Proceeding The 1st IBSC: Towards The Extended Use Of Basic Science For Enhancing Health, Environment, Energy And Biotechnology
ISBN: 978-602-60569-5-5

Analysis of Simultaneous Equation Model (SEM) on Non normally Response used the Method of Reduce Rank Vector Generalized Linear Models (RR-VGLM)

Miftahul Ulum[1], Alfian Futuhul Hadi[2], Dian Anggraeni[3]
[1]Mathematic depart,Jember University, Jember, Indonesia, ulummath10@gmail.com
[2]Mathematic depart,Jember University, Jember, Indonesia, afhadi@unej.ac.id
[3]Mathematic depart,Jember University, Jember, Indonesia,anggraeni@gmail.com

***Abstract*—**Multivariate linear regression is a statistical analysis methods used to data in the case of multiple response variables associated with several predictor variables. In this method of analysis there is an assumption of matrix coefficient regression must be full rank. In the case of simultaneous equations, full rank condition is not fulfilled. Consequently, to analyze that case is not possible because it would produce a regression coefficient that is very large so needed reduction in rank. Reduce Rank Regression (RRR) Method is an alternative method in the case where this method if there is a weak regression coefficients will be cut. However, Reduced rank regression method only applies in response which continuous and normal in econometric data analysis and others. Therefore, to overcome that problem so introduced to f analysis method of Reduce Rank Vector Generalized Linear Model (RR-VGLM). This article will discuss simultaneous equations with non-normal variable response using RR-VGLM by simulating non normal conditions.
Key word : regression coefficient, full rank, reduce rank reggession, RR-VGLM

INTRODUCTION

Often, econometric make a difficult the researcher because it needs the big data and the variable with other variables. One of solution to overcome the problem is Simultaneous equation model (sem). Sem relates a set of endogenous variables to a set of exogenous variables with error variables. In this method, endogenous variables can become exogenous variables that endogenous variables have a possibility to relate exogenous variables. For those complexities, the researcher is interested to use this method. Yet, Anderson said that there is the restriction for coefficient matrix of regression which nontrivial must be reduced. This reduce method is named as Reduce Rank Regression.

Reduce Rank Regression (RRR) is made a popular firstly by Anderson and developed by Izeman on 1975. In this developing, Izeman verified RRR model detail. The reduced rank regression model is a multivariate regression model with a coefficient matrix with reduced rank from low dimension till achieving Full Rank dimension. In the case of economy data, there is possibility a correlation between a variable with another variable in which it causes relation on both sides between correlation variable and predictor variable. Because of relation on both sides, it can cause the number of parameter of regression matrix is too big and cannot be reached. So for the case of low rank assumption, it requires the existence of reduce on the parameter of coefficient matrix of regression to achieve Full Rank condition. however, Yee said that RRR is not popular because it is only used to respond variable which has the characteristic of continually and normally.

In science development, the researcher has been discovered the new method which can do the case of respond variable which has the characteristic of un-continually and un-normally with low rank condition. This method Reduce Rank Regression for Vector Generalize Linear Model or it is known asReduce Rank VGLM (RR-VGLM). This method is very effective at all of models so the parameter can be reached at all of distributing and model. In this article, the researcher will research the study of RR-VGLMat the similarity of simultaneous nonnormally.

RESEARCH METHODOLOGY

This research uses data simulation assumed thenonnormal data. This data will be analyzed by using RR-VGLM method to make easierthe researcher on that research.

RESEARCH DATA

This research data in the form of simulation data where the data is not normal. Simulation is an imitation of the real world process or system. Simulation arise of the generation process as well as observation of the process to draw conclusions from that represented [5]. This simulation data generated with the help of program R i386 3.2.3.

IDENTIFICATION VARIABLES

Simulation data on this research using two types of variables consisting of several exogenous and endogenous variables. these two variables are interconnected between exogenous and endogenous variables with exogenous with exogenous to forge a simultaneous equation Model.

RESEARCH METHOD

The data in this study using a support package `VGAM`with function `rrvglm ()`and package `car` to discribe a dataon software R i386 3.2.3.

RESEARCH STEPS

Research steps that will be done are follow as :

Process of Simulation Data

To simplify the process of simulation data, using the assistance program R i386 3.2.3. at this stage is also carried out the transformation of data by using a uniform distribution of exogenous and endogenous variables negative binomial distribusition.

Estimasi Parameter RR-VGLM

Traditionally, the reduce rank regression model is estimated by maximum likelihood estimator or equivalent least squares methods[6]. In RR-VGLM estimated by Iteratively Reweighted Least Square (IRLS). At iteration a, one can minimize a residual sum of squares

$$\sum_{i=1}^{n}\left(z_i^{(a)} - \hat{B}_1^T x_{1i} - \hat{A}C^T x_{2i}\right)^T W_i \left(z_i^{(a)} - \hat{B}_1^T x_{1i} - \hat{A}C^T x_{2i}\right) \quad (1)$$

By fixing A and solving for $\nu = C^T$ and B_1, and then keeping ν to solve for A and B_1[3].

Determine RR-VGLM Model

RR-VGLM are defined by VGLMs. VGLMs are general class of regression models that are parameterized using M linear predictors. The explanatory variables At VGLMs affect all the parameters of the model through the linear predictor matrix so that there are constraints must be met. RRR is where a high rank matrix is approximated by a low rank matrix. When RRR is applied to a VGLM matrix of regression coefficients this result in a RR-VGLM. RR-VGLMs can also be explained using constraint matrices some of them are unknown and to be estimated. RR-VGLM model is $\eta = \mathrm{B}_1^{\mathrm{T}}\mathrm{x}_1 + \mathrm{B}_2^{\mathrm{T}}\mathrm{x}_2$, where approximate in $\mathrm{B}_2$by RRR is $\mathrm{B}_2 = CA^T$, where C and $A^T$are respectively matrix p_2 x R and M x Rwith two matrices have a lower rank so that$\eta = \mathrm{B}_1^T x_1 + A$ where $v = C^T x_2$is the vector of latent variables R[4].

DISCUSSION

Data Overview

In this research, the researcher used the data nonnormally. To describe the data, this research used box plot. It is a method to present the 5 – number summary which consist of the minimum and macimum range values, the upper and lower quartiles, and the median[7]. In addition to box plot, the data also described scatterplot in order to be seen the outlier of the data. Here, the researcher provides the data of box plot and scattlerplot integrated by Figure 1.

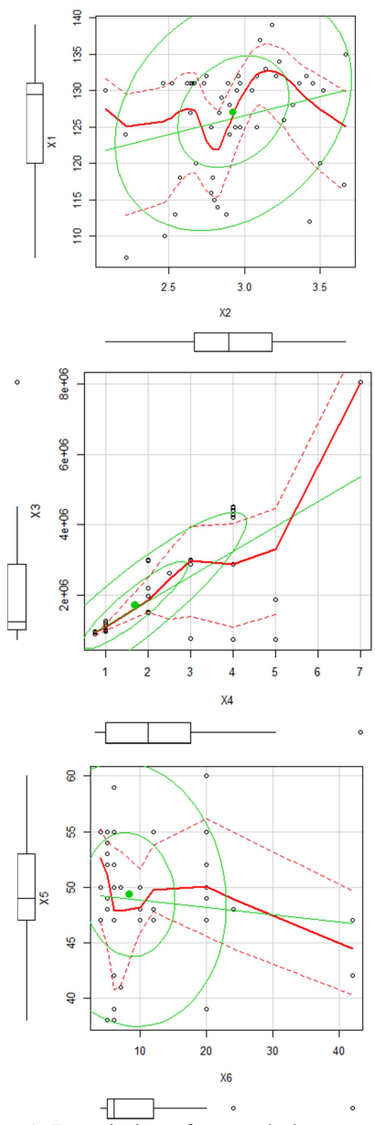

Figure 1. Description of research data

In the Figure 1, it describes the data of box plot and scatterplot. The box plot is beside the table and the scatterplot is under the table. The box plot of X1, X2, X3, X4, X5, and X6 describe that the data is nonnormal. It is described by the line which is the box unbalance. X4 and X6 is found the outlier. If X1, X2, X3, X4, X5, and X6 are seen from scatterplot, it will be found outlier at all of table. Outlier is signed with the data which is outside ellips so that the data is unnormal.

Estimasi parameter RR-VGLM

The Estimated parameter RR-VLGM has a similarity with VGLM in which it used the method of IRLS. IRLS is a very powerful and general technique for the numerical maximum likelihood estimation[3]. IRLS can minimize a residual sum of squares and residual deviance

$$\sum_{i=1}^{n}(z_i^{(a)} - \widehat{\mathrm{B}}_1^T x_{1i} - \hat{A}C^T x_{2i})^T W_i (z_i^{(a)} - \widehat{\mathrm{B}}_1^T x_{1i} - \hat{A}C^T x_{2i})$$

The process of iterasi is used by R program. Output produced R program consists of 2 fitting model on minimizing residual sum of squares like table 1 and table 2.

Tabel 1 Output IRLS in fitting model 1

Linear Loop	Residual Deviance	Alternating iteration	Convergen criterion	ResSS
1	123.01738	1	1	139.4236
		2	0.002928158	139.0165
		3	0.001892276	138.754
		4	0.001471087	138.5501
		5	0.001902354	138.2871
		6	0.002541416	137.9365
		7	0.001628632	137.7122
2	116.36072	1	1	136.7494
		2	8.600568e-05	136.7376
		3	1.003218e-05	136.7362
		4	1.405471e-06	136.736
		5	1.913326e-07	136.736
		6	2.576228e-08	136.736
3	115.72329	1	1	136.9631
		2	3.319392e-07	136.963
		3	2.680264e-08	136.963
4	115.71987	1	1	137.2254
		2	4.521947e-10	137.2254
5	115.71987			

Table 1 explains the process of IRLS. Liner loop shows the number of repeating in fitting first model. Residual deviance explained the variety of data. Alternative iteration is the number of iterasi in each looping. Convergen criterion is the criteria of the data convergen. ResSS is residual sum square. The method of IRLS aims to minimize residual sum square to get good model. Here, the researcher provides the tables to check on fitting model 1 which it is showed IRLS fitting model 2.

Tabel 2. Ouput IRLS fitting model 2

Linear Loop	Residual Deviance	Alternating iteration	Convergen criterion	ResSS
1	122.77317	1	1	138.2357
		2	0.001851332	137.9802
		3	0.001246191	137.8085
		4	0.001246594	137.6369
		5	0.0009244813	137.5098
		6	0.0003481003	137.4619
		7	7.403157e-05	137.4517
2	116.51573	1	1	135.0616
		2	4.601088e-05	135.0554
		3	1.211107e-06	135.0552
		4	1.313409e-07	135.0552
		5	1.52604e-08	135.0552
3	115.72432	1	1	136.8775
		2	1.883236e-06	136.8773
		3	2.666133e-07	136.8772
		4	4.064213e-08	136.8772
4	115.71987	1	1	137.2285

ISBN: 978-602-60569-5-5

5	115.7198 7	2	3.860883e-11	137.2285

In the table 2, the output fitting of second model is the result of checking from fitting of first model. In the Table 2, it is found the value of residual deviance which has similar value. It means that residual deviance in minimum condition on the research data.

Determine RR-VGLM Model

RR-VGLM Model is difined by

$\eta = B_1^T x_1 + CA^T x_2$

In this study, the model obtained with the help of program R. From the program R output obtained matrix A, B1, and C as follows.

$$A = \begin{bmatrix} 2.2391577 & 3.43536378 \\ -0.313160 & -0.04263375 \\ 0.2423782 & -0.50731254 \end{bmatrix}$$

$$C = \begin{bmatrix} 1.730844e-01 & 6.535991e-03 \\ 4.480700e-07 & 4.911587e-07 \\ 2.530474e-01 & -9.986844e-01 \\ -1.756726e-01 & 2.353295e-02 \end{bmatrix}$$

$$B1 = \begin{bmatrix} 0.85735158 & 0.12552290 & -0.37988421 \\ -0.05006834 & -0.05006834 & -0.05006834 \\ 0.12656251 & 0.12656251 & 0.12656251 \end{bmatrix}$$

The matrix is then obtained from the model as follows.

$$\eta = \begin{bmatrix} 0.85735158 & 0.12552290 & -0.37988421 \\ -0.05006834 & -0.05006834 & -0.05006834 \\ 0.12656251 & 0.12656251 & 0.12656251 \end{bmatrix} x_1 + \begin{bmatrix} 1.730844e-01 & 6.535991e-03 \\ 4.480700e-07 & 4.911587e-07 \\ 2.530474e-01 & -9.986844e-01 \\ -1.756726e-01 & 2.353295e-02 \end{bmatrix} \begin{bmatrix} 2.2391577 & 3.43536378 \\ -0.313160 & -0.04263375 \\ 0.2423782 & -0.50731254 \end{bmatrix}^T x_2$$

The characteristics of respondents using RRVGLM method can be known with the help of R program to obtain plots and biplot as Figure 2 and Figure 3 below,

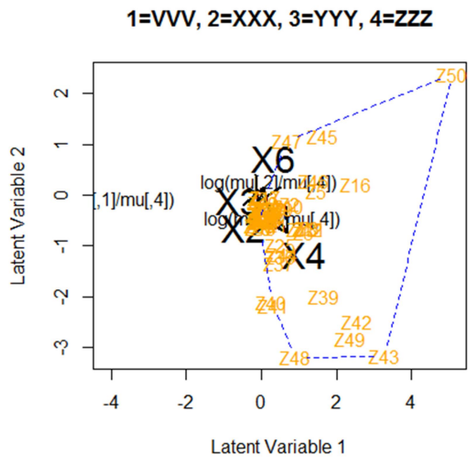

Figure 2. Plotting a respondent

Figure 2 represented a characteristic responden. A yellow line indicate characteristic of respondent that restriction the research object or variable. The most object on Figure 2 indicate that they have a similiarity characteristic of X_2, X_3, and X_4. It can be shown with the most respondent which have a group of X_2, X_3, and X_4.

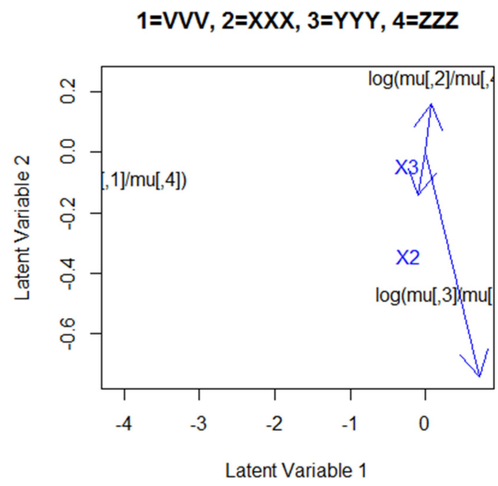

Figure 3. Biplot respondent with RRVGLM

According to Figure 3, X_2, X_3, and X_4 indicate the same direction. It mean that, X_2, X_3, and X_4 have similiarity of characteristic. Where as X_6 is in opposite with X_2, X_3, and X_4. It mean X_6 has value under averages. Other wise, X_6 have a negative correlation with X_2, X_3, and X_4 variable. In other word X_6 as know as weak variable in this research. To know the value from changing object every object, show that. It can the lenght of the vector in the biplot. Variable of big length vector. It means that the object with has variance that can be indicate can be indicate X_4 variable.

CONCLUSION

In the 2003 Yee make a research using RRVGLM method in a normal data. Yee use 111 data and get log likelihood value -41.79, residual deviance 83.58 and AIC 113.58. In other side this research, researcher use simulation data with asumption this data non normally as much as 50 respondent. This research get log likelihood -57.86, residual deviance 115.72 and AIC 145.58. based on the value of log likelihood AIC and residual deviance on the model, the result almost similar. So, the RR-VGLM method is one of the statistic method flexibility to analysis kind of the data.

REFERENCES

[1] Anderson, T.W. 2005. Reduced rank Regression for block of simutaneous equations. USA : Stanford University.

[2] Reinsel,G.C. & Velu,R.P.1998. *Multivariat Reduced-Rank Regression*. New york:Springer Science + Business Media.

[3] Yee, Thomas W. 2015. Vector Generalized Linear and Additive Models. New York :Springer.

[4] Yee, Thomas W. 2013. Reduced-rank vector generalized linear models with two linear predictors. New Zealand : Department of Statistics.

[5] Tirta, I made. 2003. Pengantar Metode Simulasi Statistika dengan Aplikasi R dan S+. Jember : universitas Jember.

[6] Schmidli,H. 1996. Bayesian Analysis of Reduce Rank Regression. Switzerland : Mathematical Applications Vol. 5 : 159-186.

[7] Potter, Kristin.___. Methods for Statistical Information :The Box Plot. Saltlake City : Univesity of Utah