumption of linear relationships between variables; thus, techniques that can model non-linear relationships are necessary. The use of artificial neural networks (ANNs) method has received increasing attention during the last two decades [30]. Therefore, the ANN model has been widely used in some studies, as it can model non-linear relationships. ANNs approach that simulates the learning process of the human brain has been widely used in many fields [31]. It usually produces a more accurate result than conventional estimation methods such as expert judgment and regression analysis that due to it can describe uncertain and non-linear relationships [32–34]. For example, [35] developed a model for forecasting the construction site cost index by using artificial neural networks (ANNs). Additionally, study [36] provided a hybrid model, which combines both multivariate regression method and ANNs, to forecast construction cost. [24] combined ANN with the principal component analysis (PCA) to develop a model for estimating the project cost of residential buildings. Although ANN can model the cost properly, the model is unstable. Moreover, the modelling process of ANN is fuzzy and more like a black box, since it cannot provide any relationship between the cost and explanatory variables.

Recently, a case-based reasoning method was used to forecast construction cost, as it can include both quantitative and qualitative variables. The case-based reasoning method copes with a new case by using previous experience and knowledge and similar cases. It is a radical departure from the methods discussed above, which only incorporate quantitative variables. Some studies [37,38] have been conducted using the case-based reasoning method. Although the case-based reasoning (CBR) method has been increasingly used in estimating construction cost, there exist some challenging issues, including parameter selection, weight assignment, and the rules of reuse case [39,40]. CBR highly depends on similar case data. If the similar data is not sufficient, the inaccurate results of new cases will be generated [41].

2.3. Gaps in the Existing Literature

Although the studies and research about cost modelling have been undertaken extensively in the literature, there are limited studies about the transfer function modelling which include the effects of other variables on the building costs. In previous studies, both univariate time series methods and multivariate regressive methods were used to develop an appropriate model for forecasting building costs. Each has its own advantages and disadvantages. Univariate time series methods can fit the changes and trends in the variables but overlook the impacts of the macro-environment. For example, the impacts of supply and demand on the building cost index are not captured in the univariate techniques. The univariate models lack explanatory ability. Multivariate causal methods incorporate explanatory variables to account for the influencing effects, in order to provide more accurate forecasts. However, the correlation between the explanatory variables may affect the accuracy of the forecasts. In this study, the transfer function method is used to combine the merits of both univariate time series methods and multivariate causal methods. By using this method, the changes and trends of the variable can be modelled, and the exogenous effects were modelled by incorporating explanatory variables. Moreover, the transfer function model used in this study incorporates explanatory variables and their past values that contain useful information for forecasting the future values of the construction cost index. It can model the time-lag relationship between building cost and explanatory variables by using CCF analysis. In addition, the existing literature about forecasting residential building costs for New Zealand is rare. This study provides the first foundation for studying the movement of building cost, which may help stakeholders in decision-making.

3. Research Methodology

3.1. The Data

The data set consists of quarterly observations for the period 2001:Q1–2018:Q4 for residential building costs of low-rise and high-rise buildings in New Zealand, and two explanatory variables including housing prices and work volume. The training sample is from 2001:Q1 to 2014:Q4, a total of 56 observations. The validation sample used the remaining 16 observations, from 2015:Q1 to 2018:Q4. The cost data is the approximate cost per building type per square meter in New Zealand.

Note that, this study categorised residential building costs into two major segments, namely, low-rise building and high-rise building costs. The low-rise residential building is believed to be the most representative of the general house type prevailing in New Zealand, while high-rise residential buildings are apartments or retirement villages. LBC and HBC are the building cost indexes of residential low-rise and high-rise buildings, respectively, which are obtained from the QV cost builder that is widely used in the construction industry of New Zealand. Cost estimation at the early stage is highly based on the use and availability of historical cost data and information [42].

As a measure of house prices (HP), the study employed house price index is quarterly data and spans from 2001:Q1 to 2018:Q4 obtained from Reserve Bank of New Zealand (RBNZ). The annual number of building consents issued is the major indicator of new building projects in New Zealand. The number of building consents is a significant indicator of construction work volume which can

be used as a proxy of construction work volume in the present study. Building consent series (BC) indicates the quarterly number of building consents awarded in New Zealand officially reported by Statistics of New Zealand, from 2001:Q1 to 2018:Q4. A complete set of descriptive statistics is available upon request.

3.2. ARIMA Model

The ARIMA model is a combination of the autoregressive (AR) model, differencing of time series, and moving average model (MA). The modelling process of the ARIMA technique was proposed by [43]. The model ARIMA (p,d,q)(P,D,Q)_L can be expressed in Equation (1). The basic steps of applying an ARIMA model include identification, estimation, diagnosis checking, and forecasting. The examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) to gain insights into the characteristics of the time series is the first stage of the approach. The series is differenced to attain stationarity. The ACF and PACF plots of the stationary series are examined to identify an appropriate model. At the estimation and diagnostic checking stages, the portmanteau test is used to evaluate the identified model. The process is repeated until a reliable model is obtained. Finally, the selected model is used to yield out-of-sample forecasts.

$$\varnothing_p(B)\varphi_P(B^L)\nabla_L^D\nabla^d y_t = \delta + \theta_q(B)\vartheta_Q(B^L)a_t, \quad (1)$$

where B is the backshift operator; L is the number of seasons in a year ($L = 4$ for quarterly data and $L = 12$ for monthly data); δ is a constant term; a_t is a random shock; \varnothing_p is a non-seasonal autoregressive parameter; φ_P is a seasonal autoregressive parameter; θ_q is a non-seasonal moving average parameter, ϑ_Q is a seasonal moving average parameter.

3.3. Bivariate Transfer Function Model

The transfer function approach is a time series model that involves more than one time series and explains explicitly the dynamic characteristics of the process [44]. The general form of the function can be written in Equation (2).

$$Y_t = v(B)X_t + N_t, \quad (2)$$

where

$$v(B) = \frac{\omega(B)}{\delta(B)}B^b, \quad (3)$$

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_s B^s, \quad (4)$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r, \quad (5)$$

where $\omega(B)$ is the numerator polynomial, $\delta(B)$ is the denominator polynomial, and X_t is the explanatory variable, N_t is the independent noise term generated by an ARIMA process. However, if the time series exhibits non-stationarity, an appropriate degree of differencing should be applied to the series to achieve stationarity. In reality, stationary dependent and explanatory variables are necessary to develop a transfer function model. As a result, the transfer function model can be more generally expressed in Equation (6).

$$y_t = \frac{\omega(B)}{\delta(B)}B^b x_t + \frac{\theta(B)\vartheta(B)}{\varphi(B)\varnothing(B)}a_t \quad (6)$$

where

$$y_t = (1 - B)^d (1 - B^L)^D Y_t, \quad (7)$$

$$x_t = (1 - B)^{d'} (1 - B^L)^{D'} X_t, \quad (8)$$

where D and d are orders of regular and seasonal differencing that transform non-stationary Y_t to stationary y_t ; D' and d' are orders of regular and seasonal differencing that transform non-stationary

X_t to stationary x_t ; l is the period of seasonality; $\frac{\theta(B)\vartheta(B)}{\varphi(B)\varnothing(B)}$ is an ARIMA process that transfer noise term N_t to white noise a_t .

According to [43], a comprehensive procedure to develop a transfer function model is displayed. The cross-correlation function (CCF) was introduced to identify the model orders such as delay order b , numerator polynomial order s , denominator polynomial order r . However, the spurious effects usually appear due to the autocorrelation in the input series. Thus, a pre-whiten filter is introduced to alleviate this spurious effect. The main principle of the pre-whiten filter transfers the input variable to a white noise process. In other words, the autocorrelation in the input variable series was removed.

The first step in pre-whitening is to identify and fit an ARIMA model to the stationary input data x_t . An appropriate ARIMA model for the input data might be written in Equation (9).

$$\varphi_p(B)\varnothing_P(B^l)x_t = \theta_q(B)\vartheta_Q(B^l)a_t, \quad (9)$$

Rewriting Equation (6), the residuals a_t are given in Equation (10).

$$\alpha_t = \frac{\varphi_p(B)\varnothing_P(B^l)}{\theta_q(B)\vartheta_Q(B^l)}x_t \quad (10)$$

After pre-whitening the input data, the next step is to pre-whiten the stationary output data y_t . This is to filter the output data through the same ARIMA model with the same model parameters fitted to the input data. In other words, x_t is substituted by y_t in Equation (10) to obtain the filtering equation for the pre-whitened output, as shown in Equation (11).

$$\beta_t = \frac{\varphi_p(B)\varnothing_P(B^l)}{\theta_q(B)\vartheta_Q(B^l)}y_t \quad (11)$$

The next step is to compute the cross-correlation function between the pre-whitened input α_t and the pre-whiten output β_t . It should be emphasised here that all the cross-correlation for negative should be insignificant. Since the model assumed that the input variable has effects on the output variable, no feedback effects exist. The first significant spike lag indicates a time delay order b . Then the damping pattern of the cross-correlation function can be used to identify the denominator polynomial order r . If the CCF fades out in an exponential decay pattern, this suggests $r = 1$. If the CCF damp down in an oscillating pattern, this indicates $r = 2$. While the numerator polynomial order s is equal to the number of lags that exist between the first spike in the CCF and the beginning of the clear damping down pattern. For example, if the first spike starts at lag 2 and the CCF exhibits decay after lag 4, this suggests $s = 4 - 2 = 2$. At this point, it should be noted that the study does not attempt to search the right values of the model orders. The models are only tentatively identified, further diagnostics checking should be performed to obtain adequate models.

The transfer function model parameters are estimated based on the identified model orders at this step. After the model has been identified, the error term should be checked. If significant autocorrelation exists in the error series, an ARIMA model should identify and fit to the error series and then be incorporated into the overall model in Equation (2). The model can be identified by examining the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the error. In this study, SPSS 23 was used to develop the transfer function models. The parameter estimates and standard errors suggest whether the parameters should be omitted from the model.

For diagnostics checking, the cross-correlation between the residuals from the developed transfer function model and the residuals of the ARIMA model that used to pre-whiten the input variable. If there is no significant spike that exists in the CCF, indicating there is no significant cross-correlation left between the input and output variables, the transfer function model is correctly identified. If that is not the case, this indicates that some information still exists in the output variable that can be explained by the input variable. Moreover, the autocorrelation function (ACF) of the residuals of the transfer function model should also be checked. If the model is appropriately identified, there should be no significant autocorrelation in the residuals.

3.4. Multivariate Transfer Function Model

The transfer function model can be further expanded to include multivariate, explaining variables [25]. If there are two or more independent variables, the cross-correlation to identify a transfer function relating the dependent variable to each independent variable. The model development process follows the strategy described by [43].

So, in this study, the two time series (input X_{1t} and X_{2t}) to estimate the dependent variable of another time series (output Y_t) is applied. This is done by modelling a linear system, which takes the form in Equation (12).

$$y_t = \frac{\omega_1(B)}{\delta_1(B)} B^{b_1} x_{1t} + \frac{\omega_2(B)}{\delta_2(B)} B^{b_2} x_{2t} + a_t \quad (12)$$

where $\omega(B)$ and $\delta(B)$ are, respectively, the polynomials of the s and r orders; $\omega_1(B)$ and $\delta_1(B)$ can be identified by examining the cross-correlation between pre-whitened y_t and x_{1t} ; $\omega_2(B)$ and $\delta_2(B)$ can be identified by examining the cross-correlation between pre-whitened y_t and x_{2t} . The transfer function in this study used the house price and work volume for the primary series (x_{1t} and x_{2t}).

4. The Forecasting Models

4.1. ARIMA Model for Building Costs

In this part, some important features of the individual time series were described and used to specify and estimate univariate ARIMA models for the building costs of residential low-rise (LBC) and the building cost of high-rise buildings (HBC). The period 2001:Q1–2014:Q4 was used to specify univariate ARIMA models for the cost series.

Prior to fitting time-series data to an ARIMA model, the data must be stationary. As explained in the preceding section, the ACF and PACF plots are used to check for stationarity. The ACF of the cost series dying down very slowly indicates that the series is non-stationary, as shown in Figure 1. Hence, the series is transformed into a stationary one using differencing. The cost series features suggest that the LBC series requires a regular and a seasonal difference to become stationary. In addition, the HBC series should be stationary after one regular difference. The ACF and PACF plots of the stationary series are used to identify the orders of the ARIMA model. The model parameters for both cost series are shown in Table 1. The ARIMA model for building cost of low-rise residential building is ARIMA (0,1,1) (0,1,1)₄ with moving average parameter $\theta_1 = 0.317$ and seasonal moving average parameter $\vartheta_1 = 0.294$. The ARIMA model for building cost of the high-rise residential building is ARIMA (0,1,0) (1,0,0)₄. The seasonal autoregressive parameter $\phi_1 = 0.594$.

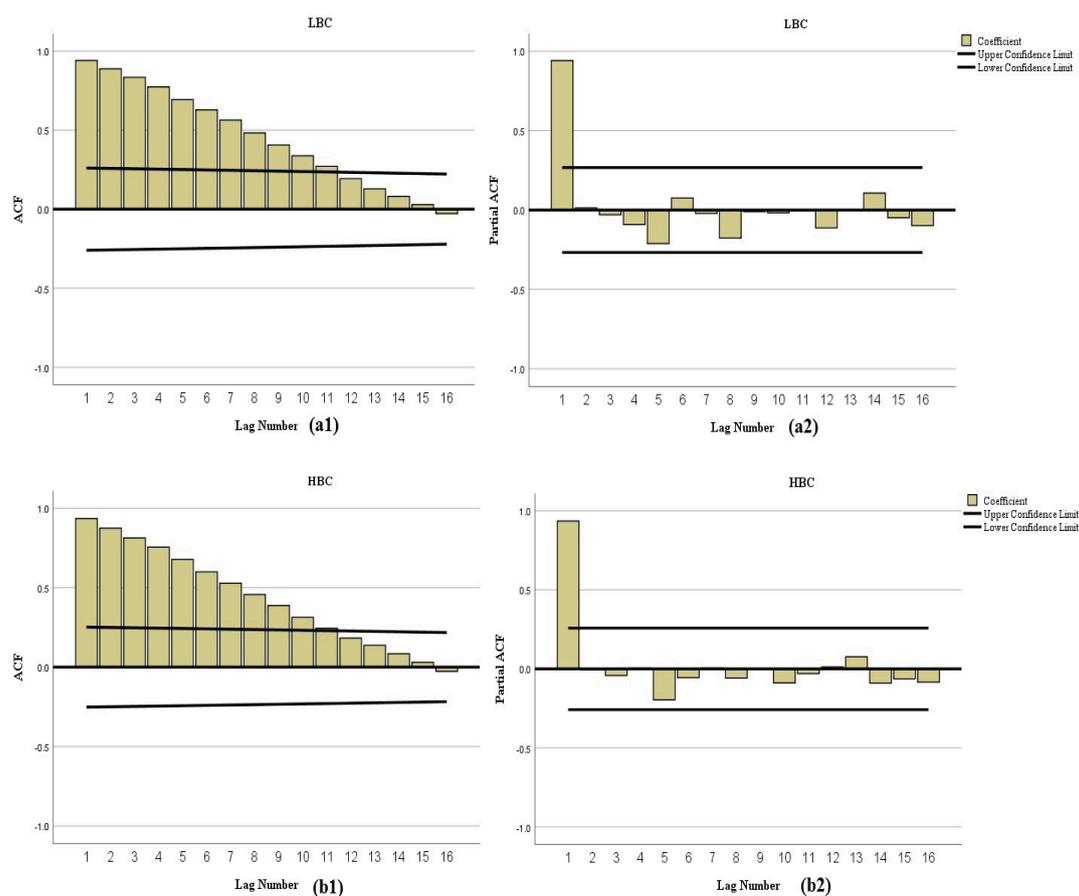


Figure 1. (a1) Autocorrelation function (ACF) plot of residential low-rise (LBC); (b1) partial auto-correlation function (PACF) plot for LBC; (a2) ACF plot of HBC; (b2) PACF plot of high-rise buildings (HBC).

Table 1. Autoregressive integrated moving average (ARIMA) models for LBC and HBC.

Cost Series	Model	Parameter	Estimate	SD	<i>t</i> -Statistics	<i>p</i> -Value
LBC	ARIMA (0,1,1) (0,1,1) ₄	θ_1	0.317	0.147	2.160	0.036 **
		ϑ_1	0.294	0.152	1.933	0.059 **
HBC	ARIMA (0,1,0) (1,0,0) ₄	ϕ_1	0.594	0.108	5.483	0.0 ***

** Significant at 0.05 level, *** Significant at 0.01 level.

The ARIMA approach only considers information related to the cost series own lagged values. However, some exogenous shocks that can marginally affect the residential building costs should also be considered, for example, house price and work volume. In order to take into account some of these effects, the transfer function models were specified.

4.2. Transfer Function Model for Building Costs

The study used the modelling process suggested by [43] to develop a transfer function model. After some initial analysis, the average house price and work volume were used as the independent variables, respectively and collaboratively. The iterative model development process, as described in the last section was adopted in the development of the transfer function models. The ARIMA models for the independent variables were used to filter the residential building costs and then the pre-whitened series are analysed by using the cross-correlation function (CCF). The lag parameter (b) can be directly identified based on CCF graph. The CCFs for the residential building costs and the input variables are shown in Figure 2. Following the Box-Jenkins transfer function modelling

strategy and the graph shown in Figure 2a1, the effect of house price on the building cost of low-rise residential building is significant at lag 3 ($b = 3$). Based on Figure 2b1, the number of building consents greatly impact the cost of a low-rise residential building at lag 4 ($b = 4$). Similarly, based on the results shown in Figure 2a2, the house price significantly affected the building cost of a high-rise residential building at lag 1 ($b = 1$) and lag 3 ($b = 3$). According to a transfer function modelling rule that only allows incorporation of the first significant lag effect into the model, the lag effect at lag 1 ($b = 1$) was modelled in this model. As shown in Figure 2b2, the number of building consents significantly impacted the cost of a high-rise residential building at lag 3 ($b = 3$). The final estimates of the bivariate and multivariate transfer function models of residential low-rise and high-rise buildings are provided in Tables 2 and 3, respectively. The bivariate transfer function model for LBC with HP involved as an independent variable is indicated as BTF-LBC1; while the bivariate transfer function model for LBC with BC involved as an independent variable is indicated as BTF-LBC2. In addition, the multivariate transfer function model for LBC with both HP and BC involved as independent variables is indicated as MTF-LBC. The bivariate transfer function model for HBC with HP involved as an independent variable is indicated as BTF-HBC1, while the bivariate transfer function model for HBC with BC involved as an independent variable is indicated as BTF-HBC2. The multivariate transfer function model for HBC with both HP and BC involved as independent variables is indicated as MTF-HBC.

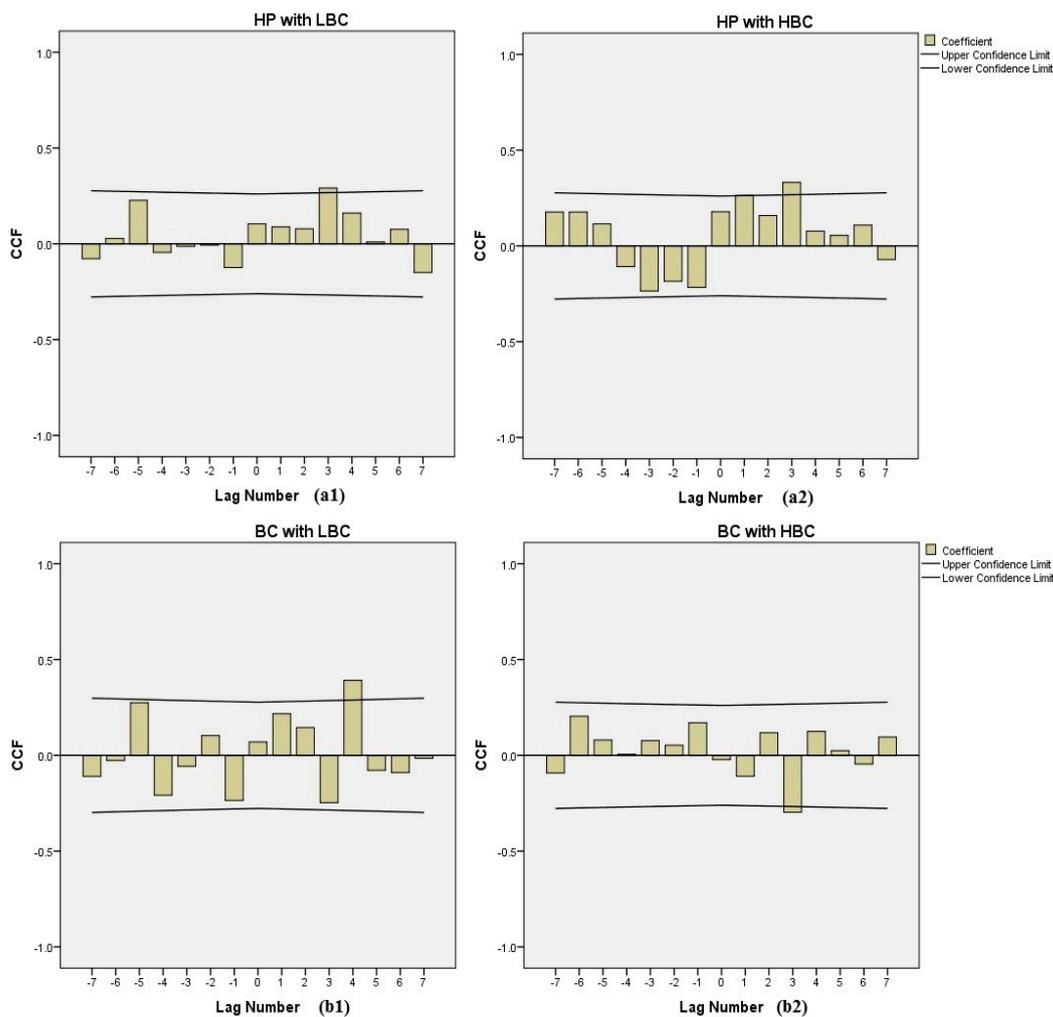


Figure 2. (a1) The cross-correlation function (CCF) of house prices (HP) and LBC; (a2) the CCF of HP and HBC; (b1) the CCF of building consent (BC) and LBC; (b2) the CCF of BC and HBC.

Table 2. Transfer function model parameters for LBC.

Model	Independent Variable	Parameter	Estimate	SD	t-Statistics	p-Value
BTF-LBC1	HP	ω_0 B	0.488 3	0.152	3.211	0.002 ***
BTF-LBC2	BC	ω_0 B	0.691 4	0.251	2.756	0.008 ***
MTF-LBC	HP & BC	ω_0^{HP} b^{HP}	0.312 3	0.141	2.210	0.032 **
		ω_0^{BC} b^{BC}	0.575 4	0.249	2.312	0.025 **

** Significant at 0.05 level; *** Significant at 0.01 level.

Table 3. Transfer function model parameters for HBC.

Model	Independent Variable	Parameter	Estimate	SD	t-Statistics	p-Value
BTF-HBC1	HP	ω_0	0.302	0.125	2.410	0.020 **
		δ_1 B	0.741 1	0.147	5.034	0.0 ***
BTF-HBC2	BC	ω_0 B	0.232 3	0.110	2.101	0.038 **
		ω_0^{HP} δ_1	0.452 0.706	0.102	4.431	0.001 ***
MTF-HBC	HP & BC	b^{HP}	1	0.302	2.338	0.027 **
		ω_0^{HP} b^{BC}	0.313 3	0.151	2.078	0.040 **

** Significant at 0.05 level; *** Significant at 0.01 level.

An examination of the plots of ACF and PACF of the residuals of the BTF-LBC1, BTF-LBC2, MTF-LBC, BTF-HBC1, BTF-HBC2, MTF-HBC models are shown in Figure 3. All the residuals fall inside the scope, which indicates that the residuals are white noise and thus the models are adequate. The performance of the models has been evaluated in terms of model evaluation parameters such as coefficients of determination (R^2), mean average percentage error (MAPE), and root mean square error (RMSE). The performance measures of the developed models for a cost series based on the data from 2001:Q1 to 2014:Q4 are shown in Table 4. An inspection of Table 4 suggests some important points. A significant improvement in the performance was observed with the inclusion of house price as an independent variable compared with the univariate ARIMA model for both building cost series. However, the inclusion of work volume series as an input does not provide a significant improvement in the performance of the model for the building cost of residential high-rise buildings (HBC). The inclusion of the work volume makes the transfer function model more complicated without much improvement in the forecasting performance, for the cost series of residential high-rise buildings (HBC).

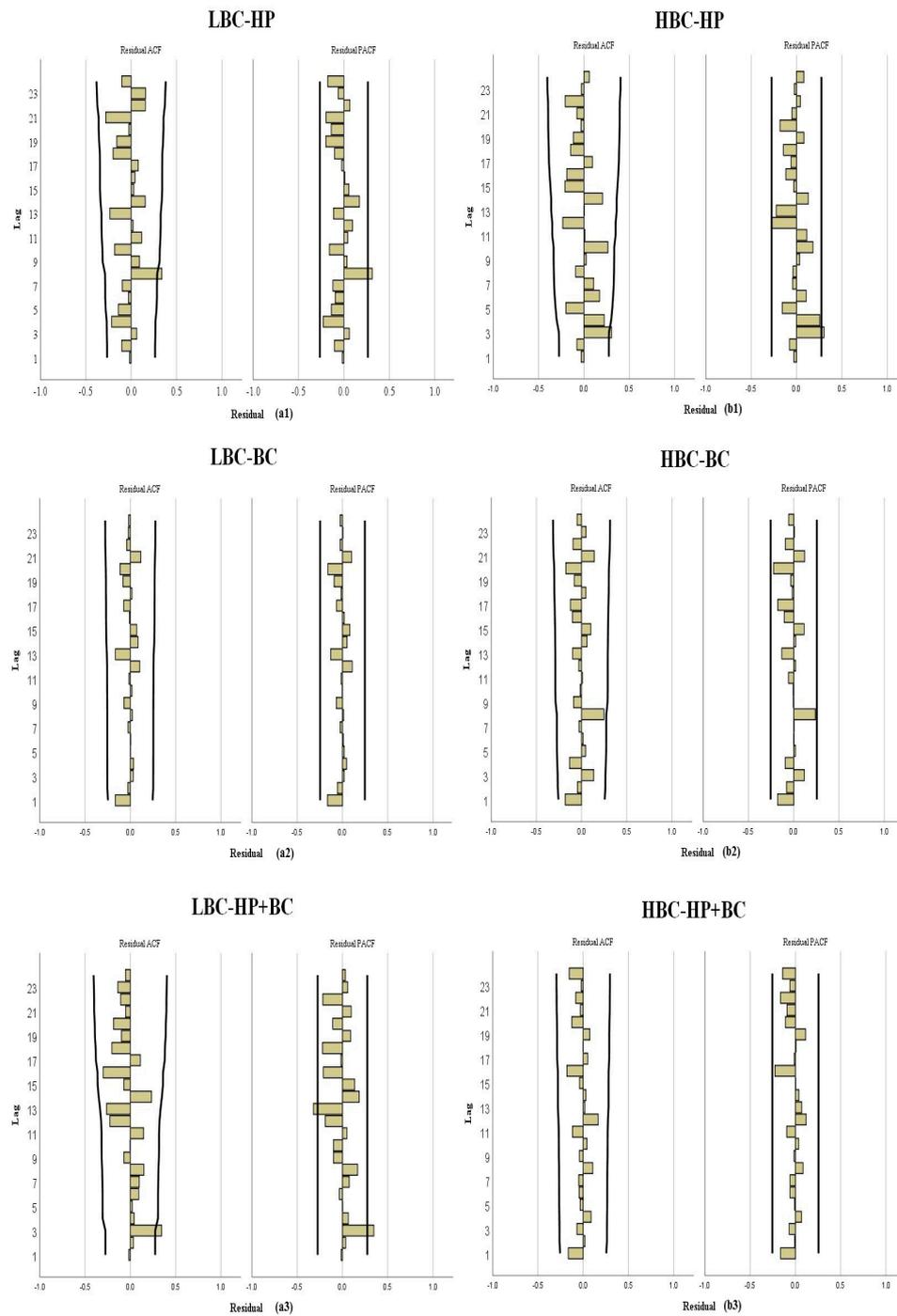


Figure 3. (a1) The ACF & PACF of residuals of bivariate transfer function (BTF)-LBC1; (a2) the ACF & PACF of residuals of BTF-LBC2; (a3) the ACF & PACF of residuals of multivariate transfer function (MTF)-LBC; (b1) the ACF & PACF of residuals of BTF-HBC1; (b2) the ACF & PACF of residuals of BTF-HBC2; (b3) the ACF & PACF of residuals of MTF-HBC.

Table 4. Model performance parameters for the fitted models.

Cost Series	Model Statistics	ARIMA	Bivariate TF		Multivariate TF
			HP	BC	HP & BC
LBC	R ²	0.955	0.935	0.934	0.942
	RMSE	41.96	36.58	34.77	33.58
	MAPE	1.846	1.814	1.776	1.755
	MAE	28.79	28.67	28.10	26.87
HBC	R ²	0.969	0.942	0.919	0.927
	RMSE	52.87	37.88	45.75	36.33
	MAPE	1.944	1.796	1.836	1.676
	MAE	37.03	34.32	35.48	32.92

5. Forecasting Results

To evaluate the forecasting performance of the transfer function model, forecasts were made for both cost series (LBC & HBC) ranging from 2015:Q1 to 2018:Q4. A summary of forecasting performance measures for the ARIMA models and transfer function models is shown in Table 5. For the building cost series of residential low-rise buildings (LBC), the lowest RMSE is given from the MTF model, for a 31% reduction over the ARIMA model. The bivariate transfer function models for LBC (BTF-LBC1 & BTF-LBC2) also perform better than the ARIMA model with an RMSE reduction of about 16% and 20%, respectively. For the LBC series, the reduction in RMSE from ARIMA to the bivariate transfer function models is not as strong as the reduction from ARIMA to the multivariate transfer function model. Again, from a dynamic and uncertain point of view, an appropriate forecasting model for the LBC series is the multivariate model and not the univariate model.

Table 5. Forecasting performance parameters for the models.

Cost Series	Model Statistics	ARIMA	Bivariate TF		Multivariate TF
			HP	BC	HP & BC
LBC	R ²	0.955	0.935	0.934	0.942
	RMSE	48.00	40.43	38.34	33.13
	MAPE	2.190	2.012	1.956	1.847
	MAE	40.00	36.49	29.88	28.01
HBC	R ²	0.969	0.942	0.919	0.927
	RMSE	56.23	42.55	47.36	41.88
	MAPE	2.159	1.861	1.942	1.795
	MAE	44.32	36.93	38.84	35.76

For the building cost series of residential high-rise buildings (HBC), the transfer function models yield RMSE improvements over the univariate ARIMA model. See the results in Table 4. The bivariate transfer function models (BTF-HBC1 & BTF-HBC2) gives a 24% and 16% RMSE reduction, respectively, and the multivariate transfer function model (MTF-HBC) gives a 26% RMSE reduction over the ARIMA model. In particular, the MTF-HBC model gives an additional RMSE reduction over the bivariate transfer function models (BTF-HBC1 & BTF-HBC2). The results shown in Table 5 indicate that the multivariate transfer function model (MTF-HBC) did not significantly reduce the out-of-sample forecasting errors relative to BTF-HBC1. The multivariate transfer function model produced lower mean square errors than the bivariate transfer function model BTF-HBC1, but this was not significant. The BTF-HBC1 model improves upon the ARIMA model. However, the BTF-HBC2 model was not improved as much as the BTF-HBC1 model. These results indicate that the building cost of a high-rise building may be more influenced by house prices than the number of building consents.

The results show that the transfer function models perform better over univariate ARIMA models for all lead times. The gain from including explaining variables in the transfer function model is obvious. For example, the MAPE and RMSE for the multivariate transfer function model (MTF-LBC)

are reduced significantly. The transfer function model including exogenous variables is expected to predict the residential building costs more accurately than the univariate ARIMA model. Owing to the inherent complexity in the construction industry and property market, the transfer function models with two independent variables is better at forecasting than the transfer function models with one independent variable.

6. Results Discussion

Both the ARIMA models and the transfer function models have been proven reliable in this study. However, transfer function models can obtain efficiency gains, especially the transfer function models with two independent variables. The results indicate that the simple models are not always the most appropriate for forecasting, especially when additional information is available. The univariate ARIMA models were primarily used as benchmarks in comparing forecasting performance. Efficiency gains can be realised when explanatory variables are available in the forms of other time series. The forecasting performance can be improved by using transfer function models. It was found that the use of the explanatory variables significantly reduces the forecasting error in building costs, relative to a univariate ARIMA model. In fact, some macro-economic indicators significantly impact building costs, but they do not have temporal effects on building cost. They usually have time-lag effects. In this study, the time-lag effects were identified by using the cross-correlation function (CCF) and transfer function model can model the time-lag effects of explanatory variables on building costs.

Furthermore, changes in the housing prices lead to changes in residential building costs of New Zealand. This result was supported by the findings of [45], which addressed that the construction activities are more sensitive to the house price. In addition, the finding of the significant correlation between residential building costs and house prices in New Zealand is consistent with the results of [46], which indicated that the upswing in residential building costs is related to the strong rise in house price. The causal relationship is also supported by the connection between the property market and construction industry and also the boom and bust cycles in the construction industry. The house price can influence the demand for housing. Increasing house prices offer the possibility of obtaining capital gains, which contributes to market expansion. This exerts a significant pulling effect on construction activity. The demand for housing is impacted by the macro-economic environment. The supply of housing is determined by the capacity of the construction industry. The gaps between supply and demand are absorbed by prices.

Moreover, an increase in the number of residential building consents indicates that more residential building activity is occurring and increased demand for resources. Under this condition, the purchase of materials and hiring of labour and equipment may cost more as compared to purchase and hiring during normal market conditions. Hence, the fluctuation in costs of resources (labour, materials, and equipment) may have a significant effect on building cost.

Residential high-rise buildings in New Zealand are usually delivered by large construction organisations because they seek to expand into areas where scale economies are present, and their advantages are the greatest. The change in work volume cannot significantly influence the large construction companies since they usually can effectively and strategically manage their supply chain with long-term relationships and partnering and collaborative procurement system. Furthermore, large construction organisations benefit from technology innovations that help them to better meet client needs and achieve competitive advantages by reducing costs and enhancing differentiation. However, small construction firms are subject to constraints on their competitiveness since significant technological lags in their deliver process and shortcomings in their management procedures. The existing literature indicated that a significant difference exists in the innovation capability and output of small construction firms compared to large construction organisations. In fact, small construction companies are more exposed to the change in market conditions than large construction companies and face fierce competition.

7. Conclusions

This study has developed models for estimating building costs of low-rise and high-rise residential buildings in New Zealand. In the quest to generate accurate cost forecasts, modelling techniques is one of the key challenges that need to overcome. In this study, explanatory variables were included in transfer function models. The house price was first involved into the model. The results indicate that the house price has a significant effect on both cost series (LBC & HBC). Additionally, the work volume was also included in the models. The results indicated that the transfer function model with two independent variables can generate better forecasts for both cost series. By including work volume, the performance of the cost model for low-rise buildings improved more than that for high-rise buildings, which indicates that work volume has more effects on the building cost of low-rise residential buildings than that on the building cost of high-rise residential buildings. When the input variables have explaining power for the dependent variable, appropriate application of the transfer function model can greatly improve forecasting accuracy.

The findings of this study contribute to the body of knowledge by exploring significant time-lag effects of house prices and work volume on building costs of low-rise and high-rise residential buildings in New Zealand. Moreover, the developed transfer function models provide an effective modelling technique for residential building costs in New Zealand. The forecasting models can be considered as contingency plans for cost estimates and management. Therefore, clients and stakeholders can use the information provided by the models to evaluate financial decisions regarding a proposed construction project. Additionally, the forecasting models can provide accurate estimates for the building costs of low-rise and high-rise residential buildings, which can help industry professionals to better understand future trends in the construction market, and thus improve their strategic planning.

The proposed models were developed for building costs of residential buildings in New Zealand, and hence its results and findings may not directly apply to other cost series or residential building costs of other countries. Future studies should focus on more building categories (commercial and industrial building costs) and employ other modelling techniques in order to obtain accurate cost forecasts. The current study can also be extended to investigate the relationships between building costs and other variables. More hidden links between the building costs and environment are waiting for further investigation.

Author Contributions: Conceptualization, L.Z., J.M.; Methodology, L.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 the China Scholarship Council, grant number 201206130069, Massey University, grant number 09166424, and Beijing University of Technology, grant number is 004000514119067. The APC was funded by these 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 the 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 contributed to this project.

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

Abbreviations

ACF	Auto-Correlation Function
ANNs	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BC	Building Consents
BTF-LBC1	Bivariate Transfer Function Model for Low-rise Residential Building Cost involved House Prices as explanatory variable
BTF-LBC2	Bivariate Transfer Function Model for Low-rise Residential Building Cost involved Building Consents as explanatory variable
BTF-HBC1	Bivariate Transfer Function Model for High-rise Residential Building Cost involved House Prices as explanatory variable
BTF-HBC2	Bivariate Transfer Function Model for High-rise Residential Building Cost involved Building Consents as explanatory variable
CBR	Case-Based Reasoning
CCF	Cross-Correlation Function
CPI	Consumer Price Index
GDP	Gross Domestic Product
HBC	High-rise Residential Building Cost
HP	House Prices
LBC	Low-rise Residential Building Cost
MAPE	Mean Absolute Percentage Error
MRA	Multiple Regression Analysis
MTF-LBC	Multivariate Transfer Function Model for Low-rise Residential Building Cost involved House Prices and Building Consents as explanatory variables
MTF-HBC	Multivariate Transfer Function Model for High-rise Residential Building Cost involved House Prices and Building Consents as explanatory variables
PACF	Partial Auto-Correlation Function
PCA	Principal Component Analysis
PPI	Producer Price Index
RBNZ	Reserve Bank of New Zealand
RMSE	Root Mean Square Error
SVM	Support Vector Machine
VEC	Vector Error Correction

References

1. Torp, O.; Klakegg, O.J. Challenges in cost estimation under uncertainty: A case study of the decommissioning of barsebäck nuclear power plant. *Adm. Sci.* **2016**, *6*, 14. [[CrossRef](#)]
2. Williams, T.; Lakshminarayanan, S.; Sackrowitz, H. Analysing bidding statistics to predict completed project cost. In Proceedings of the International Conference on Computing in Civil Engineering, Cancun, Mexico, 1 January 2005.
3. Flyvbjerg, B.; Bruzelius, N.; Rothengatter, W. *Megaprojects and Risks: An Anatomy of Ambition*; Cambridge University Press: Cambridge, UK, 2003.
4. Li, H.; Shen, Q.; Love, P.E. Cost modelling of office buildings in Hong Kong: An exploratory study. *Facilities* **2005**, *23*, 438–452. [[CrossRef](#)]
5. Goh, B.; Teo, H.P. Forecasting construction industry demand, price and productivity in Singapore: The Box-Jenkins approach. *Constr. Manag. Econ.* **2000**, *18*, 607–618.
6. Dang, T.H.G.; Low, S.P. Role of construction in economic development: Review of key concepts in the past 40 years. *Habitat Int.* **2011**, *35*, 118–125.
7. Lewis, T.M. The construction industry in the economy of Trinidad and Tobago. *Constr. Manag. Econ.* **2004**, *22*, 541–549. [[CrossRef](#)]
8. Su, C.K.; Lin, C.Y.; Wang, M.T. Taiwanese construction sector in a growing ‘maturity’ economy, 1964–1999. *Constr. Manag. Econ.* **2003**, *21*, 719–728. [[CrossRef](#)]

9. MBIE. *NZ Sector Report 2013-Construction*; Ministry of Building Innovation and Employment: Wellington, New Zealand, 2013.
10. Rafiei, M.H.; Adeli, H. Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes. *J. Constr. Eng. Manag.* **2018**, *144*. [[CrossRef](#)]
11. Ashuri, B.; Shahandashti, S.M.; Lu, J. Empirical tests for identifying leading indicators of ENR Construction Cost Index. *J. Constr. Manag. Econ.* **2012**, *30*, 917–927. [[CrossRef](#)]
12. Yorucu, V.; Keles, R. The construction boom and environmental protection in Northern Cyprus as a consequence of the Annan Plan. *Constr. Manag. Econ.* **2006**, *25*, 77–86. [[CrossRef](#)]
13. Ball, M.; Farshchi, M.; Grilli, M. Competition and the persistence of profits in the UK construction industry. *Constr. Manag. Econ.* **2000**, *18*, 733–745. [[CrossRef](#)]
14. Martin, S. *Advanced Industrial Economics*; Blackwell: Oxford, UK, 1993.
15. Onur, I.; Ozcan, R.; Tas, B.K.O. Public procurement auctions and competition in Turkey. *Rev. Ind. Organ.* **2012**, *40*, 207–223. [[CrossRef](#)]
16. Assaf, S.A.; Al-Hejji, S. Causes of delay in large construction projects. *Int. J. Proj. Manag.* **2006**, *24*, 349–357. [[CrossRef](#)]
17. Boermans, T.; Bettgenhauser, K.; Hermelink, A.; Schimschar, S. *Cost Optimal Building Performance Requirements—Calculation Methodology for Reporting on National Energy Performance Requirements on the Basis of Cost Ptimality within the Framework of the EPBD*; European Council for Energy Efficient Economy: Stockholm, Sweden, 2011.
18. Figueiredo, E.; Figueiras, J.; Park, G.; Farrar, C.R.; Worden, K. Influence of the autoregressive model order on damage detection. *Comput. Aided Civ. Infrastruct. Eng.* **2011**, *26*, 225–238. [[CrossRef](#)]
19. Moon, T.; Shin, D.H. Forecasting Construction Cost Index Using Interrupted Time-Series. *KSCE J. Civ. Eng.* **2018**, *22*, 1626–1633. [[CrossRef](#)]
20. Zhang, H.B. Research on Structural Cost Analysis at Schematic Design Stage using Multi-variant Regression Analysis. *Eng. Cost Manag.* **2015**, *2*, 49–52.
21. ElMousalami, H.H.; Elyamany, A.H.; Ibrahim, A.H. Predicting Conceptual Cost for Field Canal Improvement Projects. *J. Constr. Eng. Manag.* **2018**, *144*, 1–8. [[CrossRef](#)]
22. Kim, G.H.; Shin, J.M.; Kim, S.; Shin, Y. Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine. *J. Build. Constr. Plan. Res.* **2013**, *1*, 1–7. [[CrossRef](#)]
23. Cheng, M.Y.; Hoang, N.D.; Wu, Y.W. Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: A Taiwan case study. *Autom. Constr.* **2013**, *35*, 306–313. [[CrossRef](#)]
24. Le’sniak, A.; Juszczuk, M. Prediction of site overhead costs with the use of artificial neural network based model. *Arch. Civ. Mech. Eng.* **2018**, *18*, 973–982. [[CrossRef](#)]
25. Juszczuk, M. Application of PCA-based data compression in the ANN-supported conceptual cost estimation of residential buildings. *AIP Conf. Proc.* **2016**, *1738*. [[CrossRef](#)]
26. Juszczuk, M.; Le’sniak, A.; Zima, K. ANN Based Approach for Estimation of Construction Costs of Sports Fields. *Complexity* **2018**, *2018*, 11. [[CrossRef](#)]
27. Ashuri, B.; Lu, J. Time series analysis of ENR construction cost index. *J. Constr. Eng. Manag.* **2010**, *136*, 1227–1237. [[CrossRef](#)]
28. Wong, J.M.W.; Ng, S.T. Forecasting construction tender price index in Hong Kong using vector error correction model. *Constr. Manag. Econ.* **2010**, *28*, 1255–1268. [[CrossRef](#)]
29. Marwala, T. *Economic Modeling Using Artificial Intelligence Methods*; Springer: London, UK, 2013.
30. Chou, J.S.; Tseng, H.C. Establishing expert system for prediction based on the project-oriented data warehouse. *Expert Syst. Appl.* **2011**, *38*, 640–651. [[CrossRef](#)]
31. Kim, S.Y.; Kim, G.H.; Kang, K.I. Comparing cost prediction methods for apartment housing projects: CBR versus ANN. *J. Asian Archit. Build. Eng.* **2005**, *4*, 113–120. [[CrossRef](#)]
32. Alex, D.P.; Ai-Hussein, M.; Bouferguene, A.; Fernando, S. Artificial neural network model for cost estimation: City of Edmonton’s water and sewer installation services. *J. Constr. Eng. Manag.* **2010**, *136*, 745–756. [[CrossRef](#)]
33. Cheng, M.-Y.; Tsai, H.C.; Sudjono, E. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Syst. Appl.* **2010**, *37*, 4224–4231. [[CrossRef](#)]

34. Duran, O.; Rodriguez, N.; Consalter, L.A. Neural networks for cost estimation of shell and tube heat exchangers. *Expert Syst. Appl.* **2009**, *36*, 7435–7440. [[CrossRef](#)]
35. Juszczyk, M.; Leśniak, A. Modelling construction site cost index based on neural network ensembles. *Symmetry* **2019**, *11*, 411. [[CrossRef](#)]
36. Yazdani-Chamzini, A.; Zavadskas, E.K.; Antucheviciene, J.; Bausys, R. A model for shovel capital cost estimation, using a hybrid model of multivariate regression and neural networks. *Symmetry* **2017**, *9*, 298. [[CrossRef](#)]
37. An, S.H.; Kim, G.; Kang, K. A Case-based reasoning cost estimating model using experience by analytic hierarchy process. *Build. Env.* **2007**, *42*, 2573–2579. [[CrossRef](#)]
38. Leśniak, A.; Zima, K. Cost calculation of construction projects including sustainability factors using the case based reasoning (cbr) method. *Sustainability* **2018**, *10*, 14.
39. Ji, S.H.; Park, M.; Lee, H.S. Case adaptation method of case-based reasoning for construction cost estimation in Korea. *J. Constr. Eng. Manag.* **2012**, *138*, 43–52. [[CrossRef](#)]
40. Koo, C.W.; Hong, T.H.; Hyun, C.T.; Koo, K.J. A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects. *Can. J. Civ. Eng.* **2010**, *37*, 739–752. [[CrossRef](#)]
41. Ji, C.Y.; Hong, T.H.; Hyun, C.T. CBR revision model for improving cost prediction accuracy in multifamily housing projects. *J. Manag. Eng.* **2010**, *26*, 229–236. [[CrossRef](#)]
42. Dysert, L. Is estimate accuracy an oxymoron? *Cost Eng.* **2007**, *49*, 32–36.
43. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 1976.
44. Vandaele, W. *Applied Time Series and Box-Jenkins Models*; Academic Press: New York, NY, USA, 1993.
45. Sunde, T.; Muzindutsi, P.F. Determinants of house prices and new construction activity: An empirical investigation of the Namibian housing market. *J. Dev. Area* **2017**, *51*, 390–407.
46. Jacobsen, D.H.; Johansen, K.S.; Haugland, K. Housing investment and house prices. *Econ. Bull.* **2007**, *78*, 33–46.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).