



The Generalized Estimating Equations for the Unknown Correlation Structure of the Data

By Md. Soyebur Rahman, Shayla Naznin & Gowranga Kumar Paul

Jahangirnagar University

Abstract- In many study the data are taken different period of time and the information about them is gathered relating to an event of concern at different time periods. The data are taken different time period are correlated. Regression analysis based on the Generalized Estimating Equation (GEE) is an increasing important method of such data. The Generalized Estimating Equation is an important and widely used approach in such analysis. Since the true correlation is unknown GEE offers to take a working correlation for analysis. In our study we consider four common correlation structure namely, independent, exchangeable, pair wise, autoregressive.

In the study the data are taken from the Dhaka stock exchange (DSE) this data are highly correlated. At first we apply different methods of estimating parameter the we apply GEE for estimating the parameters. Finally we get the GEE gives better estimate than any other method.

Keywords: *generalized estimating equation, GEE, OLS, GLM.*

GJCST-G Classification : *G.1.5, G.1.6*



Strictly as per the compliance and regulations of:



The Generalized Estimating Equations for the Unknown Correlation Structure of the Data

Md. Soyebur Rahman ^α, Shayla Naznin ^σ & Gowranga Kumar Paul ^ρ

Abstract- In many study the data are taken different period of time and the information about them is gathered relating to an event of concern at different time periods. The data are taken different time period are correlated. Regression analysis based on the Generalized Estimating Equation (GEE) is an increasing important method of such data. The Generalized Estimating Equation is an important and widely used approach in such analysis. Since the true correlation is unknown GEE offers to take a working correlation for analysis. In our study we consider four common correlation structure namely, independent, exchangeable, pair wise, autoregressive.

In the study the data are taken from the Dhaka stock exchange (DSE) this data are highly correlated. At first we apply different methods of estimating parameter the we apply GEE for estimating the parameters. Finally we get the GEE gives better estimate than any other method.

Keywords: *generalized estimating equation, GEE, OLS, GLM.*

I. INTRODUCTION

The term Generalized Estimating Equations indicates that an estimating equation is not the result of a likelihood-based derivation, but that it is obtained by generalizing other estimating equation. Liang and Zeger (1986), Zeger and Liang (1986) and prentice (1988) developed a most recent method of estimating the parameters of the marginal model. They present a class of estimating equations that take the correlation into account to increase the efficiency. This class of estimating equations is called Generalized Estimating Equations (GEE) and the correlation matrix is called working correlation matrix. The name working is used in the sense that it is an approximate correlation matrix of Y_i 's (the response variable). The estimates of β obtained by GEE are consistent and in addition

consistent variance estimates can be obtained under the weak assumption the weighted average of the estimated correlation matrices converge to a fixed matrix.

II. DATA ANALYSIS

The stock exchange data are highly correlated from one company to other company.

In our study, we have used stock exchange data because, we want to compare the different method of estimation. We have taken six company and their closing price and the general index in 2009 of DSE. In our calculation we have taken general index as an explanatory variable and the different company closing price as a independent variable.

We want to estimate the following model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_6 X_6 \quad (1)$$

where,

y = DSE general index 2009

X_1 = Aims first guaranteed mutual fund

X_2 =ACI pharmaceuticals Limited

X_3 =Glasco Smith-kline Bangladesh Limited

X_4 =Beximco Pharmaceuticals Limited

X_5 =Al-Arafah Islami Bank Limited

X_6 =Bata Shoe Company (Bangladesh) Limited.

Calculation of the model (1) in the different methods as follows:

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-101.01708	244.3919221	-0.413341	0.679737	-582.5174	380.48328
X Variable 1	-27.610049	7.774982161	-3.55114	0.000464	-42.9283	-12.291798
X Variable 2	-7.917902	1.068852888	-7.40785	2.34E-12	-10.02375	-5.8120506
X Variable 3	2.2945252	0.397909548	5.766449	2.56E-08	1.510565	3.0784856
X Variable 4	13.603066	1.598677955	8.508947	2.16E-15	10.45336	16.752778
X Variable 5	3.2163438	0.279808985	11.49478	1.61E-24	2.665065	3.7676227
X Variable 6	1.6285946	0.833022998	1.955042	0.051774	-0.012625	3.2698144

Author ^α: Jahangirnagar University, Savar, Dhaka. e-mail: soyebur.rahman@gmail.com

Author ^σ: Lecturer, Department of Statistics, Mawlana Bhashani Science and Technology University.

Author ^ρ: Assistant Professor, Department of Statistics, Mawlana Bhashani Science and Technology University.

Using the GLM procedure

The GLM Procedure

Standard

Parameter	Estimate	Error	t Value	Pr > t
Intercept	-97.97753227	244.4865048	-0.40	0.6890
x1	-27.79733308	7.7704859	-3.58	0.0004
X2	-7.88169809	1.0654488	-7.40	<.0001
X3	2.29819026	0.3979271	5.78	<.0001
X4	13.56079439	1.5965023	8.49	<.0001
X5	3.20755189	0.2794746	11.48	<.0001
X6	1.63484104	0.8332389	1.96	0.0509

Using GEE procedure

The GENMOD Procedure

Analysis of Initial Parameter Estimates

Standard Wald 95% Confidence Chi-						
Parameter	DF	Estimate	Error	Limits	Square	Pr > ChiSq
Intercept	1	-97.9775	240.8947	-570.122	374.1674	0.17 0.6842
x1	1	-27.7973	7.6563	-42.8035	-12.7912	13.18 0.0003
X2	1	-7.8817	1.0498	-9.9393	-5.8241	56.37 <.0001
X3	1	2.2982	0.3921	1.5297	3.0667	34.36 <.0001
X4	1	13.5608	1.5730	10.4777	16.6439	74.32 <.0001
X5	1	3.2076	0.2754	2.6678	3.7473	135.68 <.0001
X6	1	1.6348	0.8210	0.0257	3.2440	3.97 0.0464
Scale	1	162.6532	7.4241	148.7342	177.8748	

III. SUMMARY AND CONCLUSION

Selection of correlation structure is an important issue in Generalized Estimating Equation (GEE). We have earlier stated that there are four correlation structure namely, independent, exchangeable, autoregressive and pair wise. One needs to select under which correlation structure is unknown GEE works well or provides efficient estimate in several situation.

We have concentrated on standard error as a measure of accuracy for an estimator of the parameter. So in this case, we look only on the standard error under different correlation structure of different methods of estimation procedure.

In this study we consider stock exchange data, this data are highly correlated.

In the previous situation, we see that the standard error of the parameters of the GEE method is lowest than any other method. From this study we may conclude that from different method of estimation the GEE parameters are gives the efficient estimate and best approach.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Chang, Yue-Cune. 2000. "Residual analysis of the Generalized Linear Models for Longitudinal Data." Statistics in medicine 19: 1277-1293.
2. Hardin, J. W., and J.M. Hilbe. (2003). *Generalized Estimating Equations*. U.S.A.: Chapman and Hall/CRC.
3. Liao, T., F. (1994). *Interpreting Probability Models Logit, probit, and other Generalized Linear Models*. New Delhi: Sage publication, inc.
4. Liang K-Y. and S.L., Zeger (1986). "Longitudinal Data Analysis Using Generalized Linear Models". *Biometrika*, 73:13-22.
5. Mc Cullagh, P. and J. A. Nelder, (1989), *Generalized Linear Models*, 2nd edition, Chapman and Hall, London.
6. Neter, J.; W. Wasserman; and M. H. Kutner. (1990). *Applied Linear Statistical Models* :
7. *Regression, Analysis of variance, and Experimental Designs*. 3rd edition. U.S.A.: Richard D. Irwin, Inc.
8. Shukla, (2007). *Selecting the Working Correlation Structure in Generalized Estimating Equation (GEE) with an Application to the Maternal Morbidity Data*. Unpublished thesis, Dhaka University.