Construction Management and Economics

ISSN: 0144-6193 (Print) 1466-433X (Online) Journal homepage: https://www.tandfonline.com/loi/rcme20

An integrated regression analysis and time series model for construction tender price index forecasting

S. Thomas Ng , Sai On Cheung , Martin Skitmore & Toby C.Y. Wong

To cite this article: S. Thomas Ng , Sai On Cheung , Martin Skitmore & Toby C.Y. Wong (2004) An integrated regression analysis and time series model for construction tender price index forecasting, Construction Management and Economics, 22:5, 483-493, DOI: 10.1080/0144619042000202799

To link to this article: https://doi.org/10.1080/0144619042000202799

Published online: 13 May 2010.

Construction Management and Economics (June 2004) **22**, 483–493

An integrated regression analysis and time series model for construction tender price index forecasting

S. THOMAS NG[1]*, SAI ON CHEUNG[2], MARTIN SKITMORE[3] and TOBY C.Y. WONG[1]

[1]*Department of Civil Engineering, University of Hong Kong, Pokfulam Road, Hong Kong*
[2]*Department of Building and Construction, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong*
[3]*School of Construction Management and Property, Queensland University of Technology, GPO Box 2434, Brisbane, Q4001, Australia*

Received 4 December 2002; accepted 22 January 2004

Clients need to be informed in advance of their likely future financial commitments and cost implications as the design evolves. This requires the estimation of building cost based on historic cost data that is updated by a forecasted Tender Price Index (TPI), with the reliability of the estimates depending significantly on accurate projections being obtained of the TPI for the forthcoming quarters. In practice, the prediction of construction tender price index movement entails a judgemental projection of future market conditions, including inflation. Statistical techniques such as Regression Analysis (RA) and Time Series (TS) modelling provide a powerful means of improving predictive accuracy when used individually. An integrated RA-TS model is developed and its predictive power compared with the individual RA or TS models. The accuracy of the RA-TS model is shown to outperform the individual RA and TS models in both one and two-period forecasts, with the integrated RA-TS model accurately predicting (95% confidence level) one-quarter forecasts for all the 34 holdout periods involved, with only one period not meeting the confidence limit for two-quarter forecasts.

Keywords: Cost estimate, integrated forecasting model, tender price index forecast, time series modelling, regression analysis

Introduction

The majority of contracts for construction work are let by competitive tender when the design is sufficiently advanced for tenders to be compiled. Prior to this, reasonably accurate predictions of the likely tender prices have to be made, as clients need to be informed in advance of their likely future financial commitments and cost implications as the design evolves. This requires the estimation of building cost based on historic cost data that is updated by a forecasted Tender Price Index (TPI) (Tysoe, 1981; Smith, 1995), with the reliability of the estimates depending significantly on accurate projections being obtained of the TPI for the forthcoming quarters (Fitzgerald and Akintoye, 1995) – the degree of accuracy of the projections being determined by their use and form, time horizon and data availability (O'Donovan, 1983; Bowerman and O'Connell, 1987). This is often difficult to do in practice, and entails a highly subjective prediction of future market conditions and inflation (Akintoye, 1991; Akintoye and Fitzgerald, 2000).

The need for more objective methods, and the benefits of quantitative predictive cost models in general, in the construction industry has been recognized for some time (e.g. Skitmore, 1985; Al Khalil, 1999; Al-Tabtabai *et al*., 1999; Ferry and Brandon, 1999; Li and Love, 1999; Li *et al*., 1999; Yin, 1999). As a result, a diversity of cost models of varying complexity has been devised by researchers. Apart from the fuzzy sets (Chang *et al*., 1997; Mason and Kahn, 1997) and artificial neural network approaches (Williams, 1994; Fang and Tam,

*Author for correspondence. E-mail: tstng@hkucc.hku.hk

Construction Management and Economics
ISSN 0144-6193 print/ISSN 1466-433X online
http://www.tandf.co.uk/journals
DOI: 10.1080/0144619042000202799

1999), statistical methods have also been extensively applied in TPI prediction, with regression analysis (RA) being the most popular approach (Boussabaine and Elhag, 1999). Univariate time series (TS) modelling has also received favourable attention. For instance, TS models have been developed to forecast the behaviour of property prices (Chin and Mital, 1998) and building costs (Taylor and Bowen, 1987) – later extended by Fellows (1991) for predicting movements in the TPI. Most recently, Ng *et al.* (2000) adopted discriminant analysis for predicting TPI directional changes.

Following Granger and Newbold (1986), researchers (e.g. Granger, 2001) have suggested that the integration of techniques might further enhance the predictive ability. In TPI prediction, the RA and TS models are the most highly developed, with RA establishing the relationship between the TPI and predominant economic indicators (e.g. McCaffer *et al.*, 1983; Runeson, 1988; Fellows, 1991; Hoptroff *et al.*, 1991; Fitzgerald and Akintoye, 1995; Akintoye *et al.*, 1998; Chau, 1998), while TS estimates the index trend through historic TPI data. The objectives of this paper are to outline the procedures for integrating the RA and TS models and to examine the reliability of the resulting model in generating TPI forecasts for Hong Kong construction projects.

Data Set

The RA variables comprised the TPI and nine exogenous economic indicators identified in Ng *et al.*'s (2000) similar previous Hong Kong study (see Appendix 1). The period covers a total of 76 quarters starting from 1980Q1 to 1998Q4. Despite a constant increase in the TPI, the construction tender price during that time was heavily affected by world recession (in 1982), the Gulf War (in 1991) and Asian Economic Turmoil (in 1997). Clearly, therefore, using this duration for subsequent analysis could help determine the extent to which an integrated RA-TS model will be influenced by volatile conditions. The required data was acquired from the relevant sources and publications of the HKSAR government (HKCSD, 1999).

To determine the relevancy of the suggested candidate indicators, a Pearson correlation analysis between the TPI and each of the nine indicators (with various degrees of leading and lagging) was carried out. With the exception of BLR and UR, there are strong positive correlations, indicating the trend of the TPI to be highly correlated with most of the indicators used. This is in line with previous research, except for BLR (interest rate), which has been found previously in the UK to have a strong positive correlation with the TPI (Fellows, 1988).

By comparing the correlation coefficients under different time lags, it was also found that the BLR movements led TPI movements by three quarters, GDP and GDPC by two quarters, with HSIAV, IGDPD and BCI leading by one quarter. CCPI, M3 and UR on the other hand, had no apparent leading or lagging effects.

Regression analysis

RA has been widely used for the prediction of tender trends (e.g. McCaffer *et al.*, 1983; Fellows, 1988; Runeson, 1988; Akintoye and Skitmore, 1990, 1993, 1994). Regression models provide accurate predictions of TPI movements when price levels are steady, e.g. moving constantly upward or downward (Ng *et al.*, 2000). In this study, a multivariate RA with an automated stepwise procedure was adopted to eliminate those factors with negligible effects on the TPI and provide a subset whose estimated equation produces the best fit, i.e. the minimum residual sum of squares or the maximum coefficient of determination, R^2.

Table 1 summarizes the stepwise procedure of the multivariate RA. Variables are added or removed from the regression model step by step. The partial R^2 indicates the partial potential contribution of variables

Table 1 Summary table of the stepwise procedure of multivariate regression

Step	Variable		Number in	Partial R^2	Model R^2	C(p)	F	Prob > F
	Entered	Removed						
1	BCI		1	0.9753	0.9753	443.9479	2839.3917	0.0001
2	IGDPD		2	0.0176	0.9929	79.6684	176.1098	0.0001
3	BLR		3	0.0012	0.9941	56.1908	14.5955	0.0003
4	UR		4	0.0007	0.9949	42.7815	9.9571	0.0024
5	M3		5	0.0003	0.9952	38.7747	4.0533	0.0480
6	HSIAV		6	0.0006	0.9958	27.6588	10.0248	0.0023
7		IGDPD	5	0.0000	0.9958	26.0036	0.2635	0.6094
8	CCPI		6	0.0005	0.9963	17.5085	9.0721	0.0037

Note: GDP, GDPC and IGDPD are automatically eliminated from the model due to their negligible influence in the regression model.

to the whole regression model, i.e. the greater the partial R^2, the greater the significance of the variable. In this analysis, the most important variable was BCI (partial $R^2 = 0.9753$) with the least important variable being M3 (partial $R^2 = 0.0003$). GDP, GDPC and IGDPD were automatically eliminated as they provided a negligible potential contribution to the regression function (partial $R^2 < 0.0001$).

The RA model was fitted to the lagged exogenous variables by forward stepwise variable entry, the resulting multivariate regression function being:

$$\begin{aligned} Y_{TPI} = {} & 66.6274 + 1.6115X_{BLR} + 0.4746X_{BCI} \\ & - 0.3117X_{CCPI} - 2.7375X_{UR} + 0.0932X_{M3} \\ & - 0.00215X_{HSIVA} \end{aligned} \tag{1}$$

Before Equation 1 can be used for forecasting, the future values of exogenous variables (i.e. BCI, CCPI, BLR, UR, M3 and HSIAV) have to be determined. These future values can be derived by the growth rate of the historic periods of each exogenous variable and then extrapolated for the next two quarters. For instance, if the general growth rate of exogenous variable BLR from 1980Q1 to 1989Q4 was 4% per quarter, then the estimated BLR for 1990Q1 and 1990Q2 would be 1.04 and 1.04^2 times of BLR_{1989Q4} respectively. Based on Equation 1 and the projected exogenous variables for the forthcoming one and two quarters, the TPI for the coming quarters can be forecasted.

Time series model

The simplest TS approach is exponential smoothing. This forecasting method is not based on the analysis of the entire historical TS. Rather, it uses a weighted moving average as the forecast, with the assigned weights decreasing exponentially for periods further into the past. Simple exponential smoothing is most effective as a forecasting method when cyclical and irregular influences comprise the main effects on the time series values. However, the exponential smoothing method was considered inadequate to provide an accurate model for TPI prediction, as it assumes that errors are uncorrelated, which in turn implies that the observations are uncorrelated. In practice, this assumption can rarely be met, as serial correlation is usually expected when data is collected sequentially in time.

A stochastic TS modelling technique known as Auto-Regressive Integrated Moving Average (ARIMA), however, can represent a variety of correlation structures (Yin, 1999). While Auto-Regressive (AR) estimates the stochastic process underlying a TS where the TS values exhibit non-zero autocorrelation (autocorrelation being the way observation in a TS is related to each other), Moving Average (MA) estimates the process where the current TS value is related to the random errors from previous time periods.

According to Fellows (1988), stochastic TS should be satisfactory in modelling tender price movements, as it can model the changing process and provides a class of models of the stationary stochastic processes. Therefore, the ARIMA models were adopted for model building, and Box-Jenkins identification-estimation-checking iterative procedure (Box and Jenkins, 1976) was followed when devising the TS model (see Appendix 2 for details of ARIMA and Box-Jenkins iterative procedure). Taylor and Bowen (1987) and Fellows (1987) have used this technique and found it to be satisfactory in modelling TPI movements.

The data series were found to be stationary after the *first differencing* ($p = 0.06849$) and the Ljung modification of the Box-Pierce $Q_{statistics}$ indicated the residuals to be reasonably random ($t = -4.032$, $p > 0.05$). Maximum likelihood parameter estimates were obtained. To determine the best-fit model, all models were examined by diagnostic checking. First, the estimation model with highly significant parameters (see (a) in Figure 1 for t-value), such as AR1, AR2, MA1, MA2, etc., were selected. If the absolute value $|t|$ of the last (highest order) parameter estimation (see (b) in Figure 1) is close to 1 or greater than 1, it is possible that the process is non-stationary.

Next, the residual was checked by examining the p-values of the Q-statistics (see (e) in Figure 1). In addition, as the residuals should be uncorrelated to each other, large residual autocorrelations (i.e. those very close to 1) may indicate problems with the fit of the model. The remaining candidate models was further examined by the checking the goodness-of-fit criteria through variance estimates – AIC & SBC (see [c] & [d] in Figure 1). The MA(2) model shown in Equation 2 below provided the best fit to the data, and the low autocorrelations (see (f) in Figure 1) confirm that the residuals to be random.

$$Y_t - Y_{t-1} = \varepsilon_t + 0.7312\,\varepsilon_{t-1} + 0.47\varepsilon_{t-2} \tag{2}$$

where ε_t is a random error term uncorrelated over time, typically called white noise; Y_t is the value of TPI time series in current time period t; and Y_{t-1} is the value of TPI time series in previous time period $t-1$.

Equation 3 shows the MA(2) model with the backshift notation (see (g) in Figure 1), i.e. the lagged value of the time series variable.

$$(1 - B^1)Y_t = (1 + 0.73125B^1 + 0.47B^2)\varepsilon_t \tag{3}$$

where B^1 is actually $B^{**}(1)$, which represents a first order backshift operator, e.g. $B^1Y_t = Y_{t-1}$, while $B^2Y_t = Y_{t-2}$.

Forecasts were made based on its own historic data. For example, for forecast made in one quarter ahead, let us say forecasting TPI of 01/10/96, historic data from 01/01/80 to 01/07/96 was used.

Maximum Likelihood Estimation

Parameter	Standard estimate	Error	Approx t value	Pr > |t|	Lag
MA1,1	−0.73125	−0.10436	−7.01 [a]	< 0.0001	1
MA1,2	−0.47000 [b]	−0.10566	−4.45	< 0.0001	2

Variance estimate	10.14103
Std error estimate	3.184498
AIC	379.0317 [c]
SBC	383.6126 [d]
Number of residuals	73

Correlations of Parameter Estimates

Parameter	MA1,1	MA1,2
MA1,1	1.000	0.505
MA1,2	0.505	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.32	4	0.2562 [e]	0.095	0.096	0.195 [f]	−0.008	0.107	−0.015
12	11.58	10	0.3138	0.040	0.147	0.030	0.021	0.149	−0.158
18	20.67	16	0.1914	−0.121	−0.163	−0.148	0.048	−0.160	−0.168
24	26.23	22	0.2420	−0.060	−0.078	−0.158	−0.107	−0.030	−0.074

Model for variable TPI

Data have been centred by subtracting the value 2.871918

Period(s) of differencing

No mean term in this model.

Moving average factors

Factor 1: 1 + 0.73125 B**(1) + 0.47 B**(2) [g]

Figure 1 Results of fitting the MA(2) model by SAS

Integrated regression time series model

The RA and TS models were integrated by linear combinations by considering the forecasts made by RA and TS as f_1 and f_2 respectively. From this, a new forecast of these quantities can be produced by:

$$f_3 = \lambda f_1 + (1 - \lambda) f_2 \tag{4}$$

where λ is the weighting which is restricted to the range (0,1).

Goodness-of-fit statistics assist in assessing the fit of a model. These statistics can be compared across competing models, and typically the model with goodness-of-fit statistics closest to zero provides the best fit. The mean square error (MSE) and its positive square root (RMSE) are often used to evaluate the fitness of models, as the MSE minimizes the sum of the variance and the square of the bias.

The weightings were derived by iteration (Figure 2). As illustrated in Table 2, the weight column represents the weighting for the RA model, while the MSE and RMSE are used as the goodness-of-fit statistics for comparing models. Since both the MSE and RMSE reach the minimum at a weighting of 0.5, further investigations within the range of 0.4 to 0.6 were performed to identify a weighting scheme that could generate a smaller residual, i.e. the goodness-of-fit statistics closest to zero.

Table 3 reveals that both the MSE and RMSE reach the minimum at a weighting of 0.51 (for the RA component), and hence further investigations within the range 0.50 to 0.52 was carried out, with an interval of 0.01, to obtain an optimal weighting. The results of iterative looping using a 0.001 interval indicate that the weightings of $0.512_{RA} : 0.488_{TS}$ for a TPI forecast of one quarter in advance (RA-TS_{Q1}) yield the closest-to-zero values for both MSE and RMSE. That means the RA-TS_{Q1} model is almost equivalent to the average of the RA and TS forecasts. For the TPI forecast in two quarters ahead (RA-TS_{Q2}), the weightings of $0.647_{RA} : 0.353_{TS}$ result in the closest-to-zero values for both MSE and RMSE. Unlike the RA-TS_{Q1} model, the weighting of the RA component is virtually twice as much as that for the TS part in the RA-TS_{Q2} models, indicating that the RA results are more significant in improving the accuracy of the two-quarter forecast.

Backcast testing

The holdout samples between 1/1/1990 and 1/4/98 were fitted into the RA-TS_{Q1} and RA-TS_{Q2} to examine the forecasting accuracy. Figures 3 and 4 show the actual TPI as compared to the results of the one and two-quarter forecasts based on different models, while the part results of the two-quarter forecasts are summarized in Table 4. The upper and lower 95% confident limits were used to determine the model accuracy, and the forecasts would be considered correct should the actual TPI value is within the confidence limits of the corresponding quarter. With the RA-TS_{Q1} model, no actual TPI value falls beyond the confidence limit, representing 100% accuracy. More than 75% were within ± 1 of the standard deviation (3.1845). As for the RA-TS_{Q2} model, out of the 34 holdout samples, only 1 quarter (i.e. 1997Q3) has the actual TPI value outside the prediction interval, which implies that 97% of test data lies within the prediction interval.

Percentage deviations were calculated by comparing the deviation to the half-forecast range. The actual value is out of the forecasting range if the percentage deviation is greater than 100. The smaller the percentage, the more accurate the forecast is. Tables 5 and 6 highlight the quarters with percentage deviations exceeding 100 when using the RA, TS and RA-TS models. The forecasting accuracy between the RA and TS models is similar when used for one-quarter TPI forecast, as both models have two inaccurate predictions (i.e. 1991Q2 & 1997Q4 for the RA model, and 1994Q4 & 1997Q2 for the TS model). However, the forecasts for these periods were improved (i.e.

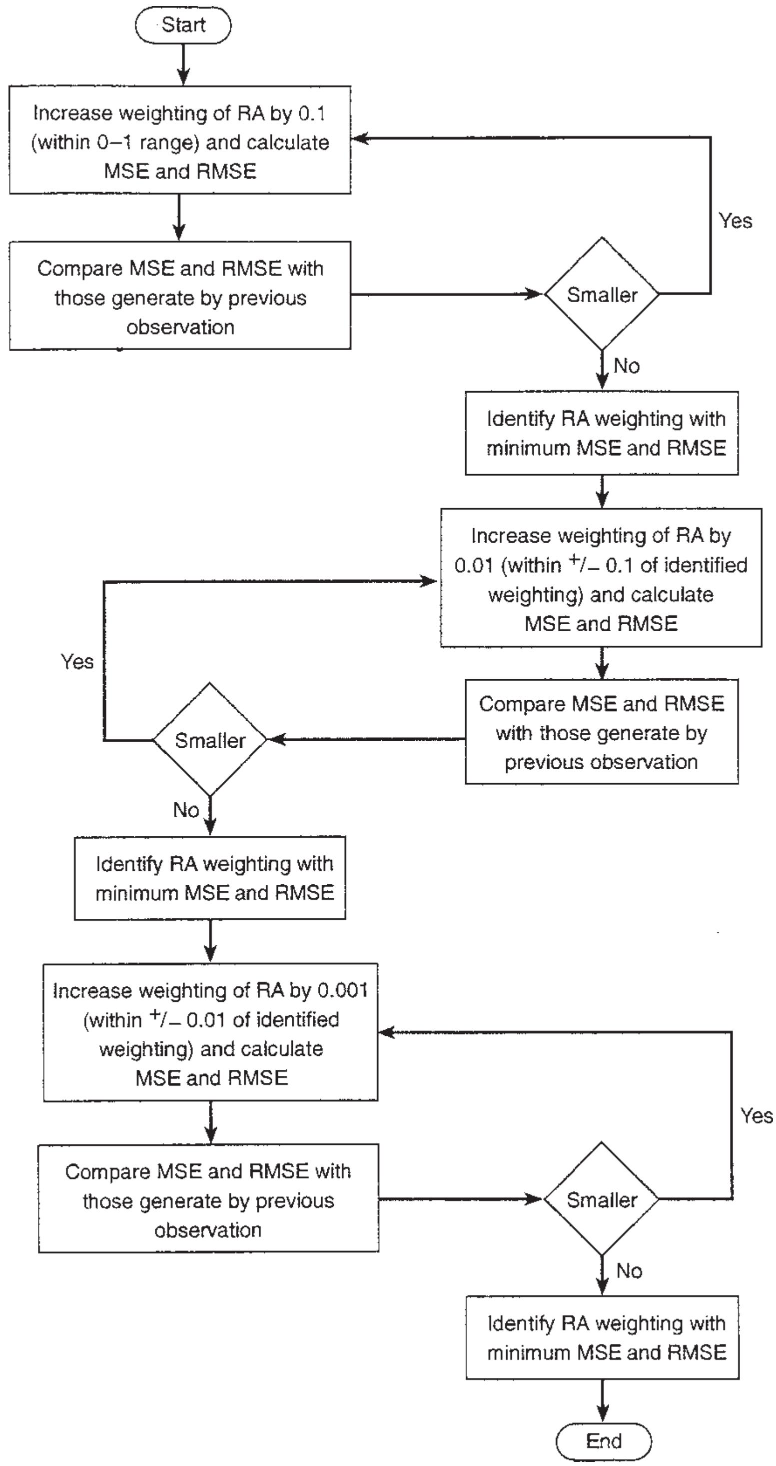

Figure 2 Iterative loop to determine the best weighting combination

percentage deviation < 100) when RA-TS_{Q1} model was adopted. For two-quarter TPI forecast, the RA model was the most accurate, while the TS model was the worst (with four inaccurate predictions: 1992Q1, 1997Q2, 1997Q3 & 1998Q2). Using the RA-TS_{Q2} model improves the forecasting accuracy by leaving only one inaccurate prediction (i.e. 1997Q3). The over-estimation of TPI between 1991 and 1992 may be caused by the launching of democratic reforms in Hong Kong at that time, resulting in sudden economic and political shocks. The under-estimation in 1997 may be due to an over-optimistic expectation for the economic

Table 2 MSE and RMSE for different weighting

Observation	Goodness of fit statistics				
	Weighting	mad_c	mper_c	mse_c	rmse_c
1	0.0	0.39955	0.11925	12.5024	3.53587
2	0.1	0.22213	0.04063	10.8270	3.29045
3	0.2	0.04471	−0.03799	9.5147	3.08458
4	0.3	−0.13270	−0.11661	8.5652	2.92664
5	0.4	−0.31012	−0.19522	7.9788	2.82467
6	0.5	−0.48754	−0.27384	7.7553*	2.78483*
7	0.6	−0.66496	−0.35246	7.8947	2.80975
8	0.7	−0.84238	−0.43107	8.3971	2.89778
9	0.8	−1.01980	−0.50969	9.2625	3.04343
10	0.9	−1.19722	−0.58831	10.4908	3.23895
11	1.0	−1.37464	−0.66692	12.0821	3.47593

Note: *MSE & RMSE closest to zero.

Table 3 MSE and RMSE for weightings from 0.40 to 0.60 at an interval of 0.01

Observation	Goodness of fit statistics				
	Weighting	mad_c	mper_c	mse_c	rmse_c
1	0.40	−0.31012	−0.19522	7.97876	2.82467
2	0.41	−0.32787	−0.20308	7.94008	2.81781
3	0.42	−0.34561	−0.21095	7.90503	2.81159
4	0.43	−0.36335	−0.21881	7.87360	2.80599
5	0.44	−0.38109	−0.22667	7.84581	2.80104
6	0.45	−0.39883	−0.23453	7.82164	2.79672
7	0.46	−0.41657	−0.24239	7.80110	2.79305
8	0.47	−0.43432	−0.25025	7.78420	2.79002
9	0.48	−0.45206	−0.25812	7.77092	2.78764
10	0.49	−0.46980	−0.26598	7.76127	2.78591
11	0.50	−0.48754	−0.27384	7.75526	2.78483
12	0.51	−0.50528	−0.28170	7.75287*	2.78440*
13	0.52	−0.52303	−0.28956	7.75411	2.78462
14	0.53	−0.54077	−0.29742	7.75898	2.78549
15	0.54	−0.55851	−0.30529	7.76748	2.78702
16	0.55	−0.57625	−0.31315	7.77961	2.78920
17	0.56	−0.59399	−0.32101	7.79537	2.79202
18	0.57	−0.61174	−0.32887	7.81476	2.79549
19	0.58	−0.62948	−0.33673	7.83778	2.79960
20	0.59	−0.64722	−0.34459	7.86443	2.80436
21	0.60	−0.66496	−0.35246	7.89471	2.80975

Note: *MSE & RMSE closest to zero.

prospect after the sovereignty of the HKSAR was returned to China. The effects of these economic and political shocks are reflected through the pattern changes of some exogenous economic indicators, such as GDP (dropped from 1998Q3), IGDPD (dropped from 1998Q3), UR (rose from 1997Q4), etc.

Conclusions

An integrated model is described for forecasting construction TPI movement. The model was derived by amalgamating the analytical power of both the RA and TS models. Hong Kong data pertinent to exogenous and endogenous variables were collected and used for model building. A multivariate regression function was derived using the five exogenous variables which have significant effects on the regression function, i.e. BLR, BCI, CCPI, IGDPD and HSIAV. The forecasting power of the RA was considered exceptional, with only two quarters exceeding 95% confidence limit (when used for one-quarter forecast). Therefore, in the absence of any sophisticated analytical model, the RA

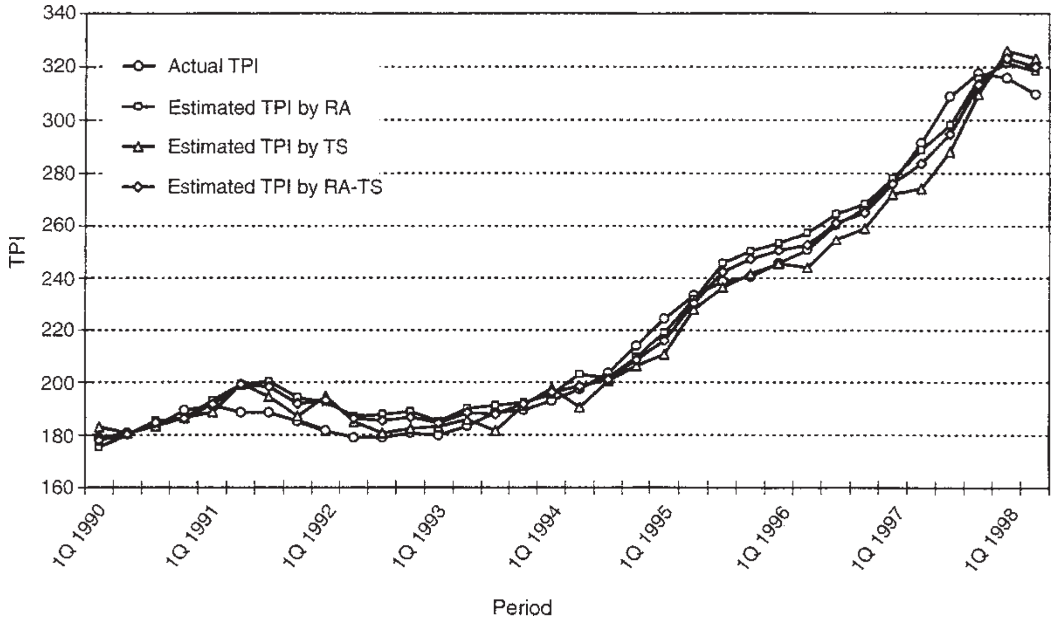

Figure 3 Actual TPI and forecast generated by various models – one-quarter forecast

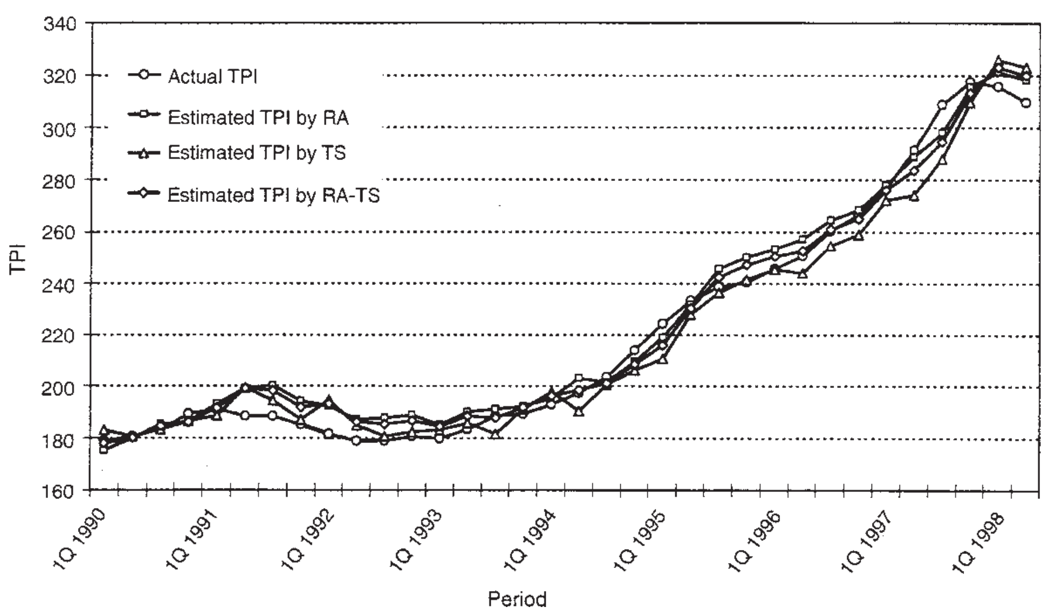

Figure 4 Actual TPI and forecast generated by various models – two-quarter forecast

should provide a reasonably reliable indication as to the TPI movement.

The derivation of the TS model was based on the stochastic ARIMA approach. Guided by the Box-Jenkins procedure for TS model development, the data was first checked for stationarity, and models with different parameters were then checked to establish the best-fit TS model. The MA(2) model was considered most suitable for the TS prediction. However, the predictive ability of the TS model alone is not as good as the RA. The percentage deviations revealed six quarters to have been inaccurately predicted by the MA(2) model (i.e. two and four for one and two-quarter forecast respectively). The TS model, therefore, may not adequately provide an accurate forecast when used in isolation with these data.

Table 4 Forecast made two-quarter ahead by RA-TS_{Q2} model

Period	Forecast TPI value	Lower 95% confidence limit	Upper 95% confidence limit	Actual TPI value	Deviation	Deviation from actual TPI (%)	Deviation from forecast range (%)
1990Q1	178.14	165.66	190.61	178.95	0.81	0.45	6.52
1990Q2	180.35	167.87	192.83	180.70	0.35	0.19	2.80
1990Q3	184.71	172.24	197.19	183.33	−1.38	−0.75	−11.10
1990Q4	186.39	173.91	198.87	189.47	3.08	1.63	24.68
1991Q1	191.60	179.12	204.07	191.23	−0.37	−0.19	−2.93
1991Q2	199.11	186.63	211.59	188.60	−10.51	−5.57	−84.22
1991Q3	198.21	185.73	210.69	188.60	−9.61	−5.10	−77.01
1991Q4	191.75	179.27	204.23	185.09	−6.66	−3.60	−53.37
1992Q1	193.16	180.68	205.64	181.58	−11.58	−6.38	−92.79
1992Q2	186.32	173.84	198.80	178.95	−7.37	−4.12	−59.06
1992Q3	185.35	172.87	197.82	178.95	−6.40	−3.58	−51.25
1992Q4	186.71	174.23	199.18	180.70	−6.01	−3.33	−48.13
1993Q1	184.55	172.07	197.03	179.82	−4.73	−2.63	−37.90
1993Q2	188.53	176.05	201.01	183.33	−5.20	−2.84	−41.68
1993Q3	187.83	175.35	200.31	188.60	0.77	0.41	6.16
1993Q4	191.88	179.40	204.36	189.47	−2.41	−1.27	−19.30
1994Q1	196.23	183.75	208.70	192.98	−3.25	−1.68	−26.01
1994Q2	198.53	186.05	211.01	197.37	−1.16	−0.59	−9.31
1994Q3	201.15	188.68	213.63	203.51	2.36	1.16	18.87
1994Q4	207.88	195.41	220.36	214.04	6.16	2.88	49.33
1995Q1	216.18	203.70	228.66	224.56	8.38	3.73	67.13
1995Q2	230.47	217.99	242.95	233.33	2.86	1.23	22.94
1995Q3	242.38	229.90	254.85	238.60	−3.78	−1.58	−30.25
1995Q4	247.14	234.66	259.62	240.35	−6.79	−2.83	−54.43
1996Q1	250.61	238.13	263.09	245.61	−5.00	−2.04	−40.06
1996Q2	252.73	240.25	265.21	250.88	−1.85	−0.74	−14.84
1996Q3	261.21	248.73	273.69	260.53	−0.68	−0.26	−5.48
1996Q4	265.18	252.70	277.66	266.67	1.49	0.56	11.95
1997Q1	275.85	263.37	288.33	276.32	0.47	0.17	3.78
1997Q2	283.38	270.90	295.86	291.23	7.85	2.70	62.92
1997Q3	294.46	281.99	306.94	308.77	14.31	4.63	114.64
1997Q4	313.36	300.88	325.84	317.54	4.18	1.32	33.51
1998Q1	322.91	310.43	335.39	315.79	−7.12	−2.25	−57.08
1998Q2	319.89	307.41	332.37	309.65	−10.24	−3.31	−82.04

Note: standard deviation = 6.2394; deviation from forecast range = deviation ÷ half of the forecast range (i.e. between upper and lower confident limits).

Table 5 Part results of percentage deviation for one-quarter forecast

Period	RA_{Q1}	TS_{Q1}	RA-TS_{Q1}
1991Q2	−110.09*	−72.89	−91.94
1994Q4	25.86	103.67*	63.83
1997Q2	41.92	138.31*	88.96
1997Q4	103.02*	−41.29	32.60

Note: * percentage deviation > 100.

Table 6 Part results of percentage deviation for two-quarter forecast

Period	RA_{Q2}	TS_{Q2}	RA-TS_{Q2}
1992Q1	−85.91	−105.40*	−92.79
1997Q2	22.24	137.47*	62.92
1997Q3	85.43	168.19*	114.64*
1998Q2	−68.31	−107.20*	−82.04

Note: * percentage deviation > 100.

The RA and TS models were then linearly combined based on the weightings of $0.512_{RA} : 0.488_{TS}$ for RA-TS_{Q1}, and $0.647_{RA} : 0.353_{TS}$ for RA-TS_{Q2}. The results of backcast testing confirmed that the integrated RA-TS model outperforms both the individual RA or TS forecasts. Only one quarter has the actual TPI value exceeding the confidence limit (based on RA-TS_{Q2}), indicating that 97% of test data lies within the prediction interval. The integrated RA-TS model, therefore, should have a high potential of improving the forecasting accuracy of TPI movement even under a rapidly changing economic environment. While the study presented in this paper was based on the Hong Kong data collected within a finite period of time, the findings should help improve our understanding on the possible problems and techniques when a predictive model for TPI forecast is developed in future.

References

Akintoye, A. and Fitzgerald, E. (2000) A survey of current cost estimating practices in the UK. *Construction Management and Economics*, **18**(2), 161–72.

Akintoye, S.A. (1991) Construction Tender Price Index: Modelling and Forecasting Trends, PhD thesis, Department of Surveying, University of Salford.

Akintoye, S.A. and Skitmore, R.M. (1990) Analysis of UK tender price level, *Transactions, 34th Annual Meeting of American Association of Cost Engineers*, Boston, MA, 7.1–7.7.

Akintoye, S.A. and Skitmore, R.M. (1993) Macro models of UK construction contract prices. *Civil Engineering Systems*, **10**(4), 279–99.

Akintoye, S.A. and Skitmore, R.M. (1994) A comparative analysis of the three macro price forecasting models. *Construction Management and Economics*, **12**(3), 257–70.

Akintoye, S.A., Bowen, P. and Hardcastle, C. (1998) Macro-economic leading indicators of construction contract prices, *Construction Management and Economics*, **16**, 159–75.

Al Khalil, M., Assaf, S., Abdul Rahman, W. and Asfoor, M. (1999) Conceptual cost estimating model for water reservoirs. *Cost Engineering*, **41**(5), 38–43.

Al-Tabtabai, H., Alex, P. and Tantash, M. (1999) Preliminary cost estimation of highway construction using neural networks. *Cost Engineering*, **41**(3), 19–24.

Boussabaine, A.H. and Elhag, T.M.S. (1999) *Statistical Analysis and Cost Models Development for Tender Price Estimation*, The Royal Institute of Chartered Surveyors, London.

Bowerman, B.L. and O'Connell, R.T. (1987) *Time Series Forecasting*, Duxbury Press, Boston.

Box, G. and Jenkins, G. (1994) *Time Series Analysis Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ.

Chang, N.B., Chen, Y.L. and Chen, H.W. (1997) Fuzzy regression analysis for the construction cost estimation of wastewater treatment plants – I: theoretical development, *Journal of Environmental Science and Health, Part A: Environmental Science and Engineering & Toxic and Hazardous Substance Control*, **32**(4), 885–99.

Chau, K.W. (1998) The implications of the difference in the growth rates of the prices of building resources and outputs in Hong Kong. *Engineering, Construction and Architectural Management*, **5**(1), 38–50.

Chin, T.C. and Mital, D.P. (1998) Time series modelling and forecasting of Singapore property price: an optimal control approach, *Proceedings: 2nd International Conference on Knowledge-Based Intelligent Electronic Systems*, 21–23 April, Adelaide, Australia, IEEE Publishing, pp. 370–5.

Fang, C.F. and Tam, C.M. (1999) Comparative cost analysis of using high-performance concrete in tall building construction by artificial neural networks. *ACI Structural Journal*, **96**(6), 927–36.

Fellows, R.F. (1988) Escalation management. PhD thesis, Department of Construction Management, University of Reading.

Fellows, R.F. (1991) Escalation management: Forecasting the effects of inflation on building projects. *Construction Management and Economics*, **9**, 187–204.

Ferry, D.J. and Brandon, P.S. (1999) *Cost Planning of Buildings*, BSP Professional Books, Oxford.

Fitzgerald, E. and Akintoye, A. (1995) The accuracy and optimal linear correction of UK construction tender price index forecasts. *Construction Management and Economics*, **13**(6), 493–500.

Granger, C.W.J. (2001) *Essays in Econometrics*, Cambridge University Press, Cambridge.

Granger, C.W.J. and Newbold, P. (1986) *Forecasting Economic Time Series*, Academic Press, Orlando.

Hamburg, M. and Young, P. (1994) *Statistical Analysis for Decision Making*, Dryden Press, Texas.

HKCSD (2000) *Hong Kong Monthly Digest of Statistics*, Census & Statistics Dept, HKSAR Government, Hong Kong.

Hoptroff, R.G., Bramson, M.J. and Hall, T.J. (1991) Forecasting economic turning points with neural nets, *Neuro Computing Application Forum*, University of Aston, Birmingham, April.

Janacek, G.J. (2001) *Practical Time Series*, Arnold, London.

Li, H. and Love, P.E.D. (1999) Combining rule-based expert systems and artificial neural networks for mark-up estimation. *Construction Management and Economics*, **17**(2), 169–76.

Li, H., Shen, L.Y. and Love, P.E.D. (1999) ANN-based mark-up estimation system with self-explanatory capacities. *Construction Engineering and Management*, ASCE, **125**(3), 185–9.

Mason, A.K. and Kahn, D.J. (1997) Estimating costs with fuzzy logic, *Proceedings of the 1997 41st Annual Meeting of AACE International*, 13–16 July, Dallas, TX.

McCaffer R., McCaffrey, M.J. and Thorpe, A. (1983) The disparity between construction cost and tender price movements. *Construction Papers*, CIB, **2**(2), 17–28.

Mendenhall, W. and Sinich, T. (1993) *A Second Course in Business Statistics: Regression Analysis*, 4th edn, Macmillan Publishing Company, New York.

Ng, S.T., Cheung, S.O., Skitmore, R.M., Lam, K.C. and Wong, L.Y. (2000) The prediction of tender price index directional changes. *Construction Management and Economics*, **18**(7), 843–52.

O'Donovan, T.M. (1983) *Short Term Forecasting: An Introduction to Box-Fenkins Approach*, John Wiley and Sons, New York.

Runeson, K.G. (1988) Methodology and method for price-level forecasting in the building industry. *Construction Management and Economics*, **6**(1), 49–55.

Skitmore, R.M. (1985) *The Influence of Professional Expertise in Construction Price Forecasts*. Department of Civil Engineering, The University of Salford.

Smith, A.J. (1995) *Estimating, Tendering and Bidding for Construction: Theory and Practice*, Macmillan, London.

Taylor, R.G. and Bowen, P.A. (1987) Building price-level forecasting: an examination of techniques and application. *Construction Management and Economics*, **5**(1), 21–44.

Tysoe, B.A. (1981) *Construction Cost and Price Indices: Description and Use*, E & FN Spon, London.

Williams, T.P. (1994) Predicting changes in construction cost indexes using neural networks. *Journal of Construction Engineering and Management*, ASCE, **120**(2), 306–20

Wong, A.L.Y. (2001) Predicting the trend of tender price index through discriminant analysis, dissertation submitted for partial fulfilment of BSc in quantity surveying, City University of Hong Kong, Hong Kong.

Yin, R. (1999) Forecasting short-term timber prices with univariate ARIMA models. *Southern Journal of Applied Forestry*, **23**(1), 53–8.

Appendix 1: data

Tender Price Index (TPI) measures both the trend of contractors' pricing strategies and the inflation of labour, plant and materials. This indicator has received empirical attention in several notable studies (e.g. Tysoe, 1981; Runeson, 1988).

Besides the TPI, nine economic factors were chosen to test their relevancy with TPI and, in turn, for forecasting the movement of TPI. These candidate indicators were selected according to literature review (e.g. Akintoye *et al.*, 1998; Fellows, 1988), the relationship between the TPI movement and the cyclical movement of each economic indicator (Akintoye, 1991), and their availability in HK (Ng *et al.*, 2000). The reasons for eliminating some important exogenous indicators from the HK studies can also be found in Wong (2001). Out of the nine chosen factors, four are domestic economic factors, three are banking sector indicators and the remaining two are construction-related factors. These should provide a comprehensive description of the economic condition of the HK.

Domestic economic indicators

Composite consumer price index (CCPI) provides a measure to reflect changes in the price level of consumer goods and services generally purchased by households. The change in CCPI is an important indicator of inflation affecting households. CCPI also relates to inflation of consumer goods and services affecting households.

Gross domestic product (GDP) could be used for analysing different aspects of economic performance. GDP refers to the net output of all producing units in an economy and is related to production activities within the economy, such as employment, productivity, industrial output, investment in equipment and structure.

Implicit gross domestic product deflator (IGDPD) measures the level of inflation, but different from CCPI, It considers inflation for the economy as a whole.

Unemployment rate (UR) compares the number of unemployed to the number of people in the work force. It is directly affected when the economy is deteriorating.

Banking sector indicators

Best lending rates (BLR): in normal situations, as the Hong Kong dollar is pegged to the US dollar; whenever the Federal Reserve moves its Federal Fund rate, HK will move its best lending rate.

Money supply definition (M3) includes the Hong Kong dollar in circulation and all kinds of deposits. It measures the HK dollar deposit in banking sector.

Hang Seng Index 100 Days Moving Average (HSIAV) is a barometer of the Hong Kong stock market. The constituent stocks are grouped under Commerce and Industry, Finance, Properties and Utilities sub-indices. HSI currently comprises 33 constituent stocks that are representative of the market. The aggregate market capitalization of these stocks accounts for about 70% of the total market capitalization on the Stock Exchange of Hong Kong Limited – an indicator of stock market performance. In this study, a 100 days moving average was used to avoid daily fluctuations.

Construction related indicators

Building cost index (BCI): referred to as the Consolidated Labour & Material Index (CLMI) in HK, this is a combination of 45% of the Labour Index and 55% of the Material Index. The Material Index and Labour Index are compiled according to the average prices of material and wages of labour figures. It is often a major indicator of building cost level.

Gross domestic product construction (GDPC) is the same as GDP but only considers net construction-related output.

Appendix 2: an overview of statistical techniques used

Regression analysis

The basic assumptions of regression analysis are:

- the independent variables are not intercorrelated;
- the predictor or independent variables are known without error;
- the prediction errors or residuals are assumed to be independent, identically normally distributed random variables and with a mean of zero;
- there is minimum autocorrelation; and
- the effect of the independent variables on the dependent variable is linear, i.e. additive and proportional to the value of each independent variable, and thus the procedure used is to solve a linear equation of the form:

$$y = \alpha_0 + \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$$

where y represents the dependent variable, v_1 etc represent the independent variables and α_0 etc., represent the regression coefficients, or weights attached to each independent variable.

There is a huge literature on RA and very many textbooks, including introductory texts. For absolute beginners, a good starting point for the practical application of RA to business problems is Mendenhall and Sinich (1993)

Auto-regressive integrated moving average

Simple exponential smoothing assumes that observations are uncorrelated. However, serial correlation can be expected if the data are collected sequentially in time. As a result, models that include the correlation structure have to be considered, and a special class of stochastic models called the Auto-Regressive Integrated Moving Average (ARIMA) models are used. ARIMA implies a variety of different correlation structures. Once the correlation structure has been appropriately modelled, it is straightforward to obtain predictions. ARIMA models can represent many stationary and non-stationary stochastic processes. A stationary stochastic process is characterized by its mean, variance, and autocorrelation function. Transformation of non-stationary data series with changing means into stationary series before time series forecast should, therefore, be carried out. Readers are recommended to refer to Janacek (2001) for the concept of ARIMA.

The Box-Jenkins procedure

The procedure suggested by Box and Jenkins (1994) for applying ARIMA models to time-series analysis, forecasting and control is selected for carrying out stochastic time-series modelling in the study. The Box-Jenkins procedure not only adequately models the changing process, but also provides a class of models of stationary stochastic processes. The main advantage of this procedure is its generality, as it can handle virtually any time-series data, partly owing to its strong theoretical foundations, and also due to its success in empirical comparisons, which have been found to be as accurate as many complex econometric models. It also allows for a wide range of possible models for the data and provides a strategy for selecting a model from that class which best represents the data.

The Box and Jenkins procedure is primarily an iterative approach of identifying a possible useful model from a general class of models. The model building strategy consists of three key stages namely: (a) model specification, (b) parameter estimation and (c) diagnostic checking. The model chosen would then be checked against the historical data to see whether it accurately describes or fits the series properly. Models presenting a good fit when the residuals between the forecasting model and the historical data are small, randomly distributed and independent. If the specified model is not satisfactory, the process is repeated until a satisfactory model (i.e. the best-fit model) is identified. Further details of the Box-Jenkins procedure can be found in standard statistical texts, such as Box and Jenkins (1994) or Hamburg and Young (1994).