THE COMBINED FORECASTING MODEL OF GRAY MODEL BASED ON LINEAR TIME-VARIANT AND ARIMA MODEL

Xia Long [1, 2], Yong Wei [2], Jie Li [3] & Zhao Long[2]
[1] School of Computer Science in Sichuan University of Science & Engineering, China
[2] Mathematics and information institute in China West Normal University, China
[3] School of Economics and Finance in Guizhou University of Finance and Economics, China

ABSTRACT
As to the established gray model based on the linear time-variant and individual prediction model of ARIMA, this article constructs the combined forecasting model based on the gray model and the time series model by means of relative error weighing. This prediction indicates that both the gray model and ARIMA model exert efficient function on the Torpedo development cost prediction, and the combination forecast model improve the precision of prediction.

Keywords: *discrete Verhulst model, ARIMA model, combined forecasting model*.

1. INTRODUCTION
Tackling data can contribute to finding out the law underlying data. Prediction plays a role in inferring the approaching variation of development of things according to the existent data. Different data process modes emerge different modeling approaches. The gray model presents in the way of cumulating the development trend of gray energy accumulation, which fully visualizes the potential law implying in the disordered original data.
However, most of data are in reality that partial message are known and partial message are unknown. Gray model is just about the process object of lean sample and small information [1-3]. The time series, such as ARIMA model [4], requires adequate data and then it might reveal the inherent regular patterns underlying the data. Thus, the gray model [6-9] is capable of making their respective advantages complementary to each other in the respect of information adequacy and so on.
This article is about the construct of discrete Verhulst model (LTDVM model）based on the linear time-variant and the building of ARIMA model , both of which are used to predict the Torpedo development cost. At last, the construct of the combined forecasting model formed from the gray model and the time series is also used for prediction and the prediction accuracy is expected to be efficient and reliable.

2. THE COMBINED FORECASTING MODEL FORMED FROM GRAY MODEL AND ARIMA MODEL
2.1 .The Discrete Verhulst Model(LTDVM Model）Based On The Linear Time –variant
Assuming the non-negative primitive sequence as $X^{(0)}=\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\}$, in which, $k=1,2,\cdots,n$; The 1-AGO sequence of $X^{(0)}$ is: $X^{(1)}=\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\}$, in which, $x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i)$, $k=1,2,\cdots,n$;
$Z^{(1)}$ is the adjoining and generative mean value sequence of $X^{(1)}: Z^{(1)}=\{z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\}$, in which, $Z^{(1)}(k)=0.5(x^{(1)}(k)+x^{(1)}(k-1)),k=2,3,\cdots,n.$

2.1.1. Gray Verhulst model [1]
There are four definitions**.**
Definition 1 Calls

$$x^{(0)}(k)+az^{(1)}(k)=b\left(z^{(1)}(k)\right)^2 \tag{1}$$

as Gray Verhulst model.
When the primitive data themselves embody the shape of S, the original data are taken as $X^{(1)}$ and the 1-AGO as $X^{(0)}$, the construct of Verhulst model is helpful for direct simulation of $X^{(1)}$.
Definition 2 Calls

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b\left(x^{(1)}\right)^2 \tag{2}$$

as a winterization equation of gray Verhulst model. From the process of solving this equation， it is not hard to see that the essence of equation solution is the shift of primitive function into backward function.
The inherent error of Gray Verhulst model derives from the approximately non-strict matching of the gray model and the winterization equation. Thus, this article firstly considers the reciprocal treatment of the primitive data and then the construct of the discrete gray prediction model.

2.1.2. Linear Time-varying Parameters Discrete Gray model [2]

Definition 3 Calls

$$x^{(1)}(k+1)=(\beta_1+\beta_2 k)x^{(1)}(k)+\beta_3 k+\beta_4 \tag{3}$$

as the discrete gray prediction model of linear time varying parameters （TDGM(1，1)model）.
In practical applications, the system behaviors sequence itself and the interaction of different behavior sequences represent complicated nonlinear relations. Besides, as time goes, the parameters and structures of the system behavior sequence may also change consistently . Therefore, it is difficult to put simulation and prediction into effect by the light of constant parameter model. For that reason, the teacher Zhang Ke builds the DGM(1,1) model of the linear time-variant parameters by means of the substitution of linear time for the constant parameter of primitive discrete gray model in the document two.
With regard to the simulation prediction model of oscillating sequence, we try to turn the primitive sequence into reciprocal and then take it into TDGM(1,1) model, equally the recombination of discrete Verhulst model and linear time-variant model, that is, putting forward a type of discrete Verhulst model based on the linear time-variant.

2.1.3. Based on the Linear Time-varying Discrete Verhulst Grey Model (LTDVM model).

Definition 4 assumes the observation value of one particular behavior characteristic sequence of the system as $X^{(0)}=\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\}$. $Y^{(0)}$ is the reciprocal sequence of $X^{(0)}$, that is $y^{(0)}(k)=\frac{1}{x^{(0)}(k)}(k=1,2,\cdots,n)$ and $Y^{(1)}$ is an additive sequence of $Y^{(0)}$,naming

$$y^{(1)}(k+1)=(\beta_1+\beta_2 k)y^{(1)}(k)+\beta_3 k+\beta_4\ ,k=1,2,\cdots,n-1 \tag{4}$$

as discrete Verhulst model based on the linear time-variant(LTDVM model).
The solution process of such a model is as follows:

I. making use of least square method to get the model parameters $\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\beta}_4:\hat{\beta}_1=\frac{B_1}{A},\hat{\beta}_2=\frac{B_2}{A},\hat{\beta}_3=\frac{B_3}{A},\hat{\beta}_4=\frac{B_4}{A}$;

II. Taking advantage of the recursion formula $\hat{y}^{(1)}(k+1)=\left(\hat{\beta}_1+\hat{\beta}_2 k\right)\hat{y}^{(1)}(k)+\hat{\beta}_3 k+\hat{\beta}_4$ to get the sequence $\hat{y}^{(1)}(k+1)$;

III. Utilizing the formula $\hat{x}^{(0)}(k+1)=\frac{1}{\hat{y}^{(0)}(k+1)}$ to get the model prediction value of primitive sequence.

2.2. ARIMA Model [4]

ARIMA model (Summation auto-regressive moving average model) is broadly used in the field of predictive control. It is classical analysis model of time series as well as common prediction model in practice. In fact, via the autocorrelation of statistics, the ARIMA model is employed in the prediction of data, the stability of time sequence, or the nonstationary time series that is trending to smooth by means of difference.
Assuming y_t as one nonstationary time series, μ_t as its residual sequence, d as difference sequency, and bringing the delay operator B, here is

$$B^k y_t = y_{t-k} \tag{5}$$

in which k is lag phase. Making d order difference operation of $\left(1-B^k\right)^d y_t$, suppose that the differentiated sequence is stable, ARIMA model can be set up. The formula is as follows:

$$\varphi(B)\left(1-B^{d}\right)y_{t}=\theta(B)u_{t}\ ,in\ which\begin{cases}\varphi(B)=1+\varphi_1 B+\varphi_2 B^2+\cdots+\varphi_P B^p\\ \theta(B)=1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\end{cases} \quad (6)$$

This is the auto-regression coefficient polynomials and smoothing coefficient polynomials of ARMA (p,q) respectively.

ARMA(p,d,q)model is actually the combination budget of ARMA(p,q) model and d order difference operation. Its model-building process is as follows:(1) testing the stationarity of observation sequence (2) If the observation sequence is not stable, making a d order difference operation for the primitive sequence.(3) according to stationary series, fitting ARMA(p,d,q)model (4) recognizing estimated parameters (5) residual test (6) according to the data, predicting.

2.3. The optimal combination of the construction of weight prediction model [5]

For the same prediction problem, there are two or more than two single forecasting methods $y_1(t), y_2(t),\cdots, y_\lambda(t)$, λ symbolizes the number of prediction formula. Assuming that the relative error of the Ath prediction model is $e_a(t)(t=1,2,\cdots,n;a=1,2,\cdots,\lambda)$ in No.t and the combined model weight is ω_a, and the combined prediction model is

$$\begin{cases}\sum_{a=1}^{\lambda}\omega_a=1\\ s.t.\min X=X(\omega_a)\end{cases} \quad (7)$$

in which, X is the objective function of optimal combined forecasting model. The objective function opts error forms. This article chooses the absolute value of relative error and minimizes it. The weight is based on the following formula:

$$\omega_a=\frac{|e_a|}{\sum_{a=1}^{\lambda}|e_a|} \quad (8)$$

in which, $|e_a|$ functions as the mean absolute difference of the relative errors.

In order to achieve the optimal prediction result, according to the principle that smaller weight is for larger relative error and larger weight is for smaller one, one optimal combined model can be constructed:

$$f(t)=\omega_1 y_1(t)+\omega_2 y_2(t)+\cdots+\omega_a y_\lambda(t) \quad (9)$$

The individual prediction model in this article is Verhulst model based on the linear time-variant and ARIMA model, either of which is represented by $y_1(t)$ and $y_2(t)$ correspondingly. Thus, the combined forecasting model based on individual prediction model shows as the following formula:

$$f(t)=\omega_1 y_1(t)+\omega_2 y_2(t) \quad (10)$$

3. APPLIED CASES: TORPEDO DEVELOPMENT COST PREDICTION [1, 3]

Torpedo development is a complicated system project which is classified into seven phases: feasibility phase, preliminary design phase, detailed design phase, trial-manufacture phase, experimental phase, validation phase and product definition phase. This kind of Torpedo development is highly volatile on account of different degrees of complexity and development cost at each stage. Generally speaking, with the development task going deeper, development cost requires more until a peak of development cost appears. Thereupon, the development cost starts to diminish down. One type of Torpedo development cost shows as the following two tables:

Table 1. one Torpedo development cost [ten-thousand yuan]

years	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
cost	496	779	1187	1025	488	255	157	110	87	79

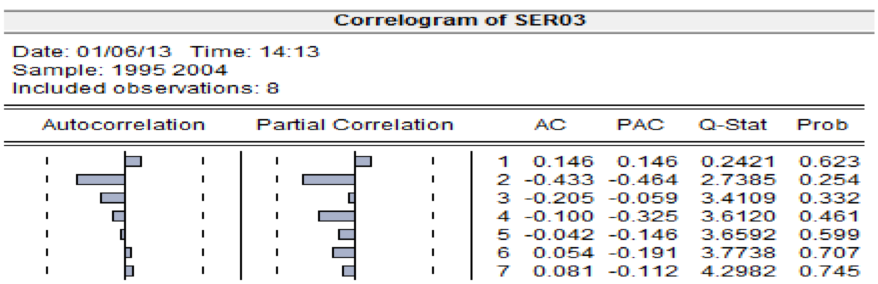
Table 2. 1-AGO of one Torpedo development cost [ten-thousand yuan]

years	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
cost	496	1275	2462	3487	3975	4230	4387	4497	4584	4663

(1) As the statistics above displays, the cumulative curve that the Torpedo development cost presents is approximative to the shape of "S". It constructs a new discrete Verhulst model (LTDVM model). Hence, the data in the Table 2 are used to make simulation and prediction according to discrete Verhulst model (LTDVM model) based on the linear time-variant. $y^{(0)}(k)=\frac{1}{x^{(0)}(k)}(k=1,2,\cdots,n)$,in which $y^{(1)}(k)=\sum_{i=1}^{k}y^{(0)}(i),k=1,2,\cdots,n$; on the basis of definition4, least square method can be taken to get model parameters $\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\beta}_4$ and then the prediction formula comes out: $y^{(1)}(k+1)=(0.3012-0.0076k)y^{(1)}(k)+0.0002k+0.002$, whose computed result is represented in Table 3.

(2)The observation value sequence in original samples is tested to be non-stable sequence by ADF by the light of eviews6.0, so the result of difference operation displays that ADF statistical magnitude is larger than marginal value and the second order difference sequence is a stable sequence. After making a second order difference, the P value in autocorrelogram conspicuously embodies the nature of 2 order truncation, that is, partial correlation coefficient is provided with remarkable characteristics of the second order difference(Fig 1). Overall, considering that this article deals with original sequence by means of the second order difference, we lean to the fitting original sequence ARIMA(2，2，2). After making a second order difference for the sequence fitting model, the consequence is as follows: $y_t=-75.2217+1.4810y_{t-1}-0.4401y_{t-2}+\varepsilon_t-0.5282\varepsilon_{t-1}+0.9949\varepsilon_{t-2}$,where $y_{t-i}(i=0,1,2)$ is the cumulative development cost of one type of Torpedo at year t-i, and $\varepsilon_{t-i}(i=0,1,2)$ is the Root Mean Squared Error of the cumulative development cost of one type of Torpedo at year t-i.

(3)


Correlogram of SER03

Date: 01/06/13 Time: 14:13
Sample: 1995 2004
Included observations: 8

	AC	PAC	Q-Stat	Prob
1	0.146	0.146	0.2421	0.623
2	-0.433	-0.464	2.7385	0.254
3	-0.205	-0.059	3.4109	0.332
4	-0.100	-0.325	3.6120	0.461
5	-0.042	-0.146	3.6592	0.599
6	0.054	-0.191	3.7738	0.707
7	0.081	-0.112	4.2982	0.745

Figure 1. Self-correlogram and partial correlogram

According to the fitting result, in the process of predicting, Theil's 0.0002, an unequal coefficient, indicates that this model has good prediction ability. Thereinto, the covariance proportion is 0.7503, which demonstrates an ideal prediction result of this model(Figure 2) . The calculation result is showed in Table 3.

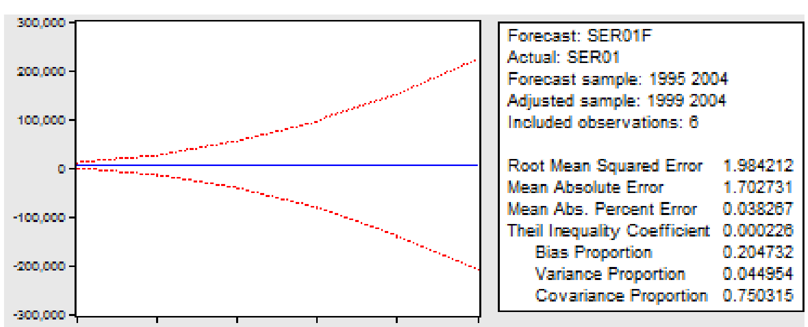


Figure 2. one Torpedo development cost in 1999-2004

(4)From the discrete Verhulst model based on the linear time-variant and the ARIMA model, we can attain a new predictive model, simply named as combination model in which $y_1(t)$ stands for the gray model and $y_2(t)$ represents

ARIMA model. Based on the principle that smaller weight is for larger relative error and larger weight is for smaller one, there are two weight models in Eq.8: $\omega_1 = \frac{0.0428}{0.1765+0.0428} = 0.1952$; $\omega_2 = \frac{0.1765}{0.1765+0.0428} = 0.8048$. On the basis of Eq.10, the new combined forecasting model is $y_3(t) = f(t) = 0.1952y_1(t) + 0.8048y_2(t)$.

Table 3. comparison of the calculated results of four types of models

years	data	Gray Verhulst model[1,3]		LTDVM model		ARIMA model		Combined forecasting model	
		Simulation value	Relative error (%)	Simulation value	Relative error (%)	Simulation value	Relative error (%)	Simulation value	Relative error (%)
1995	496								
1996	1275	1119.115	12.226	1274.8908	0.0086	1275	0	1275	0
1997	2462	2116.017	14.053	2465.0106	0.1223	2462	0	2462	0
1998	3487	3177.486	8.876	3473.1565	0.3970	3487	0	3487	0
1999	3975	3913.739	1.541	3983.2205	0.2068	3973.7932	0.0304	3975.6334	0.0159
2000	4230	4286.182	1.328	4241.625	0.2748	4226.4001	0.0851	4229.372	0.0148
2001	4387	4444.806	1.318	4387.462	0.0105	4382.4613	0.1035	4383.4375	0.0812
2002	4497	4507.361	0.230	4487.3902	0.2137	4495.4905	0.0336	4493.9094	0.0687
2003	4584	4531.277	1.150	4575.9492	0.1756	4579.0858	0.1072	4578.4736	0.1206
2004	4663	4540.311	2.631	4671.3441	0.1789	4661.791	0.0259	4663.6558	0.0141
Average relative error(%)			4.817		0.1765		0.0428		0.0350

Attention: the relative errors in the chart take absolute value.

From the calculated result in Table 3, it is easy to see that the LTDVM model and the ARIMA model displays the best simulation effect and the average relative error is 0.035%, while the Gray Verhulst model shows the worst simulation and the average relative error is 4.817% [3].

4. CONCLUSION

The prediction result indicates that both the relative error of gray model based on the linear time variation and the relative error of ARIMA model are smaller than the one of traditional Verhulst model. These two models can attain good prediction effects. Through optimizing the combined model, the error of prediction result gradually diminishes. Besides, the predictive error of combined model is smaller than the individual prediction error. Hence, the prediction accuracy is bound to be improved.

5. REFERENCES

[1]. Sifeng Liu, Yaoguo Dang ,Zhigeng Fang and Naiming Xie, The Gray System Theory and Its Application, fifth ed., science press, Bei Jing,2010.
[2]. Ke Zhang, Sifeng Liu, Systems Engineering-Theory & Practice,30(2010)1650- 1657. In Chinese.
[3]. Lizhi Cui, Sifeng Liu, Zhiping Li, Systems Engineering and Electronics,33(2011)590-593.In Chinese.
[4]. Rui Dan, Shuhua Wang, Lingling Li and Donglian Gao, Journal of Liaoning Technical University (JCR Science Edition), 1(2012)118-121. In Chinese.
[5]. Yu Dai, The optimal combination forecast model construction and application. Economic mathematics, 1(2010) 92-98. In Chinese.
[6]. Qingwei Liang, Baowei Song,Yue Jia, Journal of System Simulation, 17(2005) 257-258. In Chinese.
[7]. Zhengxin Wang,Yaoguo Dang,Sifeng Liu, System Engineering-Theory & Practice, 29(2009)138- 144. In Chinese.
[8]. Mingxi Lu, Guangle Yan, Statistics and decision, 17(2009)52-53. In Chinese.
[9]. Weiguo Li, Aiqing Zhang, Statistics and Decision, 21 (2007)21-22. In Chinese.