COMPARISON OF THE GREY MODEL AND THE BOX JENKINS MODEL IN FORECASTING MANPOWER IN THE UK CONSTRUCTION INDUSTRY

Paul H. K. Ho[1]

Division of Building Science and Technology, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, PR China

Accurate forecasts of manpower in the construction industry are very important for the government, educational institutions and individual construction firms in their manpower planning. Although many good forecasting models have been developed, they require a sufficient amount of good-quality data to produce accurate results. Grey systems theory developed by Deng has become popular due to its ability to deal with systems in which some information is unknown. Grey models require only a limited amount of data to estimate the behaviour of unknown systems. This paper compares the grey model with the Box-Jenkins model in the forecasting of manpower in the UK construction industry. A GM (1, 1) grey model is proposed to forecast construction manpower one quarter ahead using manpower data from the Construction Statistics Annual published by the Office for National Statistics covering 72 quarters from 1991 Q1 to 2008 Q4. Within this period, two sets of manpower data were used: the total manpower available for employment (labour supply) and the total number of employees in employment (labour demand). An Excel programme was formulated to execute the forecasts using the grey model. An SPSS programme was used to conduct autoregressive integrated moving average (ARIMA) forecasts (Box-Jenkins model). The minimum mean absolute percentage error (MAPE) forecasted by the grey model for the total manpower and total employees time series was 1.52% and 2.14%, respectively, whereas the MAPE forecasted by the ARIMA model was 1.61% and 2.33%, respectively. Given the small forecasting error, it is concluded that both the GM (1, 1) model and the ARIMA model can accurately forecast manpower in the UK construction sector, but that the GM (1, 1) model performs slightly better than the ARIMA model.

Keywords: manpower forecast; Grey Model; Box-Jenkins Model.

INTRODUCTION

The construction industry is characterised by a workload that fluctuates cyclically. These cyclical fluctuations in workload in turn create fluctuations in labour supply and demand. Consequently, labour supply and demand in the construction sector are rarely in exact equilibrium: there is always either a labour shortage or a labour surplus. Labour shortages results in the hiring of unskilled labour, overtime work to maintain

[1] bshkho@cityu.edu.hk.

Ho P.H.K (2012) Comparison of the Grey Model and the Box Jenkins Model in forecasting manpower in the UK construction industry *In*: Smith, S.D (Ed) *Procs 28th Annual ARCOM Conference*, 3-5 September 2012, Edinburgh, UK, Association of Researchers in Construction Management, 369-.379

original programmes, and increased wages to maintain existing labour and attract new workers. In contrast, labour surplus results from the dismissal of skilled labour. The recruitment and training of skilled workers is costly (Uwakweh and Maloney, 1991). Accurate forecasting of the workforce is thus very important in the construction industry to maintain the correct balance between labour supply and demand.

Forecasting models for construction manpower can be classified into different types or levels: national aggregate manpower models, occupational manpower models, regional manpower models, regional manpower models by occupation (Briscoe and Wilson, 1993), company manpower models and project manpower models. Each of these models requires different input data and forecasting approaches. However, the transience of manpower, the reliance on self-employment and problems with estimating the informal labour market reduce the reliability of employment statistics in the construction sector (CLR-GB, 2008). As such, accurate manpower forecasting at any level is a challenging task for government, educational institutions and individual construction firms alike.

Many good statistical models have been developed for forecasting manpower. However, most of these models require the input of a sufficient amount of suitable data. If there is insufficient data or if the data is sufficient but does not follow certain patterns, then the forecasts may not be reliable or accurate. Given the potential limitations of existing statistical models, this study introduces a grey model for forecasting national aggregate labour supply and demand in the construction industry based on a limited amount of data. To demonstrate the accuracy of the grey model, its forecasting results are compared with those forecasts obtained with the Box-Jenkins statistical model.

PREVIOUS MANPOWER FORECASTING MODELS

A wide range of statistical models of varying degrees of complexity is available for forecasting manpower. These models can be grouped into five broad approaches: the exponential smoothing approach, single-question (linear or non-linear) regression approach, simultaneous-equation regression approach, autoregressive integrated moving average (ARIMA) approach and vector autoregression (VAR) approach (Gujarati and Porter, 2009). Many studies have attempted to use these models to forecast the aggregate labour supply and demand, some of which are reviewed here.

Based on the neoclassical labour economics theory, several employment functions have been developed to forecast labour demand in the short run. The optimal employment functions so derived usually incorporate a lagged adjustment mechanism, and the inputs can only be adjusted gradually towards their desired levels. One of the most seminal employment functions is that proposed by Ball and St. Cyr (1966):

$$\log E_t = a_0 - a_1 t + a_2 \log Q_t + a_3 \log E_{t-1,} \quad (1)$$

where E = the level of employment, t = the time trend and Q = the construction output. The coefficients α_0, α_1, α_2 and α_3 are estimated by the regression method, which captures explanatory rigour and uncovers the causal links amongst the variables. However, the inclusion of a lagged dependent variable means that the parameter estimates may be biased and the sampling variance underestimated. The regression model is also unstable over time.

Econometric models also derive the supply and demand employment functions from labour economics theory. By identifying appropriate reduced-form equations, the relevant coefficients are then estimated by either the ordinary least-squares (OLS) or two-stage least-squares (2SLS) methods. For instance, Black and Kelejian (1970) employed the 2SLS method to estimate a quarterly model of US labour demand and supply and wage adjustments. The parameters estimated from an econometric model are dependent on prevailing policy at the time of estimation, and will change if there is a policy change.

The ARIMA model (or the Box-Jenkins model) focuses on analysing the probabilistic properties of time series data (i.e. it lets the data speak for itself). Unlike single- and simultaneous-equation regression models, the ARIMA model is not derived from any economics theory. Wong *et al.* (2005) applied the model to forecast five construction manpower time series in Hong Kong: employment level, productivity, unemployment rate, underemployment rate and real wages. Their results showed the mean absolute percentage error (MAPE) in the predicted employment level, productivity, unemployment rate, underemployment rate and real wage values to be 12.8%, 6.6%, 7.4%, 7.5% and 2.2%, respectively. Depending on the patterns of the data, it is well recognised that the ARIMA model achieves a better forecasting performance than single- and simultaneous-equation regression models.

The VAR model resembles simultaneous-equation modelling in that several endogenous variables are considered together, but each is explained by its lagged values and the lagged values of all other endogenous variables in the model. By examining a wide range of variables with the cointegrating regression method, Briscoe and Wilson (1991, 1993) formulated the following long-run labour demand specification for the UK engineering sector.

$$\mathrm{E_t} = \mathrm{a} + \mathrm{b}(\mathrm{Q_t}) + \mathrm{c}(\mathrm{RW_t}) + \mathrm{d}(\mathrm{H_t}) + \mathrm{e}(\mathrm{ROP_t}) + \mathrm{f}(\mathrm{BR_t}), \quad (2)$$

where E = the employment level, Q = the construction output, RW = the real wages, H = the average hours worked, ROP = the real oil prices, BR = the bank rates and t = the time trend. Due to the small degree of freedom, the over-parameterized VAR model was then simplified into a restricted error correction (EC) form.

$$\triangle \mathrm{E_t} = \mathrm{a} + \mathrm{b_0}(\triangle \mathrm{Q_t}) + \mathrm{b_1}(\triangle \mathrm{Q_{t-1}}) + \mathrm{c_0}(\triangle \mathrm{RW_t}) + \mathrm{d_0}(\triangle \mathrm{H_t}) + \mathrm{e_1}(\triangle \mathrm{ROP_{t-1}}) + \mathrm{f_0}(\triangle \mathrm{BR_t}) + \mathrm{g_1}(\triangle \mathrm{E_{t-1}}) + \mathrm{g_2}(\triangle \mathrm{E_{t-2}}) + (\mathrm{EC_{t-1}}).. \quad (3)$$

Briscoe and Wilson found that the dominant variables were output, real wages and the lagged dependent variable.

Most statistical forecasting models require a large amount of data in order to achieve a reliable conclusion. For instance, Box and Jenkins (1976) stated that 50 to 100 observations are necessary to ensure adequate power for model testing, although other studies have shown that the minimum number of observations is more likely to be between 100 and 250 (Yaffe and McGee, 1982). The foregoing literature review clearly indicates the need for a manpower forecasting model that can produce accurate forecasts with a limited amount of input data.

GREY MODEL

Grey systems theory, originally developed by Deng (1982), is a generic theory that deals with problems with small samples or poor data. It looks for realistic patterns based on the modelling of a small amount of available data. Several types of grey models have been developed over the years. Probably, the most popular predicting model is the single-variable first-order grey model, abbreviated as GM (1, 1). The GM (1, 1) modelling algorithm can be summarised in the following steps, details of which can be referred to Liu and Lin's (2006) book.

Step 1: Assume that $x^{(0)}$ is the original raw data sequence with n samples:

$$x^0 = \left[x^{(0)}(1), x^{(0)}(2), \ldots\ldots, x^{(0)}(n)\right] = [x^0(k)]; \quad n \geq 4, \tag{4}$$

where the superscript (0) represents the original data and negative data values are not allowed. The task here is to forecast $x^{(0)}(k+1)$ (i.e. one sampling time ahead).

Step 2: In the grey system, the original sequence $x^{(0)}$ is transformed into a new sequence $x^{(1)}$ using a first-order accumulated generation operator (AGO). The main function of the AGO is to discover the potential regular pattern or trend of the original data sequence through the accumulation of data so that the prediction can be more accurate. The new sequence $x^{(1)}$ is the AGO sequence of $x^{(0)}$, which is defined as follows.

$$x^1 = \mathrm{AGO} \cdot x^0 = \left[\sum_{k=1}^{1} x^0(k), \sum_{k=1}^{2} x^0(k), \ldots\ldots, \sum_{k=1}^{n} x^0(k)\ \right] \tag{5}$$

where

$$x^1(1) = x^0(1)$$

$$x^1(2) = x^1(1) + x^0(2)$$

$$x^1(3) = x^1(2) + x^0(3)$$

…………………..........

$$x^1(k) = x^1(k-1) + x^0(k) \tag{6}$$

Step 3: The sequence $x^{(1)}$ is then modelled by a first-order differential equation (whitening equation) as follows.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b, \tag{7}$$

where $dx^{(1)}/dt$ is the derivative of the function x, x is the background value, and the parameters a and b are the development coefficient and grey input, respectively.

Step 4: In order to handle the discrete data sequence which is not continuous and differential, Equation (7) is generalized into the following grey differential equation.

$$x^{(0)}(2) + az^{(1)}(2) = b$$

$$x^{(0)}(3) + az^{(1)}(3) = b$$

………………………..

$$x^{(0)}(k) + az^{(1)}(k) = b, \tag{8}$$

where $z^{(1)}(k)$ is the generated sequence of the consecutive neighbours of $x^{(1)}$ given by

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)x^{(1)}(k-1); \ \ \alpha \in [0,1] \text{ and } k = 1, 2, \ldots\ldots, n, \tag{9}$$

where α is the generation coefficient, which depends on whether $z^{(1)}(k)$ places more emphasis on new or old data. When $\alpha > 0.5$, the generation of $z^{(1)}(k)$ is said to have emphasis more on new data, whereas when $\alpha < 0.5$, $z^{(1)}(k)$ is said to have emphasis more on old data.

Step 5: In order to provide a solution for Equation (8), the parameters a and b can be solved by the traditional least-squares error method as follows.

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}^T = (B^T B)^{-1} B^T Y, \tag{10}$$

where

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ . \\ . \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \\ . & \\ . & \\ -z^{(1)}(n) & 1 \end{bmatrix}, \hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}. \quad (11)$$

Step 6: Setting the initial value $\hat{x}^{(1)}(1) = x^{(0)}(1)$ and t = 1, the solution to Equation (7) is given approximately by

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} + \frac{b}{a}; \quad k = 1, 2, 3, \dots, n, \quad (12)$$

where $\hat{x}^{(1)}(k+1)$ is the predicted value of $x^{(1)}(k+1)$ at time (k+1) and $\widehat{\ }$ denotes the forecasted value.

Step 7: By using the inverse accumulated generation operator (IAGO) which is an inverse operator of the accumulated generation, the forecasted value $\hat{x}^{(0)}(k+1)$) can be obtained by:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \quad (13)$$

Step 8: The metabolic GM (1, 1) model is used by inserting $x^{(0)}(n+1)$ and deleting $x^{(0)}(1)$ in the sequence $x^{(0)} = [x^{(0)}(2), \ldots, x^{(0)}(n), x^{(0)}(n+1)]$ because the value of new data is greater than that of old data. As the system develops, older data are deleted and newer data added so that the modelling sequence is constantly renewed to reflect the latest characteristics of the system.

Step 9: The overall accuracy of this forecasting model can be measured by its means absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n} \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} 100\%; \quad k = 1, 2, \dots, n, \quad (14)$$

where $x^{(0)}(k)$ and $\hat{x}^{(0)}(k)$ are the actual and forecasted values, respectively, and n is the number of forecasts.

RESEARCH METHODOLOGY

Based on a limited amount of data, the metabolic GM (1, 1) model is used to forecast the aggregate labour supply and demand in the UK construction sector. The Box-Jenkins (ARIMA) model is used to perform the same forecasts based on the same datasets. The results forecasted by the two approaches are then compared based on the means absolute percentage error (MAPE). Amongst several statistical models, the ARIMA model is chosen for the comparison purpose for two reasons. First, in many cases, the ARIMA model produces more reliable and accurate forecasts than the exponential smoothing model, single-question regression model and simultaneous-equation regression model. Second, same as the grey model, the ARIMA model does not rely on the input of other external data so that these two approaches can be directly compared on the same base.

The manpower data is based on the Construction Statistics Annual published by the Office for National Statistics. It contains a total of 72 quarterly datasets, covering the first quarter of 1991 to the fourth quarter of 2008. Within this period, this study examines two sets of data: the 'total manpower' time series and the 'total employees' time series. The manpower time series represents the total amount of labour available for employment (i.e. the labour supply), whereas the employees time series represents the total amount of labour in employment (i.e. the labour demand). All of the data were adjusted to reduce seasonal fluctuations. Figures were occasionally revised due to differences in the definitions and coverage from the original series. This created certain fluctuations in the time series that would affect the accuracy of the forecasts.

EMPIRICAL RESULTS AND DISCUSSION

GM (1, 1) Model

One of the major advantages of the GM (1, 1) model is that it only requires the input of a few data for model building. The sample number determines the amount of memory in the model. Generally, when a series is very random, a larger sample number produces a lower MAPE. In contrast, if the series moves smoothly up and down, then a smaller number produces a lower MAPE. The optimal sample number is that which produces the lowest MAPE. To achieve the lowest MAPE, a search method was employed to find the optimal sample size. Different sample sizes were used to build the models, forecast the number of workers one quarter ahead, compare the differences between the actual and forecasted numbers and finally calculate the absolute percentage error of each forecast and the overall MAPE. Table 1 presents the MAPE of forecasts based on sample sizes of 4, 5, 6, 7, 8 and 9.

Table 1: MAPE of forecasts based on different sample sizes

Time series	MAPE of forecasts based on					
	4 data input	5 data input	6 data input	7 data input	8 data input	9 data input
Total manpower	1.95%	1.74%	1.66%	1.75%	1.81%	1.98%
Total employees	2.44%	2.47%	2.64%	2.89%	2.95%	3.11%

As can be seen in Table 1, the results demonstrate that the grey model requires few input data and indeed, a larger amount of input data does not produce more accurate results. Based on the minimum MAPE, the optimal sample sizes for the total manpower and total employees time series are 6 (1.66%) and 4 (2.44%), respectively.

The coefficient α in Equation (9) depends on the relative reliance on new versus old date in the accumulated generation process. Again, the optimal coefficient is that which produces the lowest MAPE. Based on the identified optimal sample sizes, a search method was used to find the optimal coefficient. Table 2 presents the MAPE of forecasts based on α values ranging from 0.50 to 1.00.

Table 2: MAPE of forecasts based on different α values

Time series	MAPE of forecasts based on					
	α=0.50	α=0.60	α=0.70	α=0.80	α=0.90	α=1.00
Total manpower (6-data solution)	1.66%	1.62%	1.59%	1.56%	1.54%	1.52%
Total employees (4-data solution)	2.44%	2.35%	2.27%	2.21%	2.17%	2.14%

As shown in Table 2, using a larger α value reduces the MAPE of forecasts in both time series. This indicates that the value of new data is greater than that of old data in the forecast. The optimal coefficients for the total manpower and total employees time series are both equal to 1.

Based on the optimal sample size of 6 data for the total manpower time series and the optimal α value of 1, the first six actual values were used to build the GM (1, 1) model and to forecast the seventh value (i.e. one quarter ahead) according to Equations (4) to (13). The seventh forecasted value was compared with its actual value. The absolute percentage error was then calculated according to Equation (14). The same calculation procedure was repeated until the end of the last set of data. As each forecast required a large number of calculations, an Excel program incorporating Equations (4) to (14) was used to manipulate the calculations. The dataset comprises 72 data. After deducting the first 6 data for model building, the remaining 66 data were used to evaluate the accuracy of the ex-post forecasts. The same forecasting method was used with the total employees time series to produce similar forecasts. Table 3 shows a comparison of the actual and forecasted values for one quarter ahead, together with its absolute percentage error, for the total manpower and total employees time series.

Table 3 shows that the MAPE of the overall forecast for the total manpower and total employees time series is only 1.52% and 2.14%, respectively. The respective maximum absolute percentage error (MaxAPE) is 7.19% at 2000 Q1 and 6.98% at 2002 Q1. These relatively larger errors are due to irregular movements of the time series at those periods. With few exceptions, the error for most forecasts is very small. Therefore, the actual and forecasted values are in a very good agreement as shown in Figure 1.

Table 3: Actual and forecasted values for one quarter ahead from the grey model

Year	Quarter	Total manpower time series			Total employees time series		
		Original value	Forecasted value	Absolute % error	Original value	Forecasted value	Absolute % error
1991	Q1	1 777	-	-	1 094	-	-
	Q2	1 723	-	-	1 057	-	-
	Q3	1 667	-	-	1 019	-	-
	Q4	1 626	-	-	994	-	-
1992	Q1	1 579	-	-	962	975	1.37%
	Q2	1 533	-	-	932	948	1.75%
	Q3	1 494	1 508	0.93%	909	915	0.69%
	Q4	1 475	1 470	0.37%	890	894	0.44%
1993	Q1	1 432	1 444	0.83%	873	878	0.62%
	Q2	1 409	1 414	0.36%	852	863	1.33%
	Q3	1 403	1 391	0.87%	831	843	1.42%
	Q4	1 398	1 380	1.27%	803	820	2.12%
1994	Q1	1 394	1 377	1.20%	795	791	0.46%
	Q2	1 379	1 385	0.46%	784	782	0.23%
	Q3	1 390	1 379	0.77%	778	780	0.21%
	Q4	1 375	1 382	0.48%	766	773	0.89%
1995	Q1	1 394	1 375	1.38%	773	762	1.37%
	Q2	1 362	1 385	1.72%	749	769	2.62%
	Q3	1 362	1 373	0.78%	743	750	0.94%
	Q4	1 382	1 359	1.64%	750	732	2.38%
1996	Q1	1 392	1 371	1.54%	758	748	1.31%
	Q2	1 354	1 383	2.11%	741	762	2.77%
	Q3	1 355	1 374	1.40%	741	743	0.27%
	Q4	1 378	1 359	1.40%	739	734	0.70%
1997	Q1	1 399	1 361	2.72%	767	739	3.67%
	Q2	1 361	1 385	1.79%	754	768	1.92%
	Q3	1 384	1 384	0.01%	806	765	5.14%
	Q4	1 392	1 386	0.45%	839	805	4.03%
1998	Q1	1 447	1 386	4.21%	901	861	4.45%
	Q2	1 426	1 429	0.20%	906	918	1.28%
	Q3	1 419	1 450	2.19%	921	931	1.05%
	Q4	1 418	1 439	1.52%	912	924	1.34%
1999	Q1	1 429	1 426	0.18%	933	918	1.66%
	Q2	1 409	1 417	0.55%	899	931	3.57%
	Q3	1 418	1 414	0.27%	897	905	0.91%
	Q4	1 403	1 416	0.92%	885	882	0.31%
2000	Q1	1 514	1 405	7.19%*	930	883	5.04%
	Q2	1 491	1 477	0.96%	963	929	3.54%
	Q3	1 493	1 512	1.29%	976	983	0.69%
	Q4	1 535	1 523	0.78%	983	990	0.73%
2001	Q1	1 518	1 548	1.95%	974	989	1.53%
	Q2	1 523	1 523	0.02%	970	976	0.64%
	Q3	1 546	1 534	0.76%	946	966	2.10%
	Q4	1 557	1 546	0.67%	968	942	2.66%
2002	Q1	1 594	1 554	2.52%	1 032	960	6.98%*
	Q2	1 629	1 594	2.14%	1 048	1 045	0.27%
	Q3	1 631	1 635	0.23%	1 036	1 074	3.68%
	Q4	1 599	1 652	3.29%	994	1 042	4.80%
2003	Q1	1 619	1 632	0.81%	994	985	0.89%
	Q2	1 559	1 619	3.88%	985	976	0.91%
	Q3	1 590	1 570	1.28%	988	984	0.38%
	Q4	1 613	1 569	2.72%	1 002	984	1.75%
2004	Q1	1 659	1 596	3.80%	1 034	1 004	2.86%
	Q2	1 682	1 642	2.37%	1 049	1 042	0.66%
	Q3	1 762	1 699	3.58%	1 080	1 063	1.57%
	Q4	1 754	1 764	0.59%	1 062	1 088	2.49%
2005	Q1	1 760	1 789	1.67%	1 067	1 073	0.60%

	Q2	1 811	1 791	1.09%	1 102	1 060	3.82%
	Q3	1 830	1 818	0.68%	1 113	1 107	0.55%
	Q4	1 799	1 832	1.83%	1 089	1 128	3.58%
2006	Q1	1 827	1 831	0.20%	1 131	1 092	3.48%
	Q2	1 815	1 836	1.15%	1 103	1 125	1.98%
	Q3	1 836	1 818	1.00%	1 120	1 118	0.15%
	Q4	1 866	1 828	2.01%	1 168	1 110	4.98%
2007	Q1	1 810	1 864	3.00%	1 095	1 178	7.62%
	Q2	1 857	1 835	1.18%	1 123	1 110	1.16%
	Q3	1 914	1 851	3.27%	1 154	1 095	5.15%
	Q4	1 859	1 894	1.87%	1 096	1 168	6.53%
			MAPE	**1.52%**		**MAPE**	**2.14%**

* *Maximum absolute percentage error.*

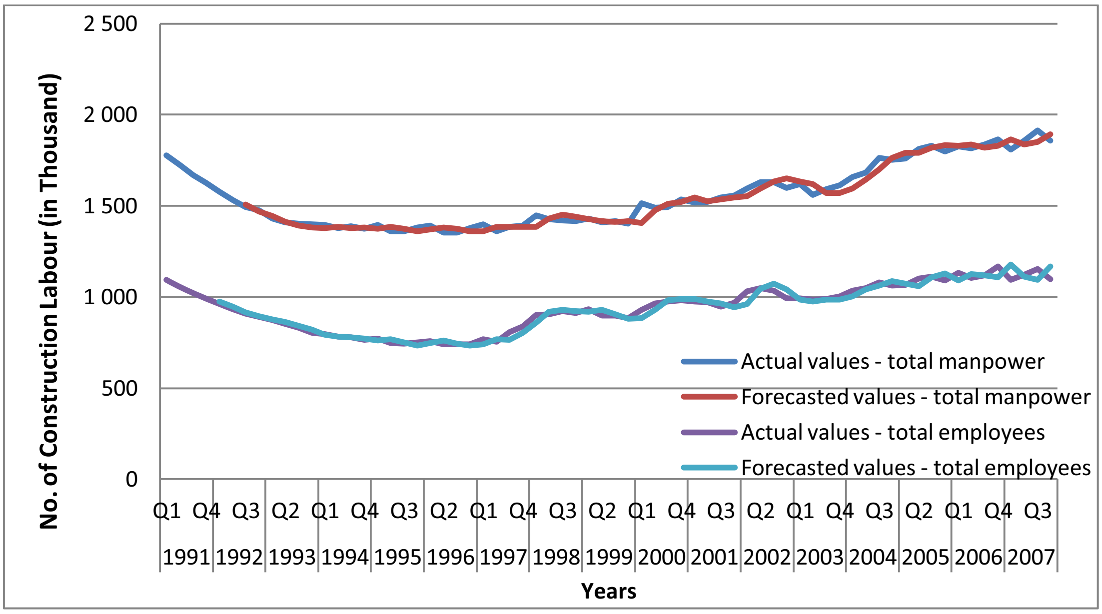

Figure 1: Actual and forecasted values for the total manpower and total employees time series from the grey model

ARIMA Model

The ARIMA approach consists of four steps: identification of the model (tentative choice of p, d, q), parameter estimation for the chosen model, diagnostic checking and forecasting, details of which can be found in Box-Jenkins' seminal book 'Time Series Analysis: Forecasting and Control'. The SPSS programme was used to find the fitted values based on the same data as that used in the grey model. The best models for the total manpower and total employees time series were found to be ARIMA (0, 1, 3) (0, 1, 1) and ARIMA (0, 1, 0) (0, 0, 0), respectively. The model parameters are shown in Table 4.

Table 4: ARIMA Model Parameters

Time series			Estimate	SE	t	Sig.
Total manpower	Difference		1			
	MA	Lag 3	0.490	0.119	4.116	.000
	Seasonal Difference		1			
	MA, Seasonal	Lag 1	0.648	0.120	5.386	.000
Total employees	Constant		0.030	3.409	0.009	.993
	Difference		1			

Table 5: Model Fit Statistics

Model	Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE
Total manpower	0.449	0.967	31.233	1.613	5.678
Total employees	-2.895E-16	0.951	27.907	2.331	6.878

According to Table 5, the MAPE of the overall forecast for the total manpower and total employees time series is 1.61% and 2.33%, respectively, whereas the respective MaxAPE is 5.68% at 2000 Q1, and 6.88% at 1998 Q1. As with the grey model, the actual and forecasted values for both time series are in very good agreement.

Comparison of the GM (1, 1) Model and ARIMA Model

Table 6 shows a comparison of the forecasting accuracy of the GM (1, 1) model and ARIMA model. The MAPE values forecasted by both models are between 1.52% and 2.33%, which is very small. Hence, both the GM (1, 1) model and the ARIMA model can be considered to be able to accurately forecast labour supply and demand in the UK construction sector.

Table 6: Comparison of the forecasting accuracy of the GM (1, 1) model and ARIMA model

Time series	MAPE of forecasts based on GM (1,1) model	MAPE of forecasts based on ARIMA model	Difference
Total manpower	1.5%	1.6%	0.1%
Total employees	2.1%	2.3%	0.2%

It is also observed from Table 6 that the GM (1, 1) model performs slightly better than the ARIMA model, although the differences are only 0.1% and 0.2% for the total manpower and total employee time series, respectively. As the ARIMA model is well recognised as one of the most reliable forecasting approaches, the results demonstrate the reliability and robustness of the grey model in forecasting short-term construction manpower.

CONCLUSIONS

As shown in Figure 1, labour supply in the construction industry has always been greater than labour demand over the past 20 years. In other words, the UK construction industry is demand led rather than supply led. Hence, it is more important to focus efforts on the forecasting of labour demand rather than labour supply.

Contrary to the general belief that it is difficult to make an accurate labour forecast, the above results empirically demonstrate that both the GM (1, 1) model and the ARIMA model can accurately forecast labour supply and demand in the UK construction sector. Besides the robustness of the models, one of the main reasons is that the data was fairly stable over the period of analysis as indicated in Figure 1. As the overall MAPE values are only around 2%, there should be little concern about the forecasting method itself.

Based on sample sizes of only 6 and 4 data for the total manpower and total employees time series respectively, the GM (1, 1) model produced a slightly more accurate forecast then the ARIMA model. This demonstrates that the GM (1, 1) model can handle a time series with a small sample size. The model may thus be equally, or even more, suitable for forecasting other detailed manpower models such as the occupational manpower model, regional manpower model and regional manpower by occupation model. This would be a promising area for further research.

REFERENCES

Ball, J. and St Cyr, E. (1966) Short-term employment functions in British manufacturing. “Review of Economic Studies”, **33**, 179-207.

Black, S.W. and Kelejian, H.H. (1970) A macro model of the US labour market. “Econometrica”, **38**(5), 712-741.

Box, G.E.P. and Jenkins, G.M. (1976) “Time series analysis: forecasting and control”. Holden Day, Oakland, CA.

Briscoe, G. and Wilson, R. (1991) Explanations of the demand for labour in the United Kingdom engineering sector. “Applied Economics”, **23**, 913-926.

Briscoe, G. and Wilson, R. (1993) “Employment forecasting in the construction industry”. Avebury Publishing Co., Hants, England.

CLR-GB (2008) “Newsletter”, European Institute for Construction Labour Research-GB Office, 2/2008.

Deng, J.L. (1982). Control problems of grey systems. “Systems and Control Letters”, **1**(5), 288-294.

Gujarati, D.N. and Porter, D.C. (2009) “Basic econometrics”. McGraw Hall, New York.

Liu, S. and Lin, Y. (2006) “Grey information: theory and practical applications”. Springer-Verlag, London.

Uwakweh, B.O. and Maloney, W. (1991) Conceptual model for manpower planning for the construction industry in developing countries. “Construction Management and Economics”, **9**, 451-465.

Wong, J.M.W., Chan, A.P.C. and Chiang, Y.H. (2005) Time series forecasts of the construction labour market in Hong Kong: the Box-Jenkins approach. “Construction Management and Economics”, **23**, 979-991.

Yaffe, R.A. and McGee, M. (2000) “Introduction to time series analysis and forecasting”. Academic Press, Inc., San Diego, California.