# A Power Study of the GFfit Statistic as a Lack-of-Fit Diagnostic

Junfei Zhu[1], Mark Reiser[2], Silvia Cagnone[3]
[1]School of Mathematical and Statistical Sciences, Arizona State University, USA
[2]School of Mathematical and Statistical Sciences, Arizona State University, USA
[3]Department of Statistical Sciences, University of Bologna, Italy

## Abstract

The Pearson and likelihood ratio statistics are commonly used to test goodness-of-fit for models applied to data from a multinomial distribution. When data are from a table formed by cross-classification of a large number of variables, the common statistics may have low power and inaccurate Type I error level due to sparseness in the cells of the table. For the cross-classification of a large number of ordinal manifest variables, it has been proposed to assess model fit by using the GFfit statistic as a diagnostic to examine the fit on two-way subtables. A new version of the GFfit statistic has been developed by decomposing the Pearson statistic from the full table into orthogonal components defined on lower-order marginal distributions and then defining the GFfit statistic as an orthogonal component. An omnibus fit statistic can be obtained as a sum of a subset of these components. In this paper, the individual components are being studied for statistical power as diagnostics to detect the source of lack of fit when the model does not fit the observed data. Simulation results for power of components to detect lack of fit along with comparisons to other diagnostics are presented.

**Key words:** sparseness, orthogonal component, latent variables, chi-square test, goodness-of-fit, GFfit statistic

# 1.INTRODUCTION

Traditionally we use the likelihood ratio ($LR$) and the Pearson chi-square ($GF$) to test goodness of fit for a model fit on cross-classified variables

$$LR = 2n\sum_{r=1}^{k} f_r \ln(\frac{f_r}{\hat{\pi}_r})$$

$$GF = n\sum_{r=1}^{k} \frac{(f_r - \hat{\pi}_r)^2}{\hat{\pi}_r}$$

Suppose we have $p$ categorical variables and the i-th variable has $c_i$ categories. Thus there are $k = \prod_{i=1}^{p} c_i$ cells, also called response patterns in the cross-classified table. Then $f_r$ is the sample proportion of the r-th response pattern and $\hat{\pi}_r$ is the estimated probability of the r-th response pattern. If the number of observations in each response pattern is large enough and under the conditions (Koehler and Larntz, 1980) that i) $H_0: \pi = \pi(\theta)$, ii) $k$ is fixed and iii) $\min_{1\le r\le k} n\pi_r \to \infty$ for $n \to \infty$, both $LR$ and $GF$ are approximately distributed $\chi^2$ with degree of freedom equal to $k - 1 -$ number of estimated parameters. However, when there is a problem of sparseness, these two statistics may not have an approximate chi-square distribution. Several statistics have been proposed using marginal distributions of the joint variables rather than the joint distribution.

Joreskog and Moustaki (2001) proposed the *GFfit* statistic as a diagnostic to help in finding the source of model lack of fit. A new version of the *GFfit* statistic is proposed by Reiser, Cagnone & Zhu (2014) by decomposing the Pearson statistic from the full table into orthogonal components defined on lower-order marginal distributions. Then the *GFfit* statistic is defined as a sum of a subset of these components. In this paper, the individual components are being studied for statistical power as diagnostics to detect the source of lack of fit when the model does not fit the observed data.

The paper is organized as follows: In Section 2 we introduce the marginal proportion and the *GFfit* orthogonal components. In Section 3 we give a discussion of the GLLVM model. In Section 4 simulation results for power of components to detect lack of fit along with comparisons to other diagnostics are presented. Finally in Section 5 we apply these orthogonal components to a real data set.

# 2. MARGINAL PROPORTIONS

A traditional method such as Pearson's statistic uses the joint frequencies to calculate goodness of fit for a model that has been fit to a cross-classified table. This section presents a transformation from joint proportions or frequencies to marginal proportions.

## 2.1 First- and Second-order Marginals

Consider the three variables, two categories case. An 8 by 3 matrix $\boldsymbol{V}$ can be used to denote the response patterns as the rows:

$$V = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Let $v_{is}$ represent element $i$ of response pattern $s$, $i = 1, \ldots, p$ and $s = 1, \ldots, k$. In this example, $p = 3$ and $k = 8$. Then, under some specific model, which we will introduce later, the first-order marginal proportion for variable $y_i$ can be defined as

$$P_i(\theta) = \text{Prob}(y_i = 1|\theta) = \sum_s v_{is}\pi_s(\theta)$$

and the true first-order marginal proportion is given by

$$P_i = \text{Prob}(y_i = 1) = \sum_s v_{is}\pi_s \ .$$

Thus the marginal proportions are linear combination of joint proportions:

$$\boldsymbol{P} = \mathbf{H}\boldsymbol{\pi}$$

The H matrix can be defined from the V matrix. For first-order marginal, $\mathbf{H}_{[1]} = \boldsymbol{V}'$.
For 3 variables with 3 categories, $\mathbf{H}_{[1]} = \boldsymbol{V}'$, where

$$V_{27*6} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Under the model, for two categories, the second-order marginal proportion for variable $y_i$ and $y_j$ can be defined as

$$P_{ij}(\boldsymbol{\theta}) = \text{Prob}\left(y_i = 1, y_j = 1\middle|\boldsymbol{\theta}\right) = \sum_s v_{is}v_{js}\pi_s(\theta) \ ,$$

and the true second-order marginal proportion is given by

$$P_{ij} = \text{Prob}\left(y_i = 1,\ y_j = 1\right) = \sum_s v_{is}v_{js}\pi_s \ .$$

If the number of categories $c$ is greater than 2, the second-order marginal proportions for $y_i$ and $y_j$ can be represented as a $c$ by $c$ table with $(c-1)^2$ proportions.

Thus for second-order marginal proportions, the rows of H are Hadamard products among the columns of V. For 3 variables with 3 categories, $\mathbf{H}_{[2]}$ is an 18 by 27 matrix:

$$\mathbf{H}_{[2]} = \begin{bmatrix} (v_1 \circ v_3)' \\ (v_1 \circ v_4)' \\ \vdots \\ (v_1 \circ v_5)' \\ (v_1 \circ v_6)' \\ \vdots \\ (v_3 \circ v_5)' \\ \vdots \\ (v_{i(c-1)} \circ v_{j(c-1)})' \end{bmatrix}$$

where $v_i$ is the column $i$ of matrix V, and $v_i \circ v_j$ is the Hadamard product of columns $i$ and $j$.

## 2.2 Test statistic

Linear combinations of $\boldsymbol{\pi}$ may be tested under the null hypothesis $H_0\colon \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ and the test statistic is

$$X^2_{[t:u]} = \boldsymbol{e}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1}\boldsymbol{e}\,,$$

$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = n^{-1}\boldsymbol{\Omega}_{\boldsymbol{e}}$ with $\boldsymbol{\Omega}_{\boldsymbol{e}}$ evaluated at the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$, and where

$$\boldsymbol{\Omega}_{\boldsymbol{e}} = \mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}'$$

$D(\boldsymbol{\pi})$ =diagnal matrix with $(s,s)$ element equal to $\pi_s(\boldsymbol{\theta})$

$$\mathbf{A} = D(\pi)^{-1/2}\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

$$\mathbf{G} = \frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

$\boldsymbol{e} = \mathbf{H}(\boldsymbol{f} - \boldsymbol{\pi})$ is the matrix form of the marginal residuals.

$\mathbf{H} = \mathbf{H}_{[1:2]}$ produces $X^2_{[1:2]}$ and $\mathbf{H} = \mathbf{H}_{[2]}$ produces $X^2_{[2]}$. It has been proven that for two categories, the distributions of $X^2_{[1:2]}$ and $X^2_{[2]}$ are chi-square distributions with degrees of freedom equal to $q(q+1)/2$ and $q(q-1)/2$ respectively. $X^2_{[1:q]} = GF$. $X^2_{[t:u]}$ is a score statistic, Reiser (1996), Reiser and Lin (1999), Cagnone and Mignani (2007), Rayner and Best (1989).

## 2.3 Orthogonal components

Consider the $k-g-1$ by $\mathrm{c}^q$ matrix $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q;-g]}$, where $g$ is the number of unknown model parameters to be estimated and $\mathbf{H}_{[1:q;-g]}$ is matrix $\mathbf{H}_{[1:q]}$ deleting $g$ rows. $\mathbf{H}^*$ has full row rank. $\boldsymbol{F}$ is the upper triangular matrix such that $\boldsymbol{F}'\boldsymbol{\Omega}_{\boldsymbol{e}}\boldsymbol{F} = \boldsymbol{I}$. $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega}_{\boldsymbol{e}}$. Premultiplication by $(\boldsymbol{C}')^{-1}$ orthonormalises the matrix $\mathbf{H}_{[1:q;-g]}$ in the matrix $D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$.

$$X^2_{PF} = X^2_{[1:q;-g]} = n\mathbf{r}'(\widehat{\mathbf{H}}^*)'\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\mathbf{H}}^* = \mathbf{H}^*(\widehat{\boldsymbol{\theta}})$, and $\mathbf{r} = (\widehat{\boldsymbol{p}} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}))$.
Define

$$\widehat{\boldsymbol{\gamma}} = n^{\frac{1}{2}}\widehat{\boldsymbol{F}}'\mathbf{H}\mathbf{r} = n^{\frac{1}{2}}\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\boldsymbol{F}}$ is the matrix $\boldsymbol{F}$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Then

$$X_{PF}^2 = \widehat{\boldsymbol{\gamma}}'\widehat{\boldsymbol{\gamma}} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{j=k-g-1}} \hat{\gamma}_j^2$$

$\widehat{\mathbf{H}}^*\mathbf{r}$ has asymptotic covariance matrix $\boldsymbol{F}'\boldsymbol{\Omega_e}\boldsymbol{F} = \boldsymbol{I_{k-g-1}}$. The elements $\hat{\gamma}_j^2$ are asymptotically independent chi-square random variables with $df = 1$ (Reiser, 2008).
Using Sequential Sum of Squares: Redefine

$$z_s = \sqrt{n}\left(\pi_s(\widehat{\boldsymbol{\theta}})\right)^{-\frac{1}{2}}(\hat{p}_s - \pi_s(\widehat{\boldsymbol{\theta}})).$$

Perform the regression of $\boldsymbol{z}$ on the columns of $\mathbf{H}'$:

$$\boldsymbol{z} = \mathbf{H}'\boldsymbol{\beta}$$

Then,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{H}\widehat{\boldsymbol{W}}\mathbf{H}')^{-\mathbf{1}}\mathbf{H}\widehat{\boldsymbol{W}}\mathbf{u}$$

where $\mathbf{u} = \sqrt{n}\mathbf{r}$, $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{D}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}} = \widehat{\boldsymbol{D}}^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}}$, and $\boldsymbol{D} = diag(\boldsymbol{\pi}(\boldsymbol{\theta}))$.

$\Sigma = \Sigma(\theta) = (\boldsymbol{I} - \boldsymbol{\pi}^{\frac{1}{2}}\left(\boldsymbol{\pi}^{\frac{1}{2}}\right)' - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-\mathbf{1}}\boldsymbol{A}')$ is idempotent.

Let $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{D}}^{\frac{1}{2}}\mathbf{H}'$. Then

$$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{M}}'\widehat{\boldsymbol{M}})^{-\mathbf{1}}\widehat{\boldsymbol{M}}'\boldsymbol{z}$$

$\hat{\gamma}_j^2, j = 1, k-g-1$ are the sequential SS from this regression. $\boldsymbol{\gamma} = \boldsymbol{C}'\boldsymbol{\beta}$ are the orthogonal coefficients.

Now define an orthogonal components version of $GFfit$:

$$GFfit_{\perp}^{(ij)} = \sum_{l=m+1}^{l=m+(c-1)^2} \hat{\gamma}_l^2$$

where $m = q + (i-1)(c-1)^2 + (j-2)(c-1)^2$, assuming $\mathbf{H} = \mathbf{H}_{[1:2]}$.

## 3. THE GENERALIZED LINEAR LATENT VARIABLE MODEL

Let $\boldsymbol{y} = (y_1, y_2, \dots, y_p)$ be the vector of p ordinal observed variables, each of them having ci categories. Thus there are $\prod_{i=1}^{p} c_i$ cells, also called response patterns in the cross-classified table. The r-th response pattern is indicated as $\boldsymbol{y_r} = (y_1 = a_1, y_2 = a_2, \dots, y_p = a_p)$, where $a_i$ is the value of the i-th observed variable($a_i = 1, \dots, c_i$ $and$ $i = 1, \dots, p$).
Let $\boldsymbol{z} = (z_1, z_2, \dots, z_p)$ be the vector of q continuous latent variables. Then the probability of the r-th response pattern $\boldsymbol{y_r}$ is given by

$$\pi_r(\theta) = \int \pi_r(\boldsymbol{z})\, h(\boldsymbol{z})\, d\boldsymbol{z}\,,$$

where $\theta$ is a vector of parameters. $h(\boldsymbol{z})$ is the density function of z, and we assume every latent variable to be distributed standard normal independently. $\pi_r(z)$ is the conditional probability of $y_r$ given z and it is a multinormial probability function

$$\pi_r(\boldsymbol{z}) = \prod_{i=1}^{p} \pi_{a_i}^{(i)}(z) = \prod_{i=1}^{p}(\tau_{a_i}^{(i)} - \tau_{a_{i-1}}^{(i)})$$

where $\tau_{a_i}^{(i)} = \pi_1^{(i)}(z) + \pi_2^{(i)}(z) + \cdots + \pi_{a_i}^{(i)}(z)$ is the probability of a response in category $a_i$

or lower on the variable $i$ and $\pi_{a_i}^{(i)}(z)$ is the probability of a response in category $a_i$ on the variable $i$.

We use logistic regression to model the interrelationship between $\tau_{a_i}^{(i)}$ and the latent variables.

$$log\left[\frac{\tau_s^{(i)}}{1-\tau_s^{(i)}}\right] = \alpha_{i0}(s) - \sum_{j=1}^{q} \alpha_{ij} z_j\,,\ \ s = 1, \dots, c_{i-1}$$

$\alpha_{i0}(s)$ and $\alpha_{ij}$ are the parameters of the model. $\alpha_{i0}(s)$ is the intercept and $\alpha_{ij}$ is the j-th slope for variable $i$. The intercepts should satisfy the condition $\alpha_{i0}(1) \leq \alpha_{i0}(2) \leq \cdots \leq \alpha_{i0}(c_i)$.

We use the E-M algorithm to calculate the maximum likelihood estimator for the parameters in the model. The integrals are approximated through the Gauss-Hermite quadrature method (Cagnone & Mignani, 2007).

## 4. MONTE CARLO SIMULATION

### 4.1 Type I error study

A simulation study was conducted using GLLVM to assess the accuracy of the Type I error rates for $GFfit_{\perp}^{(ij)}, M_{ij}$ and $X_{ij}^2$. $M_{ij}$ is the individual Joe-Maydeu chi-square statistic. $X_{ij}^2$ is similar to the Pearson's statistic. However, instead of using the joint frequencies, $X_{ij}^2$ is calculated by using the marginal frequencies. Although the Pearson's statistic is distributed approximately chi-square, $X_{ij}^2$ is not.

The design of this Monte Carlo study is described as follows

- Model GLLVM with 1 latent factor
- Number of observed variables $p = 4, p = 5, p = 6$
- Number of categories for each variable $c = 3, c = 4$
- Number of samples 500
- Sample size $n = 500$

The intercepts range from -3 to 3. The factor loadings are the following: for $p = 4, \alpha_1 = (0.0, 0.1, 0.2, 0.6)'$; for $p = 5, \alpha_1 = (0.0, 3.0, 2.0, 1.0, 2.0)'$ ; for $\mathrm{p} = 6, \alpha_1 = (0.8, 0.7, 0.5, 0.3, 0.2, 0.1)'$.

Simulation results for Type I error are shown in the following tables. The tables show empirical Type I error for nominal $\alpha = 0.05$, using a chi-square distribution for each statistic.

TABLE 1: Type I error of $GFfit_{\perp}^{(ij)}$, $M_{ij}$ and $X_{ij}^2$ , 4 variables 3 categories

| | Type I error | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.0301 | 0.048 | 0.024 |
| (13) | 0.0281 | 0.068 | 0.018 |
| (14) | 0.0481 | 0.06 | 0.024 |
| (23) | 0.0462 | 0.042 | 0.022 |

| | | | |
|---|---|---|---|
| (24) | 0.0441 | 0.07 | 0.024 |
| (34) | 0.0441 | 0.042 | 0.026 |

TABLE 2: *Type I error of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 4 variables 4 categories*

| | Type I error | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.05 | 0.048 | 0.04 |
| (13) | 0.05 | 0.068 | 0.05 |
| (14) | 0.036 | 0.036 | 0.028 |
| (23) | 0.024 | 0.04 | 0.022 |
| (24) | 0.052 | 0.05 | 0.034 |
| (34) | 0.08 | 0.076 | 0.06 |

TABLE 3: *Type I error of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 5 variables*

| | Type I error | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.052 | 0.052 | 0.034 |
| (13) | 0.07 | 0.07 | 0.052 |
| (14) | 0.054 | 0.05 | 0.046 |
| (15) | 0.064 | 0.048 | 0.03 |
| (23) | 0.042 | 0.05 | 0.034 |
| (24) | 0.042 | 0.04 | 0.032 |
| (25) | 0.042 | 0.048 | 0.026 |
| (34) | 0.042 | 0.042 | 0.03 |
| (35) | 0.044 | 0.056 | 0.04 |
| (45) | 0.032 | 0.028 | 0.022 |

TABLE 4: *Type I error of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 6 variables*

| | Type I error | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.042 | 0.034 | 0.032 |
| (13) | 0.04 | 0.04 | 0.032 |
| (14) | 0.034 | 0.026 | 0.016 |
| （15） | 0.05 | 0.04 | 0.028 |
| （16） | 0.058 | 0.05 | 0.038 |
| （23） | 0.052 | 0.056 | 0.042 |
| （24） | 0.056 | 0.048 | 0.038 |
| （25） | 0.034 | 0.046 | 0.02 |

| | | | |
|---|---|---|---|
| （26） | 0.052 | 0.058 | 0.036 |
| （34） | 0.036 | 0.048 | 0.044 |
| （35） | 0.066 | 0.056 | 0.04 |
| （36） | 0.048 | 0.04 | 0.032 |
| （45） | 0.046 | 0.048 | 0.04 |
| （46） | 0.042 | 0.05 | 0.04 |
| （56） | 0.054 | 0.054 | 0.032 |

From these tables we can see that both $GFfit_{\perp}^{(ij)}$ and $M_{ij}$ have a good Type I error when sparseness is present.

A Kolmogorov-Smirnov test has also been applied to each statistic. The p-values are shown in the following tables.

TABLE 5: *p-value of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 4 variables 3 categories*

| | | *p-value* | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | <0.001 | 0.860 | <0.001 |
| (13) | 0.003 | 0.029 | <0.001 |
| (14) | 0.259 | 0.193 | <0.001 |
| (23) | 0.876 | 0.277 | <0.001 |
| (24) | 0.498 | 0.125 | <0.001 |
| (34) | 0.488 | 0.928 | <0.001 |

TABLE 6: *p-value of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 4 variables 4 categories*

| | | *p-value* | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.006 | 0.135 | <0.001 |
| (13) | 0.216 | 0.163 | 0.003 |
| (14) | 0.334 | 0.447 | <0.001 |
| (23) | 0.633 | 0.938 | <0.001 |
| (24) | 0.474 | 0.281 | 0.011 |
| (34) | 0.257 | 0.523 | <0.001 |

TABLE 7: *p-value of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 5 variables*

| | | *p-value* | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |

| | | | |
|---|---|---|---|
| (12) | 0.851 | 0.866 | 0.009 |
| (13) | 0.228 | 0.100 | 0.633 |
| (14) | 0.523 | 0.434 | 0.081 |
| (15) | 0.189 | 0.059 | 0.636 |
| (23) | 0.065 | 0.875 | <0.001 |
| (24) | 0.903 | 0.919 | <0.001 |
| (25) | 0.150 | 0.137 | 0.010 |
| (34) | 0.602 | 0.943 | 0.018 |
| (35) | 0.309 | 0.998 | <0.001 |
| (45) | 0.269 | 0.198 | <0.001 |

TABLE 8: *p-value of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 6 variables*

| | | *p-value* | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.002 | 0.008 | <0.001 |
| (13) | 0.953 | 0.845 | 0.033 |
| (14) | 0.076 | 0.074 | <0.001 |
| （15） | 0.672 | 0.911 | <0.001 |
| （16） | 0.785 | 0.276 | <0.001 |
| （23） | 0.431 | 0.623 | 0.613 |
| （24） | 0.539 | 0.610 | 0.123 |
| （25） | 0.465 | 0.497 | 0.003 |
| （26） | 0.089 | 0.283 | 0.004 |
| （34） | 0.618 | 0.616 | 0.004 |
| （35） | 0.922 | 0.957 | <0.001 |
| （36） | 0.192 | 0.621 | 0.026 |
| （45） | 0.221 | 0.923 | <0.001 |
| （46） | 0.517 | 0.850 | 0.002 |
| （56） | 0.855 | 0.309 | 0.028 |

From these p-values we can see that in most cases the $GFfit_{\perp}^{(ij)}$ and $M_{ij}$ are distributed chi-square but clearly $X_{ij}^2$ is not.

## 4.2 Power Study

A simulation to examine power of the statistics was also conducted using GLLVM for the 4 variables 3 categories case and 6 variables 4 categories case. Pseudo data for 1000 samples were generated from a confirmatory two-factor model with all parameters fixed and then fit with a one factor model. The 4 variables case has a sample size 500 and the 6 variables case has a sample size 300. The parameters for the data generating models are the following: for p=4, $\alpha_{0(1)} = (-1.5, -0.6, 0.3,\ 1.0)', \alpha_{0(2)} = (-1.0, -0.3, 0.6, 1.5)', \alpha_1 =$

$(0.0, 1.0, 1.0, 0.0)', \alpha_2 = (2.0, 0.1, 0.2, 2.0)'$ ; for p=6, $\alpha_{0(1)} = (-3, -2.5, -2, -1.8, -1.5, -0.8)', \alpha_{0(2)} = (-1, -0.5, 0, 0.2, 0.5, 1.2)', \alpha_{0(3)} = (1, 1.5, 2, 2.2, 2.5, 3.2)', \alpha_1 = (1.6, 1.35, 1.25, 0.4, 0.5, 0.6)', \alpha_2 = (0, 0, 0, 1, 1, 1)'$. The two latent variables were specified as uncorrelated, each with variance equal to 1.0. Estimation of the one-factor GLLVM for 4 variables case converged for 981 of the 1000 samples. For 6 variables case, it converged for all 1000 samples. A chi-square distribution was used to evaluate each statistic. The power simulation results are shown below:

TABLE 9: *power of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$ *, 4 variables 3 categories*

| | Power | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.0754 | 0.0560 | 0.0438 |
| (13) | 0.1590 | 0.0479 | 0.0254 |
| (14) | 0.2538 | 0.0407 | 0.0224 |
| (23) | 0.7186 | 0.0570 | 0.894 |
| (24) | 0.0550 | 0.0570 | 0.031 |
| (34) | 0.0530 | 0.0489 | 0.0275 |

TABLE 10: *power of* $GFfit_{\perp}^{(ij)}$, $M_{ij}$ *and* $X_{ij}^2$*, 6 variables*

| | *p-value* | | |
|---|---|---|---|
| | $GFfit_{\perp}^{(ij)}$ | $M_{ij}$ | $X_{ij}^2$ |
| (12) | 0.160 | 0.046 | 0.038 |
| (13) | 0.204 | 0.039 | 0.034 |
| (14) | 0.072 | 0.046 | 0.055 |
| （15） | 0.060 | 0.059 | 0.044 |
| （16） | 0.075 | 0.042 | 0.047 |
| （23） | 0.794 | 0.046 | 0.048 |
| （24） | 0.061 | 0.051 | 0.057 |
| （25） | 0.074 | 0.049 | 0.045 |
| （26） | 0.140 | 0.053 | 0.047 |
| （34） | 0.154 | 0.054 | 0.070 |
| （35） | 0.115 | 0.055 | 0.052 |
| （36） | 0.114 | 0.058 | 0.048 |
| （45） | 0.066 | 0.039 | 0.416 |
| （46） | 0.086 | 0.047 | 0.373 |
| （56） | 0.113 | 0.029 | 0.338 |

From these tables we can see that for 4 variables case, $GFfit_{\perp}^{(23)}$ has a power of 0.7186, which shows that primarily the association between variables 2 and 3 was not adequately explained by the one-factor model. All the $M_{ij}$ 's have a very low power. Although $X_{23}^2$ has a large power of 0.894, as we demonstrated above, it is not distributed chi-square. For 6 variables case we have a similar conclusion: $GFfit_{\perp}^{(23)}$ has a power of 0.794, which shows that primarily the association between variables 2 and 3 was not adequately explained by the one-factor model. All the $M_{ij}$ 's have a very low power. Although several $X_{ij}^2$ statistics have a power of 0.3 to 0.4, they are not truly distributed chi-square.

## 5. APPLICATION

In this section, we analyze a real data set about agoraphobia. Agoraphobia is a type of anxiety disorder in which sufferers fear and often avoid places or situations that might cause panic and feeling trapped, helpless or embarrassed. Those who suffer from agoraphobia often have a hard time feeling safe in any public places, especially where crowds gather. They may even feel unable to leave their home (Wittchen, Gloster, Beesdo-Baum, Fava, Craske, 2010). This dataset consists in judgments expressed by 3305 patients about several fears. There are 5 variables in this dataset:

- Fear of tunnels or bridges
- Fear of being in a crowd
- Fear transportation
- Fear of going out of house alone
- Fear of being alone

Each variable has three categories: “yes”, “no”, “kind of”. Our goal is to study whether these five variables can be modeled by a one-factor latent variable model. The number of all the possible response patterns are k=243. However, as most of the answers are “no”, 139 response patterns are empty. Furthermore many response patterns have a frequency less than 5.

The GLLVM with one factor was fit to these data, and fit statistics were calculated, using R software. Goodness-of-fit test results are shown in Table 11.

TABLE 11: $GFfit_{\perp}^{(ij)}$*'s of the Agoraphobia Sample*

| $GFfit_{\perp}^{(ij)}$ | Value |
|---|---|
| (12) | 29.34 |
| (13) | 31.82 |
| (14) | 1.58 |
| (15) | 6.39 |
| (23) | 47.91 |
| (24) | 43.60 |
| (25) | 61.56 |
| (34) | 16.40 |
| (35) | 23.69 |
| (45) | 58.90 |

In this sample, $GFfit_{\perp}^{(ij)}$ should distribute chi-square on $(3-1)^2 = 4$ degrees of freedom independently if the model is correct. The critical value is 9.49 for Type I error equal to 0.05. We can see that 8 out of 10 $GFfit_{\perp}^{(ij)}$'s are greater than the critical value. This also indicates that the one-factor model is not appropriate. We can also see that $X_{[2]}^2 = \sum_i \sum_j GFfit_{\perp}^{(ij)} = 321.19$ on 40 df.

## 6. Conclusion

The $GFfit_{\perp}^{(ij)}$ statistics can be calculated by using sum of squares from an orthogonal regression. Monte Carlo simulations demonstrated that the $GFfit_{\perp}^{(ij)}$ statistics perform well when sparseness is present. The $GFfit_{\perp}^{(ij)}$ statistics can be used as diagnostic to assist in detecting the source of poor fit when the model specified in the null hypothesis is rejected. An application to agoraphobia symptoms showed that these data cannot be explained by a model of a single underlying factor.

## REFERENCES

Agresti, A. & Yang, M. C. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and data Analysis*, **May**, 9-21.

Bartholomew, D. J. & Knott, M. (1999). *Latent Variable Models and Factor Analysis*, Kendall's Library of statistics, London, second edition.

Cagnone, S. & Mignani S. (2007). Assessing the goodness of fit of a latent variable model for ordinal data. *Metron*, LXV, 337-361.

Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests. *Biomedical Journal*, **10**, 417-451.

Goodman, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association*, **59**, 319-385.

Goodnight, J. H. (1978). The sweep Operator: Its importance in Statistical Computing. SAS technical Report R-106, SAS Institute, Cary, NC

Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit test. *Biometrika*, **59**, 137-145

Koehler, K. J. (1986). Goodness-of-fit tests for log-linear models in sparse contingency tables. *Journal of the American Statistical Association*, **81**,483-493

Koehler, K. J. & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*, 75, 336-344.

Morris, C. (1975). Central limit theorems for multinomial sums. *Annals of Statistics*, **3**, 365-384.

Rayner, J. C. W. & Best, D, J. (1989). *Smooth Tests of Goodness of Fit.* Oxford: New York.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, **61**, 509-528.

Reiser, M. (2008). Goodness-of-fit testing using components based on marginal frequencies of multinomial data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 331-360.

Reiser, M. (2012). Limited-information statistics when the number of variables is large. Proceedings of 2012 Joint Statistical Meetings, San Diego, CA.

Reiser, M. & Lin, G. (1999), A goodness-of-fit test for the latent class model when expected frequencies are small. In M. Sobel & M. Becker (Eds), *Sociological Methodology 1999*, 81-111. Boston: Blackwell.

Reiser, M. & Vandenberg, M. (1994). Validity of the chi-square test in dichotomous variable factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology*, 47, 85-107.

Reiser, M. & Cagnone, S. & Zhu, J. (2014). An Extended *GFfit* Statistic Defined on Orthogonal Components of Pearson's Chi-Square. *In JSM Proceedings, Biometrics Section, Alexandria, VA: American Statistical Association.*

Wittchen, H. & Gloster A. T. & Beesdo-Baum, K. & Fava, G. A. & Craske, M. G. (2010). Agoraphobia: a review of the diagnostic classificatory position and criteria. *Depression and Anxiety*, 27, 113-133.