

FORECASTING CONSTRUCTION DEMAND: A COMPARISON OF BOX-JENKINS AND SUPPORT VECTOR MACHINE MODEL

[Lam, Ka Chi](#); [Oshodi, Olalekan Shamsideen](#); [Lee, Eric Wai Ming](#)

Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR, China

Abstract

The construction industry provides infrastructure which are needed to drive the process of economic development. Despite its importance, the demand for construction industry's product tends to fluctuate with changes in the economic climate. These unforeseen events have negative impact on the productive capabilities of the construction sector. In order to formulate strategies and policies to minimize the impact of such fluctuations, there is a need to develop predictive models that can reliably and accurately predict such unforeseen events. In the study reported here, two univariate modelling (support vector machine and Box Jenkins) techniques were used to predict demand in the Hong Kong Construction market. The results of predictive accuracy test suggest that the support vector machine (SVM) and the Box Jenkins are both satisfactory for forecasting construction demand. However, the SVM model achieves higher prediction accuracy than the Box Jenkins model. The findings validate the reliability of using artificial intelligence models in predicting construction demand. The findings and robust modelling techniques are valuable to both developed and developing countries when estimating future demand patterns of construction demand.

Keywords: Accuracy, Box–Jenkins model, Construction demand, Forecasting, Support vector machine

1 Introduction

Almost every publication in the field of construction economics have shown that the construction sector plays an important role in the economic development process of any country (see Low and Leong, 1992; Han and Ofori, 2001). The construction sector provides constructed space for economic activities, employment and utilizes goods and services of other sector of the economy during the production process. However, empirical evidence has shown that changes in the macro-economic environment causes fluctuations in the output of the construction industry (Goh 2005; Jiang et al. 2013). In addition, findings from Wong and Ng (2010) shows that Gross Domestic Product, Construction output and tender price are closely correlated. Therefore, the cyclic behaviour of tender price index linked to changes in the economy of Hong Kong (due to Asian Financial Crisis of 1997, Severe Acute Respiratory Syndrome of 2003, etc.) would result in the fluctuations in construction industry's output (hereafter referred to as construction output). Thus, in order to sustain the growth of the economy, constant monitoring of the construction industry is vital.

Cyclic construction output is a problem in both developed and developing countries. This cyclic behaviour creates fluctuations in construction output which has an adverse effect on the construction industry. Fluctuations in construction output leads to employee turnover, loss of

knowledge and experience gained in projects, bankruptcy of construction organisations and increased competition among contractors (Ofori et al. 1996; Lam and Oshodi 2015; Ren and Lin 1996; Soo ad Oo 2014). Also, unplanned expansion of the construction sector due to increase in government investments often fail to achieve intended outcomes. This is evident in the studies reported on Nigerian (Awotona, 1990) and Trinidad and Tobago (Lewis, 1984). Therefore, it has become increasingly important to implement forward planning policies for the construction sector so as to sustain economic growth and avoid economic waste.

Accurate and reliable forecasting of future construction output is crucial to developing strategic long term plans for the construction sector. Goh (1998) asserts that insight into future volumes of construction output will assist stakeholders (such as property developers) to optimize profits. Similarly, governments can use forecast information to develop intervention policies meant to minimize the impact of changes in construction output. Australian government's stimulus plan after the global financial crisis is a good illustration of such government interventions (Jiang et al. 2013). It is evident that reliable forecast of future levels of construction output is crucial for all stakeholders. The aim of the study reported in this paper is to apply two univariate modelling (i.e. Box Jenkins and Support Vector Machine) techniques to forecast future volumes of construction output using Hong Kong as a representative case.

2 Literature Review

The importance of an accurate and reliable construction output forecast has been established in the preceding section. Despite its importance, construction output forecasting research has been limited when compared to studies such as construction cost forecasting. Limited number of studies can be linked to lack of statistical data on construction output, problems in the data collection process and lack of statistical data on other variables that influence construction output (K'Akumu, 2007; Wong and Ng, 2010). To address this problem, studies (such as Gruneberg and Folwell, 2013) have explored the possibility of using construction component of gross fixed capital formation (GFCF) as a proxy of construction output. It was found that construction output, GFCF and the construction component of GFCF are closely related. Though, there might still be a need for further research to validate this finding. The results of Gruneberg and Folwell's study is important to researchers in most developing countries, where data on construction output is largely unavailable. This is because statistical data on GDP and GFCF are reported in most cases to meet the requirements of donor and global funding agencies.

In construction output forecasting literature, there is an evident preference for quantitative modelling techniques. This can be attributed to accuracy and the ability to reproduce forecast generated from this technique when compared with qualitative approaches. Also, construction output forecasting is largely a time series problem (i.e. the relationship between variables in the past are used for future prediction of the dependent variable). Although multivariate models have shown good predictive capability when compared with univariate models (Goh 1996), this finding does not always hold. For instance, Fan et al. (2010) demonstrated that Box-Jenkins (univariate) model outperforms multiple regression (multivariate) model in predicting construction output of Hong Kong. Poor choice of selection of independent variables and the presence of auto-correlated errors have been attributed to inaccurate forecast from multivariate methods (Akintoye and Skitmore, 1994; Killingsworth, 1990). In addition, the application of multivariate modelling techniques is subject to availability of data on explanatory variables. Thus, it is imperative to identify univariate techniques that produce reliable and accurate forecast.

A wide variety of time series modelling techniques have been applied to construction output forecasting problems. Univariate modelling techniques can be grouped into two broad

categories: econometrics (statistical) and artificial modelling techniques. Econometric techniques such as Box-Jenkins was applied to construction demand forecasting for Singapore (Goh and Teo, 2000), Hong Kong (Fan et al., 2010) and the United Kingdom (Notman et al., 1998). Similarly, artificial intelligence modelling techniques have also been applied to construction demand forecasting problems in Singapore (Goh, 2000). Evidences shows that artificial intelligence models tend to produce more reliable and accurate forecast when compared with econometric models (Goh, 1996; Goh, 1998). This can be attributed to the capability of artificial intelligence models (e.g. artificial neural network) to capture nonlinear characteristics of construction demand data. There is little published data on application of support vector machine to construction output forecasting (see Fan et al., 2007). Therefore, this study sets out to apply SVM to predict construction output in the short and medium term. The result of this forecast is compared with those of Box-Jenkins approach which is considered a benchmark as suggested in Goh and Teo (2000).

3 Model Development

3.1 Construction Output

Construction output which measures the volume of construction works by executed main contractors within a defined time frame is collected from the Census and Statistics department of Hong Kong (CSD-HK). In Hong Kong, the construction output time series data is available from CSD-HK between 1983 and 2014 is presented in Figure 1. The cyclic behaviour (i.e. fluctuations) of construction output series appears to be related with changes in the economy and market conditions which shows the effect of the Asian financial crisis of 1997, SARS outbreak of 2003 and global financial crisis experience towards the end of 2008.

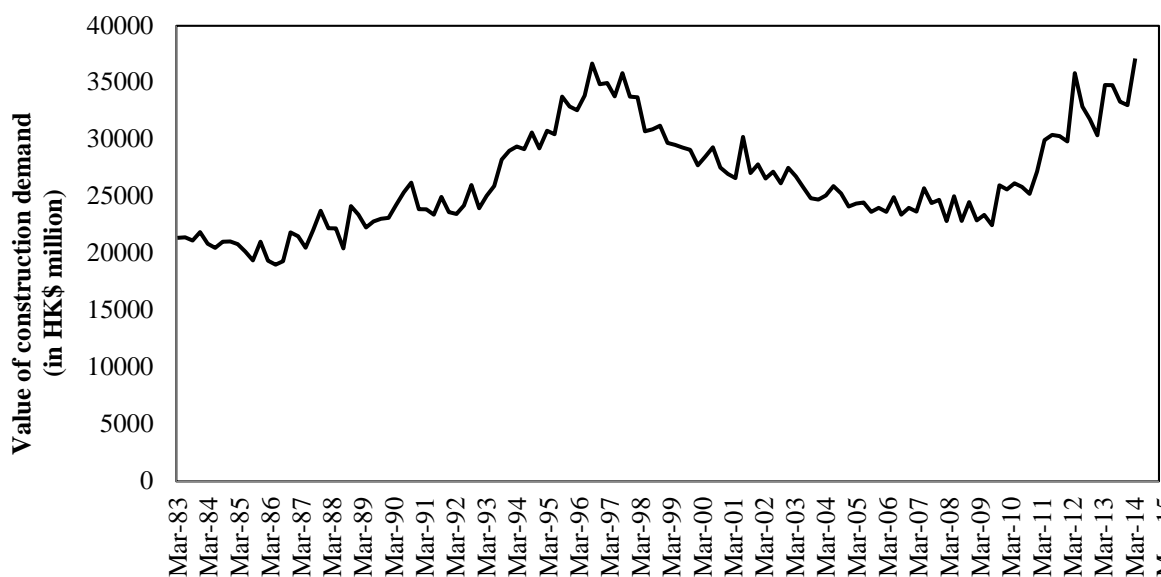


Figure 1. Value of construction demand (data from various years by Census and Statistics Department of Hong Kong)

3.2 Box Jenkins Modelling

Box Jenkins modelling technique is a combination of autoregressive (AR) and moving average (MA) model with differencing which was suggested by Box and Jenkins in 1976. This method is also often referred to as Autoregressive Integrated Moving Average (ARIMA) modelling approach. The process of applying Box Jenkins approach to time series forecasting is an iterative process which involves three major steps: identification, estimation and diagnostic checking, and application. The process is depicted in Appendix 1. The Box Jenkins model is

implemented using the ‘Arima’ code which is part of the forecast package in R programming (Hyndman et al., 2015). Also, the augmented Dickey-Fuller (ADF) and Box-Ljung (referred to as portmanteau) test used in this study are found in ‘urca’ and ‘stats’ package in R, respectively (Pfaff, 2013; R Core Team, 2015). For a detailed explanation on the procedure of fitting time series data to a Box Jenkins model (see Hyndman and Athanasopoulos, 2013).

3.3 Support Vector Machine (SVM)

SVM is an artificial intelligence modelling techniques that has a capability to capture nonlinear behaviour. The SVM algorithm was built based on theoretical foundations found in statistical learning (Vapnik, 1995). Subsequently, this method was later adopted in machine learning and statistics. In addition, SVM has been extensively applied to solving classification and regression problems (see Bin et al. 2006; Lam et al. 2009). This clearly shows that SVM model could be used as a tool for predicting construction output. This is based on the results that emanates from Goh’s (1998) study. The detailed explanation and proofs of SVM can be found in Vapnik (1995) and Vapnik (1998).

The SVM creates a binary classifier, called hyperplane, which maps the input vectors into a high-dimension feature space. Subsequently, the regression problem is solved in the new space. The regression SVM can be represented in the following mathematical form:

$$y = f(x) + b \quad (1)$$

The main task is to identify a functional form (f) which can correctly predict new cases that has not been used to train the model (i.e. test set). The function is estimated by using the Sequential Minimum Optimization (SMO) of an error function (Vapnik, 1995). With the introduction of slack variables (equation 2), the coefficients can be estimated by minimizing the error function of the SVM:

$$\frac{1}{2} w^T w + C \sum_{i=1}^N (\xi_i) + C \sum_{i=1}^N (\xi_i^*) \quad (2)$$

Subject to:

$$\begin{aligned} w^T \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - w^T \phi(x_i) - b &\leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0, i=1, \dots, N \end{aligned}$$

By introducing Lagrangian multipliers which are solvable under Karush-Kuhn-Tucker conditions, the solution of constrained optimization problem is determined. Once the Lagrange multipliers are found, the functional form of the regression SVM model can be expressed as:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (3)$$

where, $K(\cdot)$ is the kernel function.

Although other types of kernel functions (e.g. polynomial) exist, the Radial Basis Function (RBF) is used in this study. This can be attributed to its performance and frequency of use in similar previous studies (Bin et al., 2006). In this study, the SVM model was implemented using the SVM with SMO algorithm (named "SMOreg") provided in WEKA (Waikato Environment for Knowledge Analysis) software (Hall et al., 2009). For a detailed explanation on implementing SVM in WEKA, we refer readers to Witten et al. (2011). The process of fitting a SVM model to time series data in this study entailed: dividing collected data into two sets (training and test), import data into the WEKA software, initialize the parameters of the

SVM model, and adjusting of the individual parameters of the SVM algorithm until the adequate parameters are identified.

4 Model Implementation

4.1 Box Jenkins model

Stationarity is critical to applying the Box Jenkins model to time series data, ACF and PACF plots presented in Figure 2 is used to check if the data is stationary. Figure 2 shows that the data is not stationary. The ACF plots gradually die down. Therefore, the first difference is computed and the ACF of the first differenced series is presented in Figure 3. Non-presence of the dying down pattern in the ACF plot suggests that the first differenced series is stationary. ADF test is applied to validate this hypothesis. The ADF shows that the first difference series can be fitted to Box Jenkins model. The parameters of the best fit Box Jenkins model is presented in Table 3-6, the best fit model is selected based on the lowest AICc values as suggested in Hyndman and Athanasopoulos (2013). Subsequently, the residuals of the tentative model are checked for serial correlation. The ACF (Figure 4) and portmanteau test confirms that the residuals are white noise. Absence of serial correlation in the residuals of the final Box Jenkins model indicates the model passed validation test. Thus, the fitted Box Jenkins model is considered adequate. The results of the fitted Box Jenkins model is presented in Table 1. The final form of models for construction outputs is:

Construction output:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \phi_5 y_{t-5} \quad (4)$$

where y_t is $y_t - y_{t-1}$ (construction output variable first-difference); ϕ is the AR coefficient; and e is the random error term (lagged errors).

The out-of-sample forecast generated with the final Box Jenkins model (2013 Q1 -2014Q4) are given in Table 2.

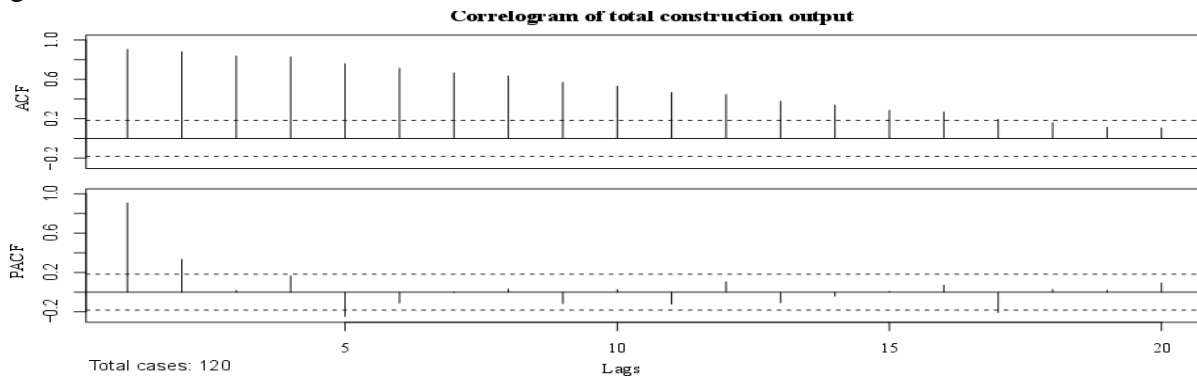


Figure 2. ACF and PACF of 'construction output'

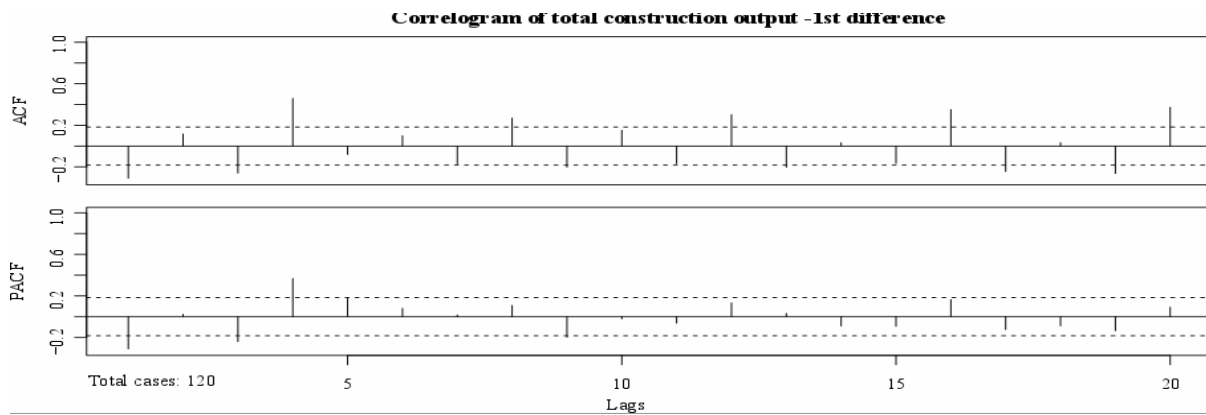


Figure 3. ACF and PACF of first differences of ‘construction output’

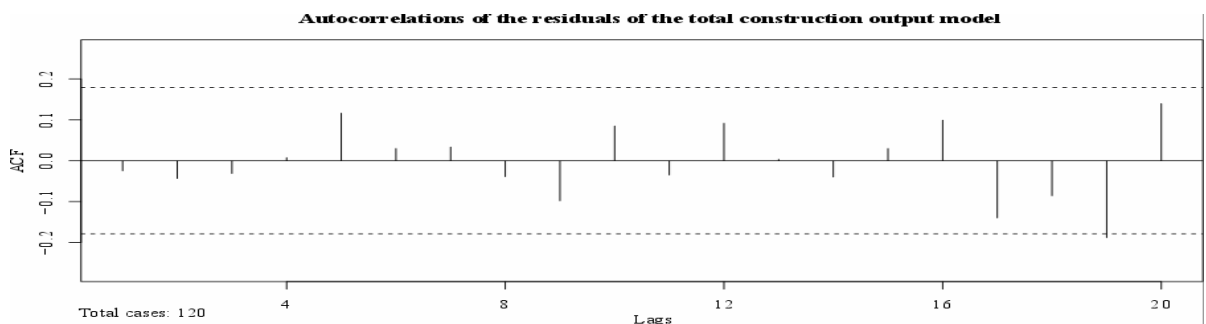


Figure 4. Autocorrelation of residuals of the fitted Box Jenkins (5,1,0) model

Table 1. Estimates of the parameter of the ARIMA (5,1,0) model for total construction output

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)
Coefficient	-0.2953	0.0003	-0.1205	0.4712	0.2324
Std. error	0.0953	0.0899	0.0889	0.0900	0.0957

4.2 SVM

The parameters of the SVM algorithm used in training the model affects its performance. Although, grid search of the parameter space could give better results, the parameters were manually tuned. This is similar to the approach used in Bell et al. (2012). The optimal parameters used in the SVM model are $(C, \gamma, \text{learning algorithm}) = (32, 0.01, \text{RegSMOImproved})$. The final SVM model is fitted to the data series to produce out-of-sample forecast (2013 Q1 -2014Q4) which served as a basis for evaluating the accuracy of forecast (Table 3).

4.3 Predictive Accuracy of Forecasting Models

Out-of-sample test is performed to evaluate the predictive reliability of the developed models SVM and conventional forecasting model (i.e. Box Jenkins), low variance between actual and predicted construction output (in the test data set) signifies better forecasting performance is achieved. In construction output forecasting studies, results of MAPE test that are lower than 10% are considered acceptable. In addition, Theil’s inequality U coefficient value closeness to zero signifies better prediction results is achieved (see Goh and Teo, 2000; Jiang and Liu, 2011 for more detailed explanation and equations).

The actual values and out-of-sample forecast value generated from the Box Jenkins and the SVM model are presented in Table 2 and 3, respectively. Three relative measures of accuracy (PE, MAPE and U coefficient) were used to evaluate the predictive accuracy of the developed

construction output models (Table 4). For both models, the values of MAPE test are less than 10% absolute error and the coefficients U are all close to 0. This indicates that the forecast generated by the models can be considered as satisfactory. In addition, the forecast generated by the SVM model achieved lower MAPE and U values show that the SVM model outperforms the conventional Box-Jenkins approach. Furthermore, the results of the evaluation of predictive accuracy test suggest that the SVM gives a more reliable and accurate forecast of construction demand.

Table 2. Out-of-sample forecast generated by the Box Jenkins model

Period	Actual	Forecast	Error
2013, Q1	32900	34935.04	-2035.04
2013, Q2	31788	35318.22	-3530.22
2013, Q3	30384	34230.43	-3846.43
2013, Q4	34796	37378.05	-2582.05
2014, Q1	34785	37371.05	-2586.05
2014, Q2	33337	37475.28	-4138.28
2014, Q3	33031	36641.74	-3610.74
2014, Q4	37132	38119.21	-987.21

Table 3. Out-of-sample forecast generated by the SVM model

Period	Actual	Forecast	Error
2013, Q1	32900	33910.8	-1010.8041
2013, Q2	31788	34178.24	-2390.2443
2013, Q3	30384	33895.57	-3511.5739
2013, Q4	34796	34831.98	-35.9831
2014, Q1	34785	33835.98	949.0157
2014, Q2	33337	33700.58	-363.5812
2014, Q3	33031	33245.3	-214.3005
2014, Q4	37132	32884.02	4247.9842

Table 4. Summarized results of evaluating predictive accuracy

Period	Box Jenkins Model	SVM model
PE for 2013, Q1	-6.19	-3.07
PE for 2013, Q2	-11.11	-7.52
PE for 2013, Q3	-12.66	-11.56
PE for 2013, Q4	-7.42	-0.10
PE for 2014, Q1	-7.43	2.73
PE for 2014, Q2	-12.41	-1.09
PE for 2014, Q3	-10.93	-0.65
PE for 2014, Q4	-2.66	11.44
MAPE	8.85	4.77
U	0.0440	0.0324

5 Discussion, Conclusion and Further Research

Construction demand forecasting is vital to the development of strategic future plans for the construction sector. Previous studies have demonstrated that the Box Jenkins model can be used to predict demand, productivity and prices in the construction market (Goh and Teo, 2000; Fan et al., 2010). However, it has been found that construction demand exhibit nonlinear characteristics. This suggests nonlinear models (such as SVM) possess the capacity to generate reliable and accurate predictions of construction demand. The main purpose of the current study is to compare two univariate modelling techniques (Box Jenkins and SVM), in order to identify and assess the predictive capability of both approaches. The out-of-sample forecast

generated between first quarter of 2013 and fourth quarter of 2014 served as a basis for evaluating the predictive performance of these two models.

The high predictive accuracy achieved by the SVM model when compared with Box Jenkins model (Box Jenkins model is considered as a benchmark for univariate model as suggested in Goh and Teo, 2000). This indicates that the SVM model can reliably and accurately forecast construction demand in Hong Kong. In addition, the finding indicates that artificial intelligence models (SVM) tend to outperform linear models. These results are in agreement with those obtained by Goh in 1998. Though previous studies (such as Goh and Teo, 2000) suggest that Box Jenkins model might not be suitable for medium and long-term forecast, one unanticipated finding was that the Box Jenkins model could generate satisfactory forecast for 8-quarters ahead. A possible explanation for this might be the relative stability of the construction demand data series in the forecast period. In general, therefore, it seems that artificial intelligence models (such as SVM) can be applied to construction demand forecasting problems under different circumstances.

The major limitation to this study is the absence of other explanatory variables in the developed models. Despite this limitation, the intended objective of the study was achieved. It is worth nothing that the univariate modelling techniques applied in the current study could be useful in cases of limited data which affects the possibility of developing large multivariate models. The findings from this study enhance the knowledge on the applicability of univariate modelling techniques to construction demand forecasting problems. Overall, the SVM model developed in this study can be used as a tool for predicting future volume of construction demand. Reliable forecast of construction demand is vital for developing and implementing strategies to minimize the adverse impact of fluctuating demand.

6 References

- Akintoye, A. and Skitmore, M. (1994). Models of UK private sector quarterly construction demand. *Construction Management and Economics*, 12(1), pp.3–13.
- Awotona, A. (1990). Nigerian government participation in housing: 1970-1980. *Habitat International*, 14(1), pp.17–40.
- Bin, Y., Zhongzhen, Y. and Baozhen, Y. (2006). Bus Arrival Time Prediction Using Support Vector Machines. *Journal of Intelligent Transportation Systems*, 10(4), pp.151–158.
- Census and Statistics Department (various years), *Report on the Quarterly Survey of Construction Output*, Census and Statistics Department, the HKSAR Government, Hong Kong.
- Fan, R.Y.C., Ng, S.T. and Wong, J.M.W. (2010). Reliability of the Box-Jenkins model for forecasting construction demand covering times of economic austerity. *Construction Management and Economics*, 28(3), pp.241–254.
- Fan, R.Y.C., Ng, T.S. and Wong, J.M.W. (2007). A study of forecasting the Hong Kong construction demand. In: W. Hughes (ed). *Proceedings of the Inaugural Construction Management and Economics 'Past, Present and Future' conference*, 16th-18th July 2007, Reading, UK, pp. 865-876.
- Goh, B.H. (1999). An evaluation of the accuracy of the multiple regression approach in forecasting sectoral construction demand in Singapore. *Construction Management and Economics*, 17(2), pp.231–241.
- Goh, B.H. (1998). Forecasting residential construction demand in Singapore: a comparative study of the accuracy of time series, regression and artificial neural network techniques. *Engineering Construction and Architectural Management*, 5(3), pp.261–275.
- Goh, B.H. (1996). Residential construction demand forecasting using economic indicators: A comparative study of artificial neural networks and multiple regression. *Construction Management and Economics*, 14(1), pp.25–34.

- Goh, B.H. (2000). Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: The case of the Singapore residential sector. *Construction Management and Economics*, 18(2), 209-217.
- Goh, B.H. (2005). The dynamic effects of the Asian financial crisis on construction demand and tender price levels in Singapore. *Building and Environment*, 40(2), pp.267–276.
- Goh, B.H. and Teo, H.P. (2000). Forecasting construction industry demand, price and productivity in Singapore: The Box-Jenkins approach. *Construction Management and Economics*, 18(5), pp.607–618.
- Gruneberg, S. and Folwell, K. (2013). The use of gross fixed capital formation as a measure of construction output. *Construction Management and Economics*, 31(4), pp.359–368.
- Han, S.S. and Ofori, G. (2001). Construction industry in China's regional economy, 1990-1998. *Construction Management and Economics*, 19(2), pp.189–205.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Peter, R. and Witten, I. H. (2009). The WEKA data mining software: *an update*. *SIGKDD Explorations*, 11(1), pp. 10-18.
- Hyndman, R.J., Athanasopoulos, G., Razbash, S., Schmidt, D., Zhou, Z., Khan, Y., Bergmeir, C. and Wang, E. (2015). *Package forecast: Forecasting functions for time series and linear models*, available at: <http://cran.r-project.org/web/packages/forecast/forecast.pdf> (accessed 09 March 2015).
- Hyndman, R.J. and Athanasopoulos, G. (2013). *Forecasting: principles and practice*, OTexts, available at: www.OTexts.com/fpp (accessed 08 January 2015).
- Hyndman, R.J. and Koehler, A.B. (2006). Another look at measures of forecast accuracy, *International Journal of Forecasting*, 22 (4), pp. 679-688.
- Jiang, H., Jin, X.-H. and Liu, C. (2013). The effects of the late 2000s global financial crisis on Australia's construction demand. *Australasian Journal of Construction Economics and Building*, 13(3), pp.65–79.
- Jiang, H. and Liu, C. (2011). Forecasting construction demand: A vector error correction model with dummy variables. *Construction Management and Economics*, 29(9), pp.969–979.
- K'Akumu, O.A. (2007). Construction statistics review for Kenya. *Construction Management and Economics*, 25(3), pp.315–326.
- Killingsworth, R.A. (1990). Preliminary investigation into formulating a demand forecasting model for industrial construction. *Cost Engineering*, 32(8), pp.11–15.
- Lam, K.C. and Oshodi, O.S. (2015). The capital budgeting evaluation practices (2014) of contractors in the Hong Kong construction industry. *Construction Management and Economics*, 33(7), pp.587–600.
- Lam, K.C., Palaneeswaran, E. and Yu, C. (2009). A support vector machine model for contractor prequalification. *Automation in Construction*, 18(3), pp.321–329.
- Lewis, T.M. (1984). A review of the causes of recent problems in the construction industry of Trinidad and Tobago. *Construction Management and Economics*, 2(1), pp.37–48.
- Low, S.P. and Leong, C.H.Y. (1992). A revisit to Turin's paradigm. Construction and development in the 1970s and 1980s. *Habitat International*, 16(3), pp.103–117.
- Notman, D., Norman, G., Flanagan, R., and Agapiou, A. (1998). A time-series analysis of UK annual and quarterly construction output data (1955-95). *Construction Management and Economics*, 16(4), pp.409–416.
- Ofori, G., Hindle, R. and Hugo, F. (1996). Improving the construction industry of South Africa: A strategy. *Habitat International*, 20(2), pp.203–220.
- Pfaff, B. (2013). *Package urca: Unit root and cointegration tests for time series data*, available at: <http://cran.r-project.org/web/packages/urca/urca.pdf> (accessed 09 March 2015).
- R Core Team (2015). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing. Vienna, Austria, available at: <http://www.R-project.org/> (accessed 02 February 2015).

Ren, H. and Lin, S.S. (1996). The UK construction industry under cyclical high inflation, high interest rates and recession. *International Journal of Project Management*, 14(5), pp.301–305.

Soo, A. and Oo, B.L. (2014). The effect of construction demand on contract auctions: an experiment. *Engineering, Construction and Architectural Management*, 21(3), pp.276–290.

Vapnik, V. (1998). *Statistical learning theory*, New York.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, New York: Springer.

Witten, I.H., Frank, E. and Hall, M.A. (2011). *Data Mining: practical machine learning tools and techniques*, 3rd edn. Morgan Kaufmann Publishers, Burlington, MA.

Wong, J.M.W. and Ng, S.T. (2010). Forecasting construction tender price index in Hong Kong using vector error correction model. *Construction Management and Economics*, 28(12), pp.1255–1268.

Appendix 1 Process of fitting a Box-Jenkins model (Adapted from Hyndman and Athanasopoulos, 2013)

