The 4th International Conference on Quality and Reliability
August 9-11th, 2005
Beijing, China

A GREY MODELLING OF COVARIATE INFORMATION IN REPAIRABLE SYSTEM

RENKUAN GUO

Department of Statistical Sciences, University of Cape Town, Private Bag, Rhodes' Gift, Rondebosh 7707
Cape Town, South Africa

Covariate modeling including system functioning environment and system operating factors is regarded as an effective approach to explore the underlying mechanism of a reparable system. The covariate information is used for monitoring state changes to the functioning of the system so that the system performance (or capacity) can be judged and decisions on maintenance actions can adjusted accordingly. Cox (1972) initiated PH model, which quickly generated medical applications during 1980s but the industrial applications of the PH model were developed somewhat later, say, Dale (1985), Love and Guo (1991), Ascher and Kobbacy (1998) etc. Andersen, Borgan, Gill, and Keiding (1993) provided a solid theoretical foundation and an applicable methodology for covariate modeling based on counting processes. Guo and Love (2003) proposed a fuzzy covariate model for repairable system which is based on fuzzy logical function and static in nature. In this paper, we will investigate the system behavior under covariate influence in terms of grey system theory, specifically, the static *GM(0,h)* and dynamic *GM(1,h)* models. An industrial data is used for illustrative purpose.

Keywords: Covariate Information, PH models, Grey System Theory, *GM(0,h)*, *GM(1,h)*.

1 Introduction

It is argued in this paper any modeling exercise represents an abstraction from reality. The value of the resultant abstraction however lies in the ability to utilize the results to usefully gain insights or predict the behavior of the underlying real system.

Proportional hazards models proposed by Cox (1972) were the milestone of covariate modeling. Since then, multiplicative and additive intensities under the framework of the point processes were deeply investigated. As a matter of reflection, conditional monitoring technique-related models such as proportional intensity (hazards) models and their extensions are one of the fastest growth areas in reliability engineering research and industrial applications.

Modern statistical methodology is in a crisis whenever facing the challenge of very small sample points (as little as four) available, because the conventional statistical estimation theory is large-sample or based. Without any doubts, reliability engineering methodology is largely relying on the maximum likelihood theory. Researchers working in the applied fields have been fully aware of the contradiction between asymptotic theory and the reality of very data available.

The methodology to solve random uncertainty is probability and statistics, i.e., to treat data in terms of statistical laws and prior (probability) laws. Since statistical laws were established via large samples the more data is the better will be the analysis. Therefore, straightforward statistical modeling is often difficult to carry on and even not approachable (due to cost consideration).

Facing the challenge of data shortage, we have to walk out from the umbrella of large-sample based statistical theory by changing our attitude and looking at the real world from a different angle. Because system dynamics can be treated from the viewpoint of the degree of information availability, it is possible for us to walk out from the umbrella of large sample statistics and develop a small-sample based the spirit of grey system theory initiated by Deng (1982), which innovatively proposed the grey differential equation concept and utilized the conventional least-squares techniques for establishing the continuous response function from discrete data sequence. The Grey System Theory is successfully used to the analysis of uncertain systems that have multi-data inputs, discrete data, and insufficient data (Wen, 2004).

In grey system theory, models reflecting a system behavior under covariate influences were proposed, investigated by Deng and others, specifically, the static *GM(0,h)* and dynamic *GM(1,h)* models, and applied successfully into a wide range of applications, for example, in regional economic programming and multivariable control.

In this paper, we intend to explore the suitability of GM(1,h) model in repair improvement context since it is noticed that the estimated response function takes exponential forms about the *n*-dimensional region where data information are available and therefore it offers the potential improvements than that the multivariate regression which is a super-plane as well as a PH model could offer.

2 Grey Models *GM(1,h)* and *GM(0,h)*

It is also necessary to admit that there are still some fundamental concepts in grey system theory not rigorous defined, for example, grey system, grey numbers, haze sets and others, which are largely remaining in a philosophical domain. Just like the concept of time, it is an undefined concept in science. However, this never affected the wide usage of time from physics, chemistry, mathematics and other displines. In the same sense, the undefined grey system, grey number and haze set does not harm the practical values of the methodologies developed, which is *at least* can be regarded as *an efficient approximation* for extracting system dynamic information.

As Deng stated (1985) grey modeling is a differential model fitting approach, or equivalently, a differential dynamic model-building approach. A brief review of GM(1,h) and GM(0,h) is given based on Liu and Guo (1993) and Wen (2004).

Definition 2.1. Given the following time-indexed data sequences

$$\left\{X_k^{(0)}(i),\ i=1,2,\cdots,N\right\},\ k=1,2,\cdots,h, \tag{1}$$

where

$$X_1^{(0)} = \left\{ X_1^{(0)}(1), X_1^{(0)}(2), \cdots, X_1^{(0)}(N) \right\} \tag{2}$$

is called the characteristic data sequence of a system (Liu and Guo, 1993) or the main data sequence of a system (Wen, 2004), and

$$X_k^{(0)} = \left\{ X_k^{(0)}(1), X_k^{(0)}(2), \cdots, X_k^{(0)}(N) \right\}, k = 2, 3, \cdots, h \tag{3}$$

are called as the relevant (Liu and Guo, 1993) or influencing (Wen, 2004) factor sequences, or simply called as covariate sequences. Furthermore, the corresponding first-order accumulating generating (*1-AGO*) data sequences are defined by

$$\left\{ X_k^{(1)}(i), i = 1, 2, \cdots, N \right\}, k = 1, 2, \cdots, h, \tag{4}$$

where

$$X_k^{(1)}(i) = \sum_{j=1}^{i} X_k^{(0)}(j), \; i = 1, 2, \cdots, N, k = 1, 2, \cdots, h, \tag{5}$$

Also, the adjacent neighboring average sequence of the 1-AGO main data sequence $\left\{ X_1^{(1)}(i)_, i = 1, 2, \cdots, N \right\}$,

$$\left\{ Z_1^{(1)}(i)_, i = 2,, 3 \cdots, N \right\}, \text{where } Z_1^{(1)}(i) = \frac{1}{2}\left(X_1^{(1)}(i) + X_1^{(1)}(i-1) \right) \tag{6}$$

Then

$$x_1^{(0)}(i) + \alpha z_1^{(1)}(i) = \sum_{j=2}^{h} \beta_j x_j^{(1)}(i), \; i = 1, 2, \cdots, N \tag{7}$$

is called a *GM(1,h)* grey differential equation.

Definition 2.2. In the *GM(1,h)* grey differential equation defined by Equation (1) – (7), term α is called the developing coefficient of the system, terms $\beta_j x_j^{(1)}, j = 2, 3, \cdots, h$ are called the driving forces, and $\beta_j, j = 2, 3, \cdots, h$ are called the driving coefficients, $\underline{\alpha} = \left[\alpha, \beta_2, \beta_3, \cdots, \beta_h \right]^T$ is called the parameter vector of the system at the first-order differential level.

Definition 2.3. Let the parameter vector of the system at the first-order differential level be $\underline{\alpha} = \left[\alpha, \beta_2, \beta_3, \cdots, \beta_h \right]^T$, then the following first-order differential equation

$$\frac{dx_1^{(1)}}{dt}+\alpha x_1^{(1)}=\sum_{j=2}^{h}\beta_j x_j^{(1)} \tag{8}$$

is called the whitened differential equation (or the shadow differential equation) of the following grey *GM(1,h)* grey differential equation.

$$x_1^{(0)}\left(i\right)+\alpha z_1^{(1)}\left(i\right)=\sum_{j=2}^{h}\beta_j x_j^{(1)}\left(i\right),\ i=1,2,\cdots,N \tag{9}$$

Theorem 2.1. Given the data sequences defined in **Definition 2.1**, then the parameter vector $\underline{\alpha}$ of the grey *GM(1,h)* grey differential equation

$$x_1^{(0)}\left(i\right)+\alpha z_1^{(1)}\left(i\right)=\sum_{j=2}^{h}\beta_j x_j^{(1)}\left(i\right),\ i=1,2,\cdots,N \tag{10}$$

has a least-square estimator $\underline{a}=\left[a,b_2,b_3,\cdots,b_b\right]^T$

$$\underline{a}=\left(\mathrm{B}^T\mathrm{B}\right)\mathrm{B}^T\underline{Y} \tag{11}$$

where

$$\mathrm{B}=\begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_h^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_h^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(N) & x_2^{(1)}(N) & \cdots & x_h^{(1)}(N) \end{bmatrix} \tag{12}$$

and

$$\mathrm{Y}=\begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(N) \end{bmatrix} \tag{13}$$

Theorem 2.2. The solution to the estimated whitened differential equation

$$\frac{dx_1^{(1)}}{dt}+ax_1^{(1)}=\sum_{j=2}^{h}b_j x_j^{(1)} \tag{14}$$

is therefore,

$$\begin{aligned} x_1^{(1)}(t) &= e^{-at}\left[x_1^{(1)}(0)+\sum_{j=2}^{h}b_j\int_0^t x_j^{(1)}(s)e^{as}ds-\sum_{j=2}^{h}b_j\int_0^t x_j^{(1)}(0)ds\right] \\ &= e^{-at}\left[x_1^{(1)}(0)-t\sum_{j=2}^{h}b_j x_j^{(1)}+\sum_{j=2}^{h}b_j\int_0^t x_j^{(1)}(s)e^{as}ds\right] \end{aligned} \tag{15}$$

where $x_1^{(1)}(0)$ is the initial condition for function $x_1^{(1)}(t)$, $t=0$, and $\underline{a}=[a,b_2,b_3,\cdots,b_h]^T$ be the least-square estimator given in **Theorem 2.1.**

Remark 2.1. When the range of changes of $\sum_{j=2}^{h}b_j x_j^{(1)}$ is small enough, $\sum_{j=2}^{h}b_j x_j^{(1)}$ can be treated as grey constant, then the approximate time response function is

$$x_1^{(1)}(k+1)=\left[x_1^{(1)}(0)-\frac{1}{a}\sum_{j=2}^{h}b_j x_j^{(1)}(k+1)\right]e^{-ak}+\frac{1}{a}\sum_{j=2}^{h}b_j x_j^{(1)}(k+1) \tag{16}$$

where $x_1^{(1)}(0) \square x_1^{(0)}(1)$ and the filtered 1-IAGO is

$$\hat{x}_1^{(0)}(k+1)=\hat{x}_1^{(1)}(k+1)-\hat{x}_1^{(1)}(k),\ k=1,2,\cdots,N \tag{17}$$

Remark 2.2. The suitability of the prediction capability of *GM(1,h)* model is questionable. Some researchers use this model for the exploratory step for the grey relational analysis between main factor and the covariates. For example, Wen (2004) simply estimated the parameter vector $\underline{a}=[a,b_2,b_3,\cdots,b_h]^T$ and seeking the ordering among the absolute value of the driving parameters $\{|b_2|,|b_3|,\cdots,|b_h|\}$. It is obvious that the largest one $\max\{|b_2|,|b_3|,\cdots,|b_h|\}$ poses the heaviest influence on $x_{(1)}$.

Remark 2.3. The static model *GM(0,h)* can be treated similarly. The formality looks like a multivariate linear regression. However, it should emphasized here the linear regression is based on the original data sequences $X_k^{(0)}=\left\{X_k^{(0)}(1),X_k^{(0)}(2),\cdots,X_k^{(0)}(N)\right\},\ k=2,3,\cdots,h$ while the GM(0,h) model is working based on the 1-AGO data sequences $\left\{X_k^{(1)}(i),\ i=1,2,\cdots,N\right\},\ k=1,2,\cdots,h$. Therefore, the fluctuation in data sequences will be effectively eliminating from AGO operation (at integral level) and the dynamics will be revealed from different view of point.

3 An Repairable System Example

The data will be used to illustrate our covariate models is the one collected from a Cement Plant Roller Mill in Canada. The original data set was used and analyzed for different type of imperfectly repaired models with covariates by Love and Guo (1990) and Guo and Love (1993, 1995, and 1996). It is noticed that the estimated response function has exponential components, like $e^{\theta y}$ and therefore, it is better to normalize all covariates first. It is noticed that in evaluating repair impacts Chen (2004 and 2005) used the recorded maximum repair cost as denominator for normalization. Following Wen (2004), for covariate *D, B, W,* we also use maximum values as normalization factors respectively. The advantage of this data generation is that all the covariate data sequences will be strictly positive. The normalized (covariates) data set will be shown in Table 1.

Table 1: Cement Plant Roller Mill Operating and Maintenance Normalized Data (11/1988 - 03/1989).

Time from Last PM	*Maintenance type*	*Cost*	*D*	*B*	*W*
54	pm	0.0468986	0.75	0.5263157	0.6153846
133	failure	0.0716086	0.8125	0.8421052	0.9230769
147	pm	0.1512859	0.9375	0.6315789	0.7692307
72	failure	0.1195159	0.75	0.7894736	0.8461538
105	failure	0	0.8125	0.8421052	0.9230769
115	pm	0.2647504	0.6875	0.6842105	0.6923076
141	pm	0.2486132	1	0.6842105	0.7692307
59	failure	0.2153303	0.5	0.8421052	0.8461538
107	pm	0.0242057	0.5625	0.5789473	0.6153846
59	pm	0.5622793	0.5	0.5263157	0.6923076
36	failure	0.1795259	0.6875	0.6842105	0.7692307
210	pm	0.1926374	0.5	0.5263157	0.6153846
45	failure	0.0186585	0.625	1	1
69	pm	0.0645486	0.75	0.7368421	0.8461538
55	failure	0.0186585	0.8125	0.9473684	0.9230769
74	pm	0.0468986	0.9375	0.6315789	0.6153846
124	failure	0.3706505	0.75	0.8947368	0.8461538
147	failure	1	0.8125	0.8421052	0.8461538
171	pm	0.1765002	0.6875	0.6842105	0.6923076
40	failure	0.0045385	0.8125	0.8421052	0.8461538
77	failure	0.6364094	0.875	0.8947368	0.8461538
98	failure	0.0716086	0.75	0.7894736	0.8461538
108	failure	0.0842158	0.75	0.7894736	0.8461538
110	pm	0.2304589	1	0.7368421	0.8461538
85	failure	0.0837115	0.5	1	1
100	failure	0.0726172	0.75	0.7894736	0.7692307
115	failure	0.0121028	0.8125	0.8421052	0.9230769
217	pm	0.2390317	0.5625	0.5789473	0.6923076
25	failure	0	0.9375	0.9473684	0.9230769
50	failure	0.3721633	0.6875	0.6842105	0.8461538
55	pm	0.0600100	0.5	0.5263157	0.6153846

The correlation matrix of normalized *cost, D, B, W* is calculated and shown in Table 4. It is obvious that The correlation between *B* and *W* are very high so that only three covariates cost, *D* and *W* will be investigated.

Table 2: Correlation matrix of normalized *cost*, *D*, *B*, and *W*.

1.00	-0.01	-0.05	-0.07
-0.01	1.00	0.26	0.25
-0.05	0.26	1.00	*0.91*
-0.07	0.25	*0.91*	1.00

Therefore, our further data analysis only contains *cost, D and W* since the high correlation between *B* and *W*.

We first use the Toolbox *GM(1,N)* as well as *GM(0,3)* provided by Wen (2004). The normalized data set is divided into four groups with 8, 8, 8, and 7 data points respectively.

Table 3: *GM(1,3)* Fitting (4 Groups) Results.

group	*1*	*2*	*3*	*4*
Development coefficient α	1.8300	1.4825	-0.2712	-0.5582
Covariate *cost*	1.3694	0.0088	0.4755	-4.1113
Covariate *B*	-0.3390	-1.1041	-0.4594	-1.4061
Covariate *W*	1.1902	1.6910	0.1853	1.6674

Table 4: *GM(0,3)* Fitting (4 Groups) Results.

group	1	2	3	4
Covariate *cost*	1.1289	-0.2302	0.2978	0.9512
Covariate *B*	-0.8364	-0.2988	0.0804	0.3218
Covariate *W*	1.1625	0.7709	0.3823	0.1107

From the viewpoint of maintenance planning, the *GM(1,3)* coefficients of covariate *B* and *W* obtained four groups gave us very consistent results in their relative scale as well as the signs of the estimators. However, the covariate *cost* in different group painted quite different pictures. In Group 1, the cost positively affected the functioning time; in Group 2, the cost has no impacts on functioning times; in Group 3, cost factor negatively affected functioning time; and in Group 4, cost factor positively affected the functioning time strongly. Therefore, the *GM(1,3)* confirms to the classical PH analysis in covariate *B* and covariate *W*. The cost-repair link does not get any confirmation from *GM(1,3)* analysis.

However, the static *GM(0,3)* results in cost factor show us different pictures, the coefficients in Group 1,3, and 4 are all positive and competitive with other covariates while in Group 2, the cost coefficient is negative and can only be treated as minor influence. It is further noticed some inconsistencies between *GM(0,3)* and *GM(1,3)* results.

Finally, as expected, *GM(h,N)* analysis is a grey relationship analysis and not for prediction purpose although it possesses certain predictive power. The observations on cost factor inevitably challenge the research results using normalized cost to measure repair improvements in the literature.

4 Conclusion

In this paper, we reviewed the basic theory in *GM(h,N)* models and applied them into an industrial example. Although the we did not obtain the definite relationship between system functioning times and repair costs, the current *GM(h,N)* analysis is still offering us some deep insights on covariate information modeling in repairable system. The research makes us realizing that a predictive grey covariate model should be explored. One of the possible improvements is to use G(1,1) model to find the grey trend equations for covariates and then combine them into equation (15) in order to increase accuracy.

References

1. D. R. Cox. Regression models and life tables. *Journal of Royal Statistical Society*, *B26*, 187-220, 1972.
2. R. Guo and C. E. Love. Reliability Modeling with Fuzzy Covariates, *International Journal of Reliability, Quality and Safety Engineering*, Vol. 10, No. 2, 131-157, 2003.
3. C. E. Love and R. Guo. Using Proportional Hazard Modeling in Plant Maintenance, *Quality and Reliability Engineering International*, Vol. 7, 7-17, 1991.
4. J. L. Deng. Control Problems of Grey Systems. *Systems and Control Letters*, Vol. 1, No 6, March, 1982.
5. L. Fu. *Grey System Theory and Applications*, Beijing: The Publishing House of Science and Technology, 1991 (in Chinese).
6. R. Guo and C. E. Love. Grey Repairable System Analysis. *Proceedings of the European Safety and Reliability Conference*, 27-30 June, 2005, Tri City, Poland, 2005[a] (Invited paper as Plenary Lecture).
7. K. L. Wen. *Grey Systems: Modelling and Prediction,* Taiwan: Yang's Scientific Research Institute, 2004.