13th Islamic Countries Conference on Statistical Sciences

PROCEEDINGS ICCS-13

Vol. 27

Editors:

Dr. Munir Ahmad & Dr. Shahjahan Khan

STATISTICS FOR BETTER LIFE

ISBN 978-969-8858-16-2

December 18-21, 2014

at

IPB International Convention Centre,

Bogor, Indonesia

Jointly organized by

Islamic Countries Society of Statistical Sciences

Lahore, Pakistan

IPB Bogor Agricultural University

Bogor, Indonesia

ORGANIZERS

44-A, Civic Centre, Sabzazar, Multan Road, Lahore, Pakistan

Email: secretary@isoss.net URL: www.isoss.net

Copyright: © 2015, Islamic Countries Society of Statistical Sciences.

Published by: ISOSS, Lahore, Pakistan.

"All papers published

in the Proceedings of

ICCS-13

were accepted after formal peer review

by the experts in the relevant field"

Dr. Munir Ahmad
Dr. Shahjahan Khan
Editors

CONTENTS

1. 033: Self-Inversion-Based Modification of Crow and Siddiqui's Coefficient of Kurtosis to Achieve Gains in Efficiency
   Saleha Naghmi Habibullah 1
2. 039: Modification in Kappa Statistics-A New Approach
   Sundus Noor Iftikhar and Nazeer Khan 9
3. 040: Development of Growth Charts of Pakistani Children Using Quantile Regression
   Sundus Iftikhar and Nazeer Khan 17
4. 041: D-Optimal Designs for Morgan Mercer Flodin (MMF) Models and its Application
   Tatik Widiharih, Sri Haryatmi and Gunardi 29
5. 042: The Performance of LS, LAD, and MLAD Regression on the Stack Loss Data
   Setyono, I Made Sumertajaya, Anang Kurnia and Ahmad Ansori Mattjik 41
6. 044: Structural Equation Modeling with Partial Least Square Estimator (Case Study in Measurement of the Level of Satisfaction Auditor of Green Audits Programs)
   Firmina Adlaida 55
7. 047: A Synthetic Control Chart Reexpression Vector Variance for Process Multivariate Dispersion
   Suwanda 63
8. 049: Bayesian Weibull Mixture Models for Dengue Fever
   Sri Astuti Thamrin, Andi Kresna Jaya and La Podje Talangko 73
9. 054: Spatial Analysis for the Distribution of Human Development Index and Regional Government Budget in East Java Province
   Vinna Rahmayanti Setyaning Nastiti, Muhammad Nur Aidi and Farit Mochammad Afendi 87
10. 061: Mobile Learning Based Flashlite in Statistics Course
    Artanti Indrasetianingsih and Permadina Kanah Arieska 97
11. 062: Comparison of Stochastic Soybean Yield Response Functions to Phosphorus Fertilizer
    Mohammad Masjkur 109
12. 065: Statistical Analysis for Non-Normal and Correlated Outcome in Panel Data
    Annisa Ghina Nafsi Rusdi, Asep Saefuddin and Anang Kurnia 119
13. 070: Multidimensional Deprivation Spectrum: An Alternate Route to Measure Poverty
    Taseer Salahuddin 131

14. 073: An Evaluation of the Performance of Local Polynomial Estimates in Generalized Poisson Regression Model
Erni Tri Astuti 139

15. 081: Determinants of Maternal Delivery-Care Services: Study from Rural Area of Bangladesh
Md. Shah Jahan, Ahm Shamsuzzoha and Hasina Akhter Chowdhury 145

16. 083: Spatial Regression Analysis to Evaluate Indonesian Presidential General Election
Alan Duta Dinasty, Asep Saefuddin and Yenni Angraini 157

17. 095: Comparison of Binary, Uniform and Kernel Gaussian Weight Matrix in Spatial Error Model (SEM) in Panel Data Analysis
M Nur Aidi, Tuti Purwaningsih, Erfiani and Anik Djuraidah 177

18. 098: Fast: A Web-Based Statistical Analysis Forum
Dwiyana Siti Meilany Dalimunthe, Debi Tomika, Eka Miftakhul Rahmawati, Muchriana Burhan, Yayan Fauzi, Muchammad Romzy, I Made Arcana, Imam Machdi, Usman Bustaman, Widyo Pura Buana and Setia Pramana1 185

19. 099: Wires: A User Friendly Spatial Analysis Software
Meidiana Pairuz, Nur'aidah, Hanik Devianingrum, Hergias Widityasari, Zumrotul Ilmiyah, Isna Rahayu, Arinda Ria, Diah Daniaty, Wahyu Hardi Puspiaji, Erma Purnatika Dewi, Pudji Ismartini, Robert Kurniawan, Sodikin Baidowi, Muchammad Romzi, Munawar Asikin, Karmaji and Setia Pramana 201

20. 105: Education Mapping in Indonesia using Geographically Weighted Regression (GWR) and Geographic Information Systems (GIS)
Robert Kurniawan and Marwan Wahyudin 213

21. 106: Row-Column Interaction Models for Zero-Inflated Poisson Count Data in Agricultural Trial
Alfian F. Hadi and Halimatus Sa'diyah 233

22. 110: Laparoscopic vs Open Repair for Incisional Hernia: Meta-analysis and systematic review of laparoscopic versus open mesh repair for elective incisional hernia
Aiman Awaiz, Foyzur Rahman, Md Belal Hossain, Rossita Mohamad Yunus, Shahjahan Khan, Breda Memon and Muhammed Ashraf Memon 245

23. 112: Testing the equality of the two intercepts for the parallel regression model
Budi Pratikno and Shahjahan Khan 263

24. 121: Exploration Interest of Jambi Community About Baitul Maal Wattamwil (BMT) by Using Regression Logistics Binary Analysis
Titin Agustin Nengsih 277

25. 126: Forecasting Spare Parts Demand: A Case Study at an Indonesian Heavy Equipment Company
Ryan Pasca Aulia, Farit Mochamad Afendi and Yenni Angraini 289

26. 046: Estimation of Demographic Statistic of Pest Aphis glycines by Leslie Matrix and Lotka-Euler Equation Based on Jackknife Resampling
Leni Marlena, Budi Susetyo and Hermanu Triwidodo 299

27. 141: Simultaneous Analysis of the Lecturers Positioning and Students Segmentation in the Selection of Thesis Supervisor
Yusma Yanti, Bagus Sartono and Farit M Afendi 311

28. 143: Classification of Dropout Student in Sulawesi with Bagging CART Methods
Dina Srikandi and Erfiani 319

29. 145: Optimizing Classification Urban / Rural Areas in Indonesia with Bagging Methods in Binary Logistic Regression
Shafa Rosea Surbakti, Erfiani and Bagus Sartono 327

30. 146: The Use of Lagrange Multiplier to Ensemble Two Return Values of Generalized Pareto and Modified Champernowne Distributions
Aji Hamim Wigena, Cici Suhaeni and Deby Vertisa 337

31. 148: Demographic Transition in Bangladesh: Evidence from Population and Housing Census 1981-2011
Md. Mashud Alam, Md. Shamsul Alam and Amjad Hossain 343

32. 151: Clusterwise Linear Regression by Least Square Clustering (LS-C) Method
Megawati Suharsono Putri, Bagus Sartono and Budi Susetyo 353

33. 152: Forecasting of Paddy Production in Indramayu Regency with Transfer Function Model
Rena Foris Windari and Erfiani 363

34. 154: E-M Algorithms Method for Estimating Missing Data in Regression Analysis
Septian Rahardiantoro and Bagus Sartono 371

35. 160: The Classification of District or City of the Poverty Data Based on Indicator of the Community Welfare in Sumatera with the Vertex Discriminant Analysis and Fisher Discriminant Analysis
Nurmaleni, I Made Sumertajaya and Bagus Sartono 375

36. 132: Institutional Framework of Poverty Reduction in Pakistan
Nadeem Iqbal and Rashda Qazi 385

37. 052: Applying SEM to Analyze the Relationship between Loyalty, Trust, Satisfaction, and Quality of Service for Students in Mathematics Study Program, State University of Makassar
Sukarna, Aswi and Sahlan Sidjara 393

38. 057: Empirical Bayes Method to Estimate Poverty Measures in a Small Area
Dian Handayani, Anang Kurnia, Asep Saefuddin and Henk Folmer 403

39. 069: Analysis of Appendectomy in Belgium Using Disease Mapping Techniques
Mieke Nurmalasari and Setia Pramana 417

40. 076: Comparison of Binary, Uniform and Kernel Gaussian Weight Matrix in Spatial Autoregressive (SAR) Panel Data Model
Tuti Purwaningsih, Dian Kusumaningrum, Erfiani and Anik Djuraidah 431

41. 096: Employee Innovation: Management Practices Affecting the Innovative Behavior at Workplace
Samiah Ahmed, Munir Ahmad and Suleman Aziz Lodhi 439

42. 140: Small Area Estimation for Non-Sampled Area Using Cluster Information and Winsorization with Application to BPS Data
Rahma Anisa, Khairil A. Notodiputro and Anang Kurnia 453

43. 150: Grouping of Public Welfare in Provinsi Aceh Pricipal Component Analysis
Winny Dian Safitri, Erfiani and Bagus Sartono 463

44. 162: Comparison of Method Classification Artificial Neural Network Back propagation, Logistic Regression, and Multivariate Adaptive Regression Splines (Mars) (Case Study Data of Unsecured Loan)
Siti Hadijah Hasanah, Kusman Sadik and Farit Mochamad Afendi 477

45. 170: Impact of Corporate Governance on Firm Financial Performance in Financial Sector of Pakistan
Nadeem Iqbal, Rashda Qazi and Nabila Khan 487

46. 048: The Dynamics Approach for Regional Disparity in Central Sulawesi After Decentralization Policy
Krismanti Tri Wahyuni 493

Proc. ICCS-13, Bogor, Indonesia
December 18-21, 2014, Vol. 27, pp. 233-244

ROW-COLUMN INTERACTION MODELS FOR ZERO-INFLATED POISSON COUNT DATA IN AGRICULTURAL TRIAL

Alfian F. Hadi[1] and **Halimatus Sa'diyah**[2]
[1] Department of Mathematics, The University of Jember, Indonesia. Email: afhadi@unej. ac.id
[2] Department of Agronomy, The University of Jember, Indonesia

ABSTRACT

Many zero observations makes some difficulties and fatal consequence in Poisson modeling and its interpretation. We consider to facilitates the analysis of two-way tables of count with many zero observations in agricultural trial. For example, in counting the pest or disease in plants. Plants that have no sign of attack, can occur because of two things, it could be resistant, or simply there is no spore disease (no endemics) or no pest attack. This is the difference between inevitable structural zero or sampling zero that is occurring according to a random process.

This paper describes a statistical framework and software for fitting row-column interaction models (RCIMs) to two-way table of count with some Zero observations. RCIMs apply some link function to the mean of a cell equaling a row effect plus a column effect plus an interaction term is modeled as a reduced-rank regression with rank of 2, then will be visualized by biplot. Therefore its potentially to be develop become AMMI models that accommodate ZIP count.

KEYWORDS

ZIP, AMMI Models, Row-Column Interaction Models, SVD Reparameteri-zation.

1. INTRODUCTION

The Poisson distribution is widely used in quality studies for count related data. Poisson regression models are basically modelling for counts. There are two strong assumptions for Poisson model to be checked: one is that events occur independently over of time or exposure period, the other is that the conditional mean and variance are equal. In practice, the Poisson with a large numbers of count, usually have greater variance than the mean are described as overdispersion. Poisson with smal value of mean, it also have small value of the variance, in this case, count data encounter with value of zero problems. This indicates that Poisson regression is not adequate. There are two common causes that can lead to overdispersion are additional variation to the mean or heterogeneity, an Negative Binomial model is often used and other cause counts with excess zeros or zero-infated. Poisson counts, since the excess zeros will give smaller conditional mean than the true value, this can be modeled by using zero-inflated Poisson (ZIP).

The proper model is needed to present a valid conclusion from the data count by the zero-inflation. Various applications related to this. For example, in calculating the pest or disease in plants. Plants that have no sign of attack, can occur because of two things, it could be as resistant to disease, or simply because there is no disease spores (no endemics) or no pest attack there. This is the difference between a structural zero, the inevitable, and zero sampling that is occurring according to a random process.

2. ZERO-INFLATED AND ITS CONSEQUENCES

Ignoring zero-inflation, especially when a sizeable proportion of the data is zero, implies that the underlying distributional assumptions will not be met. This will more than likely affect the results of an analysis, and hence lead to incorrect conclusions concerning the data. In addition to accounting for zero-inflation, we also need to consider the possibility of over-dispersion, which is variation larger than would be expected under the distributional assumptions. This is a commonly occurring phenomenon with Poisson models, and if ignored, can lead to underestimated standard errors and hence misleading inference about regression parameters (Hinde & Demetrio, [5]). Both zero-inflation and over-dispersion can occur simultaneously in a data set.

3. HANDLING ZERO INFLATED ON AN ADDITIVE MODEL

Suppose that yi is the number of occurrences of an attribute or event and xi ϵ R is a vector of covariates, both recorded for each of i = 1, . . . ,m sites. The simplest approach to modeling the relationship between yi and xi is an ordinary least squares fit of the transformed response, such as $\sqrt{yi}$ or log(yi). However, such transformations are not helpful when the data contain many zeros because the zeros are unchanged under a square-root transformation and are undefined under a logarithmic transformation. Additionally, the underlying distributional assumptions of linearity, homoscedasticity and Gaussianity do not hold so this approach is not suitable in this context.

A better approach is to fit a Poisson generalised linear model (GLM) with Log - link function. We can model over-dispersion by a proper variance function (McCullagh & Nelder [10]). However, the overall fit can be poor because there are typically many more zeros in the data than expected under a Poisson model which allows for over-dispersion.

Another approach to modeling zero-inflated count data concerns the classification of the zero observations into two different groups. Distributions which classify their zero counts in this way have been referred to as zero-modified distributions, distributions with added zeros, zero-inflated distributions or mixture distributions. In these distributions, the zeros are classified into two groups: one group, which along with the positive counts are modeled by a discrete distribution such as the Poisson or negative binomial distribution, occur with probability $1 - \omega$; and the other group, which represent the 'extra' zeros, occur with probability ω. Dietz & Bohning [3] discuss estimation of the parameter in zero-modified Poisson distributions and for illustration, they analyse counts recorded in a dental epidemiological study. Lambert [7], Welsh et al [15], and Bohning [1], also discuss applications of this approach to modeling zero-inflated count data.

4. INTRODUCING ROW-COLUMN INTERACTION MODEL FOR ZERO INFLATED

Zero-inflated models often found in additive models, but less in the model of interaction. In the study of pest/diseases, analyses concern with the interaction between genotype and environmental influences (GEI). Crops that have no sign of attack, can occur because of two things, it could be resistant, or simply there is no spore disease (no endemics) or no pest attack. This is the difference between inevitable structural zero or sampling zero that is occurring according to a random process.

In this research area of the GEI, AMMI model said to be most powerful one to analyzed the GEI, by main effects plus multiplicative interaction terms. Nowadays, AMMI model has been developed to be more generalized, named GAMMI. It can handle Poisson count as well. AMMI model is basically presents interaction through dimension reduction techniques. Here, it is very important to introduce a statistical methodology handling problems with inflation-zero count. It will be possible to model the zero count as an expression of resistance and not because chance did not affected.

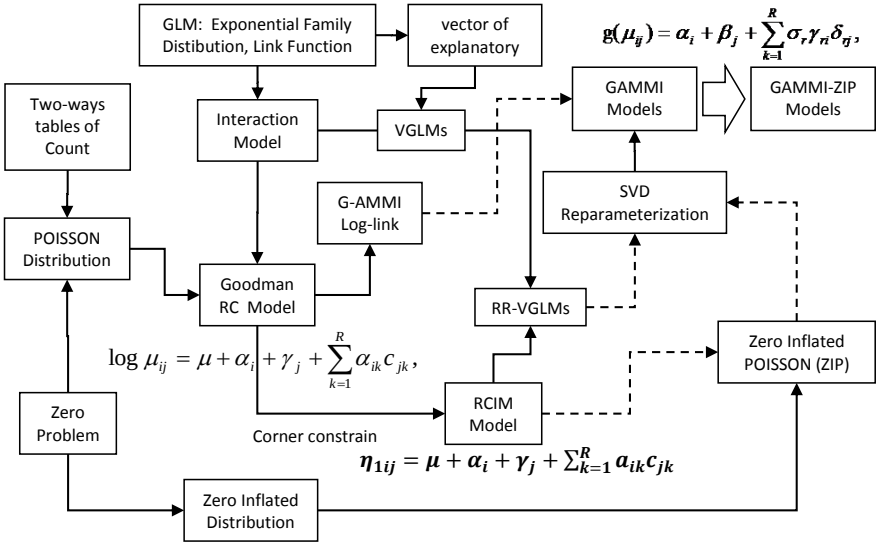

Fig. 1. TheFramework

This section will discuss the framework (Figure 1) by something's related to the development of zero-inflated in the multiplicative model. Beginning with the concept of mixture distribution of zero-inflated Poisson, and the regression framework terampat by reduction of dimensions or Reduced Rank Vector Generalized Linear Models (RR-VGLMs) which was introduced by Yee and Hastie [17]. Then the Row-Column Interaction Model (RCIM) Yee & Hadi [16].

4.1 The Zero-Inflated Poisson Distribution

A (discrete) random variable Yi is said to have a zero-inflated distribution if it has value 0 with probability ω, otherwise it has some other distribution with P(Y = 0) > 0. Hence P(Y = 0) comes from two sources, and the ω source can sometimes be thought of as a structural zero. The most famous of Zero-inflated distribution is the distribution of zero-inflated Poisson (ZIP) (Yee, [19]).

Zero-infated poisson (ZIP) model, well described by Lambert [7] is a simple mixture model for count data with excess zeros. The model is a combination of a Poisson distribution and a degenerate distribution at zero. Specifically if Yi are independent random variables having a zero-infated Poisson distribution, the zeros are assumed to arise in to ways corresponding to distinct underlyingstates. The first state occurs with probability zero ω_i and produces only zeros, whilethe other state occurs with probability $1-\omega_i$ and leads to a standard Poisson countwith mean λ and hence a chance of further zeros. In general, the zeros from the first state are called structural zeros and those from the Poisson distribution are called sampling zeros (Jansakul and Hinde, [12]). This two-state process gives a simple two-component mixture distribution with p.m.f

$$f(Y_i = y_i) = \begin{cases} \frac{(1-\omega_i)e^{-\theta_i}\theta_i{}^{y_i}}{y!}\ ;\ y = 1,2,3,\ldots \\ \omega_i + (1-\omega_i)e^{-\theta_i};\ y = 0 \end{cases}$$

$$\text{with } 0 \le \omega_i < 1 \qquad (1)$$

The mean and variance are:

$$E(Y_i) = \sum_{y_i=0}^{n} y_i[\mathrm{I}_{(y_i=0)}.\left(\omega_i + (1-\omega_i)e^{-\theta_i}\right) + \mathrm{I}_{(y_i>0)}.(1-\omega_i)e^{-\theta_i}\theta_i{}^{y_i}/y_i!]$$

$$E(Y_i) = \sum_{y_i=1}^{n} y_i\,(1-\omega_i)e^{-\theta_i}\theta_i{}^{y_i}/y_i! = \cdots = \theta_i(1-\omega_i) = \mu_i \qquad (2)$$

$$E(Y_i^2) = \sum_{y_i=0}^{n} y_i{}^2[\mathrm{I}_{(y_i=0)}.\left(\omega_i + (1-\omega_i)e^{-\theta_i}\right) + \mathrm{I}_{(y_i>0)}.(1-\omega_i)e^{-\theta_i}\theta_i{}^{y_i}/y_i!]$$

$$E(Y_i^2) = \sum_{y_i=0}^{n} y_i{}^2((1-\omega_i)e^{-\theta_i}\theta_i{}^{y_i})/y_i! = \cdots = (1-\omega_i)\left(\theta_i + \theta_i{}^2\right)$$

$$Var(Y_i) = E(Y_i^2) - \left(E(Y_i)\right)^2 = (1-\omega_i)(\theta_i + \theta_i{}^2) - \left(\theta_i(1-\omega_i)\right)^2.$$

$$Var(Y_i) = \cdots = \theta_i(1-\omega_i)(1+\omega_i\theta_i) = \cdots = \mu_i + \left(\frac{\omega_i}{1-\omega_i}\right)\mu_i^2 \qquad (3)$$

By writing $\mathrm{E}(\mathrm{Y_i}) = \mu_i$ dan $\mathrm{Var}(\mathrm{Y_i}) = \mu_i + \left(\frac{\omega_i}{1-\omega_i}\right)\mu_i^2$seen that the distribution of conditional $\mathrm{Y_i}$shows a phenomenon of over-dispersion, if $\omega_i > 0$. It is clear that this reduces to the standard Poisson model when $\omega_i = 0$. For a random sample of observations $y_1, y_2 y_3, \ldots, y_n$; the log-likelihood function is given by

$$\ell = \ell(\lambda, \omega; \mathbf{y}) = \sum_i \{ I_{(y_i=0)} \ln[\omega_i + (1-\omega_i)e^{-\lambda_i}] + I_{(y_i>0)}[\ln(1-\omega_i) - \lambda_i + y_i \ln\lambda_i - \ln(y_i!)]\},$$

where I(.) is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise. To apply the zero-infated Poisson model in practicalmodelling situations, Lambert [7] suggested the following joint models for
ω_i and λ_i as follows:

$$ln\left(\frac{\boldsymbol{\omega}}{1-\boldsymbol{\omega}}\right) = \boldsymbol{G\gamma} \text{ dan } ln(\boldsymbol{\lambda}) = \boldsymbol{XB}$$

4.2 Reduce – Rank Vector Generalized Linear Models

Suppose our data comprises(x_i, y_i) for $i = 1, \cdots, n$ where x_i denotes the vector of explanatory variables for the ith observation, and y_i is the response (possibly a vector). The first value of x_i is unity for an intercept term.VGLMs are similar to ordinary GLMs but allow for multiple linear predictors. VGLMSs handle M linear predictors (the dimension M depends on the model to be fitted) where the jth one is

$$n_j = n_j(x) = \beta_j^T x = \sum_{k=1}^{p} \beta_{(j)k}\, x_k \ , j = 1, \cdots, M \tag{4}$$

The ηj of VGLMs may be applied directly to parameters of a distribution, θj, rather than just to mean μ = E(Y) as for GLMSs, in general, $\eta_j = g_j(\theta_j)$ for some parameter link function g_j and parameter θ_j.

$$\eta = \eta(x) = \begin{bmatrix} \eta_1(x) \\ \vdots \\ \eta M(x) \end{bmatrix} = \mathrm{B}^{\mathrm{T}} x = \begin{bmatrix} \mathrm{B}_1^{\mathrm{T}} x \\ \vdots \\ \mathrm{B}_{\mathrm{M}}^{\mathrm{T}} x \end{bmatrix} \tag{5}$$

where B is a p × M matrix of (sometimes too many) regression coefficients. In many situations the regressions coefficients are related to each other. For example, some of the $\beta_{(j)k}$ may be equal, set to zero, or add up to a certain quantity. These situations may be dealt with by use of constraint matrices. VGLMs in general have

$$n_j(x) = \sum_{k=1}^{p} \mathbf{B}_k \beta^*_{(k)} x_k, \ \ j = 1,\ldots,M \tag{6}$$

where H1, H2 ,…, Hp are known full-column rank constraint matrices and $\beta^*_{(k)}$ are vectors of unknown coefficients. With no one constraint at all H1 = H2 = … =Hp = IM. then, for VGLMs,

$$\mathbf{B}^T = \left(H_1\beta^*_{(1)} \quad H_2\beta^*_{(2)} \ldots H_p\beta^*_{(p)}\right) \tag{7}$$

Partition x into $\left(x_1^T, x_2^T\right)^T$ (of dimension $p_1 + p_2 = p$) and B = $\left(\mathbf{B}_1^T, \mathbf{B}_2^T\right)^T$ if $\mathbf{B}_2$ has too many regression coefficients then we can reduce its number dramatically by a reduced-rank regression. RRVGLMs then have:

$$\eta = \mathbf{B}_1^T x_1 + \mathbf{B}_2^T x_2 \tag{8}$$

where we approximate $\mathbf{B}_2$ by a reduced-rank regression

$$\mathbf{B}_2 = \mathbf{C}\,\mathbf{A}^T. \tag{9}$$

Here, $\mathbf{C}$ and $\mathbf{A}$ are $p_2 \times R$ and $M \times R$ respectively, and they are 'thin' because the rank R is low, e.g., R = 1 or 2. Thus

$$\eta = \mathbf{B}_1^T x_1 + \mathbf{A}\nu \tag{10}$$

where $\nu = \mathbf{C}^T x_2$ is a vector of R latent variables.

To make the parameters unique, it is common to enforce corner constraint on $\mathbf{A}$. By default, the top $R \times R$ submatrix is fixed to be $\mathbf{I_R}$ and the remainder of $\mathbf{A}$ is estimated.

4.3 Row-Column Interaction Model for Data count in the RR-VGLM

We use Goodman's RC association model (Goodman, [3]) to explain what a RCIM is. For more background see Yee and Hastie [17]. How does Goodman's RC association model fit within the VGLMs framework? Suppose Y = [(yij)] be a n × M matrix of counts. Goodman's model fits a reduced-rank type model to Y by firstly assuming that Yij has a Poisson distribution, and that

$$\eta_{ij} = \log(\mu_{ij}) = \mu + \alpha_i + \gamma_j + \sum_{k=1}^{R} a_{ik} c_{jk} \tag{11}$$

where $\mu_{ij} = \mathrm{E}(Y_{ij})$ is the mean of the i-j cell. Identifiability constraint are needed in (11) for the row and column effects $\alpha_i$dan γ_j; we use corner constraints $\alpha_i = \gamma_j = 0$ in this article. The parameters a_{ik} and c_{jk} also need constraints, e.g., we use $k = 1, \ldots, R$ for $a_{ik} = c_{jk} = 0$. We can write (11) as $\log(\mu_{ij}) = \mu + \alpha_i + \gamma_j + \delta_{ij}$. Where the $n \times M$ matrix $\Delta = [(\delta_{ij})]$ of interaction terms is approximated by the reduced ran quantit y $\sum_{k=1}^{R} a_{ik} c_{jk}$. Goodman's RCassociation model fits within the VGLMs framework by letting $\eta_i = \log \mu_i$. Where $\mu_i = \mathrm{E}(Y_i)$ is the mean of ith row of Y. Then the matrix $(\eta_1, \ldots, \eta_n)^T$ fits into RR-VGLM framework as follows. From last section, we obtain

$$\mathbf{B}_1^T x_{1j} = \begin{pmatrix} \mu 1_M & \alpha_2 1_M & \cdots & \alpha_n 1_M \end{pmatrix} \begin{pmatrix} 1 \\ e_{(-1)i} \\ 1_{M-1} \end{pmatrix}$$

when a subscript "(-1)" means the first element or row is removed from the vector or matrix. This shows, for example,that the intercept and row score variables have 1M as their constraint matrices. Similiarly, because B2 is approximated by $\boldsymbol{CA}^T$, the i-th row of Δ will be approximated by $\boldsymbol{x_{2i}^T C A}^T$, or equivalently, Δ is approximated by

$$\begin{pmatrix} \boldsymbol{x_{21}} \\ \vdots \\ \boldsymbol{x_{2n}} \end{pmatrix} \boldsymbol{CA}^T.$$

The desired reduced-rank approximation of Δ can be obtained if $\boldsymbol{x}_{2\mathrm{i}} = \boldsymbol{e_i}$ so that$\boldsymbol{I}_{\mathrm{p2}} \boldsymbol{CA}^T = \boldsymbol{CA}^T$. Note that

$$\boldsymbol{\Delta} = \begin{pmatrix} \mathbf{0} & \mathbf{0}^T \\ \mathbf{0} & \widetilde{\boldsymbol{\Delta}} \end{pmatrix} \approx \boldsymbol{CA}^T = \begin{pmatrix} \mathbf{0}^T \\ \boldsymbol{C}_{(-1)} \end{pmatrix} \begin{bmatrix} \mathbf{0} & \left(\boldsymbol{A}_{(-1)}\right)^T \end{bmatrix},$$

That is, the first row of matrix of A consist of structural zeros which are 'omitted' from the reduced rank regression of $\boldsymbol{\Delta}$.

One could define RCIMs as a RR-VGLM with

$$\boldsymbol{\eta_{1ij}} = \boldsymbol{\mu} + \boldsymbol{\alpha_i} + \boldsymbol{\gamma_j} + \sum_{k=1}^{R} \boldsymbol{a_{ik} c_{jk}} \tag{12}$$

Note that (12) applies to the first linear/additive predictor; for models with M >1 one can leave $\eta_2, \ldots, \eta_M$ unchanged. Of course, choosing $\eta_1$for (12) is only for convenience. The software chooses $g_1^{-1}(\hat{\eta}_{ij})$ as the fitted values of the model and the result should be the same dimensions as the two-way table.

4.4 Zero-inflated Poisson model in the RR-VGLM

ZIP model is very powerful in dealing with count data with excess zeros than the usual Poisson distribution, partly it is because the ZIP model also handles over-dispers. Distribution as ZIP equation (1) can be mentioned that the event (Y = 0) come from two sources and in this model it with the RR-VGLM η_1 and η_2 as

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} logit\psi \\ log\mu \end{pmatrix} \tag{13}$$

There are two processes of how the data occurs, the first data is zero and the Poisson count data. Both processes are modeled respectively by η_1 and η_2.

Liu and Chan [8] gives examples of research involving surveys in which the spatio-temporal aggregation of fish catches show a positive opportunity similar to a monotonic function of the middle value.

Liu and Chan [8] proposed several new methodologies that allow ZIP to handle the linkage between the two processes. They call it COZIGAM, which is constrained zero-inflated generalized additive models. Which in the fact now, can be seen simply that this is a dimension reduction regression models ZIP or reduced-rank zero-inflated Poisson model (RR-ZIP). RR-ZIP is given by

$$logit\psi = \eta_1 = \beta_{(1)1} + \alpha_{11}.\eta_2 \tag{14}$$

$$log\mu = \eta_2 = \beta_2^T \mathcal{X} \tag{15}$$

with β(1)1 and a(1)1 is coeficient who want predictable.

Actually, because ηj here is the linear predictor as in equation (4), then equation (14) and (15) should be called COZIGLM. Connectedness can be seen, for example, if μ increases, η2 increases, and then η1 and increasing ψ if $a_{(1)1}$appreciating positive.

Equation (14) and (15) is a model RR-ZIP was rank 1 with $H_1 = I_2$ and $H_2 = \ldots = H_p = (a_{(1)1})^T$. There is a trivial complication that the constraint angle (can use other constraints) imposed on parameters that are used instead of the first two. This can be simplified if the order parameter exchanged.

4.5 SVD-Reparameterization for the GAMMI Model

An overlapping methodology is the generalized additive main effects and multiplicative interaction models, or GAMMI models, of Turner and Firth [13]. These also comprise the row and column main effects plus one or more components of the multiplicative interaction. The singular value corresponding to each multiplicative

component is often factored out, as a measure of the strength of interaction between the row and column scores, indicating the importance of the component, or axis. For cell means μ_{ij}a GAMMI-R model has the form

$$g(\mu_{ij}) = \alpha_i + \beta_j + \sum_{k=1}^{R} \sigma_k \gamma_{ki} \delta_{kj}. \quad (13)$$

Based on (13) GAMMI models appear to be identical to RCIMs. Here they apply a SVD to our AC^T. While our interaction term uses corner constraints, their SVD parameterization is quite interpretable and is related to some of the other parameterizations described in Yee and Hastie [17]. The advantage of RCIMs is that it should work for any VGAM family function, thus the family size is much bigger. It is easy to perform some post-transformations such as applying svd() to the VGAM output to obtain the SVD parameterization for GAMMI models.

5. APPLICATION: LEAF RUST DISEASE ATTACKS ON MUNG BEAN

The data comes from the Indonesian Legumes and Tuber Crops Research Institute (ILETR) Malang. This trial involved 10 genotypes and two green beans varieties which planted in 5 different environments at Probolinggo, Jombang, Jember, Rasanae, and Bolo. Experiments conducted on plots of size 4x $5m^2$ with a spacing of 40 cmx 10cm, two seeds per hole. The design in each environment is completely randomized, with 3 replications. One of the researchers' attention is on resistance to *leaf rust* disease. This disease is a not major disease. Observations done on trials field without inoculated. Theatrically it allows to be happened what is called by the term "escape", event with no attack. Statistically it is such of the zero-inflated phenomenon. Table 1 shows the amount average from three replications.

Table 1:
Count of *Leaf Rust* Disease Attacks on Mung Bean

Genotype		Locations (Environments)				
		Proboliggo	**Jember**	**Jombang**	**Bolo**	**Rasanae**
MLG	1002	0	167	100	150	150
MLG	1004	0	217	250	233	250
MLG	1021	0	200	217	183	217
MMC	74d Kp1	0	133	200	183	133
MMC	71d Kp2	0	200	200	233	367
MMC	157d Kp1	0	133	150	167	150
MMC	203d Kp5	0	50	100	67	83
MMC	205e	0	50	67	100	67
MMC	100f Kp1	0	50	83	83	83
MMC	87d Kp5	0	83	117	133	83
MURAI		0	0	50	33	33
PER KUTUT		0	67	133	117	117

Data were analyzed using the VGAM package with SVD reparameterization in the RCIM model follows what is done by Turner & Fifth[13] on the Poisson distribution with the GAMMI model proposed by VanEeuwijk [14].

Zero inflated Poisson (ZIP) model give us two results, the logit and the log-part. The logit-pat give us the probability of being zero at random, from a intercept only model. The log-part is our main attentions since it give us more information. The model used is the RCIM model with rank = 0 and SVD-reparaterization on working residuals to get the interaction with rank =2. The data were number of crops attacked by leaf-rust, so that genotype with large numbers indicate that it is vulnerable. Genotype with average count in almost allocations, said to be *stable* (in fact, it is vulnerable) and will be located close to the origin point on biplot. That is, the zero origin biplot point does not always describe the resistance of the genotype. Probolinggo are drawn close to the zero point, because of in general, the overall genotypes have the same number of zeros. However, the zeros on the observation in Probolinggo sometimes called "escape" observation, where all the columns on this row is zero. The ZIP model relies on the assumption that "zero" are as structural zero and random zero. As random zero, ZIP model will provide us the probability of cell to be zero, and the fitted value for Poisson count, as well.

Table 1
The Log-Likelihood Ratio Test

Model	DF	Log-Likelihood	G	DF-Chisq	p-value	
FullModel (Rank=4)	58	-150.2266				
GAMMI Rank=3	50	-150.2266	0	8	1.00E+00	ok
GAMMI Rank=2	40	-160.1218	19.7904	18	3.45E-01	ok
GAMMI Rank=1	28	-182.6969	64.9406	30	2.23E-04	bad
Main Effects(Rank=0)	14	-249.6574	198.8616	44	1.44E-21	bad

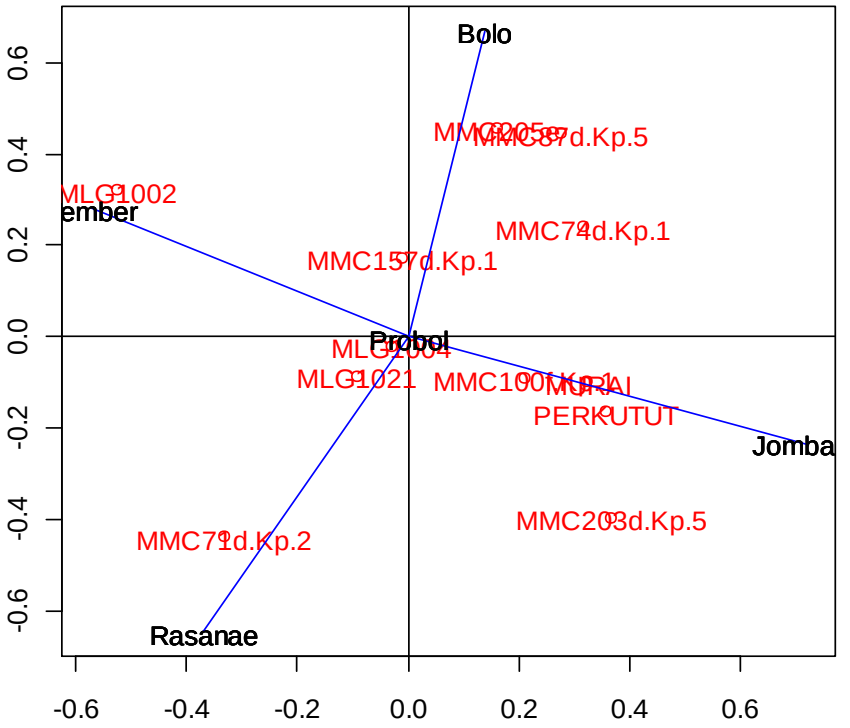

Figure 2: Biplot of the Interaction Effect on Log-Scale of Zero-Inflated Poisson

Determining the Rank 2 model, we test the existence the interaction term using Likelihood Ratio (LR) between the null model with the saturated model. In this case the LR value between Rank = 2 and Rank = 1 is equal to 19.7904 with a p-value of on Chi-square distribution (degrees of freedom = 18) which very small that is equal to 3.45e-01. With the ZIP model Rank = 2 of GAMMI-ZIP model, the biplot of it is presented in Figure 2. The variability shown by the eigen value of matrix interaction, the five root traits in a row are: 1.056, 0.807, 0.696, 0.000, 0.000. The first two eigen, explaining the total variability Biplot, hat is 72.78%.

ACKNOWLEDGEMENT

Special thanks to Thomas W. Yee, Department of Statistics, The University of Auckland for the VGAM package and for kindness during my previous visiting research. Thanks to Rudi Iswanto, Indonesian Legumes and Tuber Crops Research Institute, Malang, Indonesia for the data of agricultural trial.

REFERENCES

1. Andrews, D., Bickel, P.J., Hampel, F., Huber, P., Rogers, W. and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton, N.J.: Princeton University Press, 83-98.
2. Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16, 147-186.
3. Brich, J.B. and Myers, R.H. (1982). Robust analysis of covariance. *Biometrics*, 38, 699-713.
4. David, H.A. (1981). *Order Statistics*. 2nd ed. New York: Wiley.
5. BoÈhning, D. (1998). Zero-Inflated Poisson Models and C.A.MAN: A Tutorial Collection of Evidence. *Biometrical Journal* 40- 7, 833-843.
6. Lord, D., Washingtonb, S.P. and Ivanc, J.N. (2005) Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37, 35-46.
7. Dietz, E. and BoÈhning, D. (2000). On estimation of the Poisson parameter in zero-modified Poisson models. *Computational Statistics and Data Analysis* 34, 441-459.
8. Goodman, L.A. (1981). Interaction models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Interaction*, 76, 320-334.
9. Hadi, A.F., Mattjik, A.A. and Sumertajaa, I.M. (2010). Generalized AMMI models for assessing the endurance of soybean to leaf pest. *Jurnal. Ilmu. Dasar*., 11(2), 151-159.
10. Hinde, J. and Demetrio, C.G.B. (1998). Over dispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151-170.
11. Kaminski, R.J. and Thomas, D. (2009). Stucky Reassessing Political Explanations for Murders of Police. *Homicide Studies*, 13(1), 3-20.
12. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
13. Liu, H. and Chan, K.S. (2010). Introducing COZIGAM: An R package for unconstrained and constrained zero-inflated generalized additive model analysis. *Journal of Statistical Software*, 35:1-26. http://www.jstatsoft.org/v35/i11.
14. McCullagh, P. and Nelder, J.A. (1989). Generalized linear models, 2nd edition. London: Chapman & Hall.
15. Motta, M.R., Gianola, D., Heringstad, B. Rosa, G.J.M. and Chang, Y.M. (2007). A Zero-Inflated Poisson Model for Genetic Analysis of the Number of Mastitis Cases in Norwegian Red Cows. *J. Dairy Sci*., 90, 5306-5315.
16. Jansakul, N. and Hinde, J.P. (2002). Score tests for zero-inflated Poisson models. *Computational Statistics and Data Analysis*, 40, 75-96.
17. Turner, H. and Firth, D. (2009). Generalized nonlinear models in R: An overview of the GNM package. URL http://CRAN.R-project.org/package=gnm. R package version 0.10-0. UK: University of Warwick.
18. Van Eeuwijk, F.A. (1995). Multiplicative interaction in generalized linear models. *Biometrics*, 51, 1017-1032.
19. Welsh, A.H., Cunningham, R.B. and Chambers, R.L. (2000). Methodology for estimating the abundance of rare animals: seabird nesting on North East Herald Cay. *Biometrics*, 56, 22-30.

20. Yee, T.W. and Hadi, A.F. (2014). Row Column Interaction Models, with an R implementation. *Computational Statistics.* Online first. http://link.springer.com/article/ 10.1007/s00180-014-0499-9
21. Yee, T.W. and Hastie, T.J. (2003). Reduced-rank vector generalized linear models. *Statistical Modeling*, 3, 15-41.
22. Yee, T.W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32, 1-34.
23. Yee, T.W. (2008). VGAM Family Functions for Positive, Zero-altered and Zero-Inflated Discrete Distributions. http://www.stat.auckland.ac.nz/