Lack-Of-Fit Diagnostics Based On Standardized Residuals And Orthogonal Components Of Pearson's Chi-Square

M.M.K. Dassanayake* M. Reiser†

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 the cross classification of a large number of variables, these statistics may have low power and inaccurate Type I error rate due to sparseness. Pearson's statistic can be decomposed into orthogonal components associated with the marginal distributions of observed variables, and an omnibus fit statistic can be obtained as a sum of these components. When the statistic is a sum of components for lower-order marginals, it has good performance for Type I error rate and statistical power even when applied to a sparse table. In this study the individual components are examined as lack-of-fit diagnostics for models fit to binary cross-classified variables. Monte Carlo simulations are used to study the statistical power of individual orthogonal components to detect the source of the model lack-of-fit. The performance of orthogonal components as diagnostics is also compared to adjusted standardized residuals

Key Words: Item response model, Statistical power, Orthogonal components, Monte Carlo simulation, Standardized residuals

1. Introduction

Testing fit for a multinomial model commonly involves the null hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$, where $\boldsymbol{\pi}$ is a vector of multinomial probabilities, and $\boldsymbol{\pi}(\boldsymbol{\beta})$ is a vector of the multinomial probabilities as a function of parameters in the vector $\boldsymbol{\beta}$. When the model parameters $\boldsymbol{\beta}$ are unknown and estimated, the null hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$ is often tested with the Pearson-Fisher statistic:

$$X^2_{PF} = \sum_s z_s^2 \ ,$$

where

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

*Arizona State University, School of Mathematical and Statistical Sciences, P.O. Box 871804, Tempe, AZ 85287, USA

†Arizona State University, School of Mathematical and Statistical Sciences P.O. Box 871804, Tempe, AZ 85287, USA

and where

$$\begin{aligned}
\hat{\text{p}}_s &= \frac{n_s}{n} \text{ is element s of } \hat{\mathbf{p}},\ \text{the vector of multinomial proportions,} \\
n_s &= \text{element } s \text{ of } \mathbf{n},\ \text{the vector of observed frequencies,} \\
n &= \text{total sample size} = \sum_{s=1}^{T} n_s, \\
\hat{\boldsymbol{\beta}} &= \text{parameter estimator vector,} \\
\pi_s(\boldsymbol{\beta}) &= \text{the expected proportion for cell } s \\
\pi_s(\hat{\boldsymbol{\beta}}) &= \text{estimated expected proportion for cell s .}
\end{aligned}$$

The goodness-of-fit test based on Pearson's chi-squared statistic is sometimes considered to be an omnibus test that gives little information about the source of poor fit when the null hypothesis is rejected. It has also been recognized that the omnibus test can often be outperformed by focused or directional tests of lower order. When data are from a table formed by the cross-classification of a large number of variables, the Pearson's chi-square and the likelihood ratio statistics may have low power and inaccurate Type I error level due to sparseness. Statistics based on marginal distributions rather than the full joint distribution of the cross-classified variables can be used to remedy this issue. They have very good performance for Type I error rate and power (Reiser, 2008).

In this paper the orthogonal components related to second order marginals are examined as lack-of-fit diagnostics for models fit to binary cross-classified variables. Monte Carlo simulations will be used to study the statistical power of individual orthogonal components to detect the source of the model lack-of-fit. The performance of orthogonal components as diagnostics is also compared to adjusted standardized residual. Finally the model will be tested on the real world data.

2. Marginal Proportions

A traditional statistic such as Pearson's chi-square uses 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. Marginal proportions are used to develop test statistics presented in Section 3.2.

2.1 First- and Second-Order Marginals

The relationship between joint proportions and marginals can be shown by using zeros and 1's to code the levels of dichotomous response random variables, $Y_i, i = 1, 2, \ldots, q$, where Y_i follow the Bernoulli distribution with parameter P_i . Then, a q-dimensional vector of zeros and 1's, sometimes called a response pattern, will indicate a specific cell from the contingency table formed by the cross-classification of q response variables. For dichotomous response variables, a response pattern is a sequence of zeros and 1's with length q. The $T = 2^q$-dimensional set of response patterns can be generated by varying the levels of the q^{th} variable most rapidly, the $q^{th} - 1$ variable next, etc. Define $\boldsymbol{V}$ as the T by q matrix with response patterns as rows.

For instance when $q = 3$,

$$\boldsymbol{V} = \begin{pmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,0 \\ 0\,1\,1 \\ 1\,0\,0 \\ 1\,0\,1 \\ 1\,1\,0 \\ 1\,1\,1 \end{pmatrix} .$$

Let v_{is} represent element i of response pattern s, $s = 1, 2, \ldots, T$. Then, under the model $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$, the first-order marginal proportion for variable Y_i can be defined as

$$P_i(\boldsymbol{\beta}) = \text{Prob}(Y_i = 1|\boldsymbol{\beta}) = \sum_s v_{is}\pi_s(\boldsymbol{\beta}),$$

and the true first-order marginal proportion is given by

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

Under the model, the second-order marginal proportion for variables Y_i and Y_j can be defined as

$$P_{ij}(\boldsymbol{\beta}) = \text{Prob}(Y_i = 1, Y_j = 1|\boldsymbol{\beta}) = \sum_s v_{is}v_{js}\pi_s(\boldsymbol{\beta}),$$

where $j = 1, 2, \ldots, q-1$; $i = j+1, \ldots q$, and the true second-order marginal proportion is given by

$$P_{ij} = \text{Prob}(Y_i = 1, Y_j = 1) = \sum_s v_{is}v_{js}\pi_s .$$

2.2 Higher-Order Marginals

A general matrix $\mathbf{H}_{[t:u]}$ to obtain marginals of any order can be defined in a similar fashion by using Hadamard products among the columns of $\boldsymbol{V}$. The symbol $\mathbf{H}_{[t:u]}$, $t \leq u \leq q$, denotes the transformation matrix that would produce marginals from order t up to and including order u. Furthermore, $\mathbf{H}_{[t]} \equiv \mathbf{H}_{[t:t]}$. $\mathbf{H}_{[1:q]}$ gives a one-to-one mapping from joint proportions to the set of $(2^q - 1)$ marginal proportions:

$$\boldsymbol{P} = \mathbf{H}_{[1:q]}\boldsymbol{\pi} ,$$

where

$$\boldsymbol{P} = (P_1,\ P_2,\ P_3, \ldots P_q,\ P_{12},\ P_{13}, \ldots P_{q-1,q},\ P_{1,1,2} \ldots P_{q-2,q-1,q} \ldots P_{1,2,3\ldots q})'$$

is the vector of marginal proportions (Bartholomew, 1987).

For instance,

$$\mathbf{H}_{[3]} = \begin{pmatrix} (\boldsymbol{v}_1 \circ \boldsymbol{v}_2 \circ \boldsymbol{v}_3)' \\ (\boldsymbol{v}_1 \circ \boldsymbol{v}_2 \circ \boldsymbol{v}_4)' \\ \vdots \\ (\boldsymbol{v}_1 \circ \boldsymbol{v}_2 \circ \boldsymbol{v}_q)' \\ (\boldsymbol{v}_2 \circ \boldsymbol{v}_3 \circ \boldsymbol{v}_4)' \\ \vdots \\ (\boldsymbol{v}_2 \circ \boldsymbol{v}_3 \circ \boldsymbol{v}_q)' \\ \vdots \\ (\boldsymbol{v}_{q-2} \circ \boldsymbol{v}_{q-1} \circ \boldsymbol{v}_q)' \end{pmatrix},$$

where $\boldsymbol{v}_f$ represents column f of matrix $\boldsymbol{V}$, and $\boldsymbol{v}_f \circ \boldsymbol{v}_g \circ \boldsymbol{v}_h$ represents the Hadamard product of columns f, g and h. Thus when q=3,

$$\mathbf{H}_{[1:3]} = \begin{pmatrix} 0\,1\,0\,1\,0\,1\,0\,1 \\ 0\,0\,1\,1\,0\,0\,1\,1 \\ 0\,0\,0\,0\,1\,1\,1\,1 \\ \cdots \\ 0\,0\,0\,1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,1\,0\,1 \\ 0\,0\,0\,0\,0\,0\,1\,1 \\ \cdots \\ 0\,0\,0\,0\,0\,0\,0\,1 \end{pmatrix}.$$

2.3 Residuals

Define the unstandardized residual $\mathrm{r}_s = \hat{\mathrm{p}}_s - \pi_s(\hat{\boldsymbol{\beta}})$, and denote the vector of unstandardized residuals as **r** with element r_s .

A vector of simple residuals for marginals of any order may be defined such that

$$\boldsymbol{e} = \mathbf{H}(\hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) = \mathbf{H}\mathbf{r},$$

.

3. Testing Fit on Marginal Distributions

3.1 Linear Combinations of Joint Frequencies

A traditional composite null hypothesis for a test of fit on a multinomial model is $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$. Linear combinations of $\boldsymbol{\pi}$ may be tested under the null hypothesis $H_o\colon \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\beta})$. $\mathbf{H}$ may specify linear combinations that form marginal proportions as defined in the previous section.

3.2 Test Statistic

The use of components of Pearson's chi-square statistic has a long history dating back at least to Lancaster (1969). The motivation for components has been the possibility that a directional test would have higher power for certain alternative hypotheses than the omnibus goodness-of-fit test (Rayner & Best, 1989). Reiser(1996, 2008) and Reiser and Lin (1999) proposed statistics that can be obtained from orthogonal components defined on marginal proportions. These statistics have higher power under some circumstances, and they usually perform well when applied to sparse frequency tables.

$\sqrt{n}\,\mathbf{r}$ has asymptotic covariance matrix $\mathbf{\Omega_r}$, where

$$\mathbf{\Omega_r} = (D(\boldsymbol{\pi}(\boldsymbol{\beta})) - \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\boldsymbol{\beta})' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'),$$

and where

$$\begin{aligned} D(\boldsymbol{\pi}(\boldsymbol{\beta})) &= \text{diagonal matrix with } (s,s) \text{ element equal to } \pi_s(\boldsymbol{\beta}), \\ \mathbf{A} &= D(\boldsymbol{\pi}(\boldsymbol{\beta}))^{-1/2}\frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \\ \text{and } \mathbf{G} &= \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \ . \end{aligned}$$

See Haberman (1973). Then consider the linear combination $\boldsymbol{e} = \mathbf{Hr}$. If $\mathbf{H}$ contains $2^q - g - 1$ linearly independent rows corresponding to marginals from order 1 to q, then define the statistic

$$X^2_{[1:q]} = n\mathbf{r}'\mathbf{H}'\mathbf{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{Hr}.$$

Here the statistic is evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is now consistent and efficient for $\boldsymbol{\beta}$, such as the maximum likelihood estimator, and where $\mathbf{\Omega}_{\boldsymbol{e}} = \mathbf{H}\mathbf{\Omega_r}\mathbf{H}'$. With the added condition that the rows of $\mathbf{H}$ are linearly independent of the columns of $\mathbf{G}$, i.e., $\text{rank}(\mathbf{H}' \vdots \mathbf{G}) = T + g$, $X^2_{[1:q]}$ can be shown to be equivalent to X^2_{PF} due to the one-to-one correspondence of the joint and marginal proportions. See also Reiser (2008). To obtain orthogonal components, define the upper triangular matrix $\boldsymbol{F}$ such that $\boldsymbol{F}'\mathbf{\Omega}_{\boldsymbol{e}}\boldsymbol{F} = \boldsymbol{I}$. $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\mathbf{\Omega}_{\boldsymbol{e}}$. Then writing $\mathbf{\Omega}_{\boldsymbol{e}}$ as $\boldsymbol{CC}'$,

$$\begin{aligned} X^2_{PF} &= n\mathbf{r}'\mathbf{H}'(\hat{\boldsymbol{C}}')^{-1}\hat{\boldsymbol{C}}'(\hat{\boldsymbol{C}}\hat{\boldsymbol{C}}')^{-1}\hat{\boldsymbol{C}}(\hat{\boldsymbol{C}})^{-1}\mathbf{Hr} \\ &= n\mathbf{r}'\mathbf{H}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\mathbf{Hr} \end{aligned}$$

where $\widehat{\boldsymbol{F}}$ and $\hat{\boldsymbol{C}}$ are the matrices $\boldsymbol{F}$ and $\boldsymbol{C}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$.

Premultiplication by $(\boldsymbol{C}')^{-1}$ orthonormalizes the matrix $\mathbf{H}_{[1:q]}$ relative to the matrix $D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$. Let $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q]}$, then

$$X^2_{PF} = n\mathbf{r}'(\widehat{\mathbf{H}}^*)'\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\mathbf{H}}^* = \mathbf{H}^*(\hat{\boldsymbol{\beta}})$.

Define

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

Then

$$X^2_{PF} = \hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}} = \sum_{j=1}^{j=T-g-1} \hat{\gamma}_j^2,$$

and the elements $\hat{\gamma}_j^2$ are orthogonal components of X^2_{PF}. Since $\widehat{\mathbf{H}}^*\mathbf{r}$ has asymptotic covariance matrix $\boldsymbol{F}'\mathbf{\Omega}_{\boldsymbol{e}}\boldsymbol{F} = \boldsymbol{I}_{T-g-1}$, the elements $\hat{\gamma}_j^2$ are asymptotically independent χ^2_1 random variables.

3.3 Standardized Residuals

There are number of ways to define a residual for the first-, or second-order marginal. Some possibilities are

$$\mathbf{e} = \hat{P}_{ij}(1,1) - P_{ij}(1,1|\hat{\boldsymbol{\beta}}),\ \ \mathbf{e} = \frac{\hat{P}_{ij} - P_{ij}(1,1|\hat{\boldsymbol{\beta}})}{P_{ij}(1,1|\hat{\boldsymbol{\beta}})^{\frac{1}{2}}}\ \ and\ \ \mathbf{e} = \frac{\hat{P}_{ij} - P_{ij}(1,1|\hat{\boldsymbol{\beta}})}{P_{ij}(1,1|\hat{\boldsymbol{\beta}})},$$

$$where\ P_{ij}(1,1|\hat{\boldsymbol{\beta}}) = P(Y_i = 1, Y_j = 1|\hat{\boldsymbol{\beta}}) = h'_\ell \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}),$$
$$\hat{P}_{ij}(1,1) = \hat{P}(Y_i = 1, Y_j = 1) = h'_\ell \hat{\mathbf{p}},$$

and h'_ℓ is row ℓ of matrix $\mathbf{H}$ defined earlier. The definition given above in the middle has a similar form to the traditional standardized residuals.

Recall,

$$\boldsymbol{e} = \mathbf{H}(\hat{\mathrm{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) = \mathbf{Hr},\ n^{\frac{1}{2}}\boldsymbol{e} \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{e}}),\ and\ \boldsymbol{\Omega}_{\mathbf{e}} = \mathbf{H}\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{H}'$$

Then $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ will be a consistent estimator for the $\boldsymbol{\Omega}_{\mathbf{e}}$ where

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = n^{-1}\mathbf{H}(\mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}' \Bigg|_{\pi=\pi(\hat{\boldsymbol{\beta}}),\ \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$$

This follows from the results in section 3.2 and the results from (Reiser, 1996). Estimated standard error for the residual is obtained by taking the square root of the diagonal element of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$. The adjusted residual is obtained by $\frac{n^{1/2}e}{\hat{\sigma}_e}$.

3.4 Application to Factor Analysis

When categorical manifest variables are hypothesized to be associated with a continuous latent variable, the model is known as categorical variable factor analysis and sometimes as the item response theory model. In order to investigate the source of the model lack-of-fit, a comparison of the statistics reviewed in the previous sections will be presented using this model with one factor.

According to the categorical factor model, the probability of the response to a manifest variable, sometimes also referred to as an item, can be given by a logistic item response function:

$$P(Y_i = 1 \mid \boldsymbol{\beta}'_i, X = x) = (1 + \exp(-\beta_{i0} - \beta_{i1}x))^{-1} \tag{3.1}$$

$$\begin{aligned} &\text{where } Y_i \text{ represents the response to item } i,\\ &\beta_{i0} = \text{intercept parameter for item } i\\ &\beta_{i1} = \text{slope parameter for item } i\\ &\boldsymbol{\beta}'_i = (\beta_{0i}, \beta_{1i})\\ &x = \text{value taken on by latent random variable } X \end{aligned}$$

Since

$$P(Y_i = 0 \mid \boldsymbol{\beta}'_i, X = x) = 1.0 - \pi(Y_i = 1 \mid \boldsymbol{\beta}'_i, X = x),$$

it follows that

$$P(Y_i = y_i \mid \boldsymbol{\beta}'_i, x) = P(Y_i = 1 \mid \boldsymbol{\beta}'_i, x)^{y_i}[1.0 - P(Y_i = 1 \mid \boldsymbol{\beta}'_i, x)]^{1-y_i}$$

It is assumed that, *conditional* upon the latent variable, responses to the manifest variables are independent. Let $\boldsymbol{Y}$ represent a random vector of responses to the items, with element Y_i, and let $\mathbf{y}$ represent a realized value of $\boldsymbol{Y}$. Then

$$P(\boldsymbol{Y} = \mathbf{y} \mid \boldsymbol{\beta}, x) = \prod_{i=1}^{k} \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)^{y_i}[1 - \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)]^{1-y_i} \tag{3.2}$$

$$\text{where } \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{i1} \\ \beta_{02} & \beta_{12} \\ \beta_{03} & \beta_{13} \\ \vdots & \vdots \\ \beta_{0q} & \beta 1q \end{pmatrix}.$$

Finally, the probability of response pattern s, say, is obtained by taking the expected value of the conditional probability over the distribution of X in the population, and is sometimes called the marginal probability:

$$\pi_s(\boldsymbol{\beta}) = \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}, x) f(x) dx \tag{3.3}$$

where $f(x)$ is the density function of X in the population of respondents.

If $\boldsymbol{U}$ represents a T-dimensional multinomial random vector of frequencies associated with the response patterns, the distribution of $\boldsymbol{U}$ is given by

$$\pi(\boldsymbol{U} = \mathbf{n}) = n! \prod_{s=1}^{T} \frac{[\pi_s(\boldsymbol{\beta})]^{n_s}}{n_s!} \tag{3.4}$$

$$\begin{aligned} \text{where } \mathbf{n} &= \text{vector of observed frequencies} \\ n_s &= \text{element } s \text{ of } \mathbf{n} \\ n &= \text{total sample size} = \sum_{s=1}^{T} n_s \quad . \end{aligned}$$

4. Simulation Study

A Monte Carlo simulation study to assess components and standardized residuals was performed under following conditions : q=8 manifest variables, and β'_0 = (-2.0, -1.5, -1, -0.5 , 0.5 ,1, 1.5, 2). The first simulation study was done to check the Type I error rate of the model. 1000 data sets were generated using Monte-Carlo methods related to one factor model where β'_1 = (0.1, 0.1, 0.1, 1.2 , 1.2 , 1.2 , 0.2 , 0.2). Then a two parameter IRT model was built for each of these datasets and Type I error rate was calculated. Since each orthogonal component is distributed as chi-square with one degree of freedom, to calculate the Type I error for each component, the sum of the number of cases that exceed the chi-square critical value (at 5% significance level) with one degree of freedom was divided by the number of datasets. A similar process was used for the standardized residuals, but for the critical value the standard normal distribution was used. This process was repeated for sample sizes 300, 500 and 1000.

Some of the simulation results for the Type I error rate are shown in Table 1 to Table 4. To check the distributional assumption, Chi-square Q-Q plots were calculated for the values of each orthogonal component and some results are shown in Figures 1, 3, 5. Similarly, Normal probability plots were calculated for the values of each standardized residuals. Results are shown in Figures 2,4,6. KolmogorovSmirnov test result for normality assumption are shown in Table 5.

As the next approach, 1000 data sets were generated using Monte-Carlo simulations related to a two factor (two latent variables) model. Loadings for the first factor were (1,1,1,0,0,0,1,1) and loadings for the second factor were (0.1, 0.1, 0.1, 1.2 , 1.2 , 1.2 , 0.2 , 0.2). Then a two parameter IRT model was built for each of these datasets and power was calculated. In the simulation, the model under H_0 is misspecified with a one factor model. Since each orthogonal component is distributed as chi-square with one degree of freedom, to calculate the power for each component, sum of the number of cases that exceed the chi-square critical value (at 5% significance level) with one degree of freedom was divided by the number of datasets. A similar process was used for the standardized residuals except the critical value from standard normal distribution was used. This process was repeated for sample size 300, 500 and 1000. It is interesting to see how the power of the orthogonal components and standardized residuals change for small factor loadings. Thus, above process was repeated with factor loadings 0.6 for item 4,5 and 6. When the sample size and/or the factor loadings are too small, estimate of the standard deviations tends to become negative or very close to zero. To fix this issue shrinkage estimator was incorporated. Mixing parameter of the Shrinkage estimator was calculated based on the largest negative eigen value in magnitude.

5. Simulation Results

As shown earlier, the Pearson-Fisher statistic for a composite null hypothesis can be partitioned into orthogonal components defined on marginal distributions. When the manifest variables are binary, each of these components, γ_j^2, is distributed as an independent $\chi^2_{(1)}$ random variate. These components can be used as item diagnostics for models fit to binary cross-classified variables when the result of an omnibus test indicates that a model should be rejected.

To demonstrate this aspect simulation study was developed as described in the previous section. Table 6, in the Appendix shows the power of each second-order marginal component. Since there were 8 items there will be (8*7)/2= 28 second order marginal components. During the Monte-Carlo simulations higher weights were given to item 4,5 and 6. Thus components related to item 4,5 and 6 should indicate a higher lack of fit, hence higher power. By examining the highlighted values in Table 6, it is clear that the power of second order marginal components (4,5), (4,6) and (5,6) were significantly higher compared to other components. Thus these second order components were successful in detecting poorly fit model. This process was repeated for n=300, n=500 and n=1000 and when the sample size increases power increases and components were more successful in predicting the poor fit. However, due to page limitations, only the results related to n=500 is included in the Appendix. When misspecification is larger, or in other words, weights given to the misspecified items are lager (weights of the items 4,5 and 6 were increased from 0.6 to 1.2), the power tends to increase. Table 8 illustrate this information. Standardized residuals can also be calculated on the univariate and bivariate distributions of the manifest variables (Reiser, 1996). By examining the values in the Tables 7 and 9, it is clear that the power of second order marginal components (4,5), (4,6) and (5,6) were significantly higher compared to other components. Thus these second order residual components were also successful in detecting lack of fit.

However if the Type I error is not small compare to power then these result does not have much meaning in terms of practical applications. Thus Type I error rate were calculated for the same settings. Results (Table 1 to 4) shows that the Type I error rates are within 0.05 for all the simulations. Thus our power calculation indeed have meaningful results. As explained in previous sections each orthogonal component is distributed with chi-square one degree of freedom. To check this assumption, chi-square Q-Q plots were built for the simulation values related to each component. Some of the results are shown in the Appendix. None of the plots had significant deviations from the straight line indicating an adequate fit for the chi-square distribution. A similar process was done to check the normality assumption of the standardized residual and some of the results are shown in the Appendix.

Initially there were some Q-Q plots showing deviations from the straight line. Further investigation was done and found that the estimates produced by the PROC IRT method in the SAS were not stable for the small factor loadings. Mplus (Muthén & Muthén, 1998) parameter estimates were more stable compared to SAS. Thus for the Type I error calculation Mplus estimates were incorporated. After using Mplus estimates none of the Q-Q plots showed deviations from straight line. For further validation KolmogorovSmirnov test for normality was incorporated. Results related to n=300 are given in the table 5. Almost all the results were not significant at 0.01, which indicates the validity of the Normality assumption. For larger samples sizes (n=1000) all p-values were larger than 0.02, hence the validity of the Normality assumption.

When the sample size and/or the factor loadings are too small, the estimate of the standard error for the residuals tends to become negative or very close to zero. Thus out of 1000 simulation only around 750-850 simulations were successful. To fix this issue a shrinkage estimator was incorporated for the standard error. After incorporating the shrinkage estimator number of successful iterations increased to 970-1000.

6. Real World Application

The Epidemiologic Catchment Area (ECA) program of research was initiated in response to the 1977 report of the President's Commission on Mental Health. The purpose was to collect data on the prevalence and incidence of mental disorders and on the use of and need for services by the mentally ill. Independent research teams at five universities (Yale, Johns Hopkins, Washington University, Duke University, and University of California at Los Angeles), in collaboration with National Institute of Mental Health (NIMH), conducted the studies with a core of common questions and sample characteristics. The ECA study was mainly focused on mental disorders related to manic episode, major depressive episode, dysthymia, bipolar disorder, alcohol abuse or dependence, drug abuse or dependence, schizophrenia, schizophreniform, obsessive compulsive disorder, phobia, somatization, panic, antisocial personality, and anorexia nervosa.

For this study chosen 8 items related to the mental disorder phobia were chosen from the ECA to analyze as a real world application. The data set was limited to Johns Hopkins (Baltimore, MD) area. There were 3316 observations related to these specifications. The selected items are given below.

DIS068A - fear of heights
DIS068F - closed places
DIS068I - fear of speaking in front of close friends
DIS068J - fear of speaking to strangers
DIS068K - storms
DIS068L - water
DIS068M - spiders
DIS068N - fear of harmless animals

Results related to real world application is shown in the table 10 and table 11. According to the table 10 and 11, components (1,8) (3,4) (3,7) (3,8) related to Catchment Area Study has larger values indicating that these pairs of variables have associations not explain by the IRT model. Further, variable 3, fear of speaking in front of close friends appears in three of these large components.

7. Conclusion

In this study, second order marginals related to Orthogonal components and Standardized residual were examined as lack-of-fit diagnostics. Simulations were based on two parameter IRT model for a two latent variable model and were successful in indicating pair of variables for which the model does not fit well. When the sample size increases, ability to indicate pair of variables for which the model does not fit well increases significantly. For instance when n=300 (with lager factor loadings) power was around 0.5 and when n=1000 power was around 0.9. Even for the small sample sizes (n=300) and small factor loadings, second order marginals related to Orthogonal components and Standardized residuals were successful in detecting the variables for which the model does not fit well. However when the sample size and/or the factor loadings are too small, estimates of the standard errors tends to become negative or very close to zero. Thus out of 1000 simulation only around 750-850 simulations were successful. Shrinkage estimator was successful in fixing this error. After incorporating the shrinkage estimator number of successful iterations increased to 970-1000. Further, the estimates produced by the PROC IRT method in the SAS were not stable for the small factor loadings. Mplus parameter estimates were more stable compared to SAS.

8. Appendix

Table 1: Type I error of the Orthogonal components for n=500

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.053	0.056	0.048	0.049	0.042	0.044	0.049
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.05	0.051	0.047	0.045	0.052	0.056	0.044
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.061	0.05	0.052	0.06	0.048	0.051	0.049
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.045	0.042	0.046	0.042	0.052	0.051	0.047

Table 2: Type I error of the Standardized residuals for n=500

res(1,2)	**res(1,3)**	**res(1,4)**	**res(1,5)**	**res(1,6)**	**res(1,7)**	**res(1,8)**
0.044	0.039	0.046	0.056	0.057	0.05	0.04
res(2,3)	**res(2,4)**	**res(2,5)**	**res(2,6)**	**res(2,7)**	**res(2,8)**	**res(3,4)**
0.045	0.051	0.055	0.045	0.04	0.05	0.041
res(3,5)	**res(3,6)**	**res(3,7)**	**res(3,8)**	**res(4,5)**	**res(4,6)**	**res(4,7)**
0.053	0.049	0.051	0.046	0.056	0.048	0.053
res(4,8)	**res(5,6)**	**res(5,7)**	**res(5,8)**	**res(6,7)**	**res(6,8)**	**res(7,8)**
0.041	0.065	0.046	0.04	0.062	0.053	0.041

Table 3: Type I error of the Orthogonal components for n=1000

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.053	0.065	0.062	0.053	0.05	0.058	0.041
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.053	0.035	0.048	0.049	0.055	0.054	0.048
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.053	0.038	0.047	0.046	0.047	0.05	0.043
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.062	0.053	0.055	0.05	0.044	0.058	0.047

Table 4: Type I error of the Standardized residuals for n=1000

res(1,2)	**res(1,3)**	**res(1,4)**	**res(1,5)**	**res(1,6)**	**res(1,7)**	**res(1,8)**
0.054	0.065	0.06	0.054	0.04	0.053	0.049
res(2,3)	**res(2,4)**	**res(2,5)**	**res(2,6)**	**res(2,7)**	**res(2,8)**	**res(3,4)**
0.053	0.035	0.049	0.04	0.058	0.051	0.047
res(3,5)	**res(3,6)**	**res(3,7)**	**res(3,8)**	**res(4,5)**	**res(4,6)**	**res(4,7)**
0.043	0.043	0.052	0.047	0.056	0.045	0.058
res(4,8)	**res(5,6)**	**res(5,7)**	**res(5,8)**	**res(6,7)**	**res(6,8)**	**res(7,8)**
0.045	0.038	0.053	0.054	0.043	0.047	0.038

Table 5: KolmogorovSmirnov test result for Normality of the Standardized residuals, n=300

res(1,2)	**res(1,3)**	**res(1,4)**	**res(1,5)**	**res(1,6)**	**res(1,7)**	**res(1,8)**
0.013	>0.15	>0.15	>0.15	>0.15	0.0945	<0.01
res(2,3)	**res(2,4)**	**res(2,5)**	**res(2,6)**	**res(2,7)**	**res(2,8)**	**res(3,4)**
>0.15	>0.15	>0.15	>0.15	>0.11	>0.15	>0.15
res(3,5)	**res(3,6)**	**res(3,7)**	**res(3,8)**	**res(4,5)**	**res(4,6)**	**res(4,7)**
>0.15	0.023	>0.15	>0.15	>0.15	>0.15	>0.15
res(4,8)	**res(5,6)**	**res(5,7)**	**res(5,8)**	**res(6,7)**	**res(6,8)**	**res(7,8)**
>0.15	0.0568	>0.15	>0.15	0.0357	>0.15	0.1185

Table 6: Power of the Orthogonal components for n=500

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.044534	0.042510	0.049595	0.051619	0.058704	0.039474	0.062753
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.048583	0.052632	0.058704	0.054656	0.044534	0.052632	0.054656
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.063765	0.060729	0.055668	0.045547	0.33300	0.30567	0.048583
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.046559	0.29352	0.047571	0.049595	0.050607	0.041498	0.052632

Table 7: Power of the Standardized residuals for n=500

res(1,2)	**res(1,3)**	**res(1,4)**	**res(1,5)**	**res(1,6)**	**res(1,7)**	**res(1,8)**
0.046559	0.051619	0.053644	0.048583	0.059717	0.044534	0.038462
res(2,3)	**res(2,4)**	**res(2,5)**	**res(2,6)**	**res(2,7)**	**res(2,8)**	**res(3,4)**
0.053644	0.052632	0.061741	0.050607	0.052632	0.046559	0.042510
res(3,5)	**res(3,6)**	**res(3,7)**	**res(3,8)**	**res(4,5)**	**res(4,6)**	**res(4,7)**
0.053644	0.048583	0.042510	0.053644	0.36640	0.31680	0.044534
res(4,8)	**res(5,6)**	**res(5,7)**	**res(5,8)**	**res(6,7)**	**res(6,8)**	**res(7,8)**
0.060729	0.29858	0.047571	0.044534	0.050607	0.052632	0.050607

Table 8: Power of the Orthogonal components with higher factor loadings for n=500

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.27199	0.29424	0.078868	0.083923	0.27806	0.082912	0.099090
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.39333	0.11223	0.13953	0.34681	0.093023	0.093023	0.16886
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.22952	0.48534	0.12437	0.096057	0.58544	0.59151	0.043478
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.063701	0.79778	0.039434	0.12336	0.068756	0.066734	0.075834

Table 9: Power of the Standardized residuals with higher factor loadings for n=500

res(1,2)	**res(1,3)**	**res(1,4)**	**res(1,5)**	**res(1,6)**	**res(1,7)**	**res(1,8)**
0.26997	0.29019	0.11830	0.10313	0.11628	0.23256	0.16785
res(2,3)	**res(2,4)**	**res(2,5)**	**res(2,6)**	**res(2,7)**	**res(2,8)**	**res(3,4)**
0.36906	0.14661	0.14661	0.12639	0.24975	0.20627	0.13953
res(3,5)	**res(3,6)**	**res(3,7)**	**res(3,8)**	**res(4,5)**	**res(4,6)**	**res(4,7)**
0.16178	0.14459	0.29626	0.25379	0.72396	0.67442	0.11122
res(4,8)	**res(5,6)**	**res(5,7)**	**res(5,8)**	**res(6,7)**	**res(6,8)**	**res(7,8)**
0.075834	0.68857	0.11426	0.10111	0.092012	0.077856	0.27503

Table 10: Standardized residuals for Catchment Area study

Obs	Var_1	Var_2	Residual	Obs	Var_1	Var_2	Residual
1	1	2	1.1541	15	3	5	-2.3842
2	1	3	-1.3479	16	3	6	-1.499
3	1	4	-2.2894	17	3	7	-3.5368
4	1	5	0.6607	18	3	8	-1.8414
5	1	6	0.6195	19	4	5	-3.2363
6	1	7	-1.2162	20	4	6	-1.5218
7	1	8	-3.3259	21	4	7	-2.5327
8	2	3	-0.5642	22	4	8	-1.7878
9	2	4	-1.9678	23	5	6	-1.4999
10	2	5	-0.9038	24	5	7	0.6942
11	2	6	-0.1089	25	5	8	0.2255
12	2	7	-2.7486	26	6	7	-2.8479
13	2	8	-1.3766	27	6	8	-2.0981
14	3	4	4.7561	28	7	8	2.3423

Table 11: Orthogonal Components for Catchment Area study

Obs	Mar_1	Mar_2	Component	Obs	Mar_1	Mar_2	Component
1	3	4	32.1183	15	3	5	1.4616
2	1	8	11.8271	16	4	6	1.2813
3	3	7	10.9885	17	1	7	1.2056
4	3	8	10.0787	18	7	8	1.1582
5	6	7	9.0142	19	1	6	0.7681
6	5	8	4.4339	20	2	7	0.5693
7	5	6	3.9102	21	2	4	0.5086
8	1	4	3.7032	22	6	8	0.4736
9	2	6	3.6350	23	2	8	0.4451
10	4	7	2.7066	24	1	5	0.2769
11	4	8	2.5555	25	2	5	0.1728
12	4	5	1.8734	26	5	7	0.1708
13	1	2	1.6371	27	2	3	0.0562
14	1	3	1.5046	28	3	6	0.0005

Figure 1: QQ plots for the simulation n=300

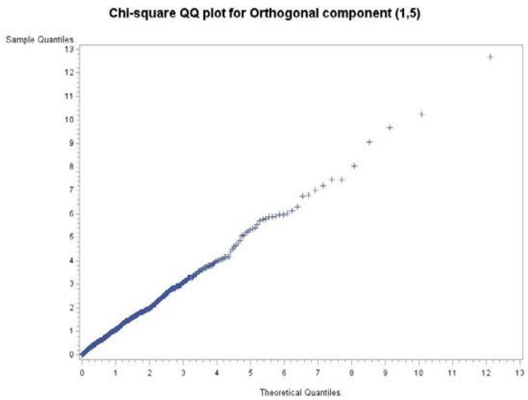


Figure 2: QQ plots for the simulation n=300

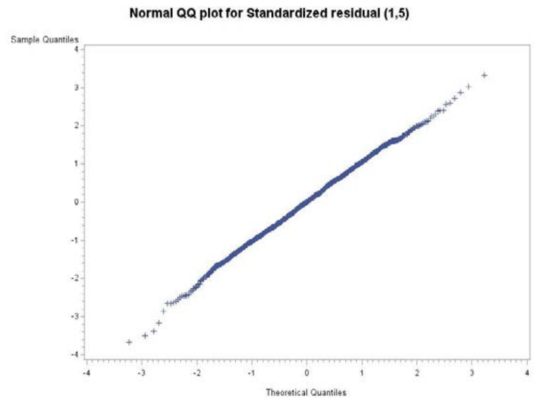


Figure 3: QQ plots for the simulation n=500

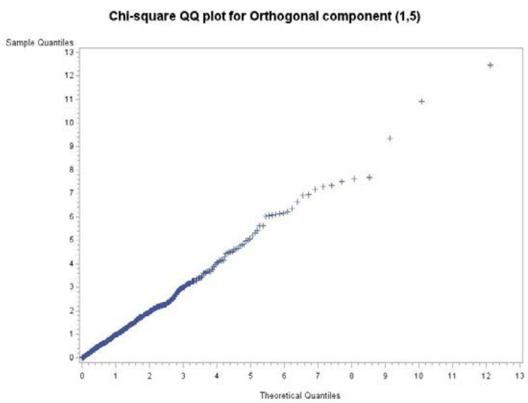


Figure 4: QQ plots for the simulation n=500

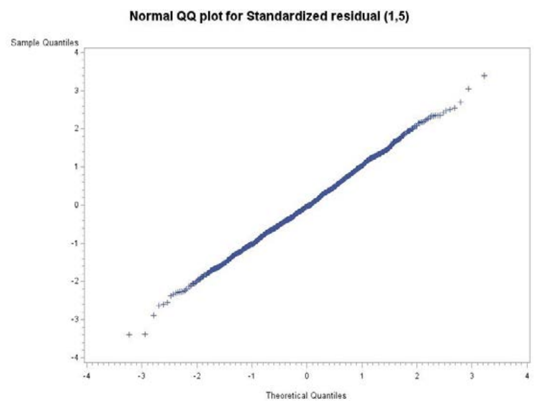


Figure 5: QQ plots for the simulation n=1000

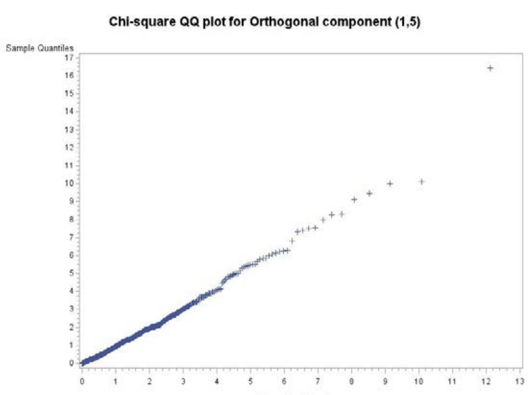
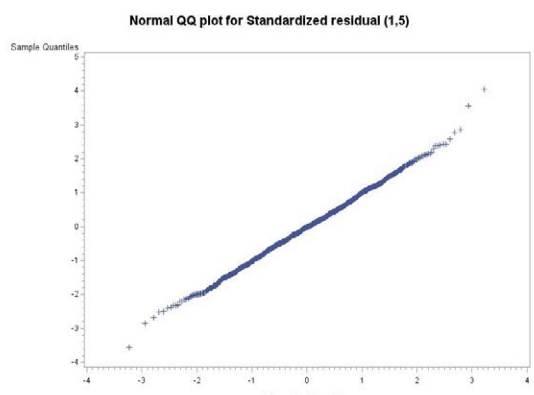


Figure 6: QQ plots for the simulation n=1000



REFERENCES

Bartholomew, D. J., & Leung, S. O. (2002). A goodness of fit test for sparse 2^P contingency tables. *British Journal of Mathematical and Statistical Psychology, 55*, 1-15.

Bartholomew, D.J., Knott, M. & Moustaki (2011). *Latent variable models and factor analysis: A unified approach, 3rd Edition.* New York: Wiley.

Birch, M. W. (1964). A new proof of the Pearson-Fisher Theorem. *Annals of Mathematical Statistics, 35*, 818-824.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179-197.

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

Cai, L., Maydeu-Olivares, A., Coffman, D.L., & Thissen, D. (2006). Limited information goodness of fit testing of item response theory models for sparse 2p tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173-194.

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

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205-220.

Joreskog & Moustaki (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of thurstonian models for paired comparison data under multiple judgement sampling. *Psychometrika, 66*, 209-228.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*, 1009-1020.

Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71, 713-732.

Maydue-Olivares, A., Garcia-forero, c., Gallardo-Pujol, D., & Renom, J. (2009). Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology: European Journal of Research Methods for the behavioral and social sciences,* 5, 131-136.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.

Muthén, L.K & Muthén, B.O. (1998-2010). Mplus User's Guide. Sixth edition. Los Angeles, CA.

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., & 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.

Takane, Y., and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 292-408.

Tollenaar, N., & Mooijaart, A. (2003). Type I errors and power of the parametric bootstrap goodness-of-fit test: Full and limited information. *British Journal of Mathematical and Statistical Psychology, 56*, 271-288.

U.S. Dept. of Health and Human Services, National Institute of Mental Health. EPIDEMIOLOGIC CATCHMENT AREA SURVEY OF MENTAL DISORDERS, WAVE I (HOUSEHOLD), 1980-1985 [Computer file]. Rockville, MD: National Institute of Mental Health [producer], 1985. Ann Arbor, MI: inter-university Consortium for Political and Social Research [distributor], 1991.