International Journal of Innovative Computing, Information and Control
Volume **9**, Number **7**, July **2013**

ICIC International ©2013 ISSN 1349-4198
pp. 2907–2918

AN OPTIMAL MULTI-VARIABLE GREY MODEL FOR LOGISTICS DEMAND FORECAST

Hongqi Hui[1], Fachao Li[1] and Yan Shi[2,*]

[1]School of Economics and Management
Hebei University of Science and Technology
No. 70, Yuhua East Road, Shijiazhuang 050018, P. R. China
huihongqi@126.com; lifachao@tsinghua.org.cn

[2]Graduate School of Science and Technology
Tokai University
9-1-1, Toroku, Kumamoto 862-8652, Japan
*Corresponding author: yshi@ktmail.tokai-u.jp

Received March 2012; revised July 2012

Abstract. *The grey system theory, which has been extensively used in many areas, is appropriate for forecasting. It is necessary to put forward new models or algorithms to improve its performance, especially for forecast accuracy. In the forecast process of grey model, the size of data sample and the number of variables can affect forecast accuracy. In this paper, we first put forward a new method to choose optimal forecast variable number and data sample size for multi-variable grey model. Then we establish an optimal multi-variable grey model, in which the goal function is the minimum fitting relative error; and one constraint is data sample constraint, and the other is variable number constraint. Finally, we give the algorithm. Case studies of logistics demand forecast indicate that the model can solve the problem of factor choice and data sample size determination with high accuracy, and can fully utilize the sample information.*

Keywords: Grey model, Optimal grey forecast, Grey incidence, Logistics demand forecast

1. **Introduction.** Statistical methods, such as auto regressive, moving average, are commonly used as time series forecast models, while artificial intelligence-based methods, such as neural network, linear regression, genetic algorithms, fuzzy systems, Markov models, are also widely applied in forecasting [1]. It is considered that the statistical models are not as accurate as the neural network-based approaches for nonlinear problems. It has been proved that the back-propagation neural network yields better outcomes than the genetic algorithm. The major shortcoming is that it needs great amount of training data and relatively long training period for robust generalization [2]. The data streams in such applications are typically large, unbounded and composed of continuous data elements. In many areas, it is difficult to obtain lots of data to meet the demands. The grey forecast model is suitable to the deficiency [3].

Grey system theory is an interdisciplinary scientific area that was first introduced by Ju-long Deng in early 1980s. Grey system theory works on unascertained systems with partially known and partially unknown information by drawing out valuable information, which generates and develops the partially known information. Grey models require only a limited amount of data to estimate the behavior of unknown systems [4]. It has been successfully applied to various systems such as social, economic, financial, scientific and technological, industrial [5-7]. Grey system theory-based approaches can achieve good performance characteristics when applied to forecasting, since grey predictors adapt their parameters to new conditions as new outputs become available. Compared with conventional methods, grey models can produce better results with respect to noise and lack of modeling information [8].

Improving the grey model preciseness is the prerequisite to obtain the forecast results with minimum error. GM(1, 1) is well-developed and has spurred extensive applications. In the grey model forecast, the size of data sample and the number of variables will influence the results. So far, few studies have given the method to determine the optimal sample size and the variable number in multi-variable grey model (MGM(1, n)) forecasting. [9] proposed in the GM(1, 1) model to choose a high fit degree of forecast samples, but it does not apply in the multi-variable forecast model. [10] discussed the GM(l, 1) model to choose a suitable constant to improve forecast accuracy. [11] proposed an improved transformed grey model based on a genetic algorithm to improve the accuracy of short-term forecast. [12] came up with an improved forecast model to enlarge the applicability of the grey forecast model in various situations. By extending the data transforming approach, this method generalized a building procedure for the grey model to grasp the data outline and information trend to improve its forecast accuracy. Numerous forecast models based on grey systems have been developed. It has always been an important research objective to improve forecast accuracy with a small amount of data. Most of the literature cares more about GM(1, 1) and the improved GM(1, 1). In economic field, MGM(1, n) may produce better forecast accuracy. The premise of the GM(1, n) model is fairly good because the system is whiten by many effective messages around its forecast origin. In this paper, we put forward an optimal model to determine the variable number and the data sample size in multi-variable grey forecast, which can also solve the factor choice of the multi-variable grey model by fully using sample information. The objective function is the smallest relative error; and the constraints include the sample size constraints and the factor number constraints. The model established can effectively solve the minimum error of MGM(1, n). Case studies show that the method has better results.

The structure of this paper is organized as follows. Section 2 provides the fundamental concepts of grey system theory, including GM(1, 1) model and MGM(1, n) model. Section 3 presents the optimal MGM(1, n) model, and describes the algorithm. In Section 4, compared with other grey models, it shows numerical results by applying the optimal models to the logistics demand forecast. Finally, Section 5 discusses the conclusions of this paper and future works.

2. **Fundamental Concepts of Grey System Theory.** In grey models, the sampling frequency of the time series is supposed to be fixed. From the point of view, grey models can be viewed as curve fitting approaches.

2.1. **Accumulated value.** Assume that $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$ is an original, non-negative data series taken in consecutive order and at equal time intervals, where n is the sample size of the data. The primitive data are subjected to an operator, called the Accumulating Generation Operator (AGO), to smooth the randomness of the data and to weaken the tendency of variation. Let $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots,\ x^{(1)}(n)\}$, where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \cdots, n$. The accumulated data series are used to set up the differential equation, which is solved to obtain the n-step ahead predicted value of the system. Using the predicted value, the Inverse Accumulating Generation Operator (IAGO) is applied to find the predicted values of original data. The future value $x^{(0)}(n+k)$ can then be predicted [13].

2.2. **GM(1, 1) model.** We can derive the sequence of the generated mean value of consecutive neighbors [14], $Z^{(1)} = (z^{(1)}(2),\ z^{(1)}(3),\ \cdots,\ z^{(1)}(n))$, where $z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) +$

$x^{(1)}(k-1))$, $k = 2, 3, \cdots, n$. For a non-negative sequence of raw data $X^{(0)}$, $X^{(1)}$ is the AGO on $X^{(0)}$. $Z^{(1)}$ is a new sequence with the application of the generated mean value of consecutive neighbors operator on $X^{(1)}$, then the following equation is the whitened equation of GM(1, 1) model.

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b. \tag{1}$$

GM(1, 1) model can only be used in positive data sequences. The method of estimating parameters applies the discrete equation, and the method of simulating and forecasting applies the continuous equation. The transition is an important cause of simulation and forecasting errors in GM(1, 1).

2.3. **Grey relational analysis.** Grey relational analysis is utilized to ascertain the primary factors that are needed to make a superiority comparison in the system. As a system analysis method, grey incidence analysis can be used to analyze the relevance degree of each factor in the system. The fundamental principle is to recognize the relevance degree among many factors, according to the similarity levels of the geometrical patterns of sequence curves. We can judge whether they are close or not according to the similar degree of sequence curve shape. The more similar the curve is, the higher the correlation degree between relative series is, and vise versa [15].

Assume sequence observation data are $x_i(k)$, $k = 1,\ 2,\ \cdots,\ n$, where k is the time sequence variable; and that $x_i = [x_i(1),\ x_i(2),\ \cdots,\ x_i(n)]$ is the data series. The initial zero image can be defined as $x_i^0 = [x_i(1) - x_i(1),\ x_i(2) - x_i(1), \cdots,\ x_i(n) - x_i(1)]$ [16].

Define ε_{0i} as the absolute degree of grey incidence between x_0 and x_i,

$$\varepsilon_{0i} = \frac{1 + |S_0| + |S_i|}{1 + |S_0| + |S_i| + |S_i - S_0|}, \tag{2}$$

where

$$|S_i| = \left| \sum_{k=1}^{n-1} x_i^0(k) + \frac{1}{2} x_i^0(n) \right| \tag{3}$$

and

$$|S_i - S_0| = \left| \sum_{k=1}^{n-1} [x_i^0(k) - x_0^0(k)] + \frac{1}{2}[x_i^0(n) - x_0^0(n)] \right|. \tag{4}$$

Given a system factor x_i, $x_i' = [x_i(1)/x_i(1),\ x_i(2)/x_i(1), \cdots,\ x_i(n)/x_i(1)]$ denotes its initial image; then define γ_{0i} as the relative degree of grey incidence between x_0 and x_i,

$$\gamma_{0i} = \frac{1 + |S_0'| + |S_i'|}{1 + |S_0'| + |S_i'| + |S_i' - S_0'|}. \tag{5}$$

According to the absolute and the relative degrees of grey incidence, given the synthetic coefficient $\theta \in [0,\ 1]$, the synthetic degree of grey incidence (ρ_{0i}) can be calculated, $\rho_{0i} = \theta\varepsilon_{0i} + (1-\theta)\gamma_{0i}$. θ can be chosen as 0.5, and if the correlation of absolute quantity is emphasized, the value of θ could be larger; conversely, if the changing speed is emphasized, the value of θ could be smaller. There are many different methods about grey relational degree application [17,18]. Different methods may produce relational degree values with minor errors, which will not affect the relational degree sequence notably. When we use the grey relational degree analysis method before setting up MGM(1, n) model, we are more concerned about the sequence rather than the values. Therefore, any kind of grey incidence methods can have similar results. In economic analysis, grey relational degree value, according to Equations (2)-(5), may be more adaptable by revising synthetic coefficient to obtain better effects.

2.4. **MGM(1, n) model.** In grey system theory, MGM(n, m) denotes a grey model, where n is the order of the differential equation and m is the number of variables. Most studies have focused on GM(1, 1) model because of its computational efficiency. Multi-variable grey model, abbreviated as MGM, cares more on multi-variables with high grey incidence, which is more suitable for economics domain. The predicted index may have great relevant to other factors rather than itself. MGM(1, n) have higher accuracy and application value [19].

MGM(1, n) model is set up by AGO. Given that $x_i^{(1)}$ is the accumulated generalization for the original series $x_i^{(0)}$ $(i = 1, \cdots, n)$, the first-order ordinary differential equation for the model is as follows:

$$\begin{cases} \frac{dx_1^{(1)}}{dt} = a_{11}x_1^{(1)} + a_{12}x_2^{(1)} + \cdots + a_{1n}x_n^{(1)} + b_1 \\ \frac{dx_2^{(1)}}{dt} = a_{21}x_1^{(1)} + a_{22}x_2^{(1)} + \cdots + a_{2n}x_n^{(1)} + b_2 \\ \qquad \vdots \\ \frac{dx_n^{(1)}}{dt} = a_{n1}x_1^{(1)} + a_{n2}x_2^{(1)} + \cdots + a_{nn}x_n^{(1)} + b_n \end{cases}, \tag{6}$$

where $x_i^{(1)}$ is the AGO of sequence $x_i^{(0)}$ $(i = 1, \cdots, n)$. After calculating parameters a_{ij} and b_i, the accumulated generalization series $x_i^{(1)}$ can be obtained, from which we obtain the fitting forecasting value of original series $x_i^{(0)}$ $(i = 1, \cdots, n)$ by subtraction.

Let

$$L = \begin{bmatrix} \frac{1}{2}(x_1^{(1)}(2) + x_1^{(1)}(1)) & \cdots & \frac{1}{2}(x_n^{(1)}(2) + x_n^{(1)}(1)) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2}(x_1^{(1)}(m) + x_1^{(1)}(m-1)) & \cdots & \frac{1}{2}(x_n^{(1)}(m) + x_n^{(1)}(m-1)) & 1 \end{bmatrix},$$

and

$$Y = [Y_1, \cdots, Y_n] = \begin{bmatrix} x_1^{(0)}(2) & \cdots & x_n^{(0)}(2) \\ \vdots & \ddots & \vdots \\ x_1^{(0)}(m) & \cdots & x_n^{(0)}(m) \end{bmatrix}.$$

We can use the least-squares method to solve Equation (6). Then, the required vector can be obtained as follows:

$$\left[\hat{a}_{i1}, \hat{a}_{i2}, \cdots, \hat{a}_{in}, \hat{b}_i\right]^T = (L^TL)^{-1}L^TY_i \quad (i = 1, \cdots, n). \tag{7}$$

The time-response equation is

$$\hat{X}^{(1)}(k) = e^{A(k-1)}X^{(0)}(1) + \hat{A}^{-1}\left(e^{A(k-1)} - I\right) \cdot \hat{B}. \tag{8}$$

According to Equation (8), we can obtain $\hat{X}^{(1)}(k) = [\hat{x}_1^{(1)}(k), \cdots, \hat{x}_n^{(1)}(k)]$. Then, according to Equation (9), the fitting forecast value of $x_i^{(0)}$ can be obtained:

$$\begin{cases} \hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \quad k = 2, \cdots, m \\ \hat{x}^{(0)}(1) = \hat{x}^{(0)}(1) \end{cases}. \tag{9}$$

The relative error $(RE(i, m))$ is as follows:

$$RE(i, m) = \sqrt{\frac{1}{m}\sum_{t=1}^{m} \frac{\left[\hat{x}_i^{(0)}(t) - x_i^{(0)}(t)\right]^2}{\left[x_i^{(0)}(t)\right]^2}} \times 100\%. \tag{10}$$

2.5. **MGM(1, n) rolling model.** MGM(1, n) rolling model is to build the GM(1, 1) based on the forward data of sequence. For instance, using $x^{(0)}(k), x^{(0)}(k+1), x^{(0)}(k+2)$ and $x^{(0)}(k+3)$, the next predicted value is $x^{(0)}(k+4)$. In the following steps, the first point is shifted to the second. It means that $x^{(0)}(k+1), x^{(0)}(k+2), x^{(0)}(k+3)$ and $x^{(0)}(k+4)$ are used to predict the value of $x^{(0)}(k+5)$. In the next steps, the latest value is always added to the sequence, and the first value is omitted. MGM(1, n) rolling model is used to predict the long continuous data sequences such as the outputs of a system, price of a specific product, trend analysis for finance statements [20].

3. **Optimal MGM(1, n) Model.** Higher grey incidence degree between forecast sequence and related sequences means the more similarity among their curves; therefore, the relevant sequence is important to the forecast sequence. Theoretically, the higher the correlation degree is, the more accurate the results will be. Therefore, when establishing MGM(1, n) model, forecast variables with high grey relational degrees are selected as much as possible, usually requiring the value of correlation is greater than 0.6. However, in the actual forecast calculation, how many forecast factors should be selected? Whether to choose all of the factors with grey incidence greater than 0.6, or just select a few? Different number of predictors will inevitably have an impact on the predicted results. In addition, whether or not the factors with grey incidence less than 0.6 are certainly inappropriate as the predictors? The answer to these problems is not clear until now.

In addition, the size of forecast sample in predicting the outcome will have a great impact. Large sample size contains more abundant information, which makes grey system white easy. However, in complex systems, such as the economic system, all data fitting the prediction error may be larger than that of part of the data. From the view of curve fitting trend, there may be some data but not all data fitting better. Therefore, on the condition of meeting the minimum requirements for sample size modeling, how to determine the optimal sample size to get better results?

The existing researches on MGM(1, n) forecasting have ignored the problem. If using enumerative algorithm, for m samples of n variables, there will be $(2^n-1)m$ calculations. Obviously, it is infeasible by enumerative algorithm. The existing applications just give the forecast factors subjectively or partially objectively and use the sample data directly. Anyway, there must be optimal results with suitable factor number and the appropriate sample size. Generally, the forecast accuracy can be measured by the relative error. The forecast factors and the sample size can both affect the relative error, which can be taken as two kinds of constraints. In view of this, we establish the optimal MGM(1, n) model, whose objective function is to fit the relative error the smallest; and it is subject to two kinds of constraints: one is the sample size constraint and the other is the factor number constraint. The model is shown as follows:

$$\begin{array}{ll} \min & z = RE(i,k) \\ \text{s.t.} & \left\{ \begin{array}{l} RE(i,k) \le \min\{RE(i+1,k), RE(i-1,k)\} \\ RE(i,k) \le \min\{RE(i,k+1), RE(i,k-1)\} \\ i = N, \cdots, 2;\ k = T, \cdots, 2 \end{array} \right. \end{array} \tag{11}$$

where $RE(i,k)$ represents the relative error, which is affected by both the factor number (i) and the sample size (k). N denotes the total factor number, and T denotes the sample size.

Essentially, the model established is a two-stage optimization problem. In the first stage we fix the sample size to determine the best factor number, whose relative error is less than that of the adjacent factor number. In the second stage we determine the best sample size. We believe that influences by factors are more important than that by sample size, for the factors are the bases of MGM(1, n). So we determine the best factor number in the first stage. From Equation (10) we can see that the relative error is affected by both the factor number and the sample size. Different sample size can produce different relative error for the same factor number. Then there exists a new problem in the first stage. When we determine optimal factor number, how should the sample size be given? In order to solve this problem, we can calculate the relative error under the total sample size first. Thus we can obtain a best factor number. Then we reduce the total sample size by one (just like the rolling model). We can obtain another best factor number. If these results are the same, the best factor number is determined. If the results are not the same, then repeat the process. We can choose the same results with the highest percentage (or weight) as the best factor number. If there are no same results at the end of the process, it means that the degree of data dispersion is high. Under this circumstance, taking the total sample size may be the best from the angle of utilizing data information. Generally speaking, data in economic domain have apparent timing characteristics. Different sample size may produce the nearly same sequence of factor number for the sake of data feature except for the value difference, which makes the algorithm simple. Equation (10) gives the relative error of each factors in Equation (6). The relative error of the forecasted factor is more emphasized. So in the process of optimization we only calculate the relative error of the forecasted factor.

Although there are various methods for modifying the residual error, e.g., Markov chain, these methods are sometimes problematic under the condition of insufficient data number. This results from the fact that only few data number is needed to model a time series. From the angle of calculations, the optimal MGM(1, n) has no distinct difference with the traditional MGM(1, n). As a result, the method established is adapted not only for few modeling data but also for fewer calculations. The model assumes that there is only one minimum relative error to meet any of constraints. By taking the variables with grey incidence and the sample size into consideration, the adaptive and high-precision forecasting model can deal with the random time series more effectively. We put forward the following algorithm to solve Equation (11).

Algorithm:

Step 1. Calculate the grey relational degree, and rank them in accordance with the smallest order.

Step 2. Determine the number of factors to meet the constraints, and define N as total factor number. First, let $i = N$, calculate the corresponding relative error, $RE(i, k)$. If $RE(i, k) \leq RE(i + 1, k)$, let $i = i - 1$. Repeat this process until $RE(i, k) \leq RE(i - 1, k)$, and then the optimal factor number is obtained.

Step 3. Determine the sample size to meet the constraints, and define T as sample size. First, let $k = T$; that is, select sample 1 to sample T to calculate the relative error, $RE(i, k)$. If $RE(i, k) \leq RE(i, k + 1)$, let $k = k - 1$, which means selecting sample j to sample T, and so on. If $RE(i, k) \leq RE(i, k - 1)$, the optimal sample size is obtained.

We seek an effective way to improve the forecast accuracy of MGM(1, n) model by using the uncertain information. The objective function is the minimum relative error; and the constraints contain both the factor number and the sample size. The optimal solution can be found; and all constraints and optimality conditions are satisfied. The method solves the problem of the factor number and the sample size choice. It is obvious that the proposed model is the most suitable short-term forecasting model, especially for the economic industry for the sake of data characteristics.

4. **An Example for Logistics Demand Forecast.**

Modern logistics industry is increasingly becoming an important support platform for economic development. Scientific decision-making based on a reasonable forecast is an effective protection for the healthy development of the logistics industry. Both macro-planning of the logistics industry and the operating decisions of logistics enterprises need the accurate forecast result. Therefore, scientific forecasts of the logistics industry have great significance for the logistics industry and the economic development. The statistical index system of logistics industry in China has not been complete; and the statistical year book has not included the logistics index. The statistical data of Logistics Association has not been perfect. Data factors have been the bottleneck to study quantitatively the logistics industry development. Under the situation that there are not long series reliable data to use, grey model is an effective method to forecast based on a small amount of data and its developing trend.

4.1. **Backgrounds.** The transportation of goods constitutes an extremely important activity within urban areas. It links suppliers with customers. It directly ensures adequate supplies to stores and places of work, as well as delivery of goods at home [21]. Logistics demand forecast is to analyze, estimate and deduce the future logistics demand according to historical data and information by apposite method. It is important to find out logistics demand for logistics industry development. And neural networks, time series prediction and hybrid system methods have been used to solve problems in demand forecasts [22].

Logistics demand can be reflected by freight scale, logistics cost, investment in fixed assets and logistics efficiency. Considering the data availability, we take freight traffic as the index to reflect logistics demand. Although logistics serves nearly each industry, involving lots of links such as transportation, warehousing, packaging, handling, distribution. Transportation among the logistics process is the basic link. Transportation amounts inevitably determine the amount of logistics. Therefore, we take freight traffic (FT) to denote logistics demand. According to the forecast of freight scale, we can obtain the logistics demand regulation and its future trends. Logistics demand is determined by economic development fundamentally, so we use key economic factors as endogenous variables to establish MGM(1, n) model. We choose the primary industry (PI), the secondary industry (SI), the tertiary industry (TI) and the total retail sales of consumer goods (TRSCG) as the economic indexes. We take Hebei province of China as an example. The statistical data are shown in Table 1 [23].

4.2. **Calculation of grey incidence degree.** Before choosing appropriate forecast factors, we first calculate the grey incidence degree. The factors with distinct lower grey

TABLE 1. Statistics data in Hebei province of China 2001∼2008

No.	Year	FT	PI	SI	TI	TRSCG
		10^4 T	10^8 RMB	10^8 RMB	10^8 RMB	10^8 RMB
1	2001	80835	913.82	2696.63	1906.31	1778
2	2002	84315	956.84	2911.69	2149.75	1968.3
3	2003	80551	1064.05	3417.56	2439.68	2177.9
4	2004	87265	1333.57	4301.73	2842.33	2576.4
5	2005	91330	1503.07	5232.50	3360.54	2952.9
6	2006	96784	1461.81	6115.01	3938.94	3397.4
7	2007	104188	1804.72	7241.80	4662.98	3986.2
8	2008	111383	2034.60	8777.42	5376.59	4880.4

Table 2. Results of grey incidence degree

	PI	SI	TI	TRSCG
Absolute	0.5214	0.6076	0.5649	0.5532
Relative	0.6778	0.6102	0.6275	0.6433
Synthetic	0.5996	0.6089	0.5962	0.5982

Table 3. Freight traffic forecast results with different factors in Hebei province

	No.	1	2	3	4	5	6	7	8	ARE (%)
	Actual value	80835	84315	80551	87265	91330	96784	104188	111383	
MGM(1, 5)	Forecast value	80835	89699	192656	2535077	52760338	1.1E+09	2.4E+10	5.3E+11	∞
MGM(1, 4)	Forecast value	80835	83127	82649	85974	91605	97801	103383	108765	1.51
MGM(1, 3)	Forecast value	80835	83195	83046	85828	90508	96595	103889	112352	1.40
MGM(1, 2)	Forecast value	80835	82274	83908	86643	90623	96022	103046	111942	1.82

relational degree will be omitted, which can reduce calculation and complexity to minimize the errors. According to Equations (2)-(5), and given synthetic coefficient $\theta = 0.5$, the grey incidence degree results are shown in Table 2.

The results in Table 2 show that the relational degree between freight traffic and the primary industry, and the secondary industry, and the tertiary industry, and the total retail sales of consumer goods respectively is not high enough. Any grey relational degree has not more distinct difference than the others. Generally, factors with grey incidence value smaller than 0.6 can be omitted. Although all the grey incidence values are not high, the factors have close relation with freight traffic from the view of economy. More added values mean more freight traffic. Therefore, we ought to have some better choices rather than abandon any of them to cause information loss and affect forecast results.

4.3. **Determination of optimal factors.** Generally, higher grey relational degree means more precise forecast results. Therefore, we arrange the grey relational degree in decreasing order. The grey incidence degree between the freight traffic and the secondary industry is 0.6089, and it is the greatest. So the secondary industry lies in the first. While the last is the tertiary industry, its grey incidence degree is the smallest, 0.5962. That is, the forecast factors' order is the secondary industry, the primary industry, the total retail sales of consumer goods and the tertiary industry.

According to Step 2, we take 4, 3, 2 and 1 factors successively and establish grey forecast model by Equations (6)-(10). After calculation, the forecast results and average relative error (ARE) are shown in Table 3.

In Table 3, MGM(1, 5) denotes that we take freight traffic, the secondary industry, the primary industry, the total retail sales of consumer goods and the tertiary industry to set up differential equation. The rest models omit the factors with the smallest grey incidence degree successively. That is, we take freight traffic, the secondary industry, the primary industry and the total retail sales of consumer goods to establish MGM(1, 4). MGM(1, 2) denotes that we take freight traffic and the secondary industry to establish grey model. Among the results, MGM(1, 3) has the smallest average relative error. So it may produce the forecast results with the minimum error to choose the two factors of the secondary industry and the primary industry to forecast. In MGM(1, 5), because forecast factors are relative more and time serious number is relative less, the forecast results have great error, especially in No.4-8. The fitting effects are shown in Figure 1.

4.4. **Determination of optimal sample size.** We select freight traffic, the secondary industry and the primary industry to establish MGM(1, 3) to calculate the forecast results

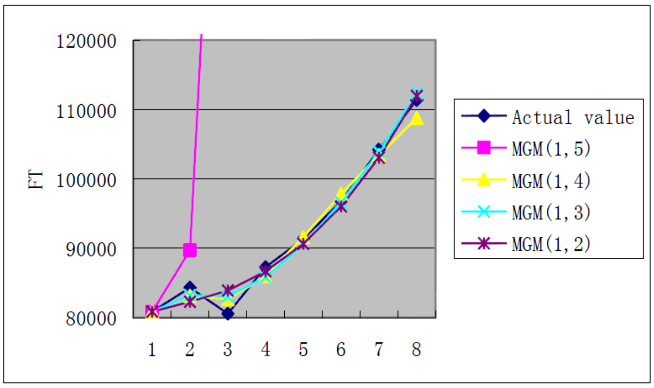

FIGURE 1. Fitting effects with different models

TABLE 4. Freight traffic forecast results with different sample size in Hebei province

	No.	1	2	3	4	5	6	7	8	ARE (%)
	Actual value	80835	84315	80551	87265	91330	96784	104188	111383	
2001∼2008	Forecast value	80835	83195	83046	85828	90508	96595	103889	112352	1.40
2002∼2008	Forecast value		84315	80770	86185	91779	97663	103956	110794	0.65
2003∼2008	Forecast value			80551	87055	91300	96996	103867	111624	0.20
2004∼2008	Forecast value				87265	91076	97064	103949	111398	0.21
2005∼2008	Forecast value					91330	113978	138374	171548	32.82

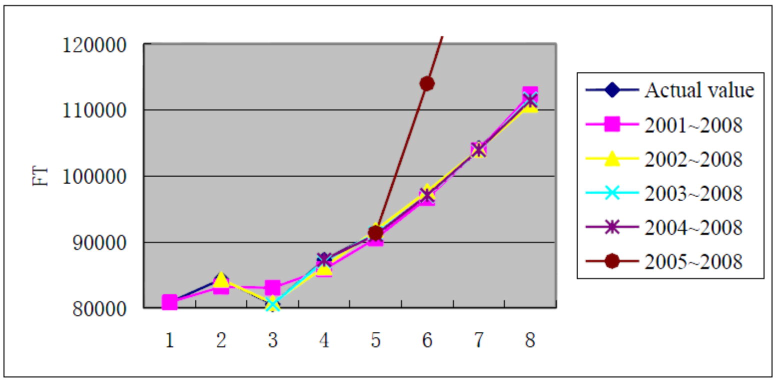

FIGURE 2. Fitting effects with different sample size

for different sample size. Considering that new samples have more influence on the forecast results than the old data, we take sample 1∼8, 2∼8, 3∼8, ... successively to calculate by Equation (10) according to Step 3. The forecast results and the average relative error are shown in Table 4.

Table 4 indicates that the relative error shows downward trend with the reduction of sample size. When the size of sample is 6, that is, the data from year 2003 to 2008, the average relative error is the smallest. As the size of sample reduces further, the relative error shows upward trend. Especially, when the size of sample is 4, the forecast results appear large deviation for the linear correlation among the variables, which leads to the rapidly increasing relative error. Therefore, 6 samples can produce better forecast results in the example. The fitting effects of different sample size are shown in Figure 2.

4.5. **Accuracy tests of the optimal model.** The accuracy tests include residual error test, relational degree test, and the posterior ratio test. The residual error test is to judge whether the model is accurate by relative error. As shown in Table 4, the average relative error of the grey model with the optimal sample size is 0.2%, that is, the average forecast precision of freight traffic is 99.8%, which indicates the model is accurate.

The relational degree test is used to calculate the similar degree between forecast value curve and the actual value curve. After calculation by Equations (2)-(4), the absolute grey incidence between fitting curve and actual series curve of freight traffic is 0.998, which indicates a good effect of the model from the curve's similar degrees.

The posterior ratio (c) and the minor error probability (p) are used to test statistical characteristics. To begin with $c = s_2/s_1$, where $s_1 = \sqrt{(\sum_{t=1}^{n}((x^{(0)}(t) - \bar{x}^{(0)})^2)/(n-1)}$ is the mean square deviation of observation values, and $\bar{x}^{(0)} = (\sum_{t=1}^{n} x^{(0)}(t))/n$ is the mean value; and $s_2 = \sqrt{(\sum_{t=1}^{n}(q^{(0)}(t) - \bar{q}^{(0)})^2)/(n-1)}$ is the mean square deviation of residual error, where $q^{(0)}(t) = x^{(0)}(t) - \hat{x}^{(0)}(t)$, $\bar{q}^{(0)} = (\sum_{t=1}^{n} q^{(0)}(t))/n$, and $p = P\{|q^{(0)}(t) - \bar{q}^{(0)}| < 0.6745 s_1\}$. The judgment standard of the forecast's accuracy is shown in Table 5.

TABLE 5. Judgment standard of forecast accuracy

p	c	Accuracy grade
> 0.95	< 0.35	good
> 0.8	< 0.5	qualification
> 0.7	< 0.65	Inadequate qualification
≤ 0.7	≥ 0.65	failed

By substituting the corresponding data, the posterior ratio of freight traffic $c = 0.021 < 0.35$ and $p = 1 > 0.95$ show the model's precision is of the first grade, which reveals a high simulation degree. The curves of forecast values and actual values are shown in Figure 3.

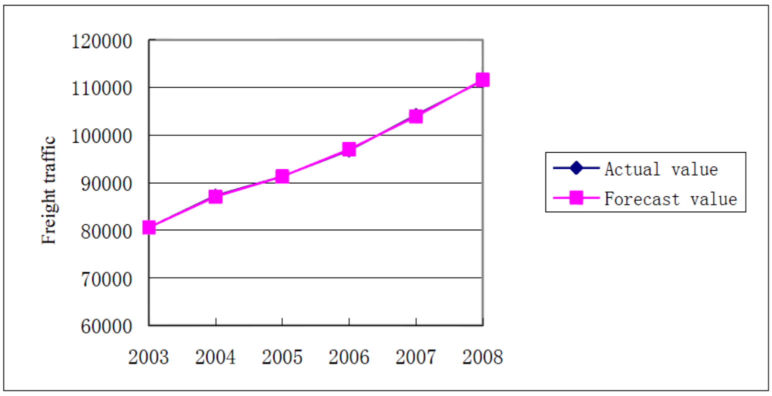

FIGURE 3. Fitting effects of the optimal grey model

5. **Conclusions and Future Works.** As the core of grey system theory, grey model is the basis of grey prediction, decision-making and control. Grey model needs less information, and prior distribution of characteristics may be unknown. According to limit accumulated generation, any discrete smooth original series can be converted to regular sequence. The grey model can reflect the actual situation of the system while maintaining the original characteristics of the system, which has important meaning in the fuzzy

system forecast. In this paper, we put forward a new optimal method to obtain best forecast results. By using information completely without abandoning low grey relational degree arbitrarily, the method can avoid subjective factor choice effectively. Case studies indicate that using factors with low grey incidence degree to forecast can also produce good forecast results, which is useful in complex economic system. The algorithm can obtain optimal forecast factors and sample size for the forecast results with the minimum error. In this paper, we assume that the local optimum is the global optimum, that is, the relative error curve is the U-shape. The distribution of relative error needs further discussion.

Acknowledgment. This research was supported by the Natural Science Foundation of China (Grant Nos. 71071049, 71210107002) and the Natural Science Foundation of Hebei Province (F2011208056). We also wish to express our gratitude to professor Bo-tang Han, Beijing Institute of Technology, for remarks and suggestions that improved this paper significantly.

REFERENCES

[1] M. Aydinalp-Koksal and V. I. Ugursal, Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector, *Applied Energy*, vol.85, no.4, pp.271-296, 2008.

[2] V. Vahidinasab, S. Jadid and A. Kazemi, Day-ahead price forecasting in restructured power systems using artificial neural networks, *Electric Power Systems Research*, vol.78, no.8, pp.1332-1342, 2008.

[3] U. Kumar and V. K. Jain, Time series models (Grey-Markov. Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India, *Energy*, vol.35, no.4, pp.1709-1716, 2010.

[4] J. L. Deng, Introduction to grey system theory, *The Journal of Grey System*, vol.1, no.1, pp.1-24, 1989.

[5] T. L. Tien, The deterministic grey dynamic model with convolution integral DGDMC(1, n), *Applied Mathematical Modeling*, vol.33, no.8, pp.3498-3510, 2009.

[6] L. M. Kung and S. W. Yu, Prediction of index futures returns and the analysis of financial spillovers – A comparison between GARCH and the grey theorem, *European Journal of Operational Research*, vol.186, no.3, pp.1184-1200, 2008.

[7] C. R. Su et al., Estimation for inner surface geometry of furnace wall using inverse process combined with grey prediction model, *International Journal of Heat and Mass Transfer*, vol.52, no.15-16, pp.3595-3605, 2009.

[8] T. L. Tien, A research on the grey prediction model GM(1, n), *Applied Mathematics and Computation*, vol.218, no.9, pp.4903-4916, 2012.

[9] W. Y. Wu and S. P. Chen, A prediction method using the grey model GMC(1, n) combined with the grey relational analysis: A case study on Internet access population forecast, *Applied Mathematics and Computation*, vol.169, no.1, pp.198-217, 2005.

[10] B. Q. Li, On three data modeling of grey theory, *University Journal of Applied Mathematics*, vol.6, no.4, pp.11-12, 1991.

[11] L. C. Hsu, Using improved grey forecasting models to forecast the output of opto-electronics industry, *Expert Systems with Applications*, vol.38, no.11, pp.13879-13885, 2011.

[12] D. C. Li et al., An extended grey forecasting model for omnidirectional forecasting considering data gap difference, *Applied Mathematical Modelling*, vol.35, no.10, pp.5051-5058, 2011.

[13] Y. H. Lin, P. C. Lee and T. P. Chang, Adaptive and high-precision grey forecasting model, *Expert Systems with Applications*, vol.36, no.6, pp.9658-9662, 2009.

[14] Y. H. Wang et al., An approach to increase prediction precision of GM(1, 1) model based on optimization of the initial condition, *Expert Systems with Applications*, vol.37, no.8, pp.5640-5644, 2010.

[15] S. F. Liu, T. B. Guo and Y. G. Dang, *Grey Systems Theory and Its Applications*, Science Press, Beijing, 1999 (in Chinese).

[16] C. Y. Kung and K. L. Wen, Applying grey relational analysis and grey decision-making to evaluate the relationship between company attributes and its financial performance – A case study of venture capital enterprises in Taiwan, *Decision Support Systems*, vol.43, no.3, pp.842-852, 2007.

[17] J. L. Deng, *Grey Prediction and Decision*, Huazhong University of Science and Technology Press, Wuhan, 1986 (in Chinese).

[18] H. Wong et al., Change-point analysis of hydrological time series using grey relational method, *Journal of Hydrology*, vol.324, no.1-4, pp.323-338, 2006.

[19] L. C. Hsu and C. H. Wang, Forecasting integrated circuit output using multivariate grey model and grey relational analysis, *Expert Systems with Applications*, vol.36, no.2, pp.1403-1409, 2009.

[20] D. Akay and M. Atak, Grey prediction with rolling mechanism for electricity demand forecasting of Turkey, *Energy*, vol.32, no.9, pp.1670-1675, 2007.

[21] T. G. Crainic, M. Gendreau and J. Y. Potvin, Intelligent freight-transportation systems: Assessment and the contribution of operations research, *Transportation Research Part C*, vol.17, no.6, pp.541-557, 2009.

[22] L. Aburto and R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing*, vol.7, no.1, pp.136-144, 2007.

[23] *Hebei Statistics Year Book*, 2002-2009.