Predicting the Talents Needed in Coal Industry in China

Xin Zhang
Austin Peay State University

Kai Tan
China University of Mining and Technology

Yingtao Shen
Austin Peay State University

China is the world's second-largest energy consuming country and the largest producer and consumer of coal. While coal production plays an important role in the Chinese economy, its coal industry lacks sufficient talents. This paper predicts the number of talents needed in the coal industry in China using multiple linear regression method and grey forecasting model. We find that the method makes reliable medium- and long-term prediction of coal talents needed in China.

INTRODUCTION

China's coal industry plays a vital role in its economy. Currently, coal accounts for approximately 70% of China's energy consumption and is expected to remain the largest source of energy for a very long period (Xu, et.al., 2015). After several decades of development, the coal industry in China has expanded its production capacity significantly. However, many problems remain, including low coal production concentration, inefficient product structure, low mining mechanization degree, low per capita efficiency, pollution of the ecological environment, insufficient safety practices and so on. These problems have hindered the further development of the coal industry in China. In addition to increasing the investment in innovation and adaptation of new technology, one key to solving these problems is to recruit more qualified personnel into the coal industry.

The development of any industry requires skilled personnel or talents. Talent resources are the most important resources and have the highest value among all resources in most industries (Li, 2006). Effective utilization of talent resources has been proven to be beneficial to the development and utilization of all other resources in an enterprise and lead to higher yield and continuous improvements.

Recently, lack of qualified talents has become a particularly predominant problem facing the Chinese coal industry (Guo, et.al., 2011). At the end of 2015, the total number of employees in coal mining and washing industry in China was about 4.424 million with trained and skilled personnel accounting for one-fourth of the total (China Coal Market Network, 2016). Although the coal industry in China has developed a sizeable talent base, both the quality and quantity of its personnel still need to be improved to achieve long-term, sustainable development.

There are different classifications of talents. Based on the characteristics of the coal industry in China, we define coal industry professional as a person who possesses certain professional knowledge or skills in the field coal production and utilization. Furthermore, we classify the professionals into three categories: management talents, professional technicians, and high-skilled talents. Management talents are those people who can control or administer all or part of an organization in coal system. Professional technicians engage in scientific research, development, innovation, or other sophisticated technical activities. High-skilled talent refers to those first-line workers working in coal production with special skills and knowledge. Reliable prediction of the quantity of talents needed in all the three categories is important to the mid-and long-term development of the Chinese coal industry which needs to meet the growing demand for coal.

We develop a quantitative method using multiple linear regression model and grey forecasting model to predict the number of coal talents needed in each category in China. This method is used to predict the total number of talents needed for each classification from the year 2015 to 2024 in the coal industry, using data collected from 30 state-owned, large- and medium-sized coal enterprises in China.

MODELS

Multiple Linear Regression Model

Regression analysis is a form of predictive modeling technique which investigates the relationships among variables. In regression terminology, the variable being predicted is called the dependent variable. The variable or variables being used to predict the value of the dependent variable are called the independent variables. When the regression analysis involves one dependent variable and two or more independent variables, it is called multiple linear regression (Freedman, 2009).

Suppose we have one dependent variable (y) and k independent variables (x_1, x_2, …, x_k). The multiple linear regression model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon, \tag{1}$$

where β_0, β_1, β_2, …, β_k are parameters, and the error term ε is a random variable that has a normal distribution with mean 0 and a constant variance. The values of β_0, β_1, β_2, …, β_k are unknown and must be estimated from sample data. A simple random sample is used to compute sample statistics b_0, b_1, …, b_k that are used as the point estimate of the parameters β_0, β_1, β_2, …, β_k. These sample statistics provide the following estimated multiple regression equation:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k, \tag{2}$$

where b_0, b_1, b_2, …, b_k are the estimates of β_0, β_1, β_2, …, β_k.

Let us suppose we have a random sample of size n selected from the population studied. Then, the multiple linear regression equation for all observations of the sample can be written as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1,\ldots,n. \tag{3}$$

The estimated multiple linear regression equation can be written as:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki}, \quad i = 1,\ldots,n. \tag{4}$$

The least squares method can be used to find the estimated multiple linear regression equation. It uses the sample data to provide the values of b_0, b_1, b_2, …, b_k that minimize the sum of squares of the deviation between the observed values of the dependent variable y_i and the predicted values of the dependent variable $\hat{y}_i$. The criterion for the least squares method is given by the following expression:

$$Min \quad \sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{5}$$

Using the first derivative from $\sum_{i=1}^{n}(y_i - \hat{y_i})^2$, we obtain $b = (X'X)^{-1}X'Y$, where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \text{ and } b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$ assuming the inverse $(X'X)^{-1}$ exists.

The linear regression method has been used extensively in practical applications because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and the statistical properties of the resulting estimators are easier to determine (Yan, 2009). One of the important applications of linear regression is to fit a predictive model to an observed data set of *Y* and *X* values. After developing such a model, if an additional value of *X* is given without its accompanying value of *Y*, the fitted model can be used to predict the value of *Y*.

Grey Forecasting Model

Grey system theory, first introduced in early 1980 (Deng, 1982), focuses on the study of problems involving small samples and incomplete information. It takes uncertain systems with partially known information as the research project, and provides a powerful technical support for forecasting. Since in the natural world, small sample and incomplete information exist commonly, grey system theory have been successfully employed in various fields, such as predicting stock price (Wang, 2002), predicting electricity demand (Zhou, et.al., 2006), forecasting financial crises for an enterprise (Chen, et.al., 2011), and forecasting annual net income of rural household in China (Zhao, et.al., 2012), etc.

Grey system theory has been developed quickly in recent years to incorporate theories in other fields of algorithm in order to enhance forecasting accuracy. For example, Li, Yamaguchi, and Nagai (Li, et.al., 2007) combined GM(1, 1) and Markov chain to predict the number of Chinese international airlines. Hsu and Wang (Hsu, et.al., 2007) used a grey model improved by the Bayesian analysis to forecast the output of integrated circuit industry. Wang and Hsu (Wang, et.al., 2008) combined genetic algorithms and grey theory to forecast high technology industrial output. Chen, Hsin, and Wu (Chen, et.al., 2010) used grey-Bernoulli model to forecast Taiwan's major stock indices.

GM (1, 1) Model

GM(1, 1) model is the main model of grey system theory and is one of the most widely used technique in the grey system. It is a single variable first order grey model, which is created with few observations (four or more) and still provides high precision results. GM(1, 1) model is fit for nonnegative raw data, which accords with exponential form and does not have quick growth rate.

In grey theory, the accumulated Generating Operation (AGO) technique is applied to reduce the randomness of the raw data. Let $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\}$ be a raw sequence, and $x^{(0)}(k) \geq 0$, $k = 1, \ldots, n$, $x^{(0)}(k)$ is the value at time k. Set $x^{(1)}(1) = x^{(0)}(1)$, by applying first-order AGO (1-AGO) on $x^{(0)}$, we have $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \ldots, x^{(1)}(n)\}$, where $x^{(1)}(k) = \sum_{i=1}^{K} x^{(0)}(i), \quad k = 1,2,\ldots,n.$

Then the following equation;

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 1,2,\ldots,n,\ldots \tag{6}$$

is a grey differential model, called GM(1, 1) as it only includes one variable $x^{(0)}$, where

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), \quad k = 1,2,\ldots,n,\ldots \tag{7}$$

a, b, are the coefficients in GM(1, 1) model. a is said to be a developing coefficient and b is the grey input. When a=0, the prediction equation fails by existence of indeterminate form of zero times infinity. $x^{(0)}(k)$ is a grey derivative which maximizes the information density for a given series.

The ordinary least squares method is used to estimate the model parameters a and b and we obtain

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B'B)^{-1}B'Y \,, \tag{8}$$

where B and Y are defined as

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}. \tag{9}$$

The model parameters a and b can also be expressed by the following parametric forms:

$$a = \frac{CD-(n-1)E}{(n-1)F-C^2}, = \frac{DF-CE}{(n-1)F-C^2}\,, \tag{10}$$

where C, D, E, and F are given by

$$C = \sum_{K=2}^{n} z^{(1)}(k), \tag{11}$$

$$D = \sum_{K=2}^{n} x^{(0)}(k), \tag{12}$$

$$E = \sum_{K=2}^{n} z^{(1)}(k)\, x^{(0)}(k), \tag{13}$$

$$F = \sum_{K=2}^{n} \left[z^{(1)}(k)\right]^2. \tag{14}$$

The whitenization equation $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ of GM(1, 1) model is solved to obtain

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}\;, \tag{15}$$
$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \tag{16}$$

Accuracy Measure of GM (1, 1) Model

GM (1, 1) model in the grey system theory is a prediction model, and thus, we need to examine how good the model is in forecasting. The forecast accuracy of a prediction model can be assessed using forecasting error, which is the difference between the forecast value and the actual value. There are different error measurement statistics. Percentage errors have the advantage of being scale-independent and easy to interpret. Therefore, we choose to use the Absolute Relative Percentage Error (ARPE) to assess the forecast error.

Relative Percentage Error (*RPE*) compares the real and forecast values at specific time k. Suppose $x^{(0)}(k)$ is the actual value at time k and $\hat{x}^{(0)}(k)$ is the forecast value at time k. *RPE* is then defined as

$$RPE = e(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, \qquad k = 1,2,\ldots,n. \tag{17}$$

The overall model forecast error for a total number of n periods can be measured by averaging the absolute value of RPE as follows:

$$ARPE = e(avg) = \frac{1}{n}\sum_{k=1}^{n}|e(k)|. \quad (18)$$

Then, the accuracy of GM (1, 1) model, as defined by P^0 , is measured using the formula below:

$$P^0 = (1 - ARPE) \times 100\%. \quad (19)$$

Larger P^0, better the model. The general requirement is $P^0 > 80\%$.

Data and Methodology

We use multiple linear regression method and grey prediction model to predict the total numbers of management talents, professional technicians, and high-skilled talents needed nationwide, respectively, in the coal industry of China from 2015 to 2024.

We collected the following information from the China Statistical Yearbook and the China Coal Industry Yearbook on a national level from year 2010 to 2014: annual coal production, annual operating revenue, annual wage of employees, overall labor productivity, and the total number of employees. We also surveyed 30 major state-owned large- and medium-sized coal enterprises in China and gathered information for the abovementioned five variables respectively for the same period. In addition, we collected data on the total numbers of management personnel, professional and technical personnel, and high-skilled personnel for these 30 enterprises.

We define the management talents needed in China as the group we are interested in studying, herein termed population. Thus, the 30 enterprises selected can be treated as a sample drawn from this population. First, using the sample data collected from 30 enterprises, we apply factor analysis to examine the relation between the number of management talents needed in the coal industry in China and its major indicators. Secondly, using those major indicators identified from the factor analysis as the independent variables and the total number of management talents needed as the dependent variable, we derive the estimated multiple linear regression equations for each year from 2010 to 2014 and calculate the predicted values for the number of management talents needed nationwide. These predicted values are the point estimates of the population values. Next, we apply Grey forecasting model (GM(1, 1)) to the predicted values obtained from multiple linear regression method and calculate the forecast values for the total number of management talents needed nationwide for the same period. We measure the accuracy of grey forecasting model using P^0 as the indicator. If grey prediction is proved efficient, this method will be applied to the prediction of the number of management talents needed in China from 2015 to 2024. The same process is also applied to predict the numbers of professional technicians and high-skilled talents needed in China coal industry.

APPLICATION

The methodology described above is applied to the annual national data collected from the China Statistical Yearbook and the China Coal Industry Yearbook, as well as annual data from the selected major state-owned large- and medium-sized coal enterprises in China. We begin by checking how suited our data is for factor analysis.

Factor analysis, first introduced by Charles Spearman in 1904 (Spearman, 1904 & 1927), is a statistical tool for explaining the structure of data by explaining the correlations between variables. It starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor. The two indicators we use to check whether our data suit factor analysis are Kaiser-Meyer-Olkin (KMO) index and Bartlett's sphericity test.

KMO measures sampling adequacy for each variable in the model and for the complete model. The statistic indicates the proportion of variance in the variables that might be caused by underlying factors. High values (close to 1.0) generally indicate that a factor analysis is useful with the data. If the value is less than 0.50, the results of the factor analysis probably will not be very useful. Bartlett's test of sphericity tests the hypothesis that the correlation matrix is an identity matrix, which would indicate that the variables are unrelated and therefore unsuitable for structure detection. Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with the data.

We apply KMO test and Bartlett's sphericity test to data collected from 30 major coal enterprises in China in 2010 and obtain the KMO index of 0.652 and the significance level from Bartlett's test of 0.000. The results show that our data are suitable for factor analysis.

We define k as the number of variables in the model and m as the number of factors needed. We desire a value for m such that $k(m+1) \leq k(k+1)/2$, i.e. $m \leq (k-1)/2$. Since we have 5 variables (k=5), we are looking for $m \leq (5-1)/2 = 2$. Our preference is to use no more than 2 factors if possible.

Figure 1 shows the scree plot of the factor analysis. Scree plot is a graph of the eigenvalues (showing on y-axis) of all the factors (showing on x-axis) where the factors are listed in decreasing order of their eigenvalues. We can observe that there is one inflection point (--component 3) in the plot. Two components to the left of component 3 have eigenvalues greater than 1 and the rest of the three components have eigenvalues less than 0.5. Figure 2 shows the proportion of variance explained by each component in factor analysis. It is clearly to see that component 1 and component 2 explained the majority of variance in the data. Thus, we choose to keep the factors corresponding to eigenvalues to the left of eigenvalue 3, i.e. the two largest eigenvalues. These two eigenvalues account for 87.29% of the total variance.

FIGURE 1
SCREE PLOT OF COMPONENTS

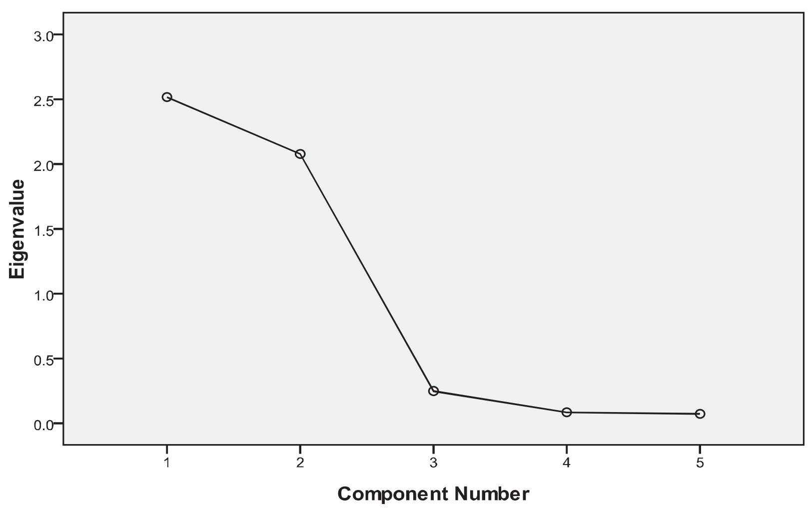


FIGURE 2
PROPORTION OF VARIANCE EXPLAINED BY EACH COMPONENT

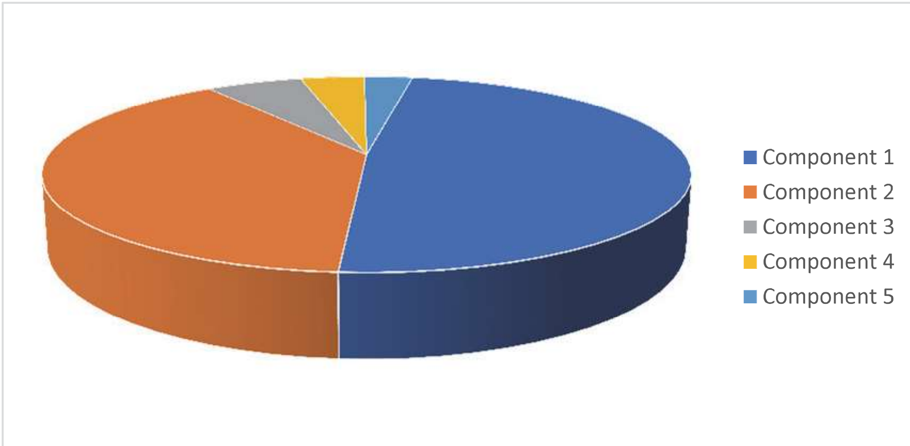


Table 1 shows the loading factors from Figure 1 and Figure 2 restricted to only the two highest common factors. The values we consider large are in boldface, using 0.6 as the cutoff. Factor 1 is correlated most strongly with coal production (0.942) and followed by operating revenue and total number of employees. We choose coal production to represent this group. Factor 2 is primarily related to wage level (0.959) and labor productivity (0.918) and we choose to use wage level to represent this group.

TABLE 1
LOADING FACTORS AND COMMUNALITIES FOR TWO FACTORS

	Factor	
Variable	1	2
Annual coal production	**0.942**	-0.057
Annual operating revenue	**0.935**	-0.073
Overall labor productivity	-0.213	**0.918**
Annual wage per employees	-0.063	**0.959**
Total number of employees	**0.818**	-0.341

Using annual coal production and annual wage per employee as the independent variables, we apply multiple linear regression method to predict the total number of management talents needed for national coal companies from 2010 to 2014, respectively. The sample estimated coefficients, standard errors, test statistics, and the corresponding p-values are reported in Table 2.

TABLE 2
PARTIAL REGRESSION INFORMATION FOR 2010-2014

Year		Coefficients	Standard Error	*t*-stat	*p*-value
2010	Intercept	10377.020	2964.694	3.503	0.002
	coal production	2.356	0.492	4.786	0.000
	wage level	-0.435	0.106	-4.100	0.000
2011	Intercept	12481.351	3064.331	4.073	0.000
	coal production	2.427	0.475	5.107	0.000
	wage level	-0.448	0.097	-4.605	0.000
2012	Intercept	12493.340	3083.470	4.051	0.000
	coal production	2.267	0.444	5.108	0.000
	wage level	-0.388	0.084	-4.618	0.000
2013	Intercept	14716.979	3702.216	3.975	0.001
	coal production	2.149	0.464	4.636	0.000
	wage level	-0.379	0.087	-4.354	0.000
2014	Intercept	18900.123	4255.003	4.442	0.000
	coal production	1.966	0.463	4.245	0.000
	wage level	-0.429	0.099	-4.325	0.000

Set the level of significance to 0.05. We observe that the *p*-values for all coefficients are smaller than it for each year, which indicates that the two independent variables coal production and wage level are both significant in predicting the value of total number of management personnel.

Figure 3 shows the histogram of standardized residuals of the regression equation for 2010. The histogram of the residual can be used to check whether the residuals (or error terms) are normally distributed. The overall distribution of residuals is bell-shaped which indicates that the normality assumption of residuals is likely to be true. In addition, Figure 4 shows the normal probability plot of residuals for 2010. The resulting plot is approximately linear, which supports the condition that the error terms are normally distributed. Thus, we may conclude that the underlying assumptions of the regression model is not violated. The histograms and normal probability plots of residuals for 2011-2014 are very similar to those in 2010.

FIGURE 3
HISTOGRAM OF RESIDUALS

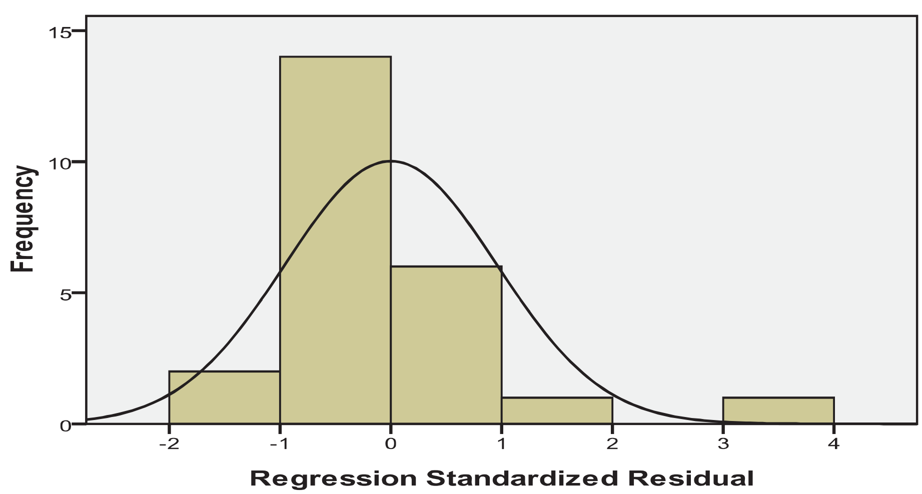
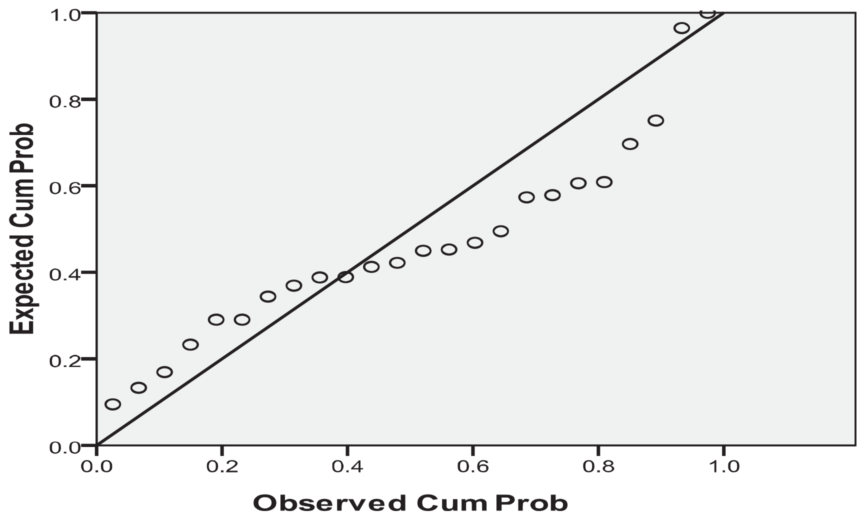


FIGURE 4
NORMAL PROBABILITY PLOT OF RESIDUALS



Using the coefficients reported in Table 2, the estimated regression equations for 2010-2014 are listed in Table 3 and the estimated values for the number of management talents needed for national coal companies are computed and reported in column 2 of Table 4. These estimated values (computed using sample data) can be considered as the point estimates of the number of management personnel needed for national coal companies.

TABLE 3
ESTIMATED REGRESSION EQUATION FROM 2010 TO 2014

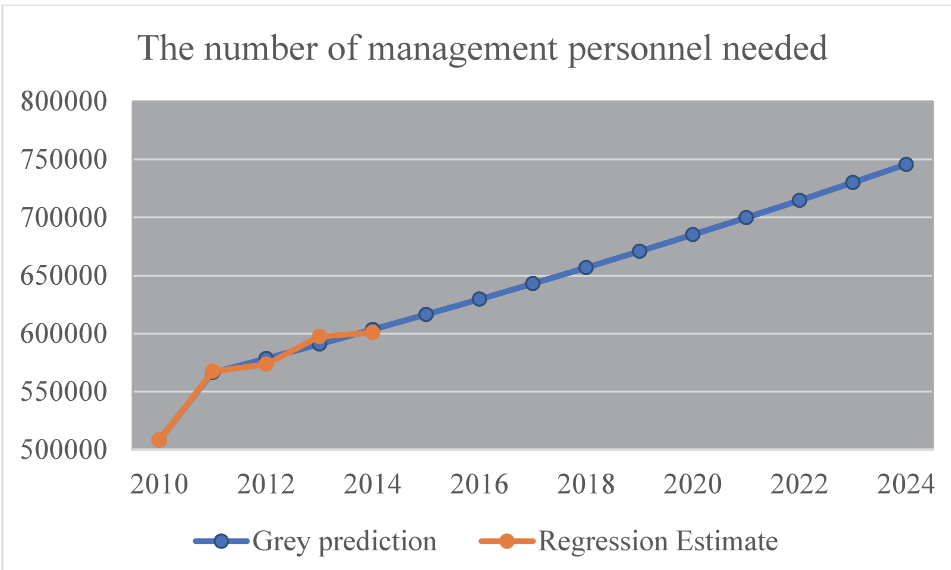
Year	Estimated Regression Equation
2010	$\hat{y} = 10377.020 + 2.356x_1 - 0.435x_2$
2011	$\hat{y} = 12481.351 + 2.427x_1 - 0.448x_2$
2012	$\hat{y} = 12493.340 + 2.267x_1 - 0.388x_2$
2013	$\hat{y} = 14716.979 + 2.149x_1 - 0.379x_2$
2014	$\hat{y} = 18900.123 + 1.966x_1 - 0.429x_2$

However, our interest is to further forecast the number of management talents needed for the next ten years, i.e., from 2015 to 2024. Using grey forecasting model (GM (1, 1)), we compute the predicted values for 2010-2014, as shown in column 3 in Table 4. The forecasting accuracy of the grey forecasting model is measured by comparing the predicted values obtained from grey forecasting model and the estimated values obtained from multiple linear regression method. Column 4 provides the forecasting errors for each year, and the ARPE and the accuracy of GM (1, 1) model (represented by P^0) are presented in columns 5 and 6, respectively. The values of P^0 from 2011 to 2014 are all above 80%, which indicates that it is appropriate to use GM(1, 1) model to predict the number of management talents needed for national coal companies from 2015 to 2024. The predicted values for the year 2015-2024 are listed in column 2 of Table 6 and plotted in Figure 5 along with the estimated values obtained from regression method.

TABLE 4
ACCURACY MEASURE OF THE GREY FORECASTING MODEL

Year	Regression Estimates	Grey Prediction	Error	ARPE	P^0
2010	508333	508333	-	-	-
2011	567599	566450	-1149	0.203%	99.80%
2012	573617	578556	4939	0.854%	99.15%
2013	597415	590920	-6495	1.099%	98.90%
2014	600934	603549	2615	0.433%	99.57%

FIGURE 5
A COMPARISON BETWEEN REGRESSION ESTIMATES AND GREY PREDICTION VALUES



Using the same methodology presented in Section "Data and Methodology", we set the number of professional technicians and the number of high-skilled talents needed as the dependent variables (y) in regression analysis, respectively, and use annual coal production (x_1) and wage level per employee (x_2) as the independent variables, we get the following estimated regression equations for years 2010 to 2014, as shown in Table 5.

TABLE 5
REGRESSION EQUATIONS OF COAL TECHNICAL PROFESSIONALS AND HIGH SKILLED TALENTS

Year	Professional technicians	High-skilled talents
2010	$\hat{y} = 11944.12 + 2.156x_1 - 0.391x_2$	$\hat{y} = -988.5 + 0.597x_1 - 0.185x_2$
2011	$\hat{y} = 12904.71 + 2.135x_1 - 0.378x_2$	$\hat{y} = -1157 + 0.592x_1 - 0.163x_2$
2012	$\hat{y} = 13902.66 + 2.075x_1 - 0.351x_2$	$\hat{y} = -893.6 + 0.602x_1 - 0.191x_2$
2013	$\hat{y} = 15717.48 + 1.953x_1 - 0.336x_2$	$\hat{y} = -1021.3 + 0.603x_1 - 0.172x_2$
2014	$\hat{y} = 20867.33 + 1.833x_1 - 0.408x_2$	$\hat{y} = -972.5 + 0.601x_1 - 0.152x_2$

Using GM(1, 1) model, we forecast the number of professional technicians and the number of high-skilled personnel needed for the years 2015 to 2024, as shown in columns 3 and 4 in Table 6, respectively.

TABLE 6
PREDICTED DATA OF MULTIPLE REGRESSION AND GREY PREDICTION METHOD

Year	Management personnel	Professional technicians	High-skilled personnel
2010	508333	467773	124311
2011	566450	501623	132353
2012	578556	527619	142585
2013	590920	545559	150203
2014	603549	564369	158292
2015	616448	588117	173698
2016	629622	611163	185355
2017	643078	635113	197795
2018	656822	660001	211069
2019	670859	685864	225234
2020	685196	712741	240350
2021	699840	740671	256480
2022	714796	769696	273693
2023	730073	799858	292061
2024	745675	831202	311662

CONCLUSIONS

Coal is the largest source of energy for China and will remain so in the future. The development of its coal industry plays a vital role in the Chinese economy. Like other industries, talents are important and in great need in the coal industry in China. Currently, China's coal industry still does not have sufficient talent resources in coal mining. One of the reasons is that coal mining is in the process of transforming from a labor-intensive industry to an integrated, talent- and technology-intensive one. This transition requires more workers with professional knowledge and special skills (Wu, et.al., 2017).

This paper utilizes multiple linear regression analysis and grey system theory to predict the number of management talents, professional technicians, and high-skilled talents needed in the coal industry in China. Using the data from China Statistical Yearbook, China Coal Industry Yearbook, and thirty major state-owned large- and medium-sized coal enterprises, we show that it is appropriate to use this method to make medium- and long-term prediction in coal talents needed in China.

REFERENCES

Analysis of the current situation and development trend of coal industry in 2016, http://www.chinamining.org.cn/index.php?m=content&c=index&a=show&catid=6&id=15976, Source: China Coal Market Network.

Chen, L.H., & Guo, T.Y. (2011). Forecasting financial crises for an enterprise by using the Grey Markov forecasting model. *Qual Quant*, 45, 911–922.

Chen, C.I., Hsin, P.H., & Wu, C.S. (2010). Forecasting Taiwan's major stock indices by the Nash nonlinear grey Bernoulli model. *Experts Systems with Applications*, 37, (12), 7557–7562.

Deng, J. (1982). Control problems of Grey systems, *Systems and Control Letters*, 5, 288-294.

Freedman D. (2009). *Statistical models: theory and practice*, Cambridge University Press.

Guo, S., & Duan, J. (2011). Current situation and prospect analysis of China coal industry development. *Future and Development*, 11.

Hsu, L.C., & Wang, C.H. (2007). Forecast the output of integrated circuit industry using a grey model improved by the Bayesian analysis. *Technological Forecasting and Social Change*, 74(6), 843–853.

Li, H. (2006). Analysis on the factors affecting the core competence of coal enterprises and the ways of improvement. *Northern Economy*, 6, 19-20.

Li, G.D., Yamaguchi, D. & Nagai, M.A. (2007). GM (1, 1)-Markov chain combined model with an application to predict the number of Chinese international airlines. *Technological Forecasting and Social Change*, 74(8), 1465–1481.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Wang, Z. (2004). Tackling the Problem of Personnel Shortage in Coal Industry, *Coal Higher Education*, 3.

Wang, Y.F. (2002). Predicting stock price using fuzzy grey prediction system. *Expert Systems with Applications*, 22(1), 33–38.

Wang, C.H. & Hsu, L.C. (2008). Using genetic algorithms grey theory to forecast high technology industrial output. *Applied Mathematics and Computation*, 195, (1), 256–263.

Wu, Y., Xiao, X., & Song, Z. (2017). Competitiveness analysis of coal industry in China: A diamond model study. *Resources Policy*, 52, 39-53.

Xu, Q. & Yang, S. (2015). Coal Mine Safety Profession Talent Demand and Talent Cultivation Analysis. *International Journal of Economics*, *Finance and Management Sciences*, 3, (5), 604-610.

Yan, X. (2009). *Linear regression analysis: theory and computing*, World Scientific.

Zhou, P. & Ang, B.W. (2006). A trigonometric grey prediction approach to forecasting electricity demand. *Energy*, 31(14), 2839–2847.

Zhao, Z., Wang, J.Z., Zhong, J.Z., & Su, Y. (2012). Using a grey model optimized by differential evolution algorithm to forecast the per capita annual net income of rural households in China. *Omega*, 40, 525–532.