A Study of Components of Pearson's Chi-Square Based on Marginal Distributions of Cross-Classified Tables for Binary Variables

by

Mudiyanselage Maduranga Kasun Dassanayake

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Mark Reiser, Chair
Ming-Hung Kao
Jeffrey Wilson
Ioannis Kamarianakis
Robert St. Louis

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

The Pearson and likelihood ratio statistics are well-known in goodness-of-fit testing and are commonly used for models applied to multinomial count data. When data are from a table formed by the cross-classification of a large number of variables, these goodness-of-fit statistics may have lower 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 dissertation, goodness-of-fit statistics using orthogonal components based on second- third- and fourth-order marginals were examined. If lack-of-fit is present in higher-order marginals, then a test that incorporates the higher-order marginals may have a higher power than a test that incorporates only first- and/or second-order marginals. To this end, two new statistics based on the orthogonal components of Pearson's chi-square that incorporate third- and fourth-order marginals were developed, and the Type I error, empirical power, and asymptotic power under different sparseness conditions were investigated. Individual orthogonal components as test statistics to identify lack-of-fit were also studied. The performance of individual orthogonal components to other popular lack-of-fit statistics were also compared. When the number of manifest variables becomes larger than 20, most of the statistics based on marginal distributions have limitations in terms of computer resources and CPU time. Under this problem, when the number manifest variables is larger than or equal to 20, the performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic, fit to confirmatory dichotomous variable factor analysis model, and the performance of Tollenaar and Mooijaart (2003) statistic were investigated.

to my

MOTHER and FATHER

with love

ACKNOWLEDGMENTS

At the very outset, I wish to place on record my deep gratitude to Dr. Mark Reiser; who gave inspiration to this dissertation and as adviser, offered guidance, suggestions and advice throughout this task. I would also like to thank my committee members, Dr. Jeffrey Wilson, Dr. Ioannis Kamarianakis, Dr. Ming-Hung Kao and Dr. Robert St Louis for reserving some of the precious time to review this work. Last but not least, I express deepest gratitude to my family and friends who were always with me in need and helped me in many ways.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Statistical modeling often involves finding a model that describes data of interest, and it is important to test the fit of the model because inferences drawn on poorly fitting models can be misleading. Since the first appearance by Pearson (1924), chi-square tests have been a common approach to test goodness of fit related to multinomial models. For a simple null hypothesis where the random sample comes from a population with completely specified cumulative distribution function F(x), the Pearson's chi-square statistic (χ_p^2) has an approximate chi-squared distribution with T-1 degrees of freedom in large samples, where T is the number of cells. On the other hand, for a composite null hypothesis where the null distribution depends on a vector of g unknown parameters $\boldsymbol{\beta} = (\beta_1, ..., \beta_g)^T$, goodness of fit can be tested using the Pearson-Fisher (PF) statistic,

$$X_{PF}^2 = \sum_s z_s^2, \tag{1.1}$$

where

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

and where, $\hat{\mathrm{p}}_s$ is element s of $\hat{\mathbf{p}}$ a vector of multinomial proportions, n is total sample size, $\hat{\boldsymbol{\beta}}$ is parameter estimator vector, $\pi_s(\boldsymbol{\beta})$ is the expected proportion for cell s and $\pi_s(\hat{\boldsymbol{\beta}})$ is the estimated expected proportion for cell s. The PF statistic is widely used in many areas of applications. Under large sample theory conditions, the PF statistic has an asymptotic chi-squared distribution with $T-g-1$ degrees of freedom where, g is the number of estimated model parameters (Koehler & Larantz, 1980).

Thus, a usual assumption for these tests is that expected cell counts become large as $n \to \infty$. This assumption is not reasonable for analyzing a sparse table. According to Agresti and Yang (1987) a contingency table said to be sparse when the ratio of the sample size to the number of cells is relatively small, but sparseness can also be produced by very skewed cell frequencies in some cases. In presence of sparse data, these Pearson's chi-square statistic may not follow the chi-square distribution even if the sample size is large. There is no universal agreement on what constitutes a small expected frequency. The most widely used rules of thumb are to consider the percentage of expected cell frequencies smaller than or equal to 1, 5 or 10 (Agresti and Yang (1987); Fisher (1941); Lancaster (1969); Agresti (2013)), and the percentage of observed zero frequencies. However, the first choice would be too insensitive to expected cell frequencies approaching 0 and the second would not be informative because the chi-square asymptotic approximation depends heavily on the expected cells which cannot be controlled for a simulation study.

Over the past years several statistics has been proposed to remedy this issue, Maydeu-Olivares and Joe (2005), Bartholomew and Leung (2002), Tollenaar and Mooijaart (2003) and Reiser (1996). Some of these statistics formed on lower-order marginals have been shown to overcome the deleterious effect of spareness. Another issue related to Pearson's chi-square test statistic is that it gives little guidance about the source of poor fit when the null hypothesis is rejected. Although several studies have been done to address the spareness issue, fewer studies have been done to find the source of poor fit related to chi-square goodness-of-fit test when the null hypothesis is rejected. Liu and Maydeu-Olivares (2014) recently proposed some methods to this end. Some of the other publications found in the literature related to identifying lack-of-fit are Cagnone and Mignani (2007), Orlando and Thissen (2000), and Reiser (1996, 2008).

I used orthogonal components of Pearson's chi-square statistic defined on lower-order marginals of the data table as a remedy to the sparseness problem. To this end, I studied three problems in my dissertation. The organization of chapters of this dissertation are as follows.

In Chapter 2 the most commonly used goodness-of-fit statistic, Pearson's chi-square, as well as various traditional goodness-of-fit statistics are discussed. Commonly used measures of sparseness in large contingency tables are also presented along with an explanation of the adverse effects of sparseness on goodness-of-fit statistics. Chapter 2 also provides a description of the decomposition of Pearson's chi-square and the focused tests based on chi-square statistic. Thereafter, it will explain statistics based on the marginal proportions. The chapter concludes with a brief review of a recent lack-of-fit statistic, $\bar{\chi}^2_{ij}$ from Liu and Maydeu-Olivares (2014).

Chapter 3 describes the mathematical details of standardized residuals and individual orthogonal components of Pearson's chi-square from Reiser (1996, 2008). It also presents the mathematical details related to two new statistics that are based on the orthogonal components of Pearson's chi-square.

Chapter 4 presents the results related to the first research problem in my dissertation. As the first problem, goodness-of-fit statistics using orthogonal components based on third-order and fourth-order marginals were studied. To this end, two new statistics based on the orthogonal components of Pearson's chi-square were developed and the Type I error, empirical power and asymptotic power of these statistics under different sparseness conditions were investigated. Performance of these statistics were also compared to other popular lack-of-fit statistics.

Chapter 5 describes the results related to the second research problem in my dissertation. As the second problem, the Type I error and power of individual orthogonal components of $\chi^2_{[2]}$ as test statistics to identify lack-of-fit were studied. In the context of this problem, both empirical and asymptotic power were investigated. The performance of individual orthogonal components of $\chi^2_{[2]}$ to other test statistics discussed in Chapter 2 was also compared.

Applications using these new statistics and lack-of-fit diagnostics are given in Chapter 6. The real-life example presented in chapter 6 is related to the data from prevalence and incidence of mental disorders in a catchment area study. The epidemiologic catchment area 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. Study was conducted by 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).

Chapter 7 presents results related to the third research problem in my dissertation. As the third problem, the statistics on lower-order marginals were extended to a large number of manifest variables. When the number of manifest variables becomes larger than 20, most of the statistics on lower-order marginals have limitations in terms of computer resources and CPU time. Under this problem, the performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic, fit to confirmatory dichotomous variable factor analysis model and the performance of Tollenaar and Mooijaart (2003) statistic when the number manifest variables is larger than or equal to 25 were investigated.

Finally, Chapter 8 includes some concluding remarks with a discussion of limitations, possible improvements, and further work on the proposed methodology.

Chapter 2

LITERATURE REVIEW

This chapter starts with a brief discussion about traditional goodness-of-fit statistics. Next, one of the most commonly used goodness-of-fit statistics, Pearson's chi-square and the directional tests related to it will be discussed. Then, adverse effects of sparseness on goodness-of-fit statistics will be explained. Thereafter, statistics based on marginal proportions will be presented.

2.1 Traditional Goodness-of-Fit Statistics

Many goodness-of-fit statistics related to testing a hypothesis about the parameters $\boldsymbol{\pi} = (\pi_1, ..., \pi_k)$ from a multinomial distribution belong to the power divergence family $\left\{I^\lambda, \lambda \in R\right\}$ and can be derived using the following definition,

$$2nI^\lambda = \frac{2}{\lambda(\lambda+1)} \sum \hat{\mathrm{p}}_s \left\{ \left(\frac{\hat{\mathrm{p}}_s}{\hat{\pi}_s} \right)^\lambda - 1 \right\} \tag{2.1}$$

where $\hat{\mathrm{p}}_s$ is the observed cell proportions and $\hat{\pi}_s$ is the estimated expected cell proportions. It can be easily seen that log likelihood ratio statistic ($\lambda = 0$, limiting case), Pearson's chi-square statistic ($\lambda = 1$), Freeman-Tukey statistic ($\lambda = -0.5$) and Neyman modified chi-square ($\lambda = -2$) are all special cases of this definition.

Cochran (1952) paper presented a comprehensive summary of the early development of Pearson's chi-square goodness-of-fit statistic. In this paper he also discusses a variety of competing tests related to goodness-of-fit as well. Amongst these competitors is the loglikelihood ratio test statistic G^2,

$$G^2 = 2\sum_{i=1}^{k} X_i log(X_i/n\pi_i) \tag{2.2}$$

where $\mathbf{X} = (X_1, X_2, ..., X_k)$ is a random vector of frequencies with $\sum X_i = n$, the sum being over $i = 1, ..., k$, and $\mathbf{E(X)} = n\boldsymbol{\pi}$, where $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_k)$ is a vector of probabilities with $\sum \pi_i = 1$ the sum being over $i = 1, ..., k$.

However, when the sample size is small, Pearson's chi-square statistic approximates a chi-squared random variable more closely than G^2 statistic for various multinomial and contingency table models (Cressie & Read, 1989). Cressie and Read illustrated this fact by comparing various enumeration and simulation studies by Upton (1978), Larntz (1978), Koehler and Larantz (1980), Lawal (1984), and Agresti and Yang (1987). The results of Larntz, Upton, and Lawal are of particular interest because they compare not only Pearson's chi-square statistic and G^2, but also the Freeman-Tukey statistic T^2,

$$T^2 = \sum_{i=1}^{k} \left(\sqrt{X_i} + \sqrt{X_i + 1} - \sqrt{4n\pi_i + 4}\right)^2. \tag{2.3}$$

This definition is sometimes referred to as the modified Freeman-Tukey statistic (Lawal & Upton, 1980). Another definition of the Freeman-Tukey statistic can be obtained by setting $\lambda = -\frac{1}{2}$ in equation 2.1:

$$F^2 = 4\sum_{i=1}^{k} \left(\sqrt{X_i} - \sqrt{n\pi_i}\right)^2. \tag{2.4}$$

Other statistics, which are special cases of the power-divergence family (2.1) can also be seen in the literature. These include, the modified loglikelihood ratio statistic or minimum discrimination information statistic ($\lambda = -1$) (Read & Cressie, 1988),

$$GM^2 = 2\sum_{i=1}^{k} n\pi_i log\left(n\pi_i/X_i\right) \tag{2.5}$$

and the Neyman-modified statistic ($\lambda = -2$) introduced by Neyman (1949)

$$NM^2 = \sum_{i=1}^{k} \frac{(X_i - n\pi_i)^2}{X_i}. \tag{2.6}$$

While these statistics have been recommended by various authors, for example, Gokhale and Kullback (1978) and Kullback and Keegel (1984), there have been no small-sample studies which indicate that they might be serious competitors to Pearson's chi-square statistic and G^2. The results of Read (1984), Larntz (1978), Lawal and Upton (1980), Lawal (1984) for T^2, and Hosmane (1987) for F^2 indicate that the exact distributions of these alternative statistics to Pearson's chi-square statistic and G^2 are less well approximated by the chi-squared distribution than are those of either Pearson's chi-square statistic or G^2 (Cressie & Read, 1989).

2.2 Pearson's Chi-Square Goodness-of-Fit Statistic

For a multi-way contingency table, the traditional Pearson's chi-square statistic is obtained by comparing observed frequencies to the expected frequencies under the null hypothesis. The general equation is given by,

$$\chi_p^2 = \sum_{s=1}^{T} \frac{(observed - expected)^2}{expected} \tag{2.7}$$

For a simple null hypothesis where the random sample comes from a population with completely specified cumulative distribution function F(x), the χ_p^2 statistic has an approximate chi-square distribution with T-1 degrees of freedom in large samples. On the other hand, a composite null hypothesis where the null distribution depends on a vector of g unknown parameters $\boldsymbol{\beta} = (\beta_1, ..., \beta_g)^T$, requires the Pearson-Fisher statistic,

$$\chi^2_{PF} = \sum_s z_s^2, \tag{2.8}$$

where

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

In the case where the $\boldsymbol{\pi}(\boldsymbol{\beta})$ depend on parameters that need to be estimated, Pearson argued that using the chi-squared distribution with T-1 degrees of freedom would still be adequate. However, Fisher (1924) gave the first derivation of the correct degrees of freedom, $T-g-1$, (hence Pearson-Fisher statistic) where, g parameters are estimated efficiently from the data.

2.3 Decomposition of Pearson's Chi-Square Statistic

Decomposing Pearson's chi-square statistic into components dates back to Lancaster (1969). By decomposing Pearson's chi-square statistic into components one can develop directional tests. Some directional tests have proven to reduce the adverse effects of sparseness and can have higher power against certain alternatives.

A well known and most widely used decomposition of the components may be associated with $T-1$ orthonormal functions $\{g_1, ..., g_{T-1}\}$ on the set $\{1, ..., T\}$. Moreover, these orthonormal functions are perpendicular to the unit function for n observations given on a set of k indicator variables of the multinomial distribution (Lancaster, 1969). Then by Parsevalś relation,

$$\chi^2_p = \sum_{j=1}^{T-1} \hat{U_{(j)}}^2 \tag{2.9}$$

$$\hat{U_{(j)}}^2 = \sum_{s=1}^{T} g_j(x_s) \tag{2.10}$$

where x_s is the observed value for the s^{th} observation and therefore necessarily in $\{1, ..., T\}$. These have a useful property of breaking the contributions of Pearson's chi-square into component pieces that may be associated with T-1 orthogonal directions corresponding to the basis functions $\{g_1, ..., g_{T-1}\}$. Note that orthogonality translates into,

$$\sum_{s=1}^{T} g_j(x_s) g_k(x_s) \hat{\pi}_s = \delta_{ik} \tag{2.11}$$

where δ_{ik} is the Kronecker delta, $\delta_{ik} = 1$ for $i = k$, and $\delta_{ik} = 0$ for $i \neq k$ and $\hat{\pi}_s$, $s = 1, ..., T$, is the estimated cell probability. Usually, the $\hat{U_{(j)}}$ are chosen so that they have interesting individual interpretations. Also, $\chi_p^2 = \sum \hat{U_{(j)}}^2$ is invariant for any choice of the set $\{g_1, ..., g_{T-1}\}$, i.e., these can be orthonormalized indicator variables, the Walsh functions, the orthogonal polynomials on T points with equal weights (Lancaster, 1969).

Another interesting approach that involves Chebyshev orthogonal polynomials were introduce by Rayner and Best (1989). However, these are computed under the equiprobable situation or ordered response patterns, which is not usually the case with large multi-way tables. This decomposition usually results in one to four large components, where the first component reasonably detects shifts in mean, the second component detects shifts in variance, the sum of the first two components detects shifts in both mean and variance, etc., which may not be useful for a multi-way contingency table with a large number of components.

According to the literature, assessing the goodness-of-fit of a hypothesized model and determining the source of misfit in poorly fitting models using an orthogonal polynomial decomposition may not be applicable as the number of multinomial categories increases. Some reasons authors indicate are that the equi-probable cells assumption might not be appropriate, the cells might not be ordered, and sparseness may be present. Another issue is that a large classification table results in many more components which might not necessarily be ordered large to small. In this case, selecting components becomes increasingly difficult. Agresti (2013) proposed an alternative partition of Pearson's chi-square statistic into independent chi-square components. This partition is not based on the orthogonal polynomial decomposition. Agresti (2013) gives the necessary conditions for determining sub-tables for which components are independent chi-square random variables. However, the sum of the chi-square values for any separate sub-tables do not sum to the overall Pearson's chi-square statistic.

There are numerous ways to decompose Pearson's chi-square statistic into orthogonal components. However, a more useful decomposition of the Pearson's chi-square statistic for extremely unbalanced non-equiprobable situations and for very sparse multinomials can be obtained by the decomposition of orthogonal components defined on lower-order marginals. Components based on these lower-order marginals are most often justified as easily interpretable because they are related to the model variables and somewhat computationally practical. Mathematical details related to obtaining these lower-order marginals are given in the Section 2.5.

As mentioned before, decomposing Pearson's chi-square statistic into components one can develop directional tests, and these directional tests can be used to reduce the adverse effects of sparseness. Next section explains the sparseness issue related to the Pearson's chi-square, in detail.

2.4 Sparseness

The sparseness issue related to Pearson's chi-square and G^2 statistics is well known. A sparse table is one where there are many cells with small counts and/or zeros. How many and how small is relative to the sample size and the table size. The large number of cells can be either due to the large number of classification variables, or small number of variables but with many levels. In this case, even with a moderate sample size, many cells may not be realized or might have small frequencies. Sparse data have an adverse effect on goodness-of-fit tests because they may invalidate using the chi-square distribution as an approximation for the distribution of Pearson's chi-square and G^2 statistics (Agresti & Yang, 1987).

Cochran (1952) points out that, when all the expectations are small and Pearson's chi-square has many degrees of freedom, the distribution of Pearson's chi-square differs substantially from the chi-square distribution, and Haldane (1939) shows that the variance of Pearson's chi-square departs noticeably from the variance of the normal approximation to the chi-squared distribution. Koehler and Larantz (1980) examined the accuracy of the chi-square and normal approximations for Pearson's chi-square and G^2 statistics via a Monte Carlo study and found that in general the chi-square approximation for Pearson's chi-square statistic is appropriate even when the expected frequencies are as low as 0.25 with $T \geq 3$, $n \geq 10$ and $n^2/T \geq 10$. On the other hand, G^2 statistic is not well approximated by a chi-square distribution when $n^2/T \geq 10$. Many suggestions have been given on how to measure sparseness in multi-way contingency table, but there is no universal definition of sparseness in the literature. The most widely used rules of thumb are to consider the percentage of expected cell frequencies smaller than or equal to 1, 5 or 10 (Agresti and Yang (1987); Fisher (1941); Lancaster (1969); Agresti (2013)), and the percentage of observed zero frequencies. The first choice would be too insensitive to expected cell frequencies approaching 0 and the second would not be informative because the chi-square asymptotic approximation depends heavily on the expected cells which cannot be controlled for a simulation study. Generally, the ratio n/T is used to measure the amount of spareness present in a table. This ratio alone is also not informative as models where a single cell has a probability near 1 with the rest approaching 0 is more likely to be sparse than an equiprobability model.

One way to overcome the adverse effects of sparseness is to use limited-information statistics. A number of authors, Knott and Tzamourani (1997), Reiser (1996), Reiser (2008), Bartholomew and Tzamourani (1999), Bartholomew and Leung (2002), Maydeu-Olivares and Joe (2005) and Maydeu-Olivares and Joe (2006), have studied limited information-statistics as a potential solution for overcoming adverse effects of sparseness. In limited-information statistics, only the information contained in suitable summary statistics of the data, typically the low-order marginals of the contingency table, is used to assess the model. This amounts to pooling cells in a systematic way so that the resulting statistics have a known asymptotic null distribution. For instance, focused test in lower-order marginal components of PF statistic is a limited-information statistic. If a cell has a small expectation, combining cells in this manner can give a more moderate expectation improving the chi-squared approximation under the null distribution. More about these focused tests will be discussed in Section 2.6.

Another method is to add a small constant to the frequency of every response pattern. According to Agresti (2013) some algorithms add 0.5 to each cell and this will have the benefit of bias reduction for a saturated model. However, this may smooth the data too much and can cause havoc with sampling distribution (Agresti, 2013). One can also do pooling cells or using resampling methods such as the parametric bootstrap. However, pooling cells after the model has been fitted often results in statistics with an unknown sampling distribution, as the procedure is data dependent. It may also lead to gross loss of information about model misfit and, as is often the case, no degrees of freedom left for testing. The use of resampling methods such as the parametric bootstrap to obtain an empirical p-value for Pearson's chi-square statistic and G^2 has become increasingly popular given todays computing power. However, this method is computationally intense since, in order to obtain a stable p-value several hundred bootstrap re-samples are needed for each model (Bartholomew & Leung, 2002).

As mentioned before, one way of remedying the problem of sparseness is to consider focused test statistics that are based on only the low-order marginals. Next section illustrates the mathematical details related to obtaining lower-order marginals. Thereafter, the focused tests based on the low-order marginals will be presented.

2.5 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. These joint proportions or frequencies can be transformed into marginal proportions and these marginal proportions can be used to define components of Pearson's chi-square.

2.5.1 First- and Second-Order Marginals

The relationship between joint proportions and marginals for a multi-way contingency table 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 $q = 3$,

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

Define, $\mathbf{H}_{[1]} = \mathbf{V}$.

Let h_{si} represent (s, i) element of $\mathbf{H}_{[1]}$, that is the element i of the 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

$$P_i(\boldsymbol{\beta}) = \text{Prob}(Y_i = 1|\boldsymbol{\beta}) = \sum_s h_{si}\pi_s(\boldsymbol{\beta}) = \mathbf{h}_i'\boldsymbol{\pi}(\boldsymbol{\beta}),$$

where $\mathbf{h}_i'$ is the vector of h_{si} elements related to the i^{th} response variable. The true first-order marginal proportion is given by

$$P_i = \text{Prob}(Y_i = 1) = \sum_s h_{si}\pi_s = \mathbf{h}_i'\boldsymbol{\pi} \ .$$

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

$$P_{ij}(\boldsymbol{\beta}) = \text{Prob}(Y_i = 1, Y_j = 1|\boldsymbol{\beta}) = \sum_s h_{si}h_{sj}\pi_s(\boldsymbol{\beta}) = (\mathbf{h}_i \circ \mathbf{h}_j)'\boldsymbol{\pi}(\boldsymbol{\beta}),$$

where $j = 1, 2, \ldots, q-1$; $i = j+1, \ldots q$ and $\mathbf{h}_i \circ \mathbf{h}_j$ represents the Hadamard product of columns $\mathbf{h}_i$ and $\mathbf{h}_j$. Thus, the true second-order marginal proportion is given by

$$P_{ij} = \text{Prob}(Y_i = 1, Y_j = 1) = \sum_s h_{si} h_{sj} \pi_s = (\mathbf{h}_i \circ \mathbf{h}_j)' \boldsymbol{\pi} \ .$$

2.5.2 Higher-Order Marginals

A general matrix $\mathbf{H}_{[t:u]}$ to obtain marginals of any order can be defined 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, $\mathbf{H}_{[t]} \equiv \mathbf{H}_{[t:t]}$ and $\mathbf{H} \equiv \mathbf{H}_{[t:u]}$. $\mathbf{H}_{[1:q]}$ gives a 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.

For example, when q=3,

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

Based on the above definition, second-order marginal proportions for variables Y_i and Y_j can also be obtained by,

$$\boldsymbol{P}_{[2]} = \mathbf{H}_{[2]}\boldsymbol{\pi} \tag{2.12}$$

where,

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

where $\boldsymbol{v}_f$ represents column f of matrix $\boldsymbol{V}$, and $\boldsymbol{v}_f \circ \boldsymbol{v}_g$ represents the Hadamard product of columns f and g.

The first column of the $\mathbf{H}$ matrix is a zero vector, so $\mathbf{H}$ is not full rank. Therefore, this zero column can be deleted and the first element of $\boldsymbol{\pi}$ vector can be deleted. This make sense because of the constraint $\pi_1 + \pi_2 + ... + \pi_T = 1$. Define, $\ddot{\mathbf{H}}_{[1:q]}$ as the matrix without the first column. The dimension of the $\ddot{\mathbf{H}}_{[1:q]}$ matrix is $T-1$ by $T-1$, and it is full rank. Then, $\ddot{\mathbf{H}}_{[1:q]}$ gives a one-to-one mapping from joint proportions to the set of $(2^q - 1)$ marginal proportions. A test of fit on marginal proportions from order 1 to order q is equivalent to a test of fit on joint proportions because marginal and joint proportions contain the same information.

2.6 Focused Tests Based on Chi-Square Statistics

As indicated in the previous sections, one way of remedying the problem of sparseness is to consider focused test statistics that are based on only the low-order marginals, which are sums of joint frequencies. Generally, the sums are not sparse for $2 * 2$ sub-tables. Any statistic formed from a sum of the components, not necessarily ones based on marginal frequencies, can be considered a focused statistic. Summing a subset of components to create a focused test statistic could increase the power against certain alternatives. Focused tests using lower-order marginals can be used in a wide variety of applications including log-linear models, categorical variable factor analysis and repeated measures on categorical variables.

Christoffersson (1975) first introduced the idea of using first- and second-order marginals for a test of fit in dichotomous variable factor analysis. Transforming to the notation in this study, this statistic can be written as,

$$\chi^2_{Ch} = \bar{\mathbf{r}}'\mathbf{H}'_{[1:2]}(D(\mathbf{p}) - \mathbf{p}\mathbf{p}')^{-1}\mathbf{H}_{[1:2]}\bar{\mathbf{r}} \tag{2.13}$$

where $\bar{\mathbf{r}} = \hat{\mathbf{p}} - \boldsymbol{\pi}(\bar{\boldsymbol{\beta}})$, $\bar{\boldsymbol{\beta}}$ is the generalized least squares estimator of $\boldsymbol{\beta}$. χ^2_{Ch} has an asymptotic chi-square distribution with $q * (q+1)/2 - g$ degrees of freedom, where g = number of model parameters to be estimated. The statistic could be generalized to include higher-order marginals, but even if marginals from first- to order q were included, this statistic would not be equivalent to the Pearson-Fisher statistic. Muthén (1978) improved χ^2_{Ch} statistic, but both used observed proportions for the calculation of covariance matrix and neither presented their test as having higher power or as a remedy for sparse data.

Reiser (1996) proposed a focused statistic, $\chi^2_{[1:2]}$, using first- and second-order marginals to test the fit of item response models when there are a large number of manifest variables and the sample size is small to moderate. Reiser and Lin (1999) developed a similar focused statistic for testing the fit of latent class models. More detailed explanations about Reiser (1996) statistic is given in Chapter 3.

2.7 Related Statistics

Bartholomew and Leung (2002) statistic Y is another important statistic that can be found in literature. Y statistic incorporates second-order marginals only:

$$Y = (\hat{\mathbf{p}} - \boldsymbol{\pi})'\mathbf{H}'_{[2]}(D(\mathbf{H}_{[2]}\boldsymbol{\pi})(\mathbf{I} - D(\mathbf{H}_{[2]}\boldsymbol{\pi}))^{-1}\mathbf{H}_{[2]}(\hat{\mathbf{p}} - \boldsymbol{\pi}). \tag{2.14}$$

Bartholomew and Leung (2002) gave a chi-square approximation for the distribution of

$$\frac{Y-a}{b}$$

on c degrees of freedom, where a, b and c are functions of the asymptotic moments of Y :

$$b = \frac{\mu_3(Y)}{4\mu_2(Y)},\ c = \frac{\mu_2(Y)}{2b^2},\ a = \mu_1(Y) - bc.$$

This statistic was presented in terms of known π, but in an application, π is replaced by probabilities estimated from the model under consideration. In the original form, the statistic is simpler to calculate because it only requires estimates for π. However, this statistic does not perform well with the degrees of freedom given by Bartholomew and Leung. Cai, Maydeu-Olivares, Coffman, and Thissen (2006) proposed a modified version of the statistic, Y_2, using both first- and second-order marginals, and revised degrees of freedom:

$$Y_2 = (\hat{\mathbf{p}} - \boldsymbol{\pi})'\mathbf{H}'_{[1:2]}(D(\mathbf{H}_{[1:2]}\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))(\mathbf{I} - D(\mathbf{H}_{[1:2]}\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})))^{-1}\mathbf{H}_{[1:2]}(\hat{\mathbf{p}} - \boldsymbol{\pi}) \tag{2.15}$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$. Since calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ is required for determination of the revised degrees of freedom, there is little computational advantage compared to the statistic $\chi^2_{[1:2]}$ introduced by Reiser.

Joe (1993) and Maydeu-Olivares and Joe (2001, 2005, 2006) proposed a class of chi-square tests for sparse dichotomous and multidimensional data with applications to the item response model, a form of categorical variable factor analysis. Their approach is closely related to that of Reiser (1996) but their focused statistic M_2 does not correspond to the same decomposition of the χ^2_{PF}. For $\mathbf{e} = \mathbf{H}_{[1:r]}\mathbf{r}$ and $\mathbf{r} = \hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$,

$$\mathbf{M}_r = \mathbf{e}'\hat{\mathbf{C}}_r\mathbf{e} \tag{2.16}$$

where $\hat{\mathbf{C}}_r = (\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1} - (\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}\mathbf{H}\hat{\mathbf{G}}(\hat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}\mathbf{H}\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}$ and $\hat{\mathbf{T}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})'$. $\mathbf{H}$ is always equal to $\mathbf{H}_{[1:r]}$ when applied to the definition of $\mathbf{M}_r$. The statistic $\mathbf{M}_r$ has an asymptotic chi-square distribution with $\sum_r \binom{q}{r} - g$ degrees of freedom, where g = number of model parameters to be estimated (Reiser, 2008).

Tollenaar and Mooijaart (2003) proposed a statistic,

$$\chi^2_{red} = n\mathbf{e}'(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})^{-1}\mathbf{e} \tag{2.17}$$

where,

$$\hat{\mathbf{T}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})'.$$

The Tollenaar and Mooijaart (2003) statistic is a reduced version of $\chi^2_{[1:2]}$ statistic (Reiser, 2008). The difference lies in the covariance matrix $\mathbf{T}$ not including the term

$\mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$, where $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\mathbf{A} = \mathbf{D}(\boldsymbol{\pi})^{-1/2}\mathbf{G}$. As indicated by Tollenaar and Mooijaart (2003), omitting this term may substantially reduce computations. Since $\chi^2_{[1:2]}$ and χ^2_{red} have different covariance matrices, the degrees of freedom are different. χ^2_{red} has an asymptotic-square distribution with $q*(q+1)/2-g$ degrees of freedom, where g = number of model parameters to be estimated. Note, this is the same degrees of freedom associated with χ^2_{Ch} and M_2.

2.8 Related Lack-of-Fit Diagnostics

According Liu and Maydeu-Olivares (2014) one of the challenges faced when trying to identify lack-of-fit in models fit to binary cross-classified variables is that some tests of interest cannot be applied owing to the lack of degrees of freedom. One way to overcome the problem of the lack of degrees of freedom is to use a large sample z statistic. Reiser (1996) suggested using bivariate z statistics to assess the source of misfit in two-way marginal subtables for binary item response data. This dissertation study will incorporate Reiser's z statistics. Mathematical details of these statistics are given in Section 3.1.1. Liu and Maydeu-Olivares, (2013) proposed a similar statistic, $R_{2,ij}$ to work-around the lack of degrees of freedom that involves a pair of item and conditions on sum score levels/groups. Drawing on the results of Joe and Maydeu-Olivares (2010), Liu and Maydeu-Olivares (2013) were able to derive the asymptotic distribution of $R_{2,ij}$. Under the null hypothesis of a correctly specified model, this statistic follows asymptotically a chi-square distribution. However, Liu and Maydeu-Olivares (2013) and Maydeu-Olivares and Liu (2012) found that statistics M_2 and $R_{2,ij}$ tend to have lower power for detecting lack-of-fit in some models.

Liu and Maydeu-Olivares (2014) proposed a statistic, $\bar{\chi}^2_{ij}$, mean and variance adjusted chi-square statistic for bi-variate distribution for variables i, j within a large table. Consider the case where Pearson's chi-square is applied to a bi-variate subtable,

$$\chi^2_{ij} = n(\mathbf{p}_{ij} - \hat{\boldsymbol{\pi}}_{ij})'\mathbf{D}_{ij}^{-1}(\mathbf{p}_{ij} - \hat{\boldsymbol{\pi}}_{ij}) \tag{2.18}$$

where $\mathbf{D}_{ij} = diag(\pi_{ij})$ is a diagonal matrix of the bivariate probabilities. Then, the mean and variance adjusted chi-square statistic, $\bar{\chi}^2_{ij}$, can be written as,

$$\bar{\chi}^2_{ij} = 2\frac{\hat{\mu_1}}{\hat{\mu_2}}\chi^2_{ij} \tag{2.19}$$

where the two asymptotic moments (μ_1, μ_2) can be obtained as below,

$$\mu_1 = tr\left(\mathbf{D}_{ij}^{-1}\boldsymbol{\Sigma}_{ij}\right) \tag{2.20}$$

$$\mu_2 = 2tr\left(\mathbf{D}_{ij}^{-1}\boldsymbol{\Sigma}_{ij}\right)^2 \tag{2.21}$$

where $\boldsymbol{\Sigma}_{ij}$ is the covariance matrix related to the residuals $n(\mathbf{p}_{ij} - \hat{\boldsymbol{\pi}}_{ij})$, for a pair of items when maximum likelihood is used to estimate the model parameters.

$\bar{\chi}^2_{ij}$ has an approximate reference chi-square distribution with degrees of freedom

$$a = \frac{2\hat{\mu_1}^2}{\hat{\mu_2}}. \tag{2.22}$$

According to Liu and Maydeu-Olivares $\bar{\chi}^2_{ij}$ has good Type I error and power behavior under certain sparseness settings. However, the simulations for $\bar{\chi}^2_{ij}$ in Liu and Maydeu-Olivares (2014) paper were limited to models with zero intercept settings. With different sparse conditions and models with skewed intercept settings, these Type I error and power results can have a different behavior. Also, the $\bar{\chi}^2_{ij}$ for different item pairs cannot be directly compared as they are on a different scale (their estimated df). Only the p-values can be directly compared across item pairs. This is undesirable in terms of actual applications because researchers have to inspect tables of p-values with a large number of decimals in order to determine the item pairs with the greatest magnitude of misfit. Also, $\bar{\chi}^2_{ij}$ requires calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. Therefore, when the number of manifest variables large, this statistic become difficult or impossible to calculate due to computer resource limitations.

Chapter 3

RESIDUALS, NEW TEST STATISTICS AND ORTHOGONAL COMPONENTS AS LACK-OF-FIT STATISTICS

This chapter will illustrative the mathematical details of standardized residuals and orthogonal components of Pearson's chi-square from Reiser (1996, 2008). It will also present mathematical details related to two new statistics that are based on the orthogonal components of Pearson's chi-square.

3.1 Testing Fit on Marginal Distributions

3.1.1 Adjusted Residuals

Reiser (1996) has shown that the traditional standardized residual may be completely inadequate for identifying response patterns (i.e. cells) that are poorly fit in a large cross-classified table and proposed adjusted residuals on marginal tables for identifying poor fit. Mathematical details about adjusted marginal residuals and how they will be incorporated for this study will be explained in the following section.

To form residuals for the cells of the multinomial vector of response patterns, there are several possible approaches, including examining $\hat{p}_s - \pi_s(\hat{\boldsymbol{\beta}})$ directly, where $\hat{p}_s = \frac{n_s}{n}$ is element s of $\hat{\mathbf{p}}$, the vector of sample proportions, $\hat{\boldsymbol{\beta}}$ is an estimator for the parameter matrix, and $\pi_s(\hat{\boldsymbol{\beta}})$ is the estimated expected proportion for cell s. For the multinomial model, it has been traditional to examine standardized residuals (Cochran, 1954). Let

$$\mathrm{r}_s = \frac{\hat{p}_s - \pi_s(\hat{\boldsymbol{\beta}})}{(\pi_s(\hat{\boldsymbol{\beta}}))^{\frac{1}{2}}} \tag{5}$$

then $n^{\frac{1}{2}}\mathrm{r_s}$ is the standardized residual. $n\sum_s \mathrm{r}_\mathrm{s}^2$ is equal to the Pearson chi-square goodness-of-fit statistic. Under some circumstances, the set of these residuals may be useful for finding cells that are not well fit by the model. However, since the distribution of $n^{\frac{1}{2}}\mathrm{r_s}$ is not necessarily $N(0,1)$, it is sometimes difficult to assess the significance of the magnitude of the standardized residual. Therefore, it is useful to divide the statistic by its standard error:

$$\frac{n^{\frac{1}{2}}\mathrm{r_s}}{\hat{\sigma}_\mathrm{s}},$$

yielding the adjusted residual, which has an approximate N(0,1) distribution in large samples. The mathematical details below represent the large sample distribution for $n^{\frac{1}{2}}\mathbf{r}$. This result will be used in Section 3.1.2 to explain the large sample distribution for the marginal residuals.

Consider the vector valued function of $\mathbf{p}$ and β:

$$\mathbf{h}(\mathbf{p},\boldsymbol{\beta}) = \mathbf{D}(\boldsymbol{\pi}(\boldsymbol{\beta}))^{-1/2}(\mathbf{p}-\boldsymbol{\pi}(\boldsymbol{\beta})),$$

where $\boldsymbol{\pi}(\boldsymbol{\beta})$= vector of multinomial probabilities as a function of β,

and $\mathbf{D}(\boldsymbol{\pi}(\boldsymbol{\beta}))$= diagonal matrix with elements (s,s) equal to $\boldsymbol{\pi}_s(\boldsymbol{\beta})$.

The $\mathbf{T}$ dimensional vector of residuals, $\mathbf{r}$, is obtained from the function $\mathbf{h}(\mathbf{p},\boldsymbol{\beta})$ when $\mathbf{p}$ takes the value $\hat{\mathbf{p}}$ and $\boldsymbol{\beta}$ takes the value $\hat{\boldsymbol{\beta}}$. Based on these settings and assuming the regularity conditions given by Birch (1964),

$$n^{\frac{1}{2}}\mathbf{r}\xrightarrow{L}N(\mathbf{0},\boldsymbol{\Omega}_\mathbf{r}), \tag{3.1}$$

where,

$\xrightarrow{L}$ indicates convergence in Law,

$\boldsymbol{\Omega}_\mathbf{r} = \mathbf{I} - \boldsymbol{\pi}^{1/2}(\boldsymbol{\pi}')^{1/2} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$,

$\mathbf{A} = \mathbf{D}(\boldsymbol{\pi})^{-1/2}\frac{\partial\boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \mathrm{vec}\boldsymbol{\beta}}$, evaluated at the true parameter values,

vec $\boldsymbol{\beta} = \binom{\boldsymbol{\beta}_1}{\boldsymbol{\beta}_2}$ with intercepts stacked on top of slopes,

$\boldsymbol{\pi}^{1/2}$ =vector with elements given by square root of true proportions.

3.1.2 Marginal Residuals

Next, I will define residuals for the marginals. In Section 2.5, I have illustrated how joint proportions can be transformed into marginal proportions using the **H** matrix. The same **H** matrix, defined in Section 2.5.2 can also be used to create residuals for marginals.

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

Then a vector of simple residuals for marginals of any order can be defined as

$$\boldsymbol{e} = \mathbf{H}(\hat{\mathrm{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) = \mathbf{H}\mathbf{u}.$$

Using unstandardized residuals will simplify the results in subsequent sections. Although these results will be based on unstandardized residuals, the results are valid for standardized residuals as well.

Extending the results in Section 3.1.1 for marginals:

$$n^{\frac{1}{2}}\ \mathbf{e} \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{e}}), \tag{3.2}$$

where,

$$\boldsymbol{\Omega}_{\mathbf{e}} = \mathbf{H}\boldsymbol{\Omega}_{\mathbf{u}}\mathbf{H}',$$

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

and

$$\mathbf{G} = \frac{\partial\boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \mathrm{vec}\boldsymbol{\beta}},$$

The validity of this result for the covariance matrix can be shown by an application of the multivariate delta method (the method of statistical differentials). It can be seen from expression 3.2 that the elements of $\mathbf{e}$ are linear combinations of the unstandardized residuals, $\mathbf{u} = \hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$, associated with the multinomial cells. The covariance matrix of $\mathbf{e}$ can be found by starting with the covariance matrix for $\mathbf{u}$, which closely resembles the matrix following expression 3.1. Define the following vector valued function of $\mathbf{p}$ and $\boldsymbol{\pi}(\boldsymbol{\beta})$:

$$\mathbf{h}(\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\beta})) = \mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\beta}).$$

Then $\mathbf{u} = \mathbf{h}(\hat{\mathbf{p}}, \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))$, and by Theorem 14.6-2 of Bishop, Fienberg, and Holland (1975),

$$\mathbf{u} \xrightarrow{L} N\left(\mathbf{0}, \left(\frac{\partial \mathbf{h}}{\partial \mathbf{p}}\right) \boldsymbol{\Sigma}_{\hat{\mathbf{p}}} \left(\frac{\partial \mathbf{h}}{\partial \mathbf{p}'}\right)\right). \tag{3.3}$$

The use of expression 3.3 requires the partial derivative of $\mathbf{h}(\mathbf{p},\boldsymbol{\pi}(\boldsymbol{\beta}))$ with respect to $\mathbf{p}$ and an expression for $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}$. Proceeding to obtain the necessary expressions,

$$n^{\frac{1}{2}}(\hat{\mathbf{p}} - \boldsymbol{\pi}) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'), \tag{3.4}$$

by Theorem 14.3.4 in Bishop, Fienberg & Holland, which gives $\boldsymbol{\Sigma}_{\hat{\mathbf{p}}}$.

The partial derivative of $\mathbf{h}(\mathbf{p},\boldsymbol{\pi}(\boldsymbol{\beta}))$ with respect to $\mathbf{p}$ follows from the chain rule:

$$\frac{\partial \mathbf{h}}{\partial \mathbf{p}} = \mathbf{I} - \frac{\partial \boldsymbol{\pi}}{\partial \mathrm{vec}\boldsymbol{\beta}} \frac{\partial \mathrm{vec}\boldsymbol{\beta}(\mathbf{p})}{\partial \mathbf{p}}.$$

$\boldsymbol{\beta}$ as a function of $\mathbf{p}$ is not known explicitly, but the existence of that function can be established by the Implicit Function Theorem. Using this approach,

$$\frac{\partial \mathrm{vec}\boldsymbol{\beta}(\mathbf{p})}{\partial \mathbf{p}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{D}(\boldsymbol{\pi})^{-1/2}.$$

Then, with $\mathbf{G} = \frac{\partial \boldsymbol{\pi}}{\partial \mathrm{vec}\boldsymbol{\beta}}$,

$$\frac{\partial \mathbf{h}}{\partial \mathbf{p}} = \mathbf{I} - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{D}(\boldsymbol{\pi})^{-1/2},$$

when evaluated at $\mathbf{p} = \boldsymbol{\pi}$, the true value.

Finally, applying these results to expression 3.3 ,

$$n^{\frac{1}{2}}\ \mathbf{u} \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{u}}),$$

where

$$\boldsymbol{\Omega}_{\mathbf{u}} = \left(\mathbf{I} - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{D}(\boldsymbol{\pi})^{-1/2}\right)\left(\mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\right)\left(\mathbf{D}(\boldsymbol{\pi})^{-1/2}\mathbf{A}\mathbf{G}'(\mathbf{A}'\mathbf{A})^{-1} - \mathbf{I}\right).$$

After multiplying, and using $\mathbf{A}'\boldsymbol{\pi}^{1/2} = \sum \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \text{vec}\boldsymbol{\beta}} = 0$, the expression simplifies as follows:

$$\boldsymbol{\Omega}_{\mathbf{u}} = \mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'. \tag{3.5}$$

Now returning to the residuals on the marginals, which are linear combinations of the elements in $\mathbf{u}$, results correspond to equation 3.2 follow from expression 3.5 and result 6a.1(ii) of Rao (1973, pg 383).

Define $\boldsymbol{\Sigma}_{\mathbf{e}}$ to be the asymptotic covariance matrix of the residuals for marginals, with estimator $\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}}$ defined by

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}} = n^{-1}\mathbf{H}(\mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}'|_{\boldsymbol{\pi}=\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}),\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\ . \tag{3.6}$$

$n\hat{\boldsymbol{\Sigma}}_{\mathbf{e}}$ is consistent for $\boldsymbol{\Omega}_{\mathbf{e}}$ when the joint table is not sparse. Sparse asymptotic results from Simonoff (1986) are applicable here. Assuming $\hat{\boldsymbol{\beta}}$ is a consistent estimator, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + O_p(n^{-\frac{1}{2}})$; if $\boldsymbol{\pi}(\boldsymbol{\beta})$ has bounded second partial derivatives with respect to $\boldsymbol{\beta}$, $sup_s \left|\pi_s(\hat{\boldsymbol{\beta}})/\pi_s - 1\right| = O_p(n^{-\frac{1}{2}})$. So, even under sparseness conditions, $\pi_s(\boldsymbol{\beta}) \xrightarrow{P} \pi_s$, $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} \boldsymbol{\pi}$, $n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \xrightarrow{P} \boldsymbol{\Omega}_{\boldsymbol{e}}$

Estimated standard errors for the residuals can be obtained by taking square roots of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}}$.

In the outset of this section it was explained that the the distribution of $n^{\frac{1}{2}}\mathrm{r_s}$ is not necessarily $N(0, 1)$, and it is useful to divide the standardized residuals by its standard error:

$$\frac{n^{\frac{1}{2}}\mathrm{r_s}}{\hat{\sigma}_{\mathrm{s}}},$$

yielding the adjusted residuals. Extending this idea, define adjusted residual k for the marginals of order l,

$$\frac{n^{\frac{1}{2}}e^{(k)}_{[l]}}{\hat{\sigma}_{e^{(k)}_{[l]}}}, \tag{3.7}$$

where $\mathbf{e}_{[l]} = \mathbf{H}_{[l]}\mathbf{u}$, $k = 1, ..., \binom{q}{l}$, $l = 1, 2, ..., q$ and q is the number of manifest variables.

Similarly, define adjusted residual k for the second-order marginals,

$$Z_{ij} = \frac{n^{\frac{1}{2}}e^{(k)}_{[2]}}{\hat{\sigma}_{e^{(k)}_{[2]}}}, \tag{3.8}$$

where $\mathbf{e}_{[2]} = \mathbf{H}_{[2]}\mathbf{u}$, $k = 1, 2, ..., \binom{q}{2}$ and ij is the item pair correspond to the adjusted residuals k of the second-order marginal.

The square roots of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}}$ can be used as estimated standard errors for calculating the adjusted residuals for the marginals.

3.1.3 Test Statistics

A traditional composite null hypothesis for a test of fit on a multinomial model is H_o: $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$. Linear combinations of $\boldsymbol{\pi}$ may be tested under the null hypothesis H_o: $\mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\beta})$. $\mathbf{H}$ may specify linear combinations that form marginal proportions as defined in the Section 2.5.

Reiser(1996, 2008) and Reiser and Lin (1999) proposed statistics for $H_0 : \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\beta})$ 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.

Consider the linear combination $\boldsymbol{e} = \mathbf{Hu}$. If $\mathbf{H}$ contains $2^q - g - 1$ linearly independent rows corresponding to marginals from order 1 to q, then define the statistic

$$\chi^2_{[1:q]} = n\mathbf{u}'\mathbf{H}'\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{Hu}. \tag{3.9}$$

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 $\boldsymbol{\Omega}_{\boldsymbol{e}} = \mathbf{H}\boldsymbol{\Omega}_{\mathbf{u}}\mathbf{H}'$. (For more details, refer Section 3.1.2). With the added condition that the rows of $\mathbf{H}$ are linearly independent of the columns of $\mathbf{G}$, i.e., rank$(\mathbf{H}'\vdots\mathbf{G}) = T + g$, $\chi^2_{[1:q]}$ can be shown to be equivalent to χ^2_{PF}. See Reiser (2008). To obtain orthogonal components, define the upper triangular matrix $\boldsymbol{F}$ 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 $\boldsymbol{\Omega}_{\boldsymbol{e}}$. Then writing $\boldsymbol{\Omega}_{\boldsymbol{e}}$ as $\boldsymbol{CC}'$,

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

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 $(D(\boldsymbol{\pi}) - \boldsymbol{\pi\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')$. Let $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q]}$, then

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

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

Define

$$\hat{\boldsymbol{\gamma}} = n^{\frac{1}{2}}\widehat{\boldsymbol{F}}'\mathbf{Hu} = n^{\frac{1}{2}}\widehat{\mathbf{H}}^*\mathbf{u}.$$

Then

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

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

χ^2_1 random variables, assuming consistent estimate for $\boldsymbol{\pi}(\boldsymbol{\beta})$ and $\Sigma_{\mathbf{e}}$. The asymptotic approximation may not be valid for components from a sparse higher-order marginal table.

By summing a subset of these components one can obtain limited-information or focused statistics. Below are some of the limited-information statistics that were discussed under the literature review. The statistic on first- and second-order marginals from Reiser (1996) is

$$\chi^2_{[1:2]} = \sum_{j=1}^{j=q(q+1)/2} \hat{\gamma}_j^2,$$

and the statistic on second-order marginals from Reiser and Lin (1999) is

$$\chi^2_{[2]} = \sum_{j=q+1}^{j=q(q+1)/2} \hat{\gamma}_j^2,$$

In general, using the matrix $\mathbf{H}_{[t:u]}$ as given above,

$$\chi^2_{[t:u]} = \sum_j \hat{\gamma}_j^2,$$

where the limits on the sum depend on t and u, the order of the selected marginals, and the statistic can also be expressed as

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

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = n^{-1}\boldsymbol{\Omega}_{\boldsymbol{e}}$, with $\boldsymbol{\Omega}_{\boldsymbol{e}}$ evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\beta}}$. However, depending on matrix $\mathbf{H}$ and the fitted model, it may be difficult to calculate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1}$ accurately due to collinearity. Direct calculation of components by weighted regression is considerably more stable.

Under the regularity conditions given by Birch (1964), the limiting distribution of $\chi^2_{[t:u]}$ as $n \to \infty$ can be shown to be the χ^2-distribution because $\boldsymbol{e}$ is a linear combination of the elements of $\mathbf{u}$, $n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \xrightarrow{P} \boldsymbol{\Omega}_{\boldsymbol{e}}$, and $\boldsymbol{e} \xrightarrow{L} MVN(\boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{e}})$. Chi-square approximation may not be valid when $\mathbf{H}$ includes sparse higher-order marginals.

The degrees of freedom for $\chi^2_{[t:u]}$ are known from theory and are determined by the rank of $\boldsymbol{\Omega_e}$, which will be equal to the number of linearly independent rows in $\mathbf{H}$, assuming rank$(\mathbf{H}' \vdots \mathbf{G}) = m + g$ where m is the rank of $\mathbf{H}$, and assuming the model $\boldsymbol{\pi}(\boldsymbol{\beta})$ is identified.

Proposed statistics for dissertation research

The statistic, $\chi^2_{[2]}$ on second-order marginals from Reiser (1996) may have a higher power for certain alternative hypotheses because it represents a test that is focused on the second-order marginals. If lack-of-fit is present in 3rd- or higher-order marginals, then a test that incorporates these higher-order marginals may have a higher power than $\chi^2_{[2]}$ against H_0. On the other hand, if the higher-order marginals (3rd, 4th and etc.) are sparse, then the asymptotic chi-square approximation may not perform well for these statistics as it did for $\chi^2_{[2]}$. To further study these issues, I created two new statistics based on the orthogonal components of Pearson's chi-square. Mathematical details related to these statistics are presented below.

Using the definitions in Section 2.5.2 and Section 3.1.3, define the following statistics for 2^q tables:

$$\chi^2_{[3]} = \sum_{j=q(q+1)/2+1}^{j=q(q^2+5)/6} \hat{\gamma}_j^2,$$

and

$$\chi^2_{[4]} = \sum_{j=q(q^2+5)/6+1}^{j=\varpi} \hat{\gamma}_j^2,$$

where,

$$\varpi = \frac{q(q+1)(q^2-3q+14)}{24}$$

Result 3.1

$\chi^2_{[3]}$ is the component of χ^2_{PF} on third-order marginals given $\chi^2_{[1:2]}$. $\chi^2_{[4]}$ is the component of χ^2_{PF} on fourth-order marginals given $\chi^2_{[1:3]}$.

Proof :

Since there are $\binom{q}{1}$ first order marginals, the dimension of $\mathbf{H}_{[1]}$ is q by 2^q. Similarly, there are $\binom{q}{2} = q*(q-1)/2$ second order marginals, $\binom{q}{3} = q*(q-1)*(q-2)/6$ third order marginals, and $\binom{q}{4} = q*(q-1)*(q-2)*(q-3)/24$ fourth order marginals. Thus, the dimension of $\mathbf{H}_{[2]}$ is $q*(q-1)/2$ by 2^q, the dimension of $\mathbf{H}_{[3]}$ is $q*(q-1)*(q-2)/6$ by 2^q, and the dimension of $\mathbf{H}_{[4]}$ is $q*(q-1)*(q-2)*(q-3)/24$ by 2^q.

Using equation 3.9,

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

Assume, $\mathbf{H} = \mathbf{H}_{[1]} \vdots \mathbf{H}_{[2]} \vdots \mathbf{H}_{[3]} \vdots \cdots \mathbf{H}_{[q]}$.

When q=3,

$$\chi^2_{[1:3]} = n(\mathbf{u}'\mathbf{H}'_{[1]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[1]}\mathbf{u} + \mathbf{u}'\mathbf{H}'_{[2]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[2]}\mathbf{u} + \mathbf{u}'\mathbf{H}'_{[3]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[3]}\mathbf{u}),$$

$$\chi^2_{[1:3]} = \chi^2_{[1]} + \chi^2_{[2|3]} + \chi^2_{[3|1,2]} \ .$$

Here the notation [.|.] is used because the orthogonal components calculate sequentially. For a instance, $\chi^2_{[3|2]}$ stands for the statistic on third-order marginals given the second-order components already calculated and $\chi^2_{[4|2,3]}$ stands for the statistics on fourth-order marginals given the second-order components followed by third-order components already calculated.

As stated in Section 3.1.3 orthogonal components, $\hat{\gamma}_j^2$, can be obtained by $\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}$ where, $\hat{\boldsymbol{\gamma}} = n^{\frac{1}{2}}\widehat{\mathbf{H}}^*\mathbf{u}$, $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q]}$ and $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega_e}$. Thus, $\mathbf{H}^*_{[1]} = \boldsymbol{F}'\mathbf{H}_{[1]}$ will be a q by 2^q matrix. Hence, $\chi^2_{[1]}$ will have q components. Similarly, $\mathbf{H}^*_{[2]} = \boldsymbol{F}'\mathbf{H}_{[2]}$ will be a $q*(q-1)/2$ by 2^q matrix. Components of $\chi^2_{[2]}$ will start from $q+1$ and ends at $q+1+q*(q-1)/2-1 = q*(q+1)/2$. Analogously, components of $\chi^2_{[3]}$ will start from $q*(q+1)/2+1$ and ends at $q*(q+1)/2+1+q*(q-1)*(q-2)/6-1 = q*(q^2+5)/6$.

Similarly,

$$\chi^2_{[1:4]} = \chi^2_{[1]} + \chi^2_{[2|3]} + \chi^2_{[3|1,2]} + \chi^2_{[4|1,2,3]}$$

Thus, components of $\chi^2_{[4]}$ will start from $q*(q^2+5)/6+1$ and ends at $q*(q^2+5)/6+1+q*(q-1)*(q-2)*(q-3)/24-1 = q*(q+1)*(q^2-3q+14)/24$.

Using those definitions I propose following statistics for my dissertation study:

statistic on second- and third-order marginals :

$$\chi^2_{[2:3]} = \sum_{j=q+1}^{j=q(q^2+5)/6} \hat{\gamma}_j^2 \ ;$$

statistic on second-, third- and fourth-order marginals :

$$\chi^2_{[2:4]} = \sum_{j=q+1}^{j=\varpi} \hat{\gamma}_j^2 \ .$$

Recall the Pearson-Fisher statistic for a cross-classified table. This statistic can be partitioned into block of components where each block represents corresponding marginals.

Result 3.2

$$\chi^2_{PF} = \chi^2_{[1]} + \chi^2_{[2|1]} + \chi^2_{[3|1,2]} + \chi^2_{[4|1,2,3]} + ... + \chi^2_{[q|1,2,...,q-1]}$$

Proof :

Note, the *df* of χ^2_{PF} is $T - g - 1$ where $T = 2^q$ for binary variables and g=number of model parameters. Therefore, the number of orthogonal components on marginals that can be obtained is equal to $T - g - 1$, and g components will be identically equal to zero. Assume $q \geq 3$. If $q = 3$ and there are two parameters for each variable in a cross-classified table then, $df = 2^3 - 2 * 3 - 1 = 1$, i.e., only one orthogonal component is possible. Also, assume $g \geq q$, with at least one parameter for each variable in a cross-classified table. For binary variables, q components of χ^2_{PF} can be fitted at zero by fixing the components of $\chi^2_{[1]}$ at zero. For instance, $\chi^2_{[1]}$ would be $\equiv 0$ for a log-linear independence model. Therefore, for a log-linear independence model, first q-order marginal components of χ^2_{PF} can be fixed at zero. If $g > q$, then $(g - q)$ components at the higher-end will be identically equal to zero, although any $(g - q)$ components could be fixed at zero.

From Reiser 2008,

$$\chi^2_{[1:q]} = \chi^2_{PF}.$$

Using above equation and substituting it in equation 3.9,

$$\chi^2_{[1:q]} = \chi^2_{PF} = n\mathbf{u}'\mathbf{H}'\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}\mathbf{u},$$

Note, $\mathbf{H} = \mathbf{H}_{[1]} \vdots \mathbf{H}_{[2]} \vdots \mathbf{H}_{[3]} \vdots \cdots \mathbf{H}_{[q]}$, thus,

$$\chi^2_{PF} = \chi^2_{[1:q]} = n\mathbf{u}'(\mathbf{H}_{[1]} \vdots \mathbf{H}_{[2]} \vdots \cdots \mathbf{H}_{[q]})'\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}(\mathbf{H}_{[1]} \vdots \mathbf{H}_{[2]} \vdots \cdots \mathbf{H}_{[q]})\mathbf{u},$$

$$\chi^2_{PF} = n(\mathbf{u}'\mathbf{H}'_{[1]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[1]}\mathbf{u} + \mathbf{u}'\mathbf{H}'_{[2]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[2]}\mathbf{u} + \cdots + \mathbf{u}'\mathbf{H}'_{[q]}\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{H}_{[q]}\mathbf{u}).$$

then,

$$\chi^2_{PF} = \chi^2_{[1]} + \chi^2_{[2|1]} + \chi^2_{[3|1,2]} + \chi^2_{[4|1,2,3]} + ... + \chi^2_{[q|1,2,...,q-1]}.$$

For example, suppose we have q=5 cross-classified variables. Then, the omnibus statistic can be partitioned into components, such as

$$\chi^2_{PF} = \chi^2_{[1]} + \chi^2_{[2|1]} + \chi^2_{[3|1,2]} + \chi^2_{[4|1,2,3]} + \chi^2_{[5|1,2,3,4]}$$

However, depending on g, the number of parameters to be estimated, the number of components will exceed degrees of freedom before the last highest-order marginal is encountered. After degrees of freedom are exhausted, additional components become identically equal to zero. For instance, a simple independence model for $q = 5$ variables would have 5 parameters with first-order marginals exactly fit, so in that case,

$$\chi^2_{PF} = \chi^2_{[2]} + \chi^2_{[3|2]} + \chi^2_{[4|2,3]} + \chi^2_{[5|2,3,4]}.$$

The model fit to the joint frequencies has 26 degrees of freedom and there would be 26 non-zero components, 10 for $\chi^2_{[2]}$, 10 for $\chi^2_{[3]}$, 5 for $\chi^2_{[4]}$ and one for $\chi^2_{[5]}$. For some other models, such as the log-linear Rasch model, some rows of **H** may be linearly dependent on columns of **G**, and each linear dependence will result in a component identically equal to zero. Calculation of components by using a method such as the unadjusted Cholesky factor would require eliminating the linear dependencies by deleting rows from **H**. The weighted regression method presented in the Section 3.2 does not require eliminating the linear dependencies.

3.2 Weighted Regression

Orthogonal components of χ^2_{PF} can be calculated using a weighted regression. The weighted regression approach is numerically more stable than the Cholesky factor method. Also, as stated in the previous section the weighted regression method does not require eliminating linear dependencies in $\mathbf{H}$ or $\mathbf{H}\vdots\mathbf{G}$. To calculate the orthogonal components using a weighted regression, the appropriate weight matrix, $\hat{\mathbf{W}}$, for the regression is given by the 2^q by 2^q matrix

$$\hat{\mathbf{W}} = (\mathbf{I} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})' - \hat{\mathbf{A}}(\hat{\mathbf{A}}'\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}') \tag{3.10}$$

where $\hat{\mathbf{A}} = \mathbf{A}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\hat{\mathbf{G}} = \mathbf{G}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, and $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$ is $\boldsymbol{\pi}(\boldsymbol{\beta})$ evaluated at MLE. Define the 2^q vector $\mathbf{z}$,

$$\boldsymbol{z} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))^{-1/2}(\hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) \ . \tag{3.11}$$

Note multiplication by $\hat{\mathbf{W}}$ can be applied to $\mathbf{z}$, but it produces no effect because $\mathbf{W}\boldsymbol{z} = \boldsymbol{z}$. To adjust the standardization of the residuals premultiply $\mathbf{H}'_{[1:q]}$ by $D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))^{-1/2}$. Then define the 2^q by $T - g - 1$ matrix $\hat{\boldsymbol{M}}$, where,

$$\hat{\boldsymbol{M}} = \hat{W} D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))^{-1/2}\mathbf{H}'_{[1:q]} \ . \tag{3.12}$$

Now fit the ordinary regression of $\mathbf{z}$ on the columns of $\hat{\boldsymbol{M}}$. Orthogonal components can be obtained as the sequential sum of squares from this regression. The sequential sum of squares can be obtained by another application of the Cholesky factor, although the SWEEP operator (Goodnight, 1979) is more stable numerically. Below is the mathematical details related to obtaining sequential sum of squares using the Cholesky factor approach.

Since, $\mathbf{z}$ is regressed on $\hat{\boldsymbol{M}}$,

$$\mathbf{z}'P_{\hat{M}}\mathbf{z} = \mathbf{z}'\hat{\boldsymbol{M}}(\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}})^{-1}\hat{\boldsymbol{M}}'\mathbf{z} \tag{3.13}$$

where $P_{\hat{M}} = \hat{\boldsymbol{M}}(\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}})^{-1}\hat{\boldsymbol{M}}'$.

Consider the Gaussian factorization of the nonsingular matrix $\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}}$ with $\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}} = \mathbf{LDL}'$, where $\mathbf{L}$ is a square nonsingular lower-traingular matrix, and $\mathbf{D}$ is a nonsingular diagonal matrix. Let $\mathbf{S} = \mathbf{LD}^{\frac{1}{2}}$. $\mathbf{S}$ is the Cholesky factor of $\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}}$, so $\hat{\boldsymbol{M}}'\hat{\boldsymbol{M}} = \mathbf{SS}'$.

Now using equation 3.13,

$$\mathbf{z}'P_{\hat{M}}\mathbf{z} = \mathbf{z}'\hat{\mathbf{K}}\hat{\mathbf{K}}'\mathbf{z}$$

where $\mathbf{K} = \hat{\boldsymbol{M}}(\mathbf{S}^{-1})'$ is 2^q by $(T-g-1)$. Then,

$$\hat{\mathbf{K}}\hat{\mathbf{K}}' = \hat{\mathbf{k}}_1\hat{\mathbf{k}}_1' + \hat{\mathbf{k}}_2\hat{\mathbf{k}}_2' + ... + \hat{\mathbf{k}}_{T-g-1}\hat{\mathbf{k}}_{T-g-1}'.$$

where $\hat{\mathbf{k}}_j$ is a column of $\hat{\mathbf{K}}$. Thus, $\mathbf{z}'\hat{\mathbf{K}}$ is 1 by $(T-g-1)$, and the sequential sum of squares for the regression are

$$SSR = \mathbf{z}'\hat{\mathbf{K}}\hat{\mathbf{K}}'\mathbf{z} = \hat{\gamma}_1^2 + \hat{\gamma}_2^2 + ... + \hat{\gamma}_{T-g-1}^2 = \hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}.$$

The $\hat{\gamma}_j^2$ are orthogonal components of χ^2_{PF}.

3.3 Asymptotic Power

In this section I will describe the theory for the calculation of asymptotic power for orthogonal components of χ^2_{PF}.

Consider the situation of testing a hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$ against alternative $H_a\colon \boldsymbol{\pi} \neq \boldsymbol{\pi}(\boldsymbol{\beta})$ using the Pearson-Fisher statistic. Suppose we have a sequence of specific alternatives $\boldsymbol{\pi}_n$ satisfying $\sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}(\boldsymbol{\beta})) \to \boldsymbol{\delta}$ for some constant matrix $\boldsymbol{\delta}$. In this approach, the best fit of the model to the population gives $\boldsymbol{\pi}_s(\boldsymbol{\beta})$ as the probability for cell s, but the true probability differs from that value by $\delta_s/\sqrt{n}$. Note the model lack-of-fit goes to zero at the rate $n^{\frac{1}{2}}$ as n approaches infinity. With this technique, Mitra (1958) shows that χ^2_{PF} has a limiting non-central chi-square distribution with non-centrality parameter λ, where

$$\lambda = \boldsymbol{\delta}' Diag[\boldsymbol{\pi}(\boldsymbol{\beta})]^{-1}\boldsymbol{\delta} \tag{3.14}$$

and $df = T - g - 1$, where $T = 2^q$ for binary variables. Using a strategy similar to Reiser (2008), it can be shown that

$$\lambda = \boldsymbol{\delta}'\mathbf{H}'\boldsymbol{\Sigma}_e^{-1}\mathbf{H}\boldsymbol{\delta} \tag{3.15}$$

where,

$$\boldsymbol{\Sigma}_{\mathbf{e}} = n^{-1}\mathbf{H}(\mathbf{D}(\boldsymbol{\pi}) - \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\boldsymbol{\beta})' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}' \ . \tag{3.16}$$

Based on the right-hand side of the expression 3.15, it is possible to decompose the noncentrality parameter into orthogonal components associated with marginals. Consider the Cholesky decomposition in Section 2.2 where $\boldsymbol{F}'\boldsymbol{\Omega_e}\boldsymbol{F} = \boldsymbol{I}$ and $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega_e}$. Using a similar decomposition, let

$$\boldsymbol{\zeta} = (\mathbf{F}')\mathbf{H}\boldsymbol{\delta} = \mathbf{H}^*\boldsymbol{\delta} \tag{3.17}$$

where $\mathbf{F}$ and $\mathbf{H}^*$ are defined as in Section 2.5. Then $\lambda = \boldsymbol{\zeta}'\boldsymbol{\zeta}$, and orthogonal components of λ are ζ_j^2, where ζ_j is an element of $\boldsymbol{\zeta}$. These components can be used to calculate the power for tests based on marginals of differing order. For example, the non-centrality parameter for $\chi^2_{[2]}$ is given by

$$\sum_{j=1}^{q(q+1)/2} \zeta_j^2 \tag{3.18}$$

As for this case, power of each orthogonal component can be calculated using the non-central chi-square distribution. The non-centrality parameter of the non-central chi-square distribution for the j^{th} component is given by ζ_j^2.

Calculation of $\boldsymbol{\delta}$:

For the purpose of power calculations under fixed, finite n, cell proportions are generated from a known model, with parameter vector $\boldsymbol{\beta}_a$. These proportions (say, $\boldsymbol{\pi}_a = \boldsymbol{\pi}(\boldsymbol{\beta}_a)$) are multiplied by a selected initial sample size such as $n_0 = 1,000$. Thereafter, the model of the null hypothesis is estimated using the resulting cell frequencies. Let, $\boldsymbol{\pi}(\boldsymbol{\beta}_a^*)$ represent the fitted proportions where, $\boldsymbol{\beta}_a^*$ are the parameter estimates that maximize the likelihood function of the model under null hypothesis. Then, the equation $\boldsymbol{\delta} = \sqrt{n}(\boldsymbol{\pi}_a - \boldsymbol{\pi}(\boldsymbol{\beta}_a^*))$ can be used to calculate $\boldsymbol{\delta}$.

3.4 Categorical Variable Factor Analysis Model

The categorical variable factor analysis model will be used for simulation and power calculations pertaining to this dissertation. These simulations and power calculations can be extended to any other models fitted to a 2^q table, such as certain log-linear models and repeated measures for categorical variables. The main motivation behind using the categorical variable factor analysis model is that I can easily find applications with large number of manifest variables. The next paragraph will illustrate the mathematical details related to the categorical variable factor analysis model.

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 (IRT) model. 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 response function:

$$P(Y_i = 1 \mid \boldsymbol{\beta}_i', X = x) = (1 + \exp(-\beta_{i0} - \beta_{i1}x))^{-1} \tag{3.19}$$

where Y_i represents the response to item i,

β_{i0} = intercept parameter for item i

β_{i1} = slope parameter for item i

$\boldsymbol{\beta}'_i = (\beta_{0i}, \beta_{1i})$

x = value taken on by latent random variable X .

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

$$P(\boldsymbol{Y} = \mathbf{y} \mid \boldsymbol{\beta}, x) = \prod_{i=1}^{q} \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)^{y_i}[1 - \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)]^{1-y_i} \tag{3.20}$$

$$\text{where } \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{11} \\ \beta_{02} & \beta_{12} \\ \beta_{03} & \beta_{13} \\ \vdots & \vdots \\ \beta_{0q} & \beta_{1q} \end{pmatrix}.$$

Finally, the probability of response pattern s, say, $\pi_s(\boldsymbol{\beta})$ 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:

$$\pi_s(\boldsymbol{\beta}) = P(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}) = \int_{-\infty}^{\infty} P(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}, x) f(x) dx \tag{3.21}$$

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

$$\pi(\boldsymbol{U} = \mathbf{n}) = n! \prod_{s=1}^{T} \frac{[\pi_s(\boldsymbol{\beta})]^{n_s}}{n_s!} \tag{3.22}$$

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

Chapter 4

RESULTS

I studied three problems related to goodness-of-fit in my dissertation. Firstly, I studied goodness-of-fit statistics using orthogonal components based on higher-order marginals, especially third-order and fourth-order marginals. To this end, I developed two new statistics based on the orthogonal components of Pearson's chi-square and studied the Type I error, empirical power and asymptotic power of these statistics under different sparseness conditions. I also compared the performance of these statistics to $\chi^2_{[2]}$, χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$ and M_r statistics. Results related to the first problem are given in Chapter 4. Secondly, I studied the Type I error and power of individual orthogonal components of $\chi^2_{[2]}$ as test statistics to identify lack-of-fit. In the context of this problem, I studied both empirical and asymptotic power. I also compared the performance of individual orthogonal components of $\chi^2_{[2]}$ to other test statistics discussed in Chapter 2. Results related to the this problem are presented in Chapter 5. Applications using these new statistics and lack-of-fit diagnostics are given in Chapter 6. Thirdly, I extended the statistics on lower-order marginals to a large number of manifest variables. When the number of manifest variables becomes larger than 20, most of the statistics on lower-order marginals have limitations in terms of computer resources and CPU time. Under this problem, I investigated the performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic, fit to confirmatory dichotomous variable factor analysis model and the performance of Tollenaar and Mooijaart (2003) statistic when the number manifest variables is larger than or equal to 25. Results related to this problem are given in Chapter 7.

4.1 Performance of $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$

I investigated the performance of $\chi^2_{[2]}$ and statistics that incorporate additional higher-order marginals. If lack-of-fit is present in 3rd- or higher-order marginals, then a test that incorporates these higher-order marginals may have a higher power than $\chi^2_{[2]}$ against H_0. On the other hand, if the higher-order marginals (3rd, 4th and etc.) are sparse, then the asymptotic chi-square approximation may not perform well for these statistics as it did for $\chi^2_{[2]}$. I further studied these issues by including the higher-order components in the test statistic. To this end, I developed two new statistics based on the orthogonal components of Pearson's chi-square:

Statistic on second- and third-order marginals :

$$\chi^2_{[2:3]} = \chi^2_{[2]} + \chi^2_{[3|2]} = \sum_{j=q+1}^{j=q(q^2+5)/6} \hat{\gamma}_j^2, \tag{4.1}$$

Statistic on second-, third- and fourth-order marginals :

$$\chi^2_{[2:4]} = \chi^2_{[2]} + \chi^2_{[3|2]} + \chi^2_{[4|2,3]} = \sum_{j=q+1}^{j=\varpi} \hat{\gamma}_j^2, \tag{4.2}$$

where,

$$\chi^2_{[3|2]} = \sum_{j=q(q+1)/2+1}^{j=q(q^2+5)/6} \hat{\gamma}_j^2,$$

and,

$$\chi^2_{[4|2,3]} = \sum_{j=q(q^2+5)/6+1}^{j=\varpi} \hat{\gamma}_j^2,$$

where,

$$\varpi = \frac{q(q+1)(q^2-3q+14)}{24}$$

Recall, the notation [.|.] is used because the orthogonal components were calculated sequentially. For instance, $\chi^2_{[3|2]}$ stands for the statistic on third-order marginals given the second-order components already calculated and $\chi^2_{[4|2,3]}$ stands for the statistics on fourth-order marginals given the second-order components followed by third-order components already calculated. $\chi^2_{[3|2]}$ and $\chi^2_{[4|2,3]}$ can be expected to have a lower power than $\chi^2_{[2]}$ when lack-of-fit is in two-way associations.

The statistic $\chi^2_{[3|2]}$ was calculated on the $2*2*2$ tables and $\chi^2_{[4|2,3]}$ was calculated on the $2*2*2*2$ tables. The $2*2*2$ and $2*2*2*2$ tables may be sparse even if the $2*2$ tables are not sparse. Mathematical details related to these statistics can be found in Section 3.1.3. Also, the $\mathbf{H}$ matrix becomes larger with additional higher-order marginals. For instance, $\chi^2_{[2]}$ had $\binom{15}{2} = 105$ components for 15 manifest variables, $\chi^2_{[3|2]}$ had $\binom{15}{3} = 455$ components, and $\chi^2_{[4|2,3]}$ had $\binom{15}{4} = 1365$ components. Thus, for 15 manifest variables, the dimension of the $\mathbf{H}$ matrix for $\chi^2_{[2:3]}$ was $560*32,768$ and the dimension of the $\mathbf{H}$ matrix for $\chi^2_{[2:4]}$ was $1,925*32,768$. Therefore, the computation required an extra 119.3 Mb to store the $\mathbf{H}$ matrix for $\chi^2_{[2:3]}$ and 477.1 Mb for $\chi^2_{[2:4]}$ than it did for $\chi^2_{[2]}$.

I started the simulation study with second and third-order marginals. I also compared the performance of these new statistics to statistics presented in the Section 2.6, namely: χ^2_{red} and M_r. Recall for $\mathbf{e} = \mathbf{H}_{[1:r]}\mathbf{r}$ and $\mathbf{r} = \hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$,

$$\mathbf{M}_r = \mathbf{e}'\hat{\mathbf{C}}_r\mathbf{e} \tag{4.3}$$

where $\hat{\mathbf{C}}_r = (\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1} - (\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}\mathbf{H}\hat{\mathbf{G}}(\hat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}\mathbf{H}\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\hat{\mathbf{T}}\mathbf{H}')^{-1}$and $\hat{\mathbf{T}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})'$. For the M_r statistic, $\mathbf{H}$ is equal to $\mathbf{H}_{[1:r]}$. Comparisons in this study were focused on second-, third- and fourth-order marginals. However, M_2, M_3 and M_4 statistics, which were used for the comparison also had first-order marginals by definition.

The Tollenaar and Mooijaart (2003) χ^2_{red} statistic is a reduced version of $\chi^2_{[1:2]}$ statistic. The difference lies in the covariance matrix $\mathbf{T}$ that does not include the term $\mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$. The generalization of χ^2_{red} to third- and fourth- order marginals is given below:

$$\chi^2_{red,[3]} = n\mathbf{e}_3'(\mathbf{H}_{[1:3]}\hat{\mathbf{T}}\mathbf{H}_{[1:3]}')^{-1}\mathbf{e}_3 \tag{4.4}$$

where,

$$\hat{\mathbf{T}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})'$$

and,

$$\chi^2_{red,[4]} = n\mathbf{e}_4'(\mathbf{H}_{[1:4]}\hat{\mathbf{T}}\mathbf{H}_{[1:4]}')^{-1}\mathbf{e}_4 \tag{4.5}$$

M_r and $\chi^2_{red,[r]}$ are often very close in value in an application. They differ by $\hat{\mathbf{C}}_r - (\mathbf{H}_{[1:r]}\hat{\mathbf{T}}\mathbf{H}_{[1:r]})^{-1}$ in the covariance matrix of the quadratic form. Note, neither M_r nor χ^2_{red} are components of χ^2_{PF}.

In this section, I introduce two new statistics, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$. Two Monte-Carlo simulation studies were performed to investigate the performance of these statistics, one with 8 manifest variables and one with 15 manifest variables. The primary purpose was to investigate the influence, if any, of sparseness in the third- and fourth-order marginals on the performance of the statistics. The simulation study was repeated for n=300, n=500 and n=1000. I also compared the performance of these two statistics to χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_2, M_3, M_4, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. Results are given in the subsequent paragraphs.

Based on the mathematical details in Chapter 2 and 3, I developed a SAS code to perform these simulations. A PROC IML macro was used to calculate the test statistics and orthogonal components. The statistics, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ were calculated by summing the appropriate orthogonal components. The weighted regression method described in the Section 3.2 was used to obtain the orthogonal components. The Sweep Operator method (Goodnight, 1979) was incorporated into the weighted regression to obtain the orthogonal components via sequential sums of squares. The SWEEP operator method is numerically more stable than the Cholesky factor method. PROC IRT method in SAS can be used for parameter estimation. However, parameter estimation from PROC IRT method was not stable for small sample sizes. Mplus (Muthn & Muthn, 2017) parameter estimates were more stable compared to SAS and therefore, Mplus estimates were used in all the calculations.

4.2 Type I Error Study

Empirical Type I error rates were examined first because a statistic may not be useful in terms of practical applications if the Type I error rate is not close to the nominal level. If a statistic does not follow the hypothesized theoretical distribution due to a condition such as sparseness, then the empirical Type I error rate may not be close to the nominal level.

Type I error simulations for $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ started with 8 manifest variables. One thousand data sets were generated using Monte-Carlo methods related to a one factor model with factor loadings (0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.2, 0.2). Similarly, to calculate the Type I error rates for 15 manifest variables, one thousand data sets were generated using Monte-Carlo methods from a known one factor model with factor loadings (0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.2, 0.2, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.2).

Savalei and Rhemtulla (2013) studied the performance of five test statistics appropriate for categorical data, and they have investigated both Type I error rate and power for different model sizes, sample sizes, numbers of categories, and threshold distributions. In their study they suggest that different pattern of intercepts may affect the power and Type I error rate. Liu and Maydeu-Olivares (2014) carried out a similar study to check the performance of several fit statistics for item pairs with known asymptotic distributions under maximum likelihood estimation of the item parameters. In their study, they suggest to investigate the performance of the test statistics when intercepts are skewed as a further research.

Therefore, I have used three different intercepts settings for my study. Simulations were repeated for each of these different intercept settings. Table 4.1 below summarizes this information for eight manifest variables and Table 4.2 below summarizes this information for 15 manifest variables :

Table 4.1: Proposed Intercepts Values - Eight Variables

Condition	Proposed Intercept Values
Symmetric	(-2.0, -1.5, -1, -0.5, 0.5, 1, 1.5, 2)
Asymmetric	(-2.1, -1.8, -1.5, -1.2, -0.9, -0.6, -0.3, 0)
Zero	(0, 0, 0, 0, 0, 0, 0, 0)

Table 4.2: Proposed Intercepts Values - Fifteen Variables

Condition	Proposed Intercept Values
Symmetric	(-3.5, -3.0, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5)
Asymmetric	(-3.5, -3.25, -3, -2.75, -2.5, -2.25, -2, -1.75, -1.5, -1.25, -1, -0.75, -0.5, -0.25, 0)
Zero	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

After generating the data, a categorical variable factor analysis model with one latent factor was estimated for each of these datasets, and empirical Type I error rates were calculated. I repeated these simulations for n=300, n=500 and n=1000. I compared the Type I error rates of $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ statistics to Type I error rates of $\chi^2_{[2]}$, χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$ and M_r statistics.

Table 4.3 below summarizes the empirical Type I error results for $q = 8$ manifest variables. The Type I error rates outside of the Monte-Carlo error interval $0.05 \pm$

$\sqrt{0.05(0.95)/1000} = (0.0365, 0.0635)$ are bolded. Type I error rates related to $\chi^2_{red,[4]}$, M_4 and $\chi^2_{[2:4]}$ were outside the Monte-Carlo error interval for both symmetric and asymmetric intercept settings. When the sample size was small (e.g. n=300) the Type I error rates related to $\chi^2_{red,[4]}$, M_4 and $\chi^2_{[2:4]}$ were considerably different from the nominal value 0.05. However, when the sample size increases, Type I error rates were improved. When n=1000, almost all the statistics had Type I error rates close to the nominal value. On the other hand, Type I error rates related to $\chi^2_{red,[3]}$, M_3 and $\chi^2_{[2:3]}$ were close to the nominal value, even for n=300. Similarly, all the Type I error rates related to χ^2_{red}, M_2 and $\chi^2_{[2]}$ were within the Monte-Carlo error interval for all the different intercept settings and sample sizes. This suggests that the $2*2*2*2$ tables were sparse when $q = 8$ but $2*2*2$ and $2*2$ tables were not sparse. However, it was interesting to see that the Type I error rates related to all the statistics were within the Monte-Carlo error interval for the zero intercept model. Thus, counts have less sparseness among cells when the intercepts are zero compared to asymmetric or symmetric intercepts.

Table 4.3: Type I Errors Rate of the Test Statistics (TS), $q = 8$

	Symmetric			**Asymmetric**			**Zero**		
TS	n=300	n=500	n=1000	n=300	n=500	n=1000	n=300	n=500	n=1000
χ^2_{red}	0.052	0.054	0.041	0.056	0.047	0.046	0.037	0.055	0.046
$\chi^2_{red,[3]}$	0.059	**0.066**	0.048	**0.075**	0.049	0.044	0.049	0.052	0.053
$\chi^2_{red,[4]}$	**0.114**	**0.095**	**0.075**	**0.092**	**0.073**	0.052	0.049	0.044	0.052
M_2	0.045	0.053	0.041	0.043	0.047	0.045	0.034	0.055	0.046
M_3	0.055	0.064	0.048	**0.071**	0.048	0.044	0.049	0.052	0.053
M_4	**0.109**	**0.093**	**0.075**	**0.087**	**0.072**	0.052	0.049	0.043	0.052
$\chi^2_{[1:2]}$	0.047	0.058	0.048	0.055	0.049	0.0365	0.041	0.047	0.045
$\chi^2_{[2]}$	0.041	0.05	0.056	0.044	0.046	0.038	0.047	0.046	0.044
$\chi^2_{[2:3]}$	0.06	**0.067**	0.057	**0.072**	0.046	0.048	0.055	0.051	0.054
$\chi^2_{[2:4]}$	**0.118**	**0.095**	**0.072**	**0.089**	**0.072**	0.051	0.045	0.046	0.054

When the number of manifest variables was extended to $q = 15$, that is 32,768 cells in the cross-classified table, Type I error rates related to $\chi^2_{red,[4]}$, M_4, $\chi^2_{[2:4]}$, $\chi^2_{red,[3]}$, M_3 and $\chi^2_{[2:3]}$ were considerably different from the nominal value 0.05 for symmetric and asymmetric intercept models. However, the Type I error rates related to χ^2_{red}, M_2 and $\chi^2_{[2]}$ were within the Monte-Carlo error interval for all the different intercept settings and sample sizes. This suggests that the $2*2*2*2$ and $2*2*2$ tables were sparse when $q = 15$ but $2*2$ tables were not sparse. However, the Type I error rates related to all the statistics were within the Monte-Carlo error interval for the zero intercept model. Therefore, the observations seem to be well distributed among cells when the intercepts are zero compared to asymmetric or symmetric intercepts even when $q = 15$.

Table 4.4: Type I Errors Rate of the Test Statistics, $q = 15$, n=500

Test Statistic	**Symmetric**	**Asymmetric**	**Zero**
χ^2_{red}	0.056	0.054	0.045
$\chi^2_{red,[3]}$	**0.13**	**0.155**	0.044
$\chi^2_{red,[4]}$	**0.279**	**0.273**	0.053
M_2	0.056	0.054	0.045
M_3	**0.129**	**0.155**	0.043
M_4	**0.279**	**0.272**	0.052
$\chi^2_{[1:2]}$	0.053	0.06	0.045
$\chi^2_{[2]}$	0.052	0.053	0.048
$\chi^2_{[2:3]}$	**0.128**	**0.157**	0.04
$\chi^2_{[2:4]}$	**0.278**	**0.274**	0.05

4.3 Power Study

As the next step, I calculated the empirical and asymptotic power of the test statistics χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_2, M_3, M_4, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ for symmetric, asymmetric and zero intercept settings. I only used $q = 8$ manifest variables because the Type I error rates related to $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_3, M_4, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ were not close to the nominal value when $q = 15$ for both symmetric and asymmetric intercept settings. However, the Type I error rates related to χ^2_{red}, M_2, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $q = 15$ were within the Monte-Carlo error interval. Therefore, I calculated the empirical and asymptotic power of these statistics for $q = 15$. Results are given in the subsequent paragraphs.

To calculate the asymptotic power under 8 manifest variables, cell proportions were generated from a known two factor model, where loadings for the first factor were (0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.2, 0.2) and loadings for the second factor were (1, 1, 1, 0, 0, 0, 1, 1). The two latent variables were specified as uncorrelated, each with variance equal to 1.0. As before, I used three intercept settings. Details related to these intercept settings are given in Table 4.1.

To calculate the asymptotic power, I used the method described in the Section 3.3. A brief description of the method is as follows. First, I generated the proportions from a two factor model with above mentioned factor loadings. I used a numerical integration method called Gauss-Hermite quadrature to generate the proportions using the model described in 3.4. Since there are two latent variables, say x_1 and x_2, equation 3.21 will now become,

$$\pi_s(\boldsymbol{\beta}) = \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}) = \int\int \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}, x_1, x_2) f(x_1, x_2) dx_1 dx_2 \qquad (4.6)$$

$$\text{where } \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{11} & \beta_{12} \\ \beta_{02} & \beta_{12} & \beta_{22} \\ \beta_{03} & \beta_{13} & \beta_{23} \\ \vdots & \vdots & \\ \beta_{0q} & \beta_{1q} & \beta_{2q} \end{pmatrix}.$$

Thereafter, these proportions were multiplied by a selected initial sample size n_0 to create the true cell frequencies under H_a. Then, the model of the null hypothesis was fitted to the resulting cell frequencies. Next, non-centrality parameters were calculated as described in equations 3.17 and 3.18. The non-centrality parameters for any other sample size, say simply n, can be approximated by using the expression $\lambda \approx \frac{n}{n_0}\lambda_0$. Once I obtained the non-centrality parameters, I used non-central chi-square distribution to calculate the asymptotic power for $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$. Note, the significance level was set to 0.05.

Simulations for the empirical power were performed with the same parameter values as in the asymptotic power study. To calculate the empirical power, one thousand data sets were generated using Monte-Carlo methods related to a two factor model. Then, a one factor model was fitted for each of these datasets and empirical power was calculated. In the simulation, the model under H_0 is misspecified with a one factor model. To calculate the empirical power for each statistic, the sum of the number of cases that exceed the chi-square critical value (at 5% significance level) under the corresponding degree of freedom of the statistic was divided by the number of datasets.

This process was repeated for sample sizes 300, 500 and 1000. Empirical and asymptotic power of $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ were also compared to the empirical and asymptotic power of χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_2, M_3, M_4, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$.

Table 4.5: Power for Symmetric Intercept Settings

Test Statistic	n=300		n=500		n=1000	
	Empirical	Asymptotic	Empirical	Asymptotic	Empirical	Asymptotic
χ^2_{red}	0.605	0.66079	0.88	0.91498	0.999	0.99944
$\chi^2_{red,[3]}$	0.318	0.35128	0.595	0.62798	0.951	0.96717
$\chi^2_{red,[4]}$	*	0.24251	*	0.44799	*	0.87271
M_2	0.604	0.66071	0.88	0.91493	0.999	0.99944
M_3	0.317	0.35122	0.595	0.62789	0.951	0.96714
M_4	*	0.24321	*	0.44937	*	0.87396
$\chi^2_{[1:2]}$	0.467	0.51813	0.796	0.81697	0.996	0.99596
$\chi^2_{[2]}$	0.531	0.57981	0.838	0.86537	0.996	0.99822
$\chi^2_{[2:3]}$	0.301	0.33193	0.567	0.59953	0.942	0.95824
$\chi^2_{[2:4]}$	*	0.23607	*	0.43582	*	0.86219

* Power is contaminated by inaccurate Type I error

Tables 4.5, 4.6 and 4.7 indicate asymptotic and empirical power results for symmetric, asymmetric and zero intercept settings, respectively. In the previous paragraphs, I have illustrated the Type I error rates related to each of these settings. Some of the Type I error rates were not close to the nominal value due to sparseness. Power results related to these cases were marked with an asterisk because if the Type I error rates are inaccurate due to sparseness then the power results do not have much validity in terms of practical applications.

Table 4.6: Power for Asymmetric Intercept Settings

Test Statistic	n=300		n=500		n=1000	
	Empirical	Asymptotic	Empirical	Asymptotic	Empirical	Asymptotic
χ^2_{red}	0.609	0.63631	0.866	0.89984	0.998	0.99909
$\chi^2_{red,[3]}$	*	0.3395	0.596	0.60979	0.941	0.96097
$\chi^2_{red,[4]}$	*	0.23473	*	0.43266	0.828	0.85811
M_2	0.608	0.63565	0.866	0.89941	0.998	0.99908
M_3	*	0.33904	0.595	0.60908	0.941	0.96071
M_4	*	0.23531	*	0.4338	0.828	0.85924
$\chi^2_{[1:2]}$	0.486	0.49429	0.779	0.79345	0.993	0.99406
$\chi^2_{[2]}$	0.518	0.55492	0.806	0.84508	0.997	0.99726
$\chi^2_{[2:3]}$	*	0.32041	0.57	0.58088	0.928	0.95062
$\chi^2_{[2:4]}$	*	0.22846	*	0.42063	0.826	0.84

* Power is contaminated by inaccurate Type I error

Table 4.7: Power for Zero Intercept Settings

Test Statistic	n=300		n=500		n=1000	
	Empirical	Asymptotic	Empirical	Asymptotic	Empirical	Asymptotic
χ^2_{red}	0.698	0.75933	0.942	0.96175	0.999	0.99994
$\chi^2_{red,[3]}$	0.396	0.42969	0.694	0.73579	0.982	0.9903
$\chi^2_{red,[4]}$	0.251	0.29583	0.512	0.54723	0.92	0.94165
M_2	0.698	0.75732	0.942	0.96101	0.999	0.99994
M_3	0.396	0.42786	0.694	0.73353	0.982	0.99
M_4	0.251	0.29545	0.512	0.54657	0.92	0.94132
$\chi^2_{[1:2]}$	0.612	0.61651	0.887	0.89623	0.999	0.99932
$\chi^2_{[2]}$	0.634	0.67971	0.902	0.93016	0.999	0.99975
$\chi^2_{[2:3]}$	0.378	0.40482	0.668	0.70591	0.977	0.98636
$\chi^2_{[2:4]}$	0.243	0.28643	0.511	0.53119	0.913	0.93383

By observing the results in Tables 4.5, 4.6 and 4.7, it is clear that the $\chi^2_{[2]}$ had higher empirical power for all the settings (symmetric, asymmetric and zero) compared to $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$. This indicates that when the lack-of-fit is in the second-order marginals, including third- or fourth-order marginals in the test statistic can dilute the test. When sample size was small (e.g. n=300) empirical power was somewhat lower than the asymptotic power. However, when the sample size increased (e.g. n=1000) empirical power was very close to the asymptotic power. This is related to how fast the empirical distribution can converge to asymptotic distribution. The lower empirical power seen in the results for small sample size is related to the empirical distribution not as close to the asymptotic distribution. The zero intercept model had the highest power. This further validates the fact that the observations are well distributed among cells in the cross-classified table when the intercepts are zero compared to asymmetric or symmetric intercept models.

When the sample size was small, M_3, M_4, $\chi^2_{red,[3]}$ and $\chi^2_{red,[4]}$ statistics each had a slightly higher power compared to $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ statistics. One can argue this might be due to the fact that $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ converge to the theoretical distribution slower than the M_3, M_4, $\chi^2_{red,[3]}$ and $\chi^2_{red,[4]}$. Note, by default M_3, M_4, $\chi^2_{red,[3]}$ and $\chi^2_{red,[4]}$ contain first-order marginals whereas $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ start from second-order marginals. It is also possible that these first-order marginals may have some contribution to the higher power of these test statistics.

4.3.1 Three-Way Association Study

For the next study, I used a different model to study the power of $\chi^2_{[2:3]}$ and $\chi^2_{[3:4]}$ where there is a higher order effect present in the model. I used a log-linear model with 3-way interactions. The log-linear version has the advantage that it is convenient to demonstrate the influence of higher-order interactions that would be found in the third-order marginals.

As before, I started the simulations with a Type I error study. If a statistic does not follow the hypothesized theoretical distribution due to a condition such as sparseness, then the empirical Type I error rate may not be close to the nominal level. To calculate the Type I error rates, I generated the data from the following log-linear model:

$$\log(m_s) = \lambda + \lambda_f^{y_1} + \lambda_g^{y_2} + ... + \lambda_m^{y_8} + \lambda_{fg}^{y_1,y_2} + \lambda_{fh}^{y_1,y_3} + ... + \lambda_{lm}^{y_7,y_8}, \qquad (4.7)$$

where, m_s is the expected count for cell s and $\lambda = 0.5$, $\lambda_1^{y_1} = \lambda_1^{y_2} = \lambda_1^{y_3} = ... = \lambda_1^{y_8} = 0.1, \lambda_{11}^{y_i,y_j} = \lambda_{00}^{y_i,y_j} = -\lambda_{01}^{y_i,y_j} = -\lambda_{10}^{y_i,y_j} = 0.2$ for $i, j = 1, 2, .., 8$.

A categorical factor model with one latent factor was fitted for the pseudo data generated from model 4.7 and Type I error rates were calculated. Results are given in the Table 4.8. The Type I error rates outside of the Monte-Carlo error interval $0.05 \pm \sqrt{0.05(0.95)/1000} = (0.0365, 0.0635)$ are bolded. Results suggest that the $2 * 2 * 2 * 2$ tables were sparse, especially when the sample size is small (e.g. n=300).

Table 4.8: Type I Error Results for Three-Way Association Study

	n=300	**n=500**	**n=1000**
χ^2_{red}	0.036	0.039	0.041
$\chi^2_{red,[3]}$	0.048	0.052	0.058
$\chi^2_{red,[4]}$	**0.118**	**0.097**	0.067
M_2	0.034	0.037	0.039
M_3	0.046	0.051	0.054
M_4	**0.111**	**0.091**	0.064
$\chi^2_{[2]}$	0.041	0.047	0.052
$\chi^2_{[2:3]}$	0.059	0.062	0.058
$\chi^2_{[2:4]}$	**0.114**	**0.093**	0.066

The degrees of freedom for M_r are not clear in this application because the pseudo data are generated from a model (4.7) where first-order marginals are exactly fit, although the first-order may not be exactly fit under the generalized linear latent variable model. Without adjusting degrees of freedom, M_r produced very low Type I error rate. Therefore, degrees of freedom for M_r were decreased by eight, because the simulation has eight manifest variables for which first-order may have been exactly fit. However, it is not known if this adjustment to the degrees of freedom is correct.

Next, I calculated empirical power using the following log-linear model which contain a three-way interaction:

$$\log(m_s) = \lambda + \lambda_f^{y_1} + \lambda_g^{y_2} + ... + \lambda_m^{y_8} + \lambda_{fg}^{y_1,y_2} + \lambda_{fh}^{y_1,y_3} + ... + \lambda_{lm}^{y_7,y_8} + \lambda_{fgh}^{y_1,y_2,y_3}, \quad (4.8)$$

where, m_s is the expected count for cell s and $\lambda = 0.5$, $\lambda_1^{y_1} = \lambda_1^{y_2} = \lambda_1^{y_3} = ... = \lambda_1^{y_8} = 0.1, \lambda_{11}^{y_i,y_j} = \lambda_{00}^{y_i,y_j} = -\lambda_{01}^{y_i,y_j} = -\lambda_{10}^{y_i,y_j} = 0.2$ for $i, j = 1, 2, .., 8$ and $\lambda_{001}^{y_1,y_2,y_3} = \lambda_{010}^{y_1,y_2,y_3} = \lambda_{100}^{y_1,y_2,y_3} = \lambda_{111}^{y_1,y_2,y_3} = -\lambda_{000}^{y_1,y_2,y_3} = -\lambda_{011}^{y_1,y_2,y_3} = -\lambda_{101}^{y_1,y_2,y_3} = -\lambda_{110}^{y_1,y_2,y_3} = 0.7$.

To study the power of $\chi^2_{[2]}$, $\chi^2_{[2:3]}$ and $\chi^2_{[3:4]}$, the one-factor categorical factor model was fitted for pseudo data generated from model 4.8. All pair-wise associations were constrained equal in the generating model 4.8 with only one three-way interaction among variables Y_1, Y_2 and Y_3. Thus, I was expecting a higher power for $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ statistics than $\chi^2_{[2]}$. I also compared the performance of these statistics to $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_3 and M_4 using a simulation. The study was repeated for n=300, 500 and 1000. Table 4.9 indicates the power results related to the three-way association study. Under the alternative hypothesis, a three-way interaction effect was present for variables Y_1, Y_2 and Y_3. Since the model of the null hypothesis does not include a three-way interaction, there is a discrepancy also in the three-way associations.

Table 4.9: Power Results for Three-Way Association Study

	n=300	**n=500**	**n=1000**
χ^2_{red}	0.234	0.269	0.390
$\chi^2_{red,[3]}$	0.567	0.928	1
$\chi^2_{red,[4]}$	*	*	1
M_2	0.214	0.264	0.388
M_3	0.528	0.914	1
M_4	*	*	0.999
$\chi^2_{[2]}$	0.194	0.575	0.954
$\chi^2_{[2:3]}$	0.369	0.817	1
$\chi^2_{[2:4]}$	*	*	0.995

* Power is contaminated by inaccurate Type I error

As can be seen from the results, power of the test based on $\chi^2_{[2:3]}$ surpasses the power of the test based on $\chi^2_{[2]}$ when there is a three-way association effect. I repeated the above simulation twice with $\lambda_{001}^{y_1,y_2,y_3} = \lambda_{010}^{y_1,y_2,y_3} = \lambda_{100}^{y_1,y_2,y_3} = \lambda_{111}^{y_1,y_2,y_3} = -\lambda_{000}^{y_1,y_2,y_3} = -\lambda_{011}^{y_1,y_2,y_3} = -\lambda_{101}^{y_1,y_2,y_3} = -\lambda_{110}^{y_1,y_2,y_3} = 0.5$, and $\lambda_{001}^{y_1,y_2,y_3} = \lambda_{010}^{y_1,y_2,y_3} = \lambda_{100}^{y_1,y_2,y_3} = \lambda_{111}^{y_1,y_2,y_3} = -\lambda_{000}^{y_1,y_2,y_3} = -\lambda_{011}^{y_1,y_2,y_3} = -\lambda_{101}^{y_1,y_2,y_3} = -\lambda_{110}^{y_1,y_2,y_3} = 0.8$. As the three-way association effect becomes larger, the power of the test based on $\chi^2_{[2]}$ rose only gradually, but the power of test based on the $\chi^2_{[2:3]}$ rose rapidly. This suggests $\chi^2_{[2:3]}$ statistic works better when there is a three-way association compared to $\chi^2_{[2]}$. Also, the power of $\chi^2_{[2:4]}$ was lower than $\chi^2_{[2:3]}$. This suggest that the lack-of-fit is in the third-order and adding additional components may dilute the test. However, the Type I error rate related to $\chi^2_{[2:4]}$ was not close to the nominal value 0.05 especially, when n=300. If Type I error rates are not close to the nominal value then the power results does not have much validity in terms of practical applications. Therefore, I replaced those power results with an asterisk mark.

To sum up, a test based on low-order marginals, $\chi^2_{[2]}$, has higher power to detect lack-of-fit located in the second-order associations when compared with a statistic that incorporates higher-order marginals such as $\chi^2_{[2:3]}$ or the $\chi^2_{[2:4]}$. $\chi^2_{[2]}$, however, would be very insensitive to a lack-of-fit that is present in the third-order marginals. Other limited-information statistics discussed in this section, χ^2_{red} and M_2, suffered the same lower power in this type of situation. In many applications of latent variable models in the social sciences, manifest variables are designed to have high bi-variate association, but sometimes it is possible to have three-way or four-way associations. In those situations, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ may out perform lower-order statistics like $\chi^2_{[2]}$. The ability to choose between different statistics $\chi^2_{[2]}$, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ in different situations can help to improve the inference and decisions made in real world applications. The SAS code I developed facilitates this approach and can help to improve the decisions made in real-word applications.

Chapter 5

PERFORMANCE OF INDIVIDUAL ORTHOGONAL COMPONENTS OF $\chi^2_{[2]}$

Results from the previous section showed that $\chi^2_{[2]}$ may have higher power for certain alternative hypothesis, especially when the lack-of-fit is in the second-order marginals. When a model fails to fit adequately, it is important to know where the model provides a good fit and where it does not. This section will illustrate the performance of individual orthogonal components of $\chi^2_{[2]}$ as test statistics to identify lack-of-fit.

According to Liu and Maydeu-Olivares (2014) paper, statistics $\bar{\chi}^2_{ij}$ (Liu & Maydeu-Olivares, 2014) and adjusted residuals (Reiser, 1996) has better Type I error rates and power compared to other test statistics explained in Section 2.6. Therefore, I included $\bar{\chi}^2_{ij}$ and adjusted residuals in my simulation and compared the performance to individual orthogonal components of $\chi^2_{[2]}$.

5.1 Simulation Study Part I: Empirical Type I Error Rates

As before, I calculated Type I error rates first because if the Type I error rates are too far from the nominal value, then the power results do not have much validity in terms of practical applications. Table 5.1 below summarizes the Type I error study:

Table 5.1: Design of Type I Error Study

Model (data generation)	categorical variable factor analysis model with one latent factor
Model (fitted)	categorical variable factor analysis model with one latent factor
Number of observed variables	q=8, q=15
Number of samples	1000
Sample size	n=300, n=500

As the the above Table indicates, I used two settings for the number of manifest variables (q). When $q = 8$ there are 256 cells in the 2^8 cross-classified table. With n=500, each cell may have, on average, two observations. When $q = 15$, there are 32768 cells in the 2^{15} cross-classified table. Even with n=500, each cell may only have, on average, 0.01 observations. Thus, the sparseness in the cross-classified table is very severe when $q = 15$. But, I'm using second-order marginals and $2 * 2$ sub-table may not be sparse even when $q = 15$ even with n=300. Therefore, I was expecting individual orthogonal components of $\chi^2_{[2]}$ to have good Type I error rates when $q = 8$ and $q = 15$.

5.1.1 Type I Error Study for Eight Variables

The first simulation included eight manifest variables. One thousand data sets were generated using Monte-Carlo methods related to a one factor model where $\beta_1' = (0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2, 0.2)$. As before, I used three intercept settings. Intercept values for each setting are given in the Table 4.1. Then, a categorical variable factor analysis model with one latent factor was estimated for each of these datasets, and empirical Type I error rates of the individual orthogonal components were calculated. Since each individual orthogonal component is distributed approximately as chi-square with one degree of freedom, to calculate the empirical Type I error rate 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. Similar process was used to calculate the Type I error rates of the adjusted residual and $\bar{\chi}^2_{ij}$. This simulation was repeated for sample sizes 300 and 500.

Table 5.2 below indicates the empirical Type I error rates for $q = 8$ manifest variables for symmetric intercept model. The Type I error rates outside of the Monte-Carlo error interval $0.05 \pm \sqrt{0.05(0.95)/1000} = (0.0365, 0.0635)$ are bolded. When n=300, Type I error rates related to orthogonal components (2,4) and (4,5) were outside the Monte-Carlo error interval. Given that there are twenty eight individual orthogonal components, it is possible that one or two components may randomly fall slightly outside the Monte-Carlo error interval. However, six components related to $\bar{\chi}^2_{ij}$ and five components related to adjusted residuals were outside the Monte-Carlo error interval. This suggests, when n=300, orthogonal components have better Type I error rates compared to $\bar{\chi}^2_{ij}$ and adjusted residuals for $q = 8$ manifest variables for symmetric intercept model. When n=500, all most all the components related to $\bar{\chi}^2_{ij}$, orthogonal components, and adjusted residuals were inside the Monte-Carlo error interval $(0.0365, 0.0635)$.

Type I error rates for $q = 8$ manifest variables for asymmetric intercept model are given in the Table A.1. As for the symmetric intercept model, when n=300, orthogonal components had better Type I error rates compared to $\bar{\chi}^2_{ij}$ and adjusted residuals. Five components related to $\bar{\chi}^2_{ij}$ and five components related to adjusted residuals were outside the Monte-Carlo error interval. Only one component related to orthogonal components was outside the Monte-Carlo error interval. Again, given that there are twenty eight individual orthogonal components, it is possible that one or two components may randomly fall slightly outside the Monte-Carlo error interval. As before, when n=500, all the components related to $\bar{\chi}^2_{ij}$, orthogonal components, and adjusted residuals were inside the Monte-Carlo error interval $(0.0365, 0.0635)$.

Type I error rates for $q = 8$ manifest variables for zero intercept model are given in the Table A.2. All the components related to $\bar{\chi}^2_{ij}$, orthogonal components, and adjusted residuals were inside the Monte-Carlo error interval $(0.0365, 0.0635)$, even for n=300. Thus, counts have less sparseness among cells when the intercepts are zero compared to asymmetric or symmetric intercepts.

As explained in the Chapter 3, each orthogonal component is distributed approximately as chi-square with one degree of freedom. To check this assumption, chi-square Q-Q plots were built for the simulation values related to each component. A similar approach was taken to check the normality assumption of the adjusted residuals. On the other hand, the $\bar{\chi}^2_{ij}$ had different degree of freedom for different item pairs. The degree of freedom of $\bar{\chi}^2_{ij}$ depends on $\boldsymbol{\Sigma}_{ij}$ where, $\boldsymbol{\Sigma}_{ij}$ is the covariance matrix related to the residuals $n(\mathbf{p}_{ij} - \hat{\boldsymbol{\pi}}_{ij})$ for a pair of items (for more information, refer Section 2.8). Thus, one thousand simulation values for a specific item pair would have one thousand different degrees of freedom. I have used an average value of these degrees of freedom to calculate the chi-square Q-Q plots for $\bar{\chi}^2_{ij}$.

Note, for q = 8 manifest variables, there are 8*7/2=28 second-order marginals. Thus, twenty eight Q-Q plots for orthogonal components, twenty eight Q-Q plots for adjusted residuals and twenty eight Q-Q plots for $\bar{\chi}^2_{ij}$ were compared. Since I repeated this simulation for n=300 and 500, I had 84*2=168 plots. It is impractical to append all these plots in to this study. Therefore, selected results are shown in the Appendix (Figures B.1 through B.12). Some of the plots related to the symmetric intercept model are given below.

<u>**Symmetric Intercept Model**</u>

Figure 5.1: Orthogonal Components, n=300

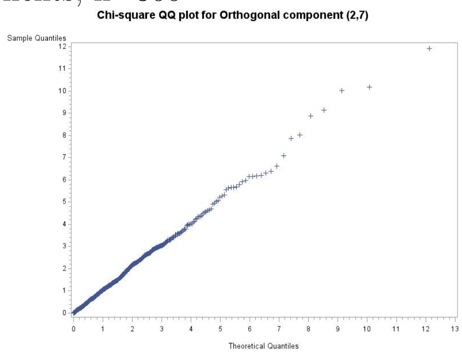


Figure 5.2: Orthogonal Components, n=500

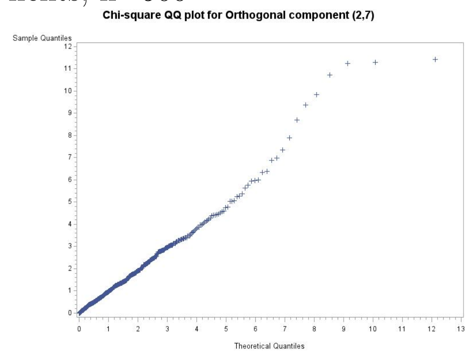


Figure 5.3: Adjusted Residuals, n=300

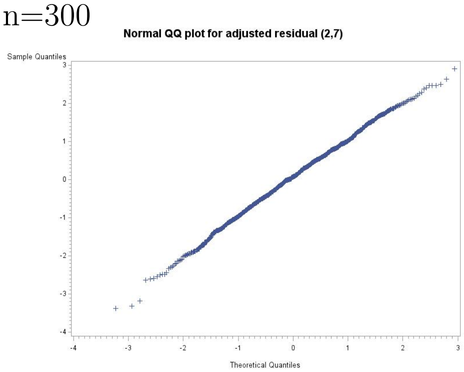


Figure 5.4: Adjusted Residuals, n=500

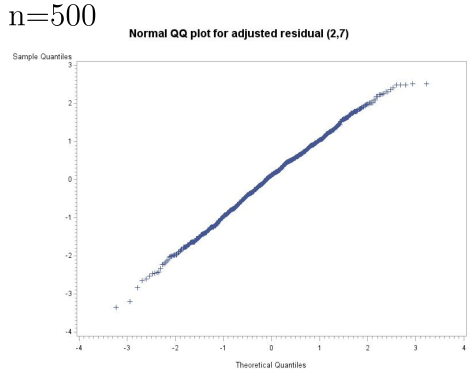


Figure 5.5: $\bar{\chi}^2_{ij}$, n=300

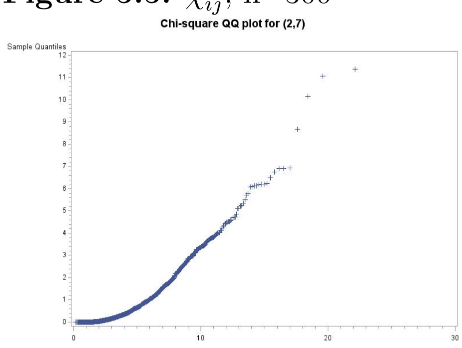
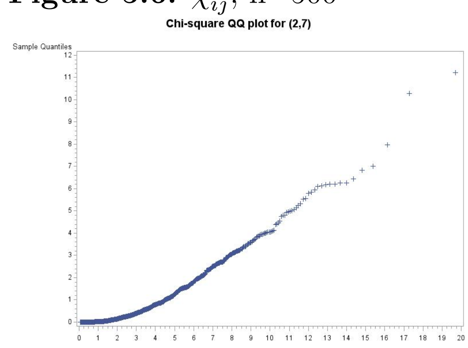


Figure 5.6: $\bar{\chi}^2_{ij}$, n=500



If the distributional assumption is attained, then the points in the Q-Q plot will approximately lie on the line $y = x$ (straight line assumption). Most of the Q-Q plots attained this assumption. There were a few Q-Q plots that showed deviations from the straight line assumption, especially when the sample size was small. Note, these Q-Q plots are very sensitive to outlier observations. When the sample size is small, some of the estimated standard errors related to the test statistics can be very small. This can results in a larger test statistic value. A few of these larger test statistic values can easily affect the pattern of the Q-Q plot.

Most of the Q-Q plots that deviated from the hypothesized distributions were related to $\bar{\chi}^2_{ij}$. This might be related to the fact that I'm using the mean value of the one thousand degrees of freedom of a particular $\bar{\chi}^2_{ij}$ to calculate the Q-Q plot for that particular $\bar{\chi}^2_{ij}$ even though each have different degree of freedom (df) under each simulation value. For example, when $n = 300$, minimum df was 0.999 and maximum df was 1.69. I recommend to further study this as a future work.

When the sample size and/or the factor loadings are too small, some of the estimated standard errors of the residuals tend to become negative or close to zero. Thus, out of 1000 simulation only around 750-850 simulations were successful in calculating the residuals. To fix this issue, a shrinkage estimator was incorporated into estimation of the covariance matrix of the residuals.

A simple version of a shrinkage estimator is constructed as follows. One considers a convex combination of the empirical estimator $\mathbf{X}$ with some suitable chosen target $\boldsymbol{Y}$, e.g., the diagonal matrix. Subsequently, the mixing parameter Δ is selected to maximize the expected accuracy of the shrunken estimator. This can be done by cross-validation, or by using an analytic estimate of the shrinkage intensity. The resulting regularized estimator, $\Delta\mathbf{X} + (1 - \Delta)\boldsymbol{Y}$, can be shown to outperform the maximum likelihood estimator for small samples.

Based on this idea, I developed a method for my simulations. First, I extracted the eigenvalues of the covariance matrix. If some eigenvalues were ≤ 0, it was fixed using the equation

$$\hat{\Sigma}^*_R = \mathbf{Diag(eigen)} + \eta * eigen_l * \mathbf{I},$$

where **Diag(eigen)** is the diagonal matrix with the eigenvalues of the covariance matrix on the diagonal, $eigen_l$ is the largest negative eigenvalue in absolute value, **I** is the identity matrix and η was chosen heuristically. The idea here is to make all the eigenvalues positive without altering too much of the underline structure of the covariance matrix of the residuals. For instance, if $\eta = 1.00005$ and the largest negative eigenvalue is -0.003 then, 0.00300015 will be added to all the eigenvalues. After incorporating this method, the number of successful iterations for the above simulations (using PROC IRT method) increased to 970-1000 out of 1000 simulations.

Mplus parameter estimates did not seems to need this shrinking estimator fix and were successful in calculating the statistics in all most all the simulations. I used Mplus parameter estimates for all the calculation in this Chapter. To compare the performance, I re-ran these simulations with PROC IRT method in SAS. Results suggest Mplus estimates were more stable in estimating parameters compared to PROC IRT method.

Table 5.2: Type I Error Study for Symmetric Intercept Model

	n=300			n=500		
Pair (i,j)	Orthgonal Comp.	Std. Residuals	$\bar{\chi}^2_{ij}$	Orthgonal Comp.	Std. Residuals	$\bar{\chi}^2_{ij}$
(1,2)	0.037	0.038	0.037	0.041	0.041	0.041
(1,3)	0.058	0.061	0.062	0.059	0.061	0.06
(1,4)	0.046	0.052	0.049	0.05	0.054	0.051
(1,5)	0.041	0.047	0.047	0.043	0.047	0.044
(1,6)	0.043	0.055	0.054	0.038	0.044	0.044
(1,7)	0.048	0.051	0.052	0.054	0.049	0.048
(1,8)	0.051	0.043	0.041	0.044	**0.034**	**0.034**
(2,3)	0.051	0.052	0.051	0.049	0.048	0.049
(2,4)	**0.029**	**0.027**	**0.03**	0.038	0.04	0.039
(2,5)	0.048	0.043	0.04	0.037	0.042	0.043
(2,6)	0.046	0.047	0.05	0.056	0.04	0.042
(2,7)	0.05	0.052	0.052	0.045	0.051	0.05
(2,8)	0.045	0.047	0.046	0.052	0.05	0.049
(3,4)	0.045	0.044	0.045	0.051	0.052	0.05
(3,5)	0.042	0.04	0.044	0.045	0.049	0.047
(3,6)	0.046	0.058	0.056	0.044	0.06	0.06
(3,7)	0.046	0.051	0.05	0.041	0.056	0.057
(3,8)	0.057	0.043	0.045	0.051	0.055	0.055
(4,5)	**0.075**	**0.079**	**0.076**	0.063	**0.066**	**0.067**
(4,6)	0.047	**0.075**	**0.081**	0.05	0.056	0.056
(4,7)	0.044	0.051	0.051	0.049	0.046	0.046
(4,8)	0.052	0.037	**0.035**	0.054	0.048	0.048
(5,6)	0.062	**0.075**	**0.073**	0.046	0.055	0.06
(5,7)	0.037	0.056	0.056	0.05	0.046	0.046
(5,8)	0.051	0.046	0.047	0.054	0.046	0.047
(6,7)	0.06	**0.069**	**0.07**	0.05	0.061	0.06
(6,8)	0.044	0.041	0.043	0.054	0.053	0.053
(7,8)	0.056	0.046	0.047	0.048	0.054	0.054

5.1.2 Type I Error Study for Fifteen Variables

Next, I increased the number of manifest variables to fifteen and carried out the same calculations and simulations as in the eight variable study. To check the Type I error rates, one thousand data sets were generated from one factor model and then a one factor model was fitted. I used a repetition of the slope parameters of one factor model for eight manifest variables as the slope parameters for the one factor model with fifteen manifest variables: (0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2, 0.2, 0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2). The idea was to make the comparison between an eight variable study and a fifteen variable study more meaningful.

As in the eight variable study, I used three intercept settings for Type I error study: symmetric, asymmetric, zero. Intercept values for each setting are given in the Table 4.2.

Table 5.3 and Table 5.6 below indicate the empirical Type I error rates for individual orthogonal components of $\chi^2_{[2]}$ for $q = 15$ manifest variables for symmetric intercept model for n=500 and n=300, respectively. The Type I error rates outside of the Monte-Carlo error interval $0.05 \pm \sqrt{0.05(0.95)/1000} = (0.0365, 0.0635)$ are bolded. When n=300, five components related to orthogonal components were outside the Monte-Carlo error interval and when n=500, four components related to orthogonal components were outside the Monte-Carlo error interval. Given that there are 105 individual orthogonal components, this is a good Type I error performance. With this many individual orthogonal components, it possible that four or five components may randomly fall slightly outside the Monte-Carlo error interval. Table 5.4 and Table 5.7 indicate Type I error rates for adjusted residuals for $q = 15$ manifest variables for symmetric intercept model when n=500 and n=300, respectively. Similarly, Table 5.5 and Table 5.8 indicate Type I error rates for $\bar{\chi}^2_{ij}$ for $q = 15$ manifest variables for symmetric intercept model when n=500 and n=300, respectively. When n=500, six components related to adjusted residuals and six components related to $\bar{\chi}^2_{ij}$ were outside the Monte-Carlo error interval. When n=300, eleven components related to adjusted residuals and fourteen components related to $\bar{\chi}^2_{ij}$ were outside the Monte-Carlo error interval. This suggests that when n=300, orthogonal components have better Type I error rates compared to $\bar{\chi}^2_{ij}$ and adjusted residuals for $q = 15$ manifest variables for symmetric intercept model. However, when n=500, orthogonal components, $\bar{\chi}^2_{ij}$ and adjusted residuals seems to have similar Type I error rates. Graphical illustration of the same information are given in Figures 5.7 and 5.8.

Type I error rates for $q = 15$ manifest variables for asymmetric intercept model for n=300 and n=500 are given in the Appendix in Tables A.4 through A.8. Graphical illustration of the same information are given in Figures B.13 and B.13. When n=300, orthogonal components, $\bar{\chi}^2_{ij}$ and adjusted residuals had somewhat similar Type I error rate performance. Eleven components related to orthogonal components, fifteen components related to $\bar{\chi}^2_{ij}$ and fourteen components related to adjusted residuals were outside the Monte-Carlo error interval. Under symmetric intercept settings, when n=300, only five components related to orthogonal components were outside the Monte-Carlo error interval. With asymmetric intercept settings this amount was increased to eleven. This indicates the 2*2 sub-table may have more sparseness under asymmetric intercept settings compared to symmetric intercept settings especially with large number of manifest variables and small sample size. Note, this did not happened with eight manifest variables (n=300 or n=500) or with fifteen manifest variables with n=500.

Table 5.3: Type I Error Study for Orthogonal Components for Symmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.052	(3,12)	0.058	(7,9)	**0.033**
(1,3)	0.039	(3,13)	0.047	(7,10)	0.05
(1,4)	0.042	(3,14)	0.056	(7,11)	0.047
(1,5)	0.046	(3,15)	0.046	(7,12)	0.059
(1,6)	0.052	(4,5)	0.06	(7,13)	0.049
(1,7)	0.052	(4,6)	0.044	(7,14)	0.056
(1,8)	0.045	(4,7)	0.06	(7,15)	0.05
(1,9)	0.05	(4,8)	0.052	(8,9)	0.051
(1,10)	0.039	(4,9)	0.047	(8,10)	0.06
(1,11)	0.035	(4,10)	0.054	(8,11)	0.043
(1,12)	0.039	(4,11)	0.059	(8,12)	0.059
(1,13)	0.057	(4,12)	0.052	(8,13)	0.046
(1,14)	0.051	(4,13)	**0.036**	(8,14)	0.056
(1,15)	0.045	(4,14)	0.044	(8,15)	0.051
(2,3)	0.039	(4,15)	0.043	(9,10)	0.048
(2,4)	0.046	(5,6)	0.046	(9,11)	0.041
(2,5)	0.06	(5,7)	0.048	(9,12)	0.042
(2,6)	0.061	(5,8)	0.048	(9,13)	0.052
(2,7)	0.048	(5,9)	0.045	(9,14)	0.043
(2,8)	0.049	(5,10)	0.058	(9,15)	0.05
(2,9)	0.042	(5,11)	0.053	(10,11)	0.061
(2,10)	0.043	(5,12)	0.046	(10,12)	0.055
(2,11)	0.053	(5,13)	0.054	(10,13)	0.049
(2,12)	0.061	(5,14)	0.062	(10,14)	0.05
(2,13)	0.051	(5,15)	0.05	(10,15)	0.045
(2,14)	0.039	(6,7)	0.05	(11,12)	0.054
(2,15)	0.055	(6,8)	0.048	(11,13)	0.043
(3,4)	0.039	(6,9)	**0.069**	(11,14)	0.047
(3,5)	0.051	(6,10)	0.057	(11,15)	0.05
(3,6)	0.056	(6,11)	0.039	(12,13)	0.049
(3,7)	0.04	(6,12)	0.045	(12,14)	0.05
(3,8)	0.055	(6,13)	0.044	(12,15)	0.049
(3,9)	0.042	(6,14)	0.048	(13,14)	0.059
(3,10)	0.057	(6,15)	0.056	(13,15)	0.05
(3,11)	0.05	(7,8)	**0.066**	(14,15)	0.046

Table 5.4: Type I Error Study for Adjusted Residuals for Symmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.052	(3,12)	0.044	(7,9)	0.039
(1,3)	0.038	(3,13)	0.053	(7,10)	0.045
(1,4)	0.042	(3,14)	0.041	(7,11)	0.052
(1,5)	0.045	(3,15)	0.04	(7,12)	0.061
(1,6)	0.047	(4,5)	0.063	(7,13)	0.043
(1,7)	0.054	(4,6)	0.043	(7,14)	**0.065**
(1,8)	0.046	(4,7)	0.057	(7,15)	0.045
(1,9)	0.048	(4,8)	0.05	(8,9)	0.051
(1,10)	0.039	(4,9)	0.049	(8,10)	0.056
(1,11)	**0.032**	(4,10)	0.058	(8,11)	0.037
(1,12)	**0.027**	(4,11)	0.056	(8,12)	0.051
(1,13)	0.051	(4,12)	0.043	(8,13)	0.043
(1,14)	**0.034**	(4,13)	0.046	(8,14)	0.046
(1,15)	0.051	(4,14)	**0.036**	(8,15)	0.048
(2,3)	0.04	(4,15)	0.044	(9,10)	0.05
(2,4)	0.045	(5,6)	0.046	(9,11)	0.043
(2,5)	0.058	(5,7)	0.046	(9,12)	0.048
(2,6)	0.062	(5,8)	0.052	(9,13)	0.045
(2,7)	0.046	(5,9)	0.051	(9,14)	0.05
(2,8)	0.052	(5,10)	0.055	(9,15)	**0.034**
(2,9)	0.04	(5,11)	0.057	(10,11)	0.061
(2,10)	0.047	(5,12)	0.049	(10,12)	0.051
(2,11)	0.055	(5,13)	0.048	(10,13)	0.056
(2,12)	0.046	(5,14)	0.055	(10,14)	0.055
(2,13)	0.046	(5,15)	0.045	(10,15)	0.047
(2,14)	0.041	(6,7)	0.046	(11,12)	0.045
(2,15)	0.053	(6,8)	0.045	(11,13)	0.05
(3,4)	0.038	(6,9)	0.06	(11,14)	0.058
(3,5)	0.05	(6,10)	0.059	(11,15)	0.042
(3,6)	0.053	(6,11)	0.037	(12,13)	0.044
(3,7)	0.04	(6,12)	0.051	(12,14)	0.039
(3,8)	0.059	(6,13)	0.053	(12,15)	0.046
(3,9)	0.038	(6,14)	0.062	(13,14)	0.048
(3,10)	0.051	(6,15)	0.043	(13,15)	0.046
(3,11)	0.057	(7,8)	0.062	(14,15)	0.038

Table 5.5: Type I Error Study for $\bar{\chi}^2_{ij}$ for Symmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.054	(3,12)	0.045	(7,9)	0.04
(1,3)	0.039	(3,13)	0.05	(7,10)	0.046
(1,4)	0.04	(3,14)	0.041	(7,11)	0.052
(1,5)	0.046	(3,15)	0.041	(7,12)	0.06
(1,6)	0.047	(4,5)	0.061	(7,13)	0.042
(1,7)	0.053	(4,6)	0.044	(7,14)	**0.065**
(1,8)	0.046	(4,7)	0.058	(7,15)	0.045
(1,9)	0.049	(4,8)	0.049	(8,9)	0.051
(1,10)	0.04	(4,9)	0.048	(8,10)	0.056
(1,11)	**0.032**	(4,10)	0.057	(8,11)	0.037
(1,12)	**0.029**	(4,11)	0.056	(8,12)	0.051
(1,13)	0.051	(4,12)	0.043	(8,13)	0.043
(1,14)	**0.036**	(4,13)	0.042	(8,14)	0.046
(1,15)	0.051	(4,14)	**0.035**	(8,15)	0.048
(2,3)	0.04	(4,15)	0.044	(9,10)	0.051
(2,4)	0.042	(5,6)	0.049	(9,11)	0.043
(2,5)	0.058	(5,7)	0.046	(9,12)	0.047
(2,6)	0.061	(5,8)	0.054	(9,13)	0.046
(2,7)	0.047	(5,9)	0.051	(9,14)	0.05
(2,8)	0.051	(5,10)	0.055	(9,15)	**0.034**
(2,9)	0.04	(5,11)	0.057	(10,11)	0.061
(2,10)	0.046	(5,12)	0.05	(10,12)	0.051
(2,11)	0.056	(5,13)	0.048	(10,13)	0.056
(2,12)	0.045	(5,14)	0.058	(10,14)	0.055
(2,13)	0.044	(5,15)	0.046	(10,15)	0.047
(2,14)	0.042	(6,7)	0.047	(11,12)	0.044
(2,15)	0.053	(6,8)	0.046	(11,13)	0.05
(3,4)	0.037	(6,9)	0.061	(11,14)	0.059
(3,5)	0.049	(6,10)	0.059	(11,15)	0.042
(3,6)	0.053	(6,11)	0.037	(12,13)	0.044
(3,7)	0.04	(6,12)	0.05	(12,14)	0.039
(3,8)	0.059	(6,13)	0.052	(12,15)	0.046
(3,9)	0.038	(6,14)	0.058	(13,14)	0.05
(3,10)	0.05	(6,15)	0.045	(13,15)	0.046
(3,11)	0.057	(7,8)	0.062	(14,15)	0.039

Table 5.6: Type I Error Study for Orthogonal Components for Symmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.049049	(3,12)	0.0550551	(7,9)	0.039039
(1,3)	0.04004	(3,13)	0.04004	(7,10)	0.039039
(1,4)	0.044044	(3,14)	0.046046	(7,11)	0.0630631
(1,5)	**0.031031**	(3,15)	0.0540541	(7,12)	0.0510511
(1,6)	0.038038	(4,5)	0.0500501	(7,13)	0.049049
(1,7)	0.0520521	(4,6)	0.046046	(7,14)	0.0590591
(1,8)	0.0510511	(4,7)	0.0570571	(7,15)	0.0560561
(1,9)	0.0560561	(4,8)	0.044044	(8,9)	0.044044
(1,10)	0.043043	(4,9)	0.045045	(8,10)	0.0600601
(1,11)	**0.033033**	(4,10)	0.041041	(8,11)	0.048048
(1,12)	0.045045	(4,11)	0.0550551	(8,12)	0.0580581
(1,13)	0.0500501	(4,12)	0.041041	(8,13)	0.041041
(1,14)	0.0510511	(4,13)	0.048048	(8,14)	0.041041
(1,15)	0.044044	(4,14)	0.047047	(8,15)	0.0570571
(2,3)	0.042042	(4,15)	0.038038	(9,10)	0.0550551
(2,4)	0.042042	(5,6)	0.047047	(9,11)	0.0500501
(2,5)	0.039039	(5,7)	**0.0650651**	(9,12)	0.0500501
(2,6)	0.0570571	(5,8)	0.037037	(9,13)	0.039039
(2,7)	0.0500501	(5,9)	0.0530531	(9,14)	0.041041
(2,8)	0.0630631	(5,10)	0.049049	(9,15)	0.0590591
(2,9)	0.042042	(5,11)	0.049049	(10,11)	0.044044
(2,10)	0.0530531	(5,12)	0.046046	(10,12)	0.046046
(2,11)	0.048048	(5,13)	0.044044	(10,13)	0.049049
(2,12)	0.045045	(5,14)	0.0510511	(10,14)	0.049049
(2,13)	0.037037	(5,15)	0.0550551	(10,15)	0.046046
(2,14)	0.038038	(6,7)	0.04004	(11,12)	0.0540541
(2,15)	0.0570571	(6,8)	0.047047	(11,13)	0.049049
(3,4)	0.037037	(6,9)	0.048048	(11,14)	0.039039
(3,5)	0.048048	(6,10)	0.0560561	(11,15)	0.045045
(3,6)	0.042042	(6,11)	0.049049	(12,13)	0.043043
(3,7)	0.043043	(6,12)	0.049049	(12,14)	0.0600601
(3,8)	0.0520521	(6,13)	0.0520521	(12,15)	0.0510511
(3,9)	**0.032032**	(6,14)	0.0500501	(13,14)	0.0520521
(3,10)	0.0550551	(6,15)	0.0560561	(13,15)	0.047047
(3,11)	**0.032032**	(7,8)	0.0550551	(14,15)	0.038038

Table 5.7: Type I Error Study for Adjusted Residuals for Symmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.0500501	(3,12)	0.0520521	(7,9)	0.042042
(1,3)	0.041041	(3,13)	0.048048	(7,10)	0.039039
(1,4)	0.044044	(3,14)	**0.031031**	(7,11)	0.0600601
(1,5)	**0.033033**	(3,15)	0.045045	(7,12)	0.0620621
(1,6)	0.046046	(4,5)	0.0530531	(7,13)	0.038038
(1,7)	0.0500501	(4,6)	0.0600601	(7,14)	0.0530531
(1,8)	0.0570571	(4,7)	0.0560561	(7,15)	0.0550551
(1,9)	0.0510511	(4,8)	0.046046	(8,9)	0.048048
(1,10)	0.044044	(4,9)	0.047047	(8,10)	0.0570571
(1,11)	**0.034034**	(4,10)	0.041041	(8,11)	0.049049
(1,12)	0.042042	(4,11)	0.0600601	(8,12)	0.0570571
(1,13)	0.041041	(4,12)	0.0520521	(8,13)	0.049049
(1,14)	0.037037	(4,13)	0.042042	(8,14)	0.044044
(1,15)	0.044044	(4,14)	0.047047	(8,15)	0.049049
(2,3)	0.042042	(4,15)	0.039039	(9,10)	0.0550551
(2,4)	0.043043	(5,6)	0.038038	(9,11)	0.0500501
(2,5)	0.037037	(5,7)	0.0590591	(9,12)	0.0520521
(2,6)	0.0500501	(5,8)	0.042042	(9,13)	0.038038
(2,7)	0.0500501	(5,9)	0.0520521	(9,14)	0.043043
(2,8)	0.0630631	(5,10)	0.0570571	(9,15)	0.041041
(2,9)	0.04004	(5,11)	0.048048	(10,11)	0.047047
(2,10)	0.0510511	(5,12)	0.047047	(10,12)	0.048048
(2,11)	0.046046	(5,13)	0.043043	(10,13)	0.0560561
(2,12)	0.039039	(5,14)	0.042042	(10,14)	0.041041
(2,13)	**0.032032**	(5,15)	0.049049	(10,15)	**0.033033**
(2,14)	0.038038	(6,7)	0.042042	(11,12)	0.0510511
(2,15)	0.0530531	(6,8)	0.044044	(11,13)	0.0570571
(3,4)	0.041041	(6,9)	0.046046	(11,14)	0.0530531
(3,5)	0.043043	(6,10)	0.049049	(11,15)	**0.035035**
(3,6)	0.046046	(6,11)	0.0520521	(12,13)	0.041041
(3,7)	0.044044	(6,12)	0.0560561	(12,14)	0.0550551
(3,8)	0.046046	(6,13)	**0.0680681**	(12,15)	0.046046
(3,9)	**0.031031**	(6,14)	0.045045	(13,14)	0.045045
(3,10)	0.0510511	(6,15)	0.0510511	(13,15)	**0.036036**
(3,11)	**0.035035**	(7,8)	0.046046	(14,15)	**0.028028**

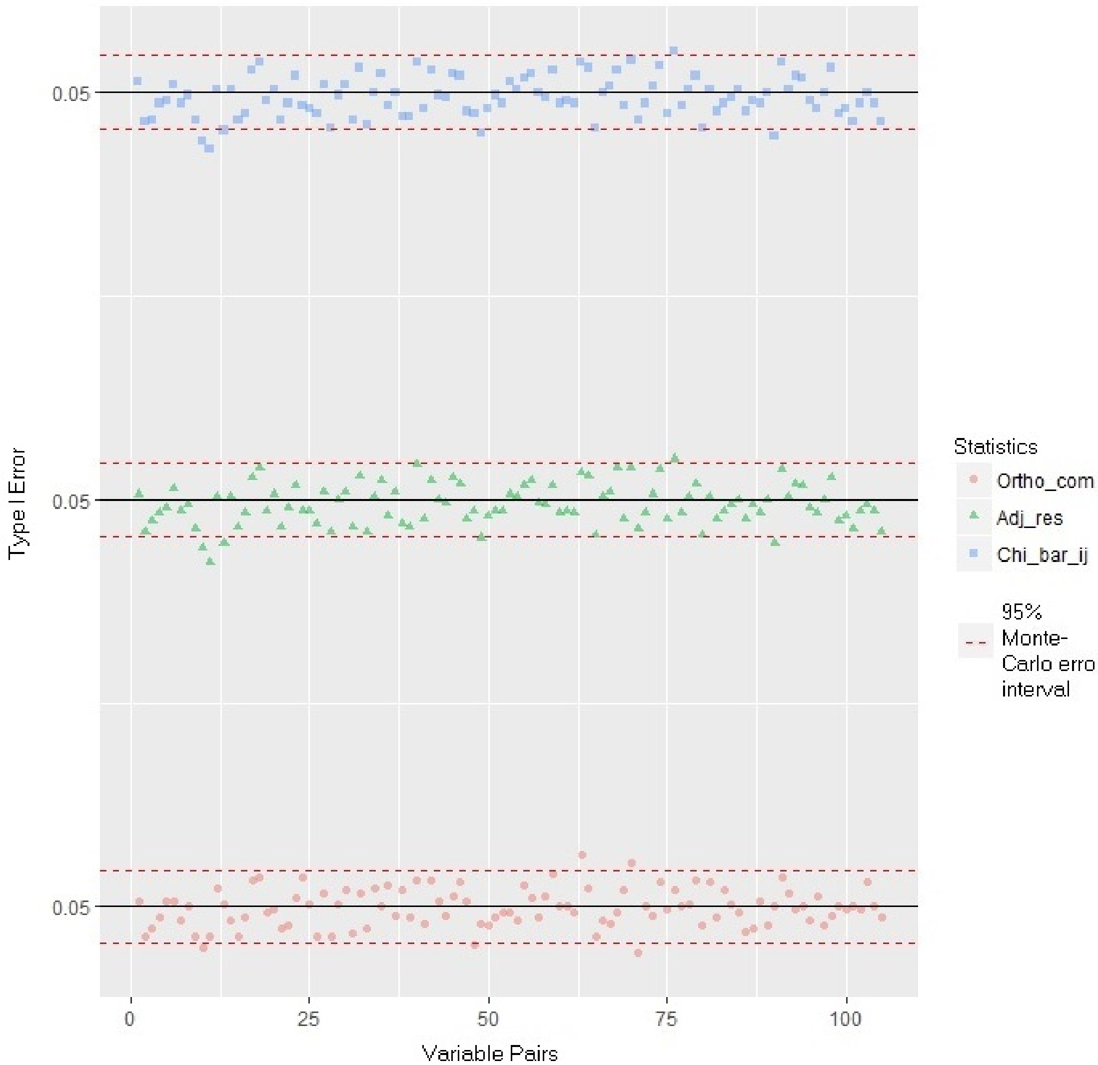
Table 5.8: Type I Error Study for $\bar{\chi}^2_{ij}$ for Symmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.049	(3,12)	0.048	(7,9)	0.041
(1,3)	0.04	(3,13)	0.049	(7,10)	0.039
(1,4)	0.044	(3,14)	**0.029**	(7,11)	0.061
(1,5)	**0.031**	(3,15)	0.045	(7,12)	0.06
(1,6)	0.043	(4,5)	0.055	(7,13)	0.038
(1,7)	0.052	(4,6)	0.059	(7,14)	0.054
(1,8)	0.056	(4,7)	0.054	(7,15)	0.054
(1,9)	0.052	(4,8)	0.047	(8,9)	0.048
(1,10)	0.045	(4,9)	0.048	(8,10)	0.057
(1,11)	**0.033**	(4,10)	0.042	(8,11)	0.049
(1,12)	0.039	(4,11)	0.059	(8,12)	0.058
(1,13)	0.038	(4,12)	0.051	(8,13)	0.049
(1,14)	**0.033**	(4,13)	0.039	(8,14)	0.044
(1,15)	0.044	(4,14)	0.047	(8,15)	0.049
(2,3)	0.042	(4,15)	0.038	(9,10)	0.054
(2,4)	0.041	(5,6)	0.041	(9,11)	0.05
(2,5)	**0.036**	(5,7)	0.058	(9,12)	0.051
(2,6)	0.048	(5,8)	0.042	(9,13)	0.039
(2,7)	0.048	(5,9)	0.049	(9,14)	0.041
(2,8)	0.062	(5,10)	0.057	(9,15)	0.041
(2,9)	0.04	(5,11)	0.048	(10,11)	0.047
(2,10)	0.051	(5,12)	0.048	(10,12)	0.048
(2,11)	0.047	(5,13)	0.039	(10,13)	0.058
(2,12)	**0.035**	(5,14)	0.046	(10,14)	0.041
(2,13)	**0.03**	(5,15)	0.05	(10,15)	**0.033**
(2,14)	**0.035**	(6,7)	0.045	(11,12)	0.051
(2,15)	0.052	(6,8)	0.045	(11,13)	0.057
(3,4)	0.038	(6,9)	0.047	(11,14)	0.053
(3,5)	0.042	(6,10)	0.051	(11,15)	**0.035**
(3,6)	0.043	(6,11)	0.052	(12,13)	0.044
(3,7)	0.044	(6,12)	0.054	(12,14)	0.059
(3,8)	0.046	(6,13)	**0.07**	(12,15)	0.046
(3,9)	**0.031**	(6,14)	0.044	(13,14)	0.047
(3,10)	0.052	(6,15)	0.052	(13,15)	0.037
(3,11)	**0.035**	(7,8)	0.046	(14,15)	**0.028**

Type I error rates for $q = 15$ manifest variables for n=300 and n=500 for zero intercept model are given in Tables A.9 through A.14. Graphical illustration of the same information are given in Figures 5.9 and 5.10 below. When n=500, two components related to adjusted residuals, orthogonal components and $\bar{\chi}^2_{ij}$ were outside the Monte-Carlo error interval. Given that there are 105 second-order marginals, it is possible one or two will randomly fall slightly outside the Monte-Carlo error interval. However, when n=300, three components related to orthogonal components, eight components related to adjusted residuals, and seven components related to $\bar{\chi}^2_{ij}$ were outside the Monte-Carlo error interval. This suggests, when n=300, orthogonal components have better Type I error rates compared to $\bar{\chi}^2_{ij}$ and adjusted residuals for $q = 15$ manifest variables for zero intercept model.

In this section I presented the simulation results related to Type I error performance of individual orthogonal components. I have compared the results to adjusted residuals and $\bar{\chi}^2_{ij}$. I used two settings for the number of manifest variables: $q = 8$ and $q = 15$ and three settings for the intercepts of the model: symmetric, asymmetric and zero. Based on the results it is clear that the orthogonal components have better performance compared to adjusted residuals and $\bar{\chi}^2_{ij}$ when $n = 300$. However, when the sample size increases (e.g. n=500), Type I error rate performance were similar between orthogonal components, adjusted residuals and $\bar{\chi}^2_{ij}$, especially for the zero intercept setting. However, this was not the case for symmetric and asymmetric intercept cases, especially when the $q = 15$. Thus, observations seems to be well distributed among cells when the intercepts are zero compared to asymmetric or symmetric intercepts. Overall, individual orthogonal components had better Type I error rates even when the cross-classified table was very sparse. Next, I compared empirical and asymptotic power of orthogonal components under these different sparse conditions.

Figure 5.7: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Symmetric Intercept Model, q=15, n=500



* Index 1 thought 105 in the x-axis of the above plot is correspond to the variable pairs (1,2), (1,3),......, (14,15), respectively. Note, there are $15 * 14/2 = 105$ variable pairs.

Figure 5.8: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Symmetric Intercept Model, q=15, n=300

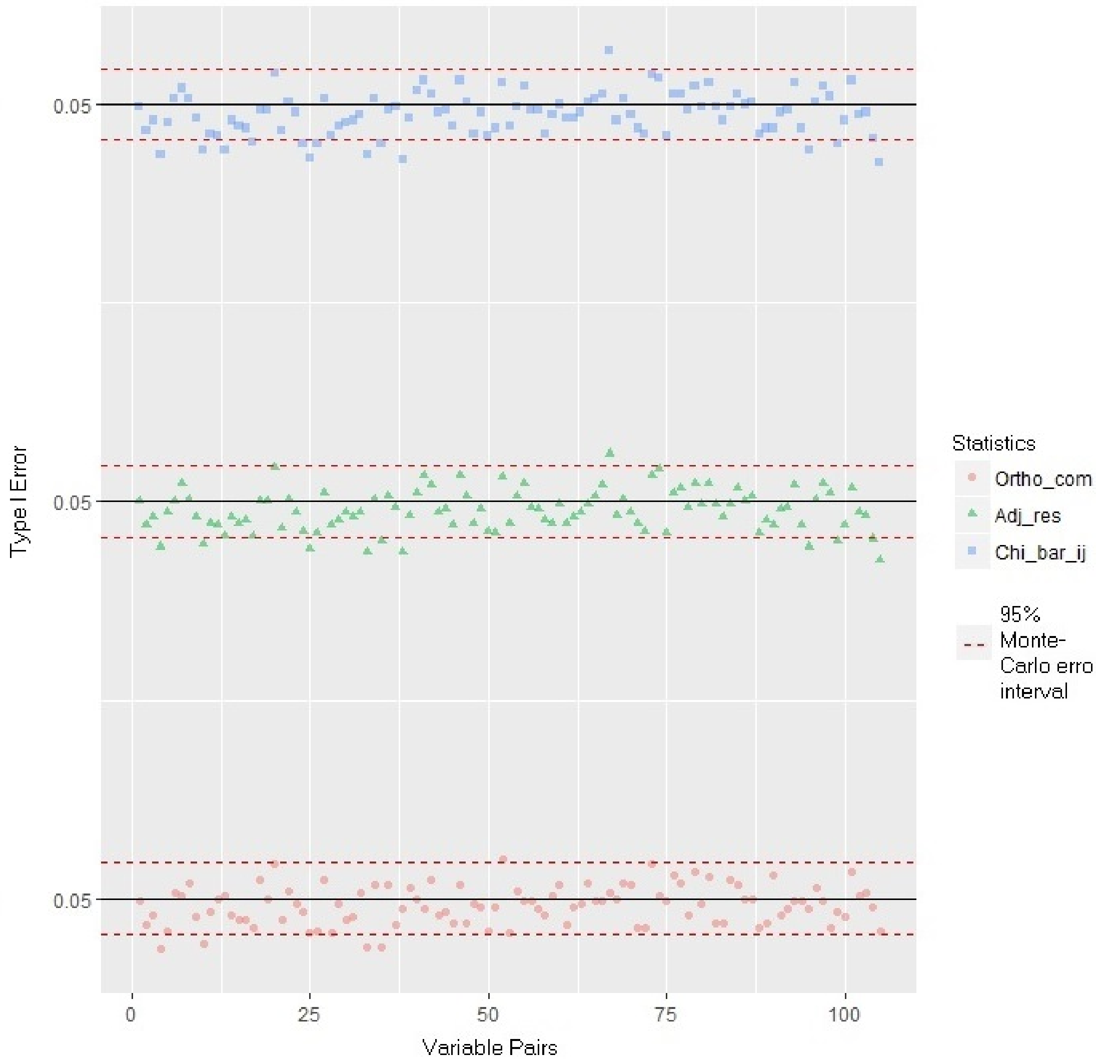


Figure 5.9: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Zero Intercept Model, q=15, n=500

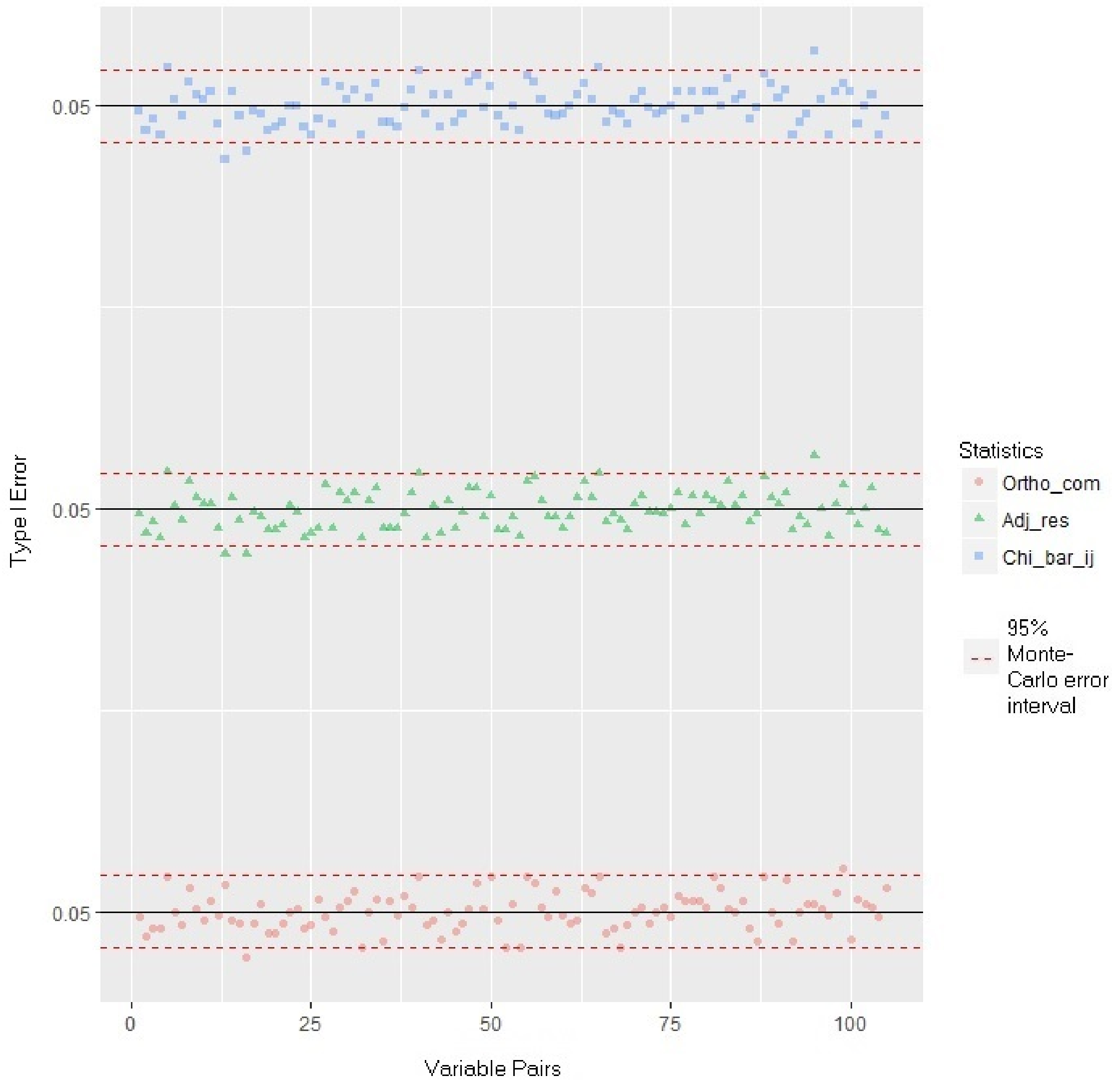
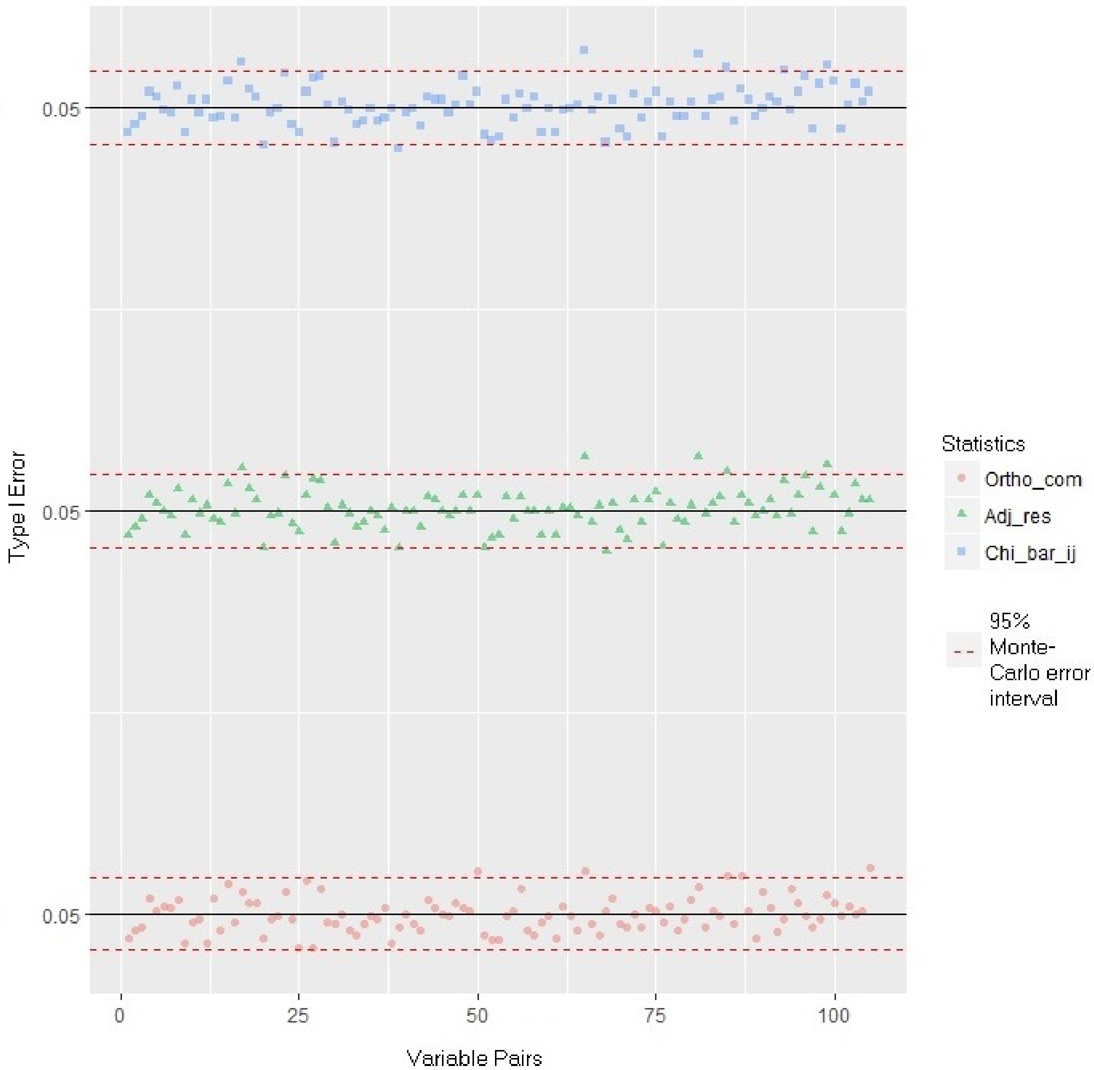


Figure 5.10: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Zero Intercept Model, q=15, n=300



5.2 Simulation Study Part II: Comparison of Empirical Power and Asymptotic Power of Individual Orthogonal Components of $\chi^2_{[2]}$

As the next approach to study empirical power of the statistics, 1000 data sets were generated using Monte-Carlo methods related to a two factor (two latent variables) model. A one factor model was fitted for each of these datasets and empirical power was calculated. In the simulation, the model under H_0 is misspecified with a one factor model. Since each orthogonal component is distributed approximately as chi-square with one degree of freedom, to calculate the empirical power 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 to calculate the empirical power of adjusted residual and $\bar{\chi}^2_{ij}$. This simulation was repeated for sample size 300 and 500.

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 approximately 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 investigate this idea of detecting item mis-fit, higher factor loadings were assigned to item 4, 5 and 6 of the data generation model. Loadings for the first factor were kept at (0.1, 0.1, 0.1, 1.2 , 1.2 , 1.2 , 0.2 , 0.2) where items 4,5 and 6 have higher factor loadings compared to other items. Loadings for the second factor were kept at (1, 1, 1, 0, 0, 0, 1, 1). Since the mis-fit is related to items 4, 5 and 6, it was expected to have higher empirical power for the components related to these second-order marginals.

As before, I used three intercept settings: symmetric, asymmetric, zero. Table 4.1 and 4.2 summarize these information. Reasons for using these different intercepts settings were explained in the Section 4.1. The design of power study is as follows:

Table 5.9: Design of Power Study

Model (data generation)	categorical variable factor analysis model with two latent factors
Model (fitted)	categorical variable factor analysis model with one latent factor
Number of observed variables	q=8, q=15
Number of samples	1000
Sample size	n=300, n=500

I also compared the empirical power to asymptotic power for individual orthogonal components of $\chi^2_{[2]}$, adjusted residuals and $\bar{\chi}^2_{ij}$. Calculation for the asymptotic power was performed with the same parameter values as in the empirical power simulation. To calculate the asymptotic power I used the method described in the Section 3.3. First, I generated the proportions from two factor categorical factor model with above mentioned factor loadings. A numerical integration method called Gauss-Hermite quadrature was used to generate the proportions. Thereafter, these proportions were multiplied by a selected initial sample size n_0 to create the true cell frequencies under H_a. Then, the model of the null hypothesis was fitted to the resulting cell frequencies. Next, the non-centrality parameters were calculated as described in equations 3.17 and 3.18. The non-centrality parameters for any other sample size, say simply n, can be approximated by using the expression $\lambda \approx \frac{n}{n_0}\lambda_0$. Once I obtained the non-centrality parameters, I used non-central chi-square distribution with one degree of freedom to calculate the asymptotic power for each orthogonal component of $\chi^2_{[2]}$. The significance level, α was set to 0.05.

5.2.1 Power Study for Eight Variables

Asymptotic and empirical power comparison for symmetric, zero and asymmetric intercept models are given in the Tables 5.10, A.15 and A.16, respectively. I have allocated higher values to items 4, 5, and 6 on a second factor and I'm expecting higher power for components related to those item pairs. By examining the highlighted values in Table 5.10, A.15 and A.16, it is clear that the empirical power of second order marginal components (4,5), (4,6) and (5,6) are significantly higher compared to other components. Thus, these second order components were successful in detecting the source of a poorly fit model. This process was repeated for n=300 and n=500. By the results in these tables, it is clear that the empirical power will increase with the sample size and the components were more successful in detecting the lack-of-fit for larger sample sizes. However, when n=300, empirical power results were somewhat lower compared to asymptotic power results. This indicates when sample size is smaller empirical distribution may not close to the hypothesized theoretical distribution. When n=500, empirical power results and asymptotic power results were fairly close. This indicates when sample size increases the empirical distribution approaches hypothesized theoretical distribution. Also, zero intercept model had higher power results compared to models with symmetric and asymmetric intercept settings, and the empirical power results were more close to asymptotic power results too. Note, zero intercept model had better Type I error rates compared to models with symmetric and asymmetric intercept settings. Thus, observations seems to be well distributed among cells when the intercepts are zero compared to asymmetric or symmetric intercepts. I think this is the reason behind the better power results under zero intercept model.

Table 5.10: Asymptotic and Empirical Power Comparison for Symmetric Intercept Model

	n=300				n=500			
Pair (i,j)	Orth. Comp.	Adj. Res.	$\bar{\chi}_{ij}^2$	Asym. Power	Orth. Comp.	Adj. Res.	$\bar{\chi}_{ij}^2$	Asym. Power
(1,2)	0.1773859	0.189834	0.179	0.0615	0.12651	0.1455823	0.134	0.06924
(1,3)	0.1960581	0.2022822	0.189	0.07908	0.13956	0.1425703	0.134	0.0989
(1,4)	0.0622407	0.0985477	0.091	0.08634	0.069277	0.0933735	0.097	0.1112
(1,5)	0.0954357	0.1151452	0.11	0.09531	0.10944	0.1054217	0.105	0.12641
(1,6)	0.1462656	0.1058091	0.1	0.1035	0.21285	0.1174699	0.116	0.1403
(1,7)	0.0798755	0.1224066	0.119	0.05	0.075301	0.1586345	0.153	0.05
(1,8)	0.0798755	0.1026971	0.088	0.05003	0.086345	0.1716867	0.169	0.05005
(2,3)	0.2417012	0.2251037	0.218	0.12393	0.17169	0.1706827	0.159	0.17493
(2,4)	0.0736515	0.1037344	0.096	0.12207	0.10241	0.1134538	0.116	0.17178
(2,5)	0.0871369	0.1016598	0.096	0.13787	0.15261	0.126506	0.129	0.19847
(2,6)	0.2095436	0.0954357	0.092	0.15344	0.2741	0.1315261	0.13	0.22461
(2,7)	0.0757261	0.159751	0.154	0.05003	0.080321	0.1817269	0.175	0.05005
(2,8)	0.0684647	0.129668	0.127	0.05089	0.089357	0.1726908	0.168	0.05149
(3,4)	0.1026971	0.0985477	0.098	0.2381	0.19478	0.12249	0.131	0.36253
(3,5)	0.1618257	0.1141079	0.111	0.27698	0.21285	0.12249	0.125	0.42264
(3,6)	0.3246888	0.1120332	0.103	0.31787	0.36145	0.1405622	0.143	0.48323
(3,7)	0.1058091	0.1919087	0.19	0.05702	0.10643	0.186747	0.178	0.06173
(3,8)	0.0798755	0.1358921	0.139	0.05032	0.091365	0.189759	0.188	0.05054
(4,5)	0.530083	0.6659751	0.647	0.80393	0.77811	0.8684739	0.855	0.95304
(4,6)	0.530083	0.6317427	0.615	0.82562	0.80924	0.8865462	0.876	0.96246
(4,7)	0.0466805	0.0819502	0.082	0.05087	0.031124	0.0983936	0.105	0.05146
(4,8)	0.0497925	0.0788382	0.074	0.05002	0.044177	0.1174699	0.123	0.05004
(5,6)	0.6732365	0.6410788	0.619	0.91976	0.9508	0.8714859	0.868	0.9914
(5,7)	0.0363071	0.1016598	0.097	0.05	0.03012	0.1024096	0.103	0.05
(5,8)	0.0881743	0.0829876	0.081	0.05026	0.066265	0.1074297	0.115	0.05043
(6,7)	0.0705394	0.0840249	0.08	0.05007	0.057229	0.1004016	0.104	0.05012
(6,8)	0.0684647	0.0829876	0.081	0.05031	0.070281	0.1144578	0.118	0.05052
(7,8)	0.0809129	0.1721992	0.169	0.05019	0.083333	0.1997992	0.198	0.05032

* Asymptotic power was calculated only for the orthogonal components.

5.2.2 Power Study for Fifteen Variables

Next, I increased the number of manifest variables to fifteen and carried out the same power calculations as in the eight variable study. I used a repetition of the slope parameters of two factor model for eight manifest variables as the slope parameters for the two factor model with fifteen manifest variables. So the factor loadings for the first factor was set to (0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2, 0.2, 0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2) and the factor loadings for the second factor was set to (1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1). The idea was to make the comparison between an eight variable study and a fifteen variable study more meaningful.

In the section 5.1.2, Type I error rates related to orthogonal components, adjusted residual and $\bar{\chi}^2_{ij}$ were studied for models with fifteen manifest variables under three different intercept settings. According to the results, the zero intercept model had better Type I error rates compared to asymmetric and symmetric intercept settings. However, a symmetric intercept model is more applicable in a real-world application rather than a zero intercept model. Therefore, I extended fifteen variable power study to both symmetric and zero intercept settings.

Asymptotic and empirical power comparison for symmetric and zero intercept models for q=15 are given in the Tables A.17 and A.18, respectively. Graphical illustration of the same information are given in Figures 5.11 and 5.12 below. I have allocated higher weights to items 4, 5, 6, 12, 13 and 14 and I'm expecting higher power for components related to these item pairs. By examining the highlighted values in Tables A.17 and A.18, it is clear that the empirical power of second order marginal components (4,5), (4,6), (4,12), (4,13), (4,14), (5,6), (5,12), (5,13), (5,14), (6,12), (6,13), (6,14), (12,13), (12,14) and (13,14) are significantly higher compared to other components. Thus, orthogonal components related to second were successful in detecting item lack-of-fit even when the cross-classified table was very sparse. Looking at the results in Figures 5.11 and 5.12 it is clear that the Empirical power results are also close to the asymptotic power results.

Figure 5.11: Power Comparison of Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Symmetric Intercept Model, q=15, n=500

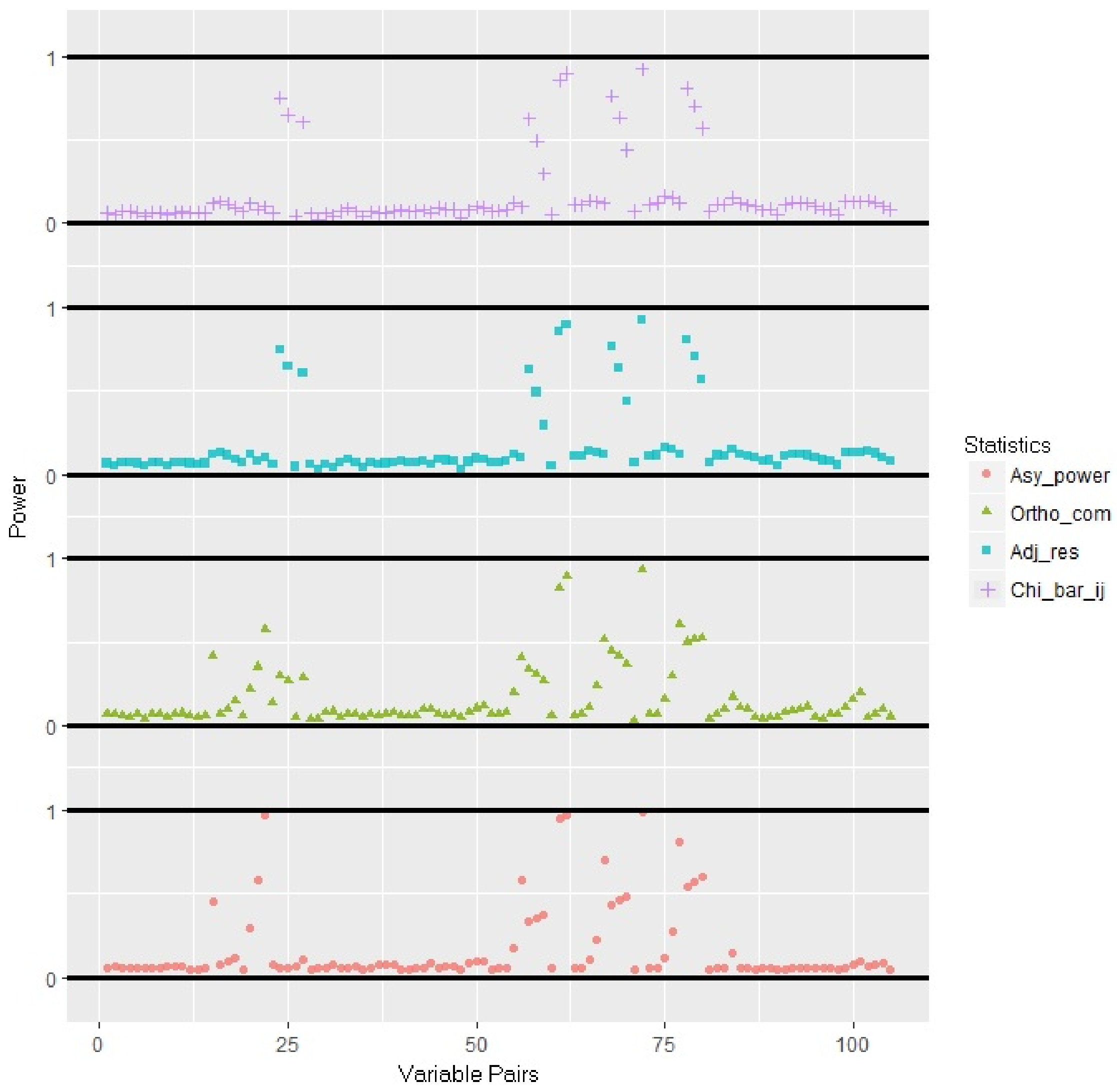
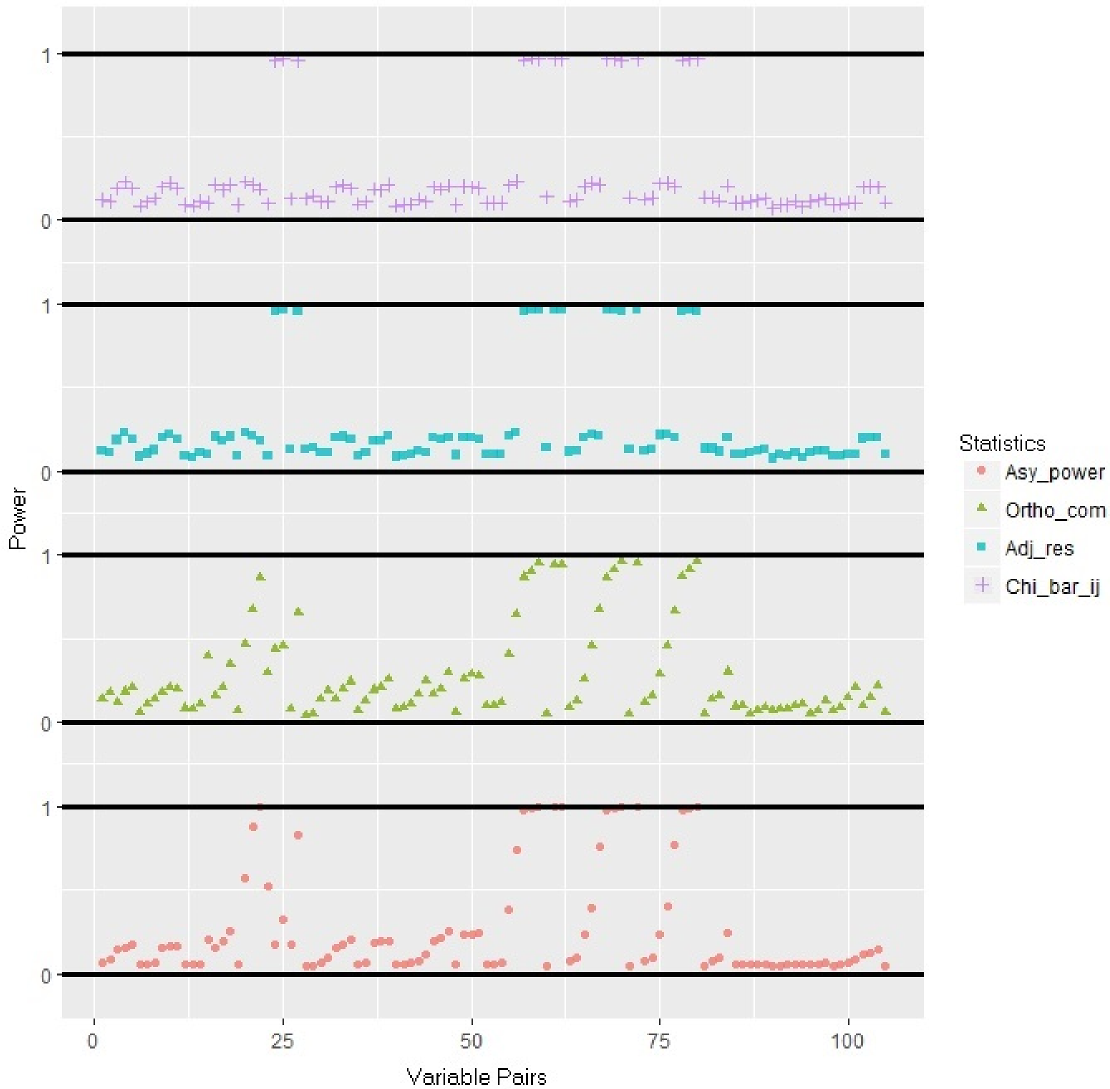


Figure 5.12: Power Comparison of Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Zero Intercept Model, q=15, n=500



In this section, second order marginals related to orthogonal components were examined as lack-of-fit diagnostics. Simulations were based on a two factor 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. The Asymptotic power results tally with empirical power results but the empirical power for orthogonal components was much closer to the asymptotic power when intercepts are zero compared to symmetric or asymmetric intercept settings. This shows that when the $2 * 2$ tables are balanced and when there is less bias in parameter estimation, the empirical power is close to the asymptotic power. Looking at the power results for both q=8 and q=15, there is very little change in asymptotic power between zero, symmetric and asymmetric intercept settings. This is because the asymptotic power depends on the slopes and not the intercepts. However, it seems the empirical power is affected by sparseness and bias of estimator for intercept and slope.

Chapter 6

REAL WORLD APPLICATIONS

Proposed limited-information test statistics based on orthogonal components defined on marginal frequencies in this research were applied to two real-life data sets. The main focus of these applications were to assess how well the proposed statistics perform with respect to detecting the lack-of-fit when model under the null hypothesis is rejected.

6.1 Application I - Data on Mental Disorder Phobia

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, eight items related to the mental disorder phobia were chosen from the ECA to analyze as a real world application. The dataset was limited to Johns Hopkins (Baltimore, MD) area.

The selected items are given below.

1) DIS068A - fear of heights

2) DIS068F - fear of closed places

3) DIS068I - fear of speaking in front of close friends

4) DIS068J - fear of speaking to strangers

5) DIS068K - storms

6) DIS068L - water

7) DIS068M - spiders

8) DIS068N - fear of harmless animals

There were 3316 observations related to these specifications. Each variable has two categories: 'yes' or 'no'. Thus, there are $2^8 = 256$ response patterns. However, as most of the answers are 'no', 165 response patterns are empty. Furthermore, many response patterns have a cell count less than five. The detailed cell counts are given in Table 6.1.

Table 6.1: Number of Response Patterns with Small Frequencies

Cell Count	Number of Response Patterns	Number of Cases
0	165	0
1	39	39
2	10	20
3	11	33
4	5	20
5	4	20
> 5	22	3184
Total	256	3316

A categorical variable factor analysis model with one latent factor was fitted to the data. The statistics χ^2_{PF}, χ^2_{red}, M_2, $\chi^2_{[2]}$, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ and the p-values are shown in the Table 6.2 below. All statistics are large and the p-values are almost 0. This indicates that the one factor model is not a good fit to the data.

Table 6.2: Goodness-of-Fit Tests for ECA Phobia Study

Statistic	Value	DF	P-value
χ^2_{PF}	488.95	239	0
χ^2_{red}	77.12	20	1.20E-08
M_2	71.99	20	8.59E-08
$\chi^2_{[2]}$	108.53	28	2.01E-11
$\chi^2_{[2:3]}$	192.9	84	1.51E-10
$\chi^2_{[2:4]}$	304.57	154	5.82E-12

Since most response patterns have a cell count less than five, it is possible that the overall table is sparse. If the overall table is sparse then χ^2_{PF} may not be valid. However, $2 * 2$ tables may not be sparse. Next, I have used individual orthogonal components of $\chi^2_{[2]}$ as test statistics to identify lack-of-fit.

When the number of variables is large, a very large number of components is produced, and a multiple decision rule should be used to determine which components are significantly large relative to the reference chi-square distribution. With a large number of variables, the traditional Bonferroni method becomes very conservative. Because the orthogonal components are independent random variates, it is possible to take advantage of the False Discovery Rate (FDR) procedure for independent tests (Benjamini & Hochberg, 1995). Under the FDR procedure for independent tests, consider testing $H_o : \gamma_j^2 = 0$ for m orthogonal components, so testing $H_1,\ H_2,\ \cdots H_m$ based on the corresponding p-values $p_1,\ p_2\ \cdots p_m$. Let $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$, be the ordered p-values, and denote by $H_{(j)}$ the null hypothesis corresponding to $p_{(j)}$. For this Bonferroni-type procedure with α^* = the false discovery rate, let k be the largest j for which

$$p_{(j)} \leq \frac{j}{m}\alpha^*,$$

and then reject all $H_{(j)}$ for $j = 1,\ 2,\ \cdots,\ k$.

The orthogonal components for second-order marginals are shown in Table 6.3 along with the raw p-values and the adaptive FDR p-values. According to the Table 6.3, components (1,8), (3,4), (3,7), (3,8) and (6,7) related to Catchment Area study have significant FDR p-values indicating that these pairs of variables have associations not explain by the one factor. Results related to orthogonal components, adjusted residual and $\bar{\chi}^2_{ij}$ are consistent with each other. Further, variable 3, 'fear of speaking in front of close friends' appears in three of these large components and variable 7, 'spiders' and variable 8, 'fear of harmless animals' appears in two of these large components. It is important to further investigate the associations between these variable-pairs. Since one factor model did not fit well, I recommend to consider other models such as, a two-factor model or a log-linear model.

Table 6.3: Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$ for ECA Study

Pair (i,j)	Orthogonal Component	Standardized Residual	$\bar{\chi}^2_{ij}$	Raw P-value	A FDR P-value
(1,2)	1.6376	1.1533	1.2858	0.200655076	0.423424734
(1,3)	1.5041	-1.3505	1.905	0.220041586	0.423424734
(1,4)	3.7028	-2.2899	5.0195	0.054321238	0.176143041
(1,5)	0.277	0.658	0.2935	0.598674299	0.698453349
(1,6)	0.7687	0.6177	0.2562	0.380619518	0.560912974
(1,7)	1.2047	-1.2205	3.8028	0.272384319	0.438190337
(1,8)	11.8253	-3.3281	11.1323	0.000584313	0.008180382
(2,3)	0.0564	-0.5664	0.2788	0.812279126	0.842363538
(2,4)	0.5087	-1.9678	3.5069	0.475702158	0.614440508
(2,5)	0.1727	-0.9056	0.9037	0.677723296	0.731854435
(2,6)	3.6338	-0.1099	0.0376	0.056617406	0.176143041
(2,7)	0.5689	-2.7526	9.1937	0.450696358	0.614440508
(2,8)	0.445	-1.3783	1.7317	0.504718989	0.614440508
(3,4)	32.1095	4.7529	24.2571	1.45723E-08	4.08E-07
(3,5)	1.4606	-2.3876	5.7302	0.226834679	0.423424734
(3,6)	0.0005	-1.5024	2.2625	0.982160245	0.982160245
(3,7)	10.9848	-3.5433	13.5719	0.000918622	0.008573805
(3,8)	10.0776	-1.8442	3.2138	0.001500827	0.010505789
(4,5)	1.8726	-3.2382	10.168	0.171177595	0.399414388
(4,6)	1.2812	-1.5231	2.104	0.257676035	0.438190337
(4,7)	2.706	-2.5367	6.6029	0.099971378	0.279844366
(4,8)	2.5551	-1.7898	2.8238	0.109938858	0.279844366
(5,6)	3.9108	-1.5036	2.7344	0.047976754	0.176143041
(5,7)	0.1706	0.6859	0.5016	0.679579118	0.731854435
(5,8)	4.4371	0.2215	0.1152	0.035165933	0.164107687
(6,7)	9.0146	-2.8555	11.4293	0.002678315	0.014998564
(6,8)	0.4754	-2.1014	4.3357	0.490513349	0.614440508
(7,8)	1.1589	2.3356	6.1359	0.281693788	0.438190337

6.2 Application II - Data on Mental Depression

To further demonstrate the use of orthogonal components for lack-of-fit diagnosis, a one-factor model with slope constrained equal was fit to responses given to 20 questions about the psychiatric condition of mental depression.

The responses to the questions were collected as part of the Epidemiological Catchment Area Study (ECA) of 1980-1985. More information about the ECA was given in the previous section. The Baltimore sample included 3,481 adults sampled from the Baltimore catchment area. The data used in this example consists of the responses from 3,187 adults who had complete data records for responses to the 20 questions. Missing data are assumed to be missing completely at random. The depression symptoms included in the survey are shown in Table 6.5. The responses were coded into two categories: (1) symptom present at a clinical level and (2) symptom not present. Even with sample size 3,187, the 2^{20} cross-classified table is very sparse with at least 1,045,389 cells that have count equal to zero.

Goodness-of-fit test results are shown in Table 6.4. The results indicate that the model of one underlying factor does not fit well for the depression symptoms. The chi-square approximation for the full Pearson statistic should not be considered valid because of the high degree of sparseness in the data table. The statistic $X^2_{[2]}$, as well as M_2, and X^2_{red}, indicate that the model should be rejected.

The 30 largest orthogonal components for second-order marginals are shown in Table 6.6 along with the raw p-values and the adaptive FDR p-values. More details about the FDR procedure is given in the previous section. Applying the FDR method to the 190 orthogonal components obtained from the example data set yields over 30 null hypotheses $H_o : \gamma_j^2 = 0$ rejected. The largest orthogonal component for this application is found for the association between variables 3 and 4 which are questions about loss of appetite and loss of weight. The second-largest component is found for the association between variables 4 and 5, and question 5 asks about gain of weight. The large magnitude of these components indicates that the somatic symptoms indicate an additional dimension of depression in addition to the affect dimension. Another large component is found for variables 17 and 18, thought of suicide and attempted suicide. These two variables have a higher association than can be explained by a single latent variable. Therefore, I recommend to consider other models such as, a two-factor model or a log-linear model.

Table 6.4: Goodness-of-Fit Tests for ECA Depression Symptoms

	Value	DF	p-value
X^2_{PF}	1455338.69	1048535.00	.
X^2_{red}	1302.85	170	< 0.0001
M_2	1293.25	170	< 0.0001
$X^2_{[2]}$	1337.66	190	< 0.0001

Table 6.5: ECA Depression Symptoms

Item	Description	Marginal Frequency	Percent
1	Two weeks dysphoria in lifetime	939	26.98
2	Two years or more of dysphoria	217	6.23
3	Lost appetite for two weeks	295	8.47
4	Loss of weight	332	9.54
5	Gain weight	485	13.93
6	Insomnia for two weeks	525	15.08
7	Sleep too much	228	6.55
8	Felt tired for two weeks	445	12.78
9	Moved slowly for two weeks	165	4.74
10	Moving all the time, two weeks	194	5.57
11	Lost interest in sex, two weeks	224	6.43
12	Felt worthless, two weeks	257	7.38
13	Trouble concentrating, two weeks	279	8.01
14	Slow thinking, two weeks	244	7.01
15	Thought of death, two weeks	729	20.94
16	Want to die, two weeks	230	6.61
17	Thought suicide, two weeks	266	7.64
18	Attempt suicide	115	3.30
19	Headaches	343	9.85
20	Crying spells	529	15.20

Table 6.6: Largest Second-Order Components for ECA Depression Symptoms

Obs	Var 1	Var 2	Component	Raw P-value	A FDR P-value
1	18	17	171.083	< 2.4425E-15	< 2.4425E-15
2	17	16	106.425	< 2.4425E-15	< 2.4425E-15
3	4	3	62.686	2.4425E-15	9.6885E-14
4	14	13	52.814	3.666E-13	1.0906E-11
5	16	15	44.876	2.0991E-11	4.9959E-10
6	9	8	41.931	9.4541E-11	.000000002
7	20	14	36.121	.000000002	.000000028
8	20	17	36.108	.000000002	.000000028
9	20	9	32.463	.000000012	.000000161
10	20	13	27.873	.000000130	.000001541
11	18	14	27.657	.000000145	.000001567
12	8	7	27.036	.000000200	.000001981
13	18	13	24.959	.000000586	.000005362
14	15	1	23.773	.000001084	.000009213
15	20	11	22.512	.000002088	.000016568
16	18	11	21.690	.000003204	.000023832
17	16	12	19.125	.000012242	.000085695
18	18	8	17.638	.000026726	.000176687
19	17	9	15.934	.000065582	.000410748
20	17	13	14.632	.000130704	.000777690
21	18	4	14.053	.000177766	.001007339
22	15	9	12.945	.000320822	.001735358
23	20	10	12.736	.000358723	.001856000
24	16	1	12.105	.000502862	.002493359
25	17	15	11.644	.000643941	.003065159
26	2	1	11.544	.000679808	.003111429
27	20	8	10.872	.000976232	.004302653
28	12	1	10.720	.001059659	.004503551
29	18	9	10.617	.001120497	.004597902
30	17	8	10.310	.001322970	.005246824

Chapter 7

GOODNESS-OF-FIT STATISTICS WHEN THE NUMBER OF VARIABLES IS LARGE

7.1 Feasibility of χ^2_{red} Statistic When the Number of Manifest Variables is Large

Some popular limited-information statistics have been discussed in Section 2.6. However, when the manifest variables exceed 20, most of these statistics will become difficult or impossible to calculate due to computer resource limitations. Among these statistics, calculation of χ^2_{Ch} is fairly straightforward since the covariance matrix, $\Sigma_{\bar{\mathbf{r}}} = D(\mathbf{p}) - \mathbf{p}\mathbf{p}'$ can be calculated from the observed counts or proportions. Simulations reported by Reiser and VandenBerg (1994) show that chi-square approximation for the distribution of χ^2_{Ch} is valid only up to 8 to 10 variables for typical sample sizes. For larger number of variables the data table becomes very sparse and then $\hat{\Sigma}_{\bar{\mathbf{r}}} = D(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}'$ is not a consistent estimator for the covariance matrix. On the other hand, $\chi^2_{[t:u]}$ tends to perform well under commonly encountered sparse situations, and has been calculated for up to 20 variables. However, calculating $\chi^2_{[t:u]}$ requires calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ which requires $2 * 2^{q+1}$ integrals, where q is the number of manifest variables to be evaluated by numerical quadrature for the factor analysis model. Using SAS PROC IML, these calculations can be accomplished in random access memory for 20 manifest variables if 6 to 8 GB of RAM are available, for $\mathbf{G}, \mathbf{H}, \mathbf{A}, \boldsymbol{\pi}(\boldsymbol{\beta})$ and $\hat{\mathbf{p}}$, in approximately 4 minutes of CPU time (Reiser, 2012). If the calculations are done using virtual memory, reading and writing to disk, then processing time for 20 variables is on the order of 30 hours. With 25 manifest variables, these calculations can take up to 64 GB of RAM. On the other hand, Tollenaar and Mooijaart (2003) statistic, stated in Section 2.6 does not require calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. The Tollenaar and Mooijaart (2003) statistic

$$\chi^2_{red} = n\mathbf{e}'(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})^{-1}\mathbf{e} \tag{7.1}$$

where,

$$\hat{\mathbf{T}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})'$$

is a reduced version of $\chi^2_{[1:2]}$ statistic. It is a statistic for simple null hypothesis but with adjusted degrees of freedom for estimated parameters. The difference lies in the covariance matrix $\hat{\mathbf{T}}$, which does not include the term $\mathbf{G}(\hat{\mathbf{A}}'\hat{\mathbf{A}})^{-1}\mathbf{G}'$ in the χ^2_{red} statistic. This term represents variance due to estimating model parameters $\boldsymbol{\beta}$. As indicated by Tollenaar and Mooijaart (2003), omitting this term may substantially reduce computations. For instance, if the categorical factor model is fitted to 20 manifest variables, it requires $8 * 2^{20} * 40$ bytes or 0.335 GB to store just the $\mathbf{G}$ matrix in SAS. With 25 variables, this amount will increase to $8 * 2^{25} * 50$ bytes or approximately 13.4 GB. Note that the categorical factor model contains both intercept and slope parameters, thus, it requires to take derivatives with respect to both intercept and slope. Hence, the $\mathbf{G}$ matrix will have $2q$ rows if fitting one factor model. The memory requirement when both the $\mathbf{A}$ matrix and $\mathbf{G}$ matrix are in memory is approximately 2*13.4 = 26.8 GB for 25 manifest variables. After calculation of the term $(\mathbf{A}'\mathbf{A})^{-1}$, the $\mathbf{A}$ matrix can be be discarded from the memory, which will save around 13.4 GB.

While χ^2_{red} does not require the term $\mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$, it still requires the $\mathbf{H}_{[1:2]}$ matrix, which becomes very large with a large number of manifest variables. For instance, with 20 manifest variables, it requires $8 * 2^{20} * 210$ bytes or approximately 1.76 GB to store $\mathbf{H}_{[1:2]}$ matrix. With 25 manifest variables this amount will increase up to 87.24 GB. This is a huge memory requirement for just one matrix, even with modern computer standards. One way to remedy this problem is to replace matrix operations with loops over vectors that consists of the rows of $\mathbf{H}$. Another technique that maybe useful for calculating the entire $\mathbf{H}$ matrix is sparse matrix operations. There are two aspects to sparse matrix techniques, namely, sparse matrix storage and sparse matrix computations. Typically, computer programs represent an M by N matrix in a dense form as an array of size M by N, making row-wise and column-wise arithmetic operations particularly efficient to compute. However, if many of these M by N numbers are zeros, then correspondingly many of these operations are unnecessary or trivial. Sparse matrix techniques exploit this fact by representing a matrix not as a complete array, but as a set of nonzero elements and their location (row and column) within the matrix. This will be ideal for my study since not only observed proportions are sparse but also the $\mathbf{H}$ matrix is sparse. By combining these techniques I have created a program to calculate the χ^2_{red} statistic that can be used for a larger number of manifest variables. This program will not store the $\mathbf{H}_{[1:2]}$ matrix but rather generate the rows of $\mathbf{H}_{[1:2]}$ matrix at each element of $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$. Therefore, to calculate the term $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$ of the χ^2_{red} statistic, this program only need to store two columns of $\mathbf{V}$ matrix to generate the second-order marginal $\mathbf{H}_{(2,i)}$ and another two columns of $\mathbf{V}$ matrix to generate the second-order marginal $\mathbf{H}_{(2,j)}$, where $j \geq i$ and i,j=1,...,q*(q-1)/2. Note, it also need to store the fitted proportions $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$ and the vectors $\mathbf{H}_{(2,i)}$ and $\mathbf{H}_{(2,j)}$. Hence, by using this method for 25 manifest variables, it will only require $7 * 2^{25}$ bytes or approximately 0.2348 GB to generate the elements of $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$. This is huge memory saving compared to the 87.24 GB that is required to store just the $\mathbf{H}_{[1:2]}$ matrix for 25 manifest variables, but there will be a very large increase in number of loops. A brief description of the steps of this program are given as follows:

1. For each l and m create two corresponding columns of the $\mathbf{V}$ matrix, $l = 1, ..., q$

and $m = l+1, ..., q$.

2. Do a element-wise multiplication of those two columns to obtain the second-order $\mathbf{H}_{(2,i)}$.

3. Use an embedded loop and create column l and n of the $\mathbf{V}$ matrix, where $n \geq m$.

4. Do a element-wise multiplication of the two columns in Step 3 to obtain second-order marginal $\mathbf{H}_{(2,j)}$, where $j \geq i$ and i,j=1,...,q*(q-1)/2.

5. Then, use the equation

$$\Sigma_{vec_p} = \sum_{i,j} ((\mathbf{H}_{(2,i)} \circ \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}) \circ \mathbf{H}_{(2,j)}) - \mathbf{H}'_{(2,i)} * \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}) * \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})' * \mathbf{H}_{(2,j)})$$

   to generate the p^{th} element of the Σ_{vec} where, Σ_{vec} is the covariance matrix $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$ in vector form.

6. Use another loop over rows of $\mathbf{H}$ to obtain the vector $\mathbf{e}$ using the equation $\boldsymbol{e} = \mathbf{H}_{[1:2]}(\hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))$, where $\hat{\mathbf{p}}$ is the observed proportions. As in the Step 1 and 2, the loop is used to reduce the memory requirement of the $\mathbf{H}_{[1:2]}$ matrix. Calculation of the rows of the $\mathbf{H}_{[1:2]}$ matrix is similar to Step 1 and 2. For each element, rows of $\mathbf{H}_{[1:2]}$ will be multiplied by the vector $(\hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}))$ to create the r^{th} element of the vector $\mathbf{e}$, $r = 1, ..., q * (q-1)/2$.

7. Use SQRVECH function in SAS to transform Σ_{vec} into a symmetric square matrix, say $\hat{\Sigma}_{\chi^2_{red}}$.

8. Finally, use the equation $\chi^2_{red} = n\mathbf{e}'(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})^{-1}\mathbf{e} = n\mathbf{e}'(\hat{\Sigma}_{\chi^2_{red}})^{-1}\mathbf{e}$ to calculate the χ^2_{red} statistic.

The table below shows results for given observed and fitted probabilities for calculating χ^2_{red} in SAS using this method. Note, these results are for only one pseudo data set.

Table 7.1: Time and Memory Requirements for χ^2_{red}

No. of variables	Real time	User CPU time	System CPU time	Memory
15 variables	8.32 sec.	6.81 sec.	1.51 sec.	0.0037 GB
20 variables	13 min 3 sec	10 min 28 sec	2 min 35 sec	0.0996 GB
25 variables	19 min 21 sec	14 min 5 sec	5 min 16 sec	3.15 GB

Next, a Monte-Carlo simulation study was performed to test the performance of χ^2_{red} for 25 manifest variables. Due to the time limitations only Type I error study was performed. Empirical power study is recommended as a future work.

The design of Type I error study is as follows:

Model (data generation)	categorical variable factor analysis model with one latent factor
Model (fitted)	categorical variable factor analysis model with one latent factor
Number of observed variables	q=25
Number of simulation samples	500
Sample size	n=500

For the Monte-Carlo simulation study, data was generated from one factor model. For the slope parameters of the model, pattern $(.1, .1, .1, 2.4, 2.4, 2.4, .2, .2)$ was repeated. Intercepts of the model were kept at zero. Result related to the simulation is given in the table below.

Table 7.2: Type I Error Results

No. varaibles	Type I error rates
25 var	0.066

When $q = 25$, there are 33,554,432 cells in the 2^{25} cross-classified table. With n=500, each cell may only have, on average, 0.00001 observations. Thus, the sparseness in the cross-classified table is very severe when $q = 25$. But, I'm using second-order marginals and $2 * 2$ sub-table may not be sparse even when $q = 25$ with n=500. Therefore, I was expecting χ^2_{red} to have good Type I error rates even when $q = 25$.

According to the results in the Table 7.2, the empirical Type I error rates are within the Monte-Carlo error interval $0.05 \pm 1.96 * \sqrt{0.05 * 0.95/500}$. Thus, the χ^2_{red} has good performance for Type I error rate even when the number of manifest variables are as large as 25.

Due to time limitations, I only extended the simulations up to $q = 25$. But, I would expect χ^2_{red} to have good Type I error rates even when $q = 50$. As a future work, I recommend extending the simulations for $q = 30$, $q = 40$ and $q = 50$.

7.2 Bootstrap Method

This section will introduce a bootstrap method to obtain p-values for Pearson-Fisher statistic, fit to confirmatory dichotomous variable factor analysis model when the number of manifest variables is large.

When there are 25 manifest variables, the cross-classified table has 2^{25}, or 33,554,432 cells. If the sample size for testing the fit of a model is a few hundred observations, then the data table will be sparse and many cells will have counts of zero or 1. As discussed in the previous sections, when the data are sparse, the asymptotic chi-square approximation for the distribution of the Pearson and likelihood ratio statistics may not be valid. Extensive simulations have also shown that p-values obtained from the chi-square distribution for a test of the categorical factor analysis model on a sample of size 1000 start to become unreliable at about 6 to 8 manifest variables, depending on the skew of distribution of the frequencies (Reiser and VandenBerg, 1994).

Not only sparseness, but also computer resources become an issue when the number of manifest variables exceeds 20. There are limits on individual objects statistical software can store. For example, having 30 manifest variables would require approximately $8 * 30 * 2^{30}$ bytes or 257.6 GB to store the $\mathbf{H}$ matrix in R or SAS assuming double precision storage. If the interest is to store only observed probabilities and fitted probabilities, with 30 manifest variables it will only require approximately 16 GB. Due to these reasons most of the simulations found in the literature are limited to 20 manifest variables. But, in an application such as educational testing, the number of manifest variables could be 50 or more, and with 50 manifest variables, it will require $8 * 2^{50}$ bytes or 9,007,199.25 GB to store the fitted probabilities.

I introduce the following method using the omnibus χ^2_{PF} statistic to overcome these issues. Calculation of the Pearson statistic itself does not necessarily encounter memory limits for large number of manifest variables because the contribution of each cell can be calculated individually and cumulated. Processing requirements of χ^2_{PF} are not a concern for 30 or more variables because calculation of $\boldsymbol{\pi}_s(\hat{\boldsymbol{\beta}})$ is required only for the cells where $n_s > 0$, and even with a large number of manifest variables, the number of cells where $n_s > 0$ can be no more than the sample size. The contribution for the cells with $n_s = 0$ is equal to $n \sum_s I(n_s = 0)\pi_s(\hat{\boldsymbol{\beta}})$ and can be obtain by subtraction since,

$$\sum_s I(n_s > 0)\pi_s(\hat{\boldsymbol{\beta}}) + \sum_s I(n_s = 0)\pi_s(\hat{\boldsymbol{\beta}}) = 1 \qquad (7.2)$$

where, I is the indicator function. Calculation of $\chi^2_{[2]}$, for example, requires much more storage. Since computational requirements may not present a barrier, obtaining p-values for χ^2_{PF} by using the parametric bootstrap may be feasible even for a very large number of variables. The theory of the parametric bootstrap is quite similar to that of the nonparametric bootstrap, the only difference is that instead of simulating bootstrap samples that are independent and identically distributed (iid) from the empirical distribution (the nonparametric estimate of the distribution of the data) the parametric bootstrap procedure simulates bootstrap samples that are iid from the estimated parametric model.

The method that is introduce here will require only the observed patterns and hence less memory requirement. A brief description of the steps of this method are given as follows:

1. Assume $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$ is true. The model $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$ could be any categorical variable model.
2. Treat the fitted proportions $\boldsymbol{\pi}_s(\hat{\boldsymbol{\beta}})$ under the model as population proportions.
3. Draw random samples from the multinomial distribution with these fitted proportions as parameters of the distribution.
4. For each sample, estimate the categorical variable model used in Step 1. For a instance, if the categorical variable model with one factor was used in Step 1 to get $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}})$ then, categorical variable model with one factor will be estimated for each sample from Step 3.
5. If $n_s > 0$, use multivariate Gaussian quadrature to obtain the expected proportions and calculate $\chi^2_{PF_{n_s>0}}$.
6. If $n_s = 0$, use the equation 7.2 to obtain $\chi^2_{PF_{n_s=0}}$.
7. Sum $\chi^2_{PF_{n_s=0}}$ and $\chi^2_{PF_{n_s>0}}$ to obtain χ^2_{PF}.
8. Repeat step 5,6 and 7 for each sample.
9. Obtain p-value by calculating the proportion of χ^2_{PF} values from bootstrap samples that are greater than the χ^2_{PF} value from the original sample.

In order to evaluate the performance of this method, Type I error study was performed. Note, the χ^2_{PF} is an omnibus test that gives little guidance of the source of poor fit and can be outperformed by focused or directional tests of lower-order.

The design of Type I error study is as follows:

Model (data generation)	categorical variable factor analysis model with one latent factor
Model (fitted)	categorical variable factor analysis model with one latent factor
Number of observed variables	q=8, q=15, q=18, q=20
Number of simulation samples	1000
Sample size	n=500
Number of bootstrap samples	500

Monte-Carlo simulation studies were performed with the information described in the Table above. One thousand datasets were generated from the one factor model. For the slope parameters of the one factor model, the pattern $(.1, .1, .1, 2.4, 2.4, 2.4, .2, .2)$ was repeated. Intercepts of the model were kept at zero. After generating the data, a one factor model was estimated for each of these datasets. To calculate the p-value correspond to the χ^2_{PF} for each dataset, five hundred bootstrap samples were obtained using the steps 1-9 above. The p-value for each dataset was obtained by calculating the proportion of χ^2_{PF} values from bootstrap samples that are greater than the χ^2_{PF} value from the original sample. This process was repeated for all one thousand datasets. The type I error rate was obtained by dividing the number of datasets that had p-value less than 0.05 by 1,000. Results related to 8, 15, and 20 variables are given in the Table 7.3 below. The 'Mplus(MonteCarlo)' column in the Table 7.3 corresponds to the Type I error rates calculated using the theoretical distribution. Note, the one thousand datasets mentioned above was generated using Mplus. For each of these datasets, χ^2_{PF} and the corresponding p-value under the theoretical distribution was calculated using Mplus. Then, the type I error rate based on the theoretical distribution was obtained by dividing the number of datasets that had p-value less than 0.05 by 1,000.

Table 7.3: Type I Error Rates Comparison for χ^2_{PF}

No. varaibles	Bootstrap Method	Mplus(MonteCarlo)
8 var	0.046	0.042
15 var	0.044	0.161
20 var	0.342	0.380

Table 7.4: Time Requirements for the Bootstrap Method

No. varaibles	Time (in sec)
8 var	29
15 var	68
20 var	360

* No. of bootstrap samples = 500

According to the results in the Table 7.3, for moderately large number of manifest variables, the bootstrap method performed well in terms of Type I error rates. When the number of manifest variables exceeds 20, the Type I error rates started to inflate. However, I believe the Type I error rates can be improved by increasing the number of bootstrap samples. Due to the time limitations I had to restrict my simulations to 500 bootstrap samples.

In this chapter, I have investigated two methods to check the feasibility of goodness-of-fit statistics when the number of manifest variables is large. Firstly, I have investigated performance of the Tollenaar and Mooijaart (2003) χ^2_{red} statistics when the number manifest variables is large. Results indicate χ^2_{red} has good performance for

Type I error rate even when the number of manifest variables is as large as 25. One of the other goals was to create memory and time efficient program to calculate goodness-of-fit statistics for large number of variables. The program that I have created improved the memory requirement. The largest amount of RAM the program consumed during the calculation of the Tollenaar and Mooijaart (2003) statistics was 3.15 GB for 25 variables. However, the number of loops this program require, and thus the computer time increased rapidly with q. For instance, 15 manifest variables would require $105 * (106/2) = 5,565$ loops to calculate components of the matrix $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$ and $15 * (14/2) = 105$ loops to calculate the $\mathbf{e}$ vector. Similarly, 20 manifest variables would require $20 * (19/2) + 190 * (191/2) = 18,335$ loops and 25 manifest variables would require $25 * (24/2) + 300 * (301/2) = 45,450$ loops. Note, heavy mathematical calculations also happening inside each of these loops. Therefore, the drawback of the this method is the large number of loops and cpu time.

Secondly, performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic was investigated. For moderately large number of manifest variables, the bootstrap method performed well in terms of Type I error rates. When the number of manifest variables exceeds 20, the Type I error rates started to inflate. This might be due to the small number of bootstrap samples used in the simulation study. Therefore, as a future work, I suggest to increase the number of bootstrap samples to 2,000 or more. The main issue that I encountered with the bootstrap method is it requires 2^q expected probabilities to generate the bootstrap samples. When the number of manifest variables increases this may cause computer resource limitations.

Chapter 8

DISCUSSION

The goodness-of-fit test is one of the most common tests in statistics. If the data table contains cell counts that are small, common test statistics such as Pearson's chi-square and likelihood ratio may not follow the usual theoretical distribution. Over the past years several statistics has been proposed to remedy this issue. Some of these statistics formed on lower-order marginal have been shown to overcome the deleterious effect of spareness. I used orthogonal components of Pearson's chi-square statistic defined on lower-order marginals of the data table as a remedy to this sparseness problem. To this end, I studied three problems in my dissertation. As my first problem, I studied goodness-of-fit components using second-order, third-order and fourth-order marginals. I developed two new statistics, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$, and studied the Type I error, empirical power and asymptotic power of these statistics under different sparseness conditions. I also compared the performance of these statistics to $\chi^2_{[2]}$, χ^2_{red}, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$ and M_r statistics. When the sample size was small (e.g. n=300) the Type I error rates related to $\chi^2_{red,[4]}$, M_4 and $\chi^2_{[2:4]}$ were considerably different from the nominal value 0.05. However, when the sample size increases, Type I error rates were improved. When n=1000, almost all the statistics had Type I error rates close to the nominal value. On the other hand, Type I error rates related to $\chi^2_{red,[3]}$, M_3 and $\chi^2_{[2:3]}$ were close to the nominal value, even for n=300. Similarly, all the Type I error rates related to χ^2_{red}, M_2 and $\chi^2_{[2]}$ were within the Monte-Carlo error interval for all the different intercept settings and sample sizes. This suggests that the $2*2*2*2$ tables were sparse when $q = 8$ but $2*2*2$ and $2*2$ tables were not sparse. When the number of manifest variables was extended to $q = 15$, Type I error rates related to $\chi^2_{red,[4]}$, M_4, $\chi^2_{[2:4]}$, $\chi^2_{red,[3]}$, M_3 and $\chi^2_{[2:3]}$ were considerably different from the nominal value 0.05 for symmetric and asymmetric intercept models. However, the Type I error rates related to χ^2_{red}, M_2 and $\chi^2_{[2]}$ were within the Monte-Carlo error interval for all the different intercept settings and sample sizes. This suggests that the $2*2*2*2$ and $2*2*2$ tables were sparse when $q = 15$ but $2*2$ tables were not sparse. However, it was interesting to see that the Type I error rates related to all the statistics were within the Monte-Carlo error interval for the zero intercept model for both $q = 8$ and $q = 15$. Thus, the observations seem to be well distributed among cells when the intercepts are zero compared to asymmetric or symmetric intercepts even when $q = 15$. This might also be related to bias in the parameter estimates. Based on the power results it is clear that a test based on second-order marginals, $\chi^2_{[2]}$ has higher power to detect lack-of-fit located in the second-order associations when compared to a statistic that incorporates higher-order marginals such as $\chi^2_{[2:3]}$ or the $\chi^2_{[2:4]}$. The $\chi^2_{[2]}$ statistic, however, would be insensitive to a lack-of-fit that is present in the third-order marginals. When I used a log-linear model with 3-way interactions, power of the test based on $\chi^2_{[2:3]}$ surpassed the power of the test based on $\chi^2_{[2]}$. As the three-way association effect becomes larger, the power of the test based on $\chi^2_{[2]}$ rose only gradually, but the power of test based on the $\chi^2_{[2:3]}$ rose rapidly. This suggests that $\chi^2_{[2:3]}$ statistic has better performance when there is a three-way association compared to $\chi^2_{[2]}$. Also, the power of $\chi^2_{[2:4]}$ was lower than $\chi^2_{[2:3]}$. This suggests that when the lack-of-fit is in the third-order, adding additional components may dilute the test. When the three-way associations were present in the model, $\chi^2_{red,[3]}$, $\chi^2_{red,[4]}$, M_3 and M_4 had somewhat of a lower power compared to $\chi^2_{[2:3]}$. The $\chi^2_{[2:3]}$ statistic seems to outperform the other statistics in this situation.

In many applications of latent variable models in the social sciences, manifest variables are designed to have high bi-variate association, but sometimes it is possible to have three- or four-way associations. If lack-of-fit for the model is in third- or fourth-order components, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ may have higher power than a lower-order statistics like $\chi^2_{[2]}$. However, it is not possible to know the location of the lack-of-fit in advance. To protect against the possibility of failing to detect a departure from the null hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$ in higher-order marginals one can examine the residual $\chi^2_{PF} - \chi^2_{[2]}$. Since $\chi^2_{[2]}$ is a component of χ^2_{PF}, a large residual relative to the df may indicate the need of inclusion of higher-order marginals in the test statistic. This can be carried out in a sequential manner by starting with $\chi^2_{[2]}$ then $\chi^2_{[3|2]}$, $\chi^2_{[4|3,2]}$ until you reach the statistic that includes marginals up to q^{th} order which is χ^2_{PF}. Note, beyond $\chi^2_{[2:3]}$ the sub-tables can be sparse, especially when the sample size is small (e.g. n=300). If that is the case, then a method like bootstrap would be needed to find p-value. The α level for the tests would need to be adjusted for multiple testing.

The ability to choose between different statistics $\chi^2_{[2]}$, $\chi^2_{[2:3]}$ and $\chi^2_{[2:4]}$ in various situations can help to improve the inference and decisions made in real world applications. The SAS code I developed facilitates this approach and can help to improve the decisions made in real world applications.

As indicated before, manifest variables are designed to have high bi-variate associations in many applications of latent variable models in the social sciences. In these situations, $\chi^2_{[2]}$ may feature higher power for a certain alternative hypothesis especially, when the lack-of-fit is in the second-order marginals. When a model fails to fit adequately, it is important to know where the model provides a good fit and where it does not. Thus, as my second problem, I checked the performance of individual orthogonal components of $\chi^2_{[2]}$ as statistics to identify lack-of-fit. In the context of this problem, I studied both empirical and asymptotic power. I also compared the performance of individual orthogonal components of $\chi^2_{[2]}$ to other test statistics discussed in Chapter 2 and 3: adjusted residuals and $\bar{\chi}^2_{ij}$. I used two settings for the number of manifest variables: $q = 8$ and $q = 15$ and three settings for the intercepts of the model: symmetric, asymmetric and zero. Based on the results it is clear that the orthogonal components exhibit better Type I error performance compared to adjusted residuals and $\bar{\chi}^2_{ij}$ when $n = 300$. However, when the sample size increases (e.g. $n = 500$), Type I error rate performance was similar between orthogonal components, adjusted residuals and $\bar{\chi}^2_{ij}$, especially for the zero intercept setting. However, this was not the case for symmetric and asymmetric intercept cases, especially when the $q = 15$. It seems counts have less sparseness among cells when the intercepts are zero compared to asymmetric or symmetric intercepts. However, this might also be related to bias in the parameter estimates. Overall, individual orthogonal components had better Type I error rates than adjusted residuals and $\bar{\chi}^2_{ij}$ even when the cross-classified table was very sparse.

Note, each orthogonal component is distributed in large samples approximately as chi-square with one degree of freedom (df). To check this assumption, chi-square Q-Q plots were built for the simulation values related to each component. A similar approach was taken to check the normality assumption of the adjusted residuals. On the other hand, the $\bar{\chi}^2_{ij}$ featured a different df for different item pairs. The df of $\bar{\chi}^2_{ij}$ depends on $\mathbf{\Sigma}_{ij}$ where, $\mathbf{\Sigma}_{ij}$ is the covariance matrix related to the residuals $n(\mathbf{p}_{ij} - \hat{\boldsymbol{\pi}}_{ij})$ for a pair of items. Thus, an average value of these df was used to calculate the chi-square Q-Q plots for $\bar{\chi}^2_{ij}$. Most of the Q-Q plots attained the distributional assumption. There were a few Q-Q plots that showed deviations from the straight line assumption, especially when the sample size was small. Note, these Q-Q plots are very sensitive to outlier observations. When the sample size is small, some of the estimated standard errors related to the test statistics can be very small. This can result in a larger test statistic value. A few of these larger test statistic values can easily affect the pattern of the Q-Q plot.

Most of the Q-Q plots that deviated from the hypothesized distributions were related to $\bar{\chi}^2_{ij}$. This might be related to the fact that I am using the mean value of the one thousand df of a particular $\bar{\chi}^2_{ij}$ to calculate the Q-Q plot for that particular $\bar{\chi}^2_{ij}$ even though each have a different df under each simulation value. I recommend a further study of this as a future work.

As shown earlier, the Pearson-Fisher statistic for a composite null hypothesis can be partitioned into $T-g-1$ orthogonal components defined on marginal distributions. When the manifest variables are binary, each of these components, $\hat{\gamma}^2_j$, is distributed approximately as an independent $\chi^2_{(1)}$ random variate in large sample. These components can be used as item diagnostics for model fitting when the result of an omnibus test indicates that a model should be rejected. Thus, the second order marginals related to orthogonal components were examined as lack-of-fit diagnostics. Simulations were based on a categorical factor model for a two latent variable model. To calculate the empirical power, the model under the null hypothesis was misspecified with one factor model. Empirical power was also compared to the asymptotic power. Based on the results, orthogonal components were successful in indicating a pair of variables for which the model does not fit well. When the sample size increases, the ability to indicate a pair of variables for which the model does not fit well increases significantly. The $\bar{\chi}^2_{ij}$ and the adjusted residual had some what higher power for some variable pairs compared to orthogonal components when $n = 300$. For example, When $q = 8$ and $n = 300$ the empirical power related to item pairs (4,5) and (4,6) were higher for the $\bar{\chi}^2_{ij}$ and the adjusted residuals compared to orthogonal components. However, the empirical power related to item (5,6) was higher for orthogonal components compared to $\bar{\chi}_{ij}^2$ and adjusted residual. Note, orthogonal components had better Type I error rates even when the cross-classified table was very sparse compared to $\bar{\chi}_{ij}^2$ and adjusted residual. In addition, the $\bar{\chi}_{ij}^2$ for different item pairs cannot be directly compared as they are on a different scale (their estimated df). Only the p-values can be directly compared across item pairs. This is undesirable in terms of actual applications because researchers have to inspect tables of p-values with a large number of decimals in order to determine the item pairs with the greatest magnitude of misfit.

The asymptotic power results tally with the empirical power results for the zero intercept condition. However, when the intercepts move away from zero, and the sample size is small, there were some discrepancies between asymptotic and empirical power. Looking at the power results for both $q = 8$ and $q = 15$, there is very little change in asymptotic power between zero, symmetric and asymmetric intercept settings. However, the empirical power can differ sometimes by a substantial amount. This is because the empirical power is affected by sparseness and bias of estimator for intercept and slope. This was also evident in the Type I error study.

Different software packages use different methods for parameter estimation and optimization. Sometimes, the same software may have different options. For example, the default optimization technique in PROC IRT to obtain maximum likelihood estimates is dual quasi-Newton optimization. But, it also allows you to select other optimization methods like EM optimization, Newton-Raphson optimization with ridging and conjugate-gradient optimization. These different methods can have different effects on parameter estimation and hence, the goodness-of-fit statistics based on those estimations. For instance, I used PROC IRT with dual quasi-Newton optimization for my initial simulations. But, the parameter estimates were not stable for small sample size and/or factor loadings. When I used the EM optimization in PROC IRT, parameter estimates were more stable. Mplus parameter estimates were more stable compared to PROC IRT estimates. Hence, I used Mplus parameter estimates for all my calculations. As a future work, I recommend comparing the performance of orthogonal components under these different parameter estimation methods.

As my third problem, I extended the statistics on lower-order marginals to a larger number of manifest variables. When the number of manifest variables exceeds 20, most of the statistics on lower-order marginals have limitations in terms of computer resources and CPU time. Under this problem, I investigated the performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic, fit to confirmatory dichotomous variable factor analysis model and the performance of Tollenaar and Mooijaart (2003) statistic when the number manifest variables is larger than or equal to 25.

Results indicate χ^2_{red} has good performance for Type I error rate even when the number of manifest variables is as large as 25. One of the other goals of this research was to create memory and time efficient program to calculate goodness-of-fit statistics for large number of variables. The program that I have created improved the memory requirement. The largest amount of RAM the program consumed during the calculation of the Tollenaar and Mooijaart (2003) statistics was 3.15 GB for 25 manifest variables. However, the number of loops this program require thus the computer time increased rapidly with q. For instance, 15 manifest variables would require $105 * (106/2) = 5,565$ loops to calculate components of the matrix $(\mathbf{H}_{[1:2]}\hat{\mathbf{T}}\mathbf{H}'_{[1:2]})$ and $15 * (14/2) = 105$ loops to calculate the $\mathbf{e}$ vector. Similarly, 20 manifest variables would require $20 * (19/2) + 190 * (191/2) = 18,335$ loops and 25 manifest variables would require $25 * (24/2) + 300 * (301/2) = 45,450$ loops. Therefore, the drawback of this method is the large number of loops and CPU time.

Performance of a bootstrap based method to obtain p-values for Pearson-Fisher statistic was also investigated when the number of manifest variables is large. For a moderately large number of manifest variables, the bootstrap method performed well in terms of Type I error rates. When the number of manifest variables exceeds 20, the Type I error rates started to inflate. This might be due to the small number of bootstrap samples used in the simulations study. Therefore, as a future work, I suggest increasing the number of bootstrap samples to 2000 or more. The main issue that I encountered with the bootstrap method is it requires 2^q expected probabilities to generate the bootstrap samples. When the number of manifest variables increases this may cause computer resource limitations.

REFERENCES

Agresti, A. (2013). *Categorical data analysis.* New York: Wiley.

Agresti, A., & Yang, M. (1987). An empirical investigation of some effects of sparsness in contingency tables. *Computational Statistics and Data Analysis*, *5*, 9–21.

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., & Tzamourani, P. (1999). The goodness of fit of latent trait models in attitude measurement. *Sociological Methods and Research*, *27*(4), 525–546.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Birch, M. W. (1964). A new proof of the pearson-fisher theorem. *Annals of Mathematical Statistics*, *35*, 818–824.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). Discrete multivariate analysis cambridge.

Cagnone, S., & Mignani, S. (2007). Assessing the goodness of fit of a latent variable model for ordinal data. *Metron*, *65*(3), 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*(1), 173–194.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*, 5–32.

Cochran, W. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, *23*, 315–345.

Cressie, N., & Read, T. R. C. (1989). Pearson's χ^2 and the loglikelihood ratio statistic g2: A comparative review. *International Statistical Institute*, *57*, 19–43.

Fisher, R. A. (1924). The conditions under which chi square measures the discrepancy between observation and hypothes. *Royal Statistical Society*, *87*, 442–450.

Fisher, R. A. (1941). *Statistical methods for researxh workers.* Edinburgh: Oliver and Boyd, Ltd.

Gokhale, D. V., & Kullback, S. (1978). *The information in contingency table.* New York: Marcel Dekker.

Goodnight, J. H. (1979). A tutorial on the sweep operator. *The American Statistician*, *33*(3), 149–158.

Haldane, J. B. S. (1939). The mean and variance of χ^2, when used a test of homogeneity, when expectations are small. *Biometrika*, *31*, 346–355.

Joe, H. (1993). Tests of uniformity for sets of lotto numbers. *Statistics and Probability Letters*, *16*, 181–185.

Knott, M., & Tzamourani, P. (1997). Fitting a latent trait model for missing observations to racial prejudice data. *In Applications of Latent Trait and Latent Class Models in the Social Sciences*, 224-252.

Koehler, K., & Larantz, K. (1980). An emperical investigation of goodness-offit statistics for sparse multinomials. *Journal of the American Statistical Association*, *75*, 336–344.

Kullback, S., & Keegel, J. C. (1984). Categorical data problems using information theoretic approach. *In Handbook of Statistics*, 831–87.

Lancaster, H. (1969). *The chi-squared distribution.* New York: Wiley.

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association*, *73*, 253–263.

Lawal, H. (1984). Comparisons of χ^2, y2, freeman-tukey and william's improved G^2 test statistics in small samples of one-way multinomial. *Biometrika*, *71*, 415–458.

Lawal, H., & Upton, G. (1980). An approximation to the distribution of the χ^2 goodness-of-fit statistic for use with small expectations. *Biometrika*, *67*, 447–453.

Liu, Y., & Maydeu-Olivares, A. (2014). Identifying the source of misfit in item response theory models. *Multivariate behavioral research*, *49*(4), 354–371.

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.

Mitra, S. (1958). On the limiting power function of the frequency chi-square test. *Annals of Statistics*, *29*, 1221–1233.

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

Muthn, L., & Muthn, B. .-. (2017). Mplus users guide. *Eighth Edition. Los Angeles, CA: Muthen and Muthen.*

Neyman, J. (1949). Contribution to the theory of the χ^2 test. *Proc. 1st Berkeley Symp.*, 239–273.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50–64.

Pearson, K. (1924). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Series*, *50*, 157–172.

Rayner, J. C., & Best, D. J. (1989). *Smooth tests of goodness of fit.* New York: Oxford.

Read, T. R. C. (1984). Small-sample comparisons for the power divergence goodness-of-fit statistics. *Am. Statist. Assoc*, *79*, 929–935.

Read, T. R. C., & Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag.

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. *Sociological methodology*, *29*, 81–111.

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, *61*, 201–223.

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.

Upton, G. (1978). *The analysis of cross-tabulated data.* New York: Wiley.

APPENDIX A

TABLES

Table A.1: Type I Error Study for Asymmetric Intercept Model

	n=300			n=500		
Pair (i,j)	Orthgonal Comp.	Std. Residuals	$\bar{\chi}_{ij}^2$	Orthgonal Comp.	Std. Residuals	$\bar{\chi}_{ij}^2$
(1,2)	0.039	0.042	0.041	0.037	0.038	0.037
(1,3)	0.041	0.045	0.043	0.04	0.04	0.04
(1,4)	0.052	0.055	0.058	0.055	0.054	0.05
(1,5)	0.04	0.057	0.052	0.04	0.049	0.047
(1,6)	0.041	0.053	0.051	0.053	0.059	0.057
(1,7)	0.039	0.046	0.046	0.041	0.054	0.053
(1,8)	0.062	0.053	0.051	0.064	0.059	0.059
(2,3)	0.057	0.06	0.059	0.057	0.058	0.058
(2,4)	0.049	0.059	0.052	0.056	0.055	0.051
(2,5)	0.046	0.063	0.055	0.054	0.055	0.056
(2,6)	0.038	0.054	0.048	0.049	0.061	0.061
(2,7)	0.047	0.043	0.043	0.045	0.034	0.032
(2,8)	0.046	0.045	0.043	0.058	0.05	0.049
(3,4)	0.044	0.059	0.058	0.043	0.042	0.043
(3,5)	0.06	0.047	0.044	0.054	0.044	0.041
(3,6)	0.051	0.059	0.058	0.049	0.05	0.05
(3,7)	0.049	0.04	0.039	0.042	0.042	0.042
(3,8)	0.046	0.043	0.043	0.042	0.042	0.043
(4,5)	0.046	**0.087**	**0.087**	0.059	0.056	0.061
(4,6)	0.061	**0.072**	**0.074**	0.046	0.062	0.062
(4,7)	0.055	**0.067**	**0.066**	0.039	0.053	0.053
(4,8)	0.05	**0.066**	**0.068**	0.037	0.04	0.04
(5,6)	0.063	**0.07**	**0.072**	0.057	0.056	0.056
(5,7)	0.042	0.055	0.054	0.041	0.046	0.048
(5,8)	0.062	0.057	0.056	0.049	0.04	0.039
(6,7)	0.038	0.045	0.048	0.054	0.052	0.052
(6,8)	0.055	0.046	0.044	0.061	0.04	0.039
(7,8)	**0.066**	0.06	0.061	0.047	0.055	0.055

Table A.2: Type I Error Study for Zero Intercept Model

	n=300			n=500		
Pair (i,j)	Orthgonal Comp.	Std. Residuals	$\bar{\chi}^2_{ij}$	Orthgonal Comp.	Std. Residuals	$\bar{\chi}^2_{ij}$
(1,2)	0.053	0.053	0.053	0.043	0.042	0.043
(1,3)	0.058	0.057	0.056	0.055	0.054	0.054
(1,4)	0.042	0.043	0.043	0.055	0.056	0.057
(1,5)	0.057	0.052	0.054	0.057	0.053	0.053
(1,6)	0.054	0.051	0.053	0.05	0.06	0.06
(1,7)	0.045	0.045	0.045	0.047	0.047	0.047
(1,8)	0.043	0.051	0.052	0.06	0.04	0.04
(2,3)	0.06	0.059	0.058	0.053	0.052	0.052
(2,4)	0.045	0.044	0.046	0.045	0.045	0.044
(2,5)	0.055	0.055	0.055	0.053	0.053	0.056
(2,6)	0.055	0.05	0.048	0.045	0.056	0.056
(2,7)	0.04	0.044	0.046	0.036	0.044	0.043
(2,8)	0.042	0.051	0.051	0.056	0.05	0.049
(3,4)	0.047	0.052	0.052	0.053	0.051	0.053
(3,5)	0.051	0.038	0.039	0.058	0.044	0.045
(3,6)	0.063	0.058	0.057	0.058	0.056	0.059
(3,7)	0.045	0.051	0.051	0.045	0.048	0.048
(3,8)	0.047	0.054	0.054	0.048	0.054	0.054
(4,5)	0.047	0.052	0.055	0.043	0.047	0.05
(4,6)	0.059	0.06	0.057	0.048	0.06	0.059
(4,7)	0.042	0.036	0.037	0.052	0.054	0.054
(4,8)	0.061	0.049	0.05	0.044	0.054	0.055
(5,6)	0.05	0.059	0.06	0.058	0.06	0.054
(5,7)	0.043	0.046	0.049	0.04	0.035	0.036
(5,8)	0.054	0.049	0.049	0.056	0.051	0.051
(6,7)	0.052	0.042	0.043	0.058	0.042	0.042
(6,8)	0.043	0.052	0.051	0.05	0.039	0.04
(7,8)	0.05	0.043	0.043	0.045	0.056	0.056

Table A.3: Type I Error Study for Orthogonal Components for Asymmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.055	(3,12)	0.055	(7,9)	0.051
(1,3)	0.045	(3,13)	0.047	(7,10)	0.047
(1,4)	0.052	(3,14)	0.046	(7,11)	0.045
(1,5)	0.053	(3,15)	0.052	(7,12)	0.053
(1,6)	0.06	(4,5)	0.056	(7,13)	0.057
(1,7)	0.046	(4,6)	0.043	(7,14)	0.046
(1,8)	0.048	(4,7)	0.046	(7,15)	0.046
(1,9)	0.044	(4,8)	0.043	(8,9)	0.042
(1,10)	0.046	(4,9)	0.049	(8,10)	0.047
(1,11)	0.046	(4,10)	0.044	(8,11)	0.045
(1,12)	0.063	(4,11)	0.063	(8,12)	**0.065**
(1,13)	0.048	(4,12)	0.056	(8,13)	0.055
(1,14)	0.04	(4,13)	0.045	(8,14)	0.048
(1,15)	0.047	(4,14)	0.053	(8,15)	0.048
(2,3)	0.056	(4,15)	0.057	(9,10)	0.055
(2,4)	0.052	(5,6)	0.051	(9,11)	0.052
(2,5)	0.051	(5,7)	0.051	(9,12)	0.052
(2,6)	0.055	(5,8)	0.055	(9,13)	0.055
(2,7)	0.05	(5,9)	0.053	(9,14)	0.048
(2,8)	0.049	(5,10)	0.046	(9,15)	0.047
(2,9)	0.05	(5,11)	0.049	(10,11)	0.049
(2,10)	**0.036**	(5,12)	0.038	(10,12)	**0.036**
(2,11)	0.046	(5,13)	0.049	(10,13)	0.05
(2,12)	0.051	(5,14)	0.044	(10,14)	0.044
(2,13)	0.045	(5,15)	0.056	(10,15)	0.056
(2,14)	0.059	(6,7)	0.061	(11,12)	0.061
(2,15)	0.054	(6,8)	0.045	(11,13)	0.044
(3,4)	**0.034**	(6,9)	0.041	(11,14)	0.042
(3,5)	0.049	(6,10)	0.054	(11,15)	0.053
(3,6)	0.06	(6,11)	0.058	(12,13)	0.057
(3,7)	0.04	(6,12)	0.051	(12,14)	0.052
(3,8)	0.038	(6,13)	0.044	(12,15)	0.044
(3,9)	0.038	(6,14)	0.044	(13,14)	0.045
(3,10)	0.052	(6,15)	0.051	(13,15)	0.049
(3,11)	0.055	(7,8)	0.053	(14,15)	0.052

Table A.4: Type I Error Study for Adjusted Residuals for Asymmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.052	(3,12)	0.055	(7,9)	0.053
(1,3)	0.047	(3,13)	0.047	(7,10)	0.047
(1,4)	**0.036**	(3,14)	0.038	(7,11)	0.035
(1,5)	0.04	(3,15)	0.037	(7,12)	0.037
(1,6)	0.045	(4,5)	0.044	(7,13)	0.043
(1,7)	0.043	(4,6)	0.045	(7,14)	0.045
(1,8)	0.04	(4,7)	0.04	(7,15)	0.04
(1,9)	**0.034**	(4,8)	0.034	(8,9)	0.034
(1,10)	0.041	(4,9)	0.043	(8,10)	0.043
(1,11)	0.042	(4,10)	0.041	(8,11)	0.042
(1,12)	0.056	(4,11)	0.057	(8,12)	0.058
(1,13)	**0.065**	(4,12)	0.058	(8,13)	0.06
(1,14)	0.051	(4,13)	0.048	(8,14)	0.047
(1,15)	0.044	(4,14)	0.045	(8,15)	0.042
(2,3)	0.038	(4,15)	0.038	(9,10)	0.039
(2,4)	**0.035**	(5,6)	**0.035**	(9,11)	0.036
(2,5)	0.037	(5,7)	0.039	(9,12)	0.04
(2,6)	0.059	(5,8)	0.056	(9,13)	0.053
(2,7)	0.04	(5,9)	0.041	(9,14)	0.04
(2,8)	0.042	(5,10)	0.043	(9,15)	0.041
(2,9)	0.039	(5,11)	0.039	(10,11)	0.039
(2,10)	0.056	(5,12)	0.057	(10,12)	0.056
(2,11)	0.05	(5,13)	0.05	(10,13)	0.048
(2,12)	0.046	(5,14)	0.043	(10,14)	0.045
(2,13)	0.048	(5,15)	0.05	(10,15)	0.047
(2,14)	0.053	(6,7)	0.06	(11,12)	0.059
(2,15)	**0.036**	(6,8)	0.048	(11,13)	0.05
(3,4)	0.039	(6,9)	0.038	(11,14)	0.038
(3,5)	0.045	(6,10)	0.047	(11,15)	0.044
(3,6)	0.06	(6,11)	0.061	(12,13)	0.06
(3,7)	0.045	(6,12)	0.042	(12,14)	0.043
(3,8)	0.042	(6,13)	0.044	(12,15)	0.044
(3,9)	0.048	(6,14)	0.048	(13,14)	0.047
(3,10)	0.062	(6,15)	0.059	(13,15)	0.06
(3,11)	0.042	(7,8)	0.042	(14,15)	0.042

Table A.5: Type I Error Study for $\bar{\chi}^2_{ij}$ for Asymmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.053	(3,12)	0.051	(7,9)	0.05
(1,3)	0.047	(3,13)	0.047	(7,10)	0.05
(1,4)	**0.035**	(3,14)	0.045	(7,11)	0.052
(1,5)	**0.036**	(3,15)	0.053	(7,12)	0.04
(1,6)	0.043	(4,5)	0.057	(7,13)	0.043
(1,7)	0.045	(4,6)	0.046	(7,14)	0.045
(1,8)	0.04	(4,7)	0.046	(7,15)	0.046
(1,9)	**0.034**	(4,8)	0.042	(8,9)	0.059
(1,10)	0.043	(4,9)	0.047	(8,10)	0.039
(1,11)	0.042	(4,10)	0.045	(8,11)	0.058
(1,12)	0.058	(4,11)	0.064	(8,12)	0.053
(1,13)	0.06	(4,12)	0.055	(8,13)	0.05
(1,14)	0.047	(4,13)	0.048	(8,14)	0.055
(1,15)	0.042	(4,14)	0.048	(8,15)	0.062
(2,3)	0.039	(4,15)	0.055	(9,10)	0.06
(2,4)	**0.036**	(5,6)	0.052	(9,11)	0.059
(2,5)	0.04	(5,7)	0.052	(9,12)	0.044
(2,6)	0.053	(5,8)	0.055	(9,13)	0.055
(2,7)	0.04	(5,9)	0.048	(9,14)	0.057
(2,8)	0.041	(5,10)	0.047	(9,15)	0.053
(2,9)	0.039	(5,11)	0.049	(10,11)	0.059
(2,10)	0.056	(5,12)	**0.036**	(10,12)	0.055
(2,11)	0.048	(5,13)	0.05	(10,13)	0.056
(2,12)	0.045	(5,14)	0.044	(10,14)	0.055
(2,13)	0.047	(5,15)	0.056	(10,15)	0.05
(2,14)	0.059	(6,7)	0.061	(11,12)	0.053
(2,15)	0.05	(6,8)	0.044	(11,13)	0.052
(3,4)	0.038	(6,9)	0.042	(11,14)	0.045
(3,5)	0.044	(6,10)	0.053	(11,15)	0.059
(3,6)	0.06	(6,11)	0.057	(12,13)	0.05
(3,7)	0.043	(6,12)	0.052	(12,14)	0.04
(3,8)	0.044	(6,13)	0.044	(12,15)	0.039
(3,9)	0.047	(6,14)	0.045	(13,14)	0.055
(3,10)	0.06	(6,15)	0.049	(13,15)	0.061
(3,11)	0.042	(7,8)	0.052	(14,15)	0.049

Table A.6: Type I Error Study for Orthogonal Components for Asymmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.054	(3,12)	0.055	(7,9)	0.038
(1,3)	0.05	(3,13)	0.05	(7,10)	0.055
(1,4)	**0.035**	(3,14)	0.047	(7,11)	0.043
(1,5)	0.047	(3,15)	0.045	(7,12)	**0.03**
(1,6)	0.037	(4,5)	0.045	(7,13)	0.049
(1,7)	0.042	(4,6)	0.047	(7,14)	0.048
(1,8)	**0.034**	(4,7)	0.05	(7,15)	0.064
(1,9)	0.041	(4,8)	0.041	(8,9)	0.041
(1,10)	**0.035**	(4,9)	0.051	(8,10)	**0.034**
(1,11)	0.043	(4,10)	0.049	(8,11)	0.055
(1,12)	0.057	(4,11)	0.05	(8,12)	0.054
(1,13)	0.057	(4,12)	0.056	(8,13)	0.056
(1,14)	0.052	(4,13)	0.044	(8,14)	0.046
(1,15)	0.051	(4,14)	0.05	(8,15)	0.052
(2,3)	0.05	(4,15)	0.053	(9,10)	0.058
(2,4)	**0.029**	(5,6)	0.042	(9,11)	0.051
(2,5)	**0.034**	(5,7)	0.054	(9,12)	0.044
(2,6)	0.038	(5,8)	0.057	(9,13)	0.048
(2,7)	**0.036**	(5,9)	0.045	(9,14)	0.063
(2,8)	0.042	(5,10)	0.052	(9,15)	0.046
(2,9)	0.043	(5,11)	0.039	(10,11)	0.052
(2,10)	0.047	(5,12)	0.038	(10,12)	0.038
(2,11)	0.036	(5,13)	0.058	(10,13)	0.058
(2,12)	0.056	(5,14)	0.05	(10,14)	0.052
(2,13)	0.054	(5,15)	0.057	(10,15)	0.058
(2,14)	0.052	(6,7)	0.057	(11,12)	0.062
(2,15)	0.039	(6,8)	0.059	(11,13)	0.05
(3,4)	0.041	(6,9)	0.037	(11,14)	0.053
(3,5)	**0.036**	(6,10)	0.056	(11,15)	0.045
(3,6)	0.05	(6,11)	0.038	(12,13)	0.055
(3,7)	**0.034**	(6,12)	0.053	(12,14)	0.05
(3,8)	0.04	(6,13)	0.057	(12,15)	0.052
(3,9)	**0.036**	(6,14)	0.038	(13,14)	0.052
(3,10)	0.048	(6,15)	0.053	(13,15)	0.053
(3,11)	0.043	(7,8)	0.047	(14,15)	0.049

Table A.7: Type I Error Study for Adjusted Residuals for Asymmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.055	(3,12)	0.05	(7,9)	**0.036**
(1,3)	0.051	(3,13)	0.055	(7,10)	0.058
(1,4)	**0.035**	(3,14)	0.042	(7,11)	0.037
(1,5)	0.045	(3,15)	0.048	(7,12)	0.044
(1,6)	**0.035**	(4,5)	0.045	(7,13)	0.048
(1,7)	0.045	(4,6)	0.051	(7,14)	0.059
(1,8)	**0.034**	(4,7)	0.054	(7,15)	0.056
(1,9)	0.04	(4,8)	0.043	(8,9)	0.045
(1,10)	**0.032**	(4,9)	0.053	(8,10)	**0.036**
(1,11)	0.042	(4,10)	0.048	(8,11)	0.06
(1,12)	0.048	(4,11)	0.048	(8,12)	0.059
(1,13)	0.054	(4,12)	0.059	(8,13)	0.059
(1,14)	0.04	(4,13)	0.056	(8,14)	0.043
(1,15)	0.056	(4,14)	0.043	(8,15)	0.053
(2,3)	0.05	(4,15)	**0.067**	(9,10)	0.057
(2,4)	**0.029**	(5,6)	0.046	(9,11)	0.051
(2,5)	**0.036**	(5,7)	0.057	(9,12)	0.037
(2,6)	0.038	(5,8)	0.057	(9,13)	0.044
(2,7)	**0.034**	(5,9)	0.046	(9,14)	0.054
(2,8)	0.042	(5,10)	0.051	(9,15)	0.054
(2,9)	0.041	(5,11)	0.042	(10,11)	0.056
(2,10)	0.048	(5,12)	0.044	(10,12)	0.048
(2,11)	0.04	(5,13)	0.045	(10,13)	0.056
(2,12)	0.055	(5,14)	0.06	(10,14)	0.053
(2,13)	0.05	(5,15)	0.047	(10,15)	0.059
(2,14)	0.048	(6,7)	0.057	(11,12)	0.058
(2,15)	0.038	(6,8)	0.058	(11,13)	0.045
(3,4)	0.038	(6,9)	**0.033**	(11,14)	0.059
(3,5)	0.037	(6,10)	0.05	(11,15)	0.055
(3,6)	0.054	(6,11)	0.04	(12,13)	0.062
(3,7)	**0.035**	(6,12)	0.058	(12,14)	0.038
(3,8)	0.039	(6,13)	0.054	(12,15)	0.043
(3,9)	**0.035**	(6,14)	0.044	(13,14)	0.052
(3,10)	0.047	(6,15)	0.04	(13,15)	**0.067**
(3,11)	0.042	(7,8)	0.043	(14,15)	0.056

Table A.8: Type I Error Study for $\bar{\chi}^2_{ij}$ for Asymmetric Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.053	(3,12)	0.052	(7,9)	**0.036**
(1,3)	0.05	(3,13)	0.054	(7,10)	0.057
(1,4)	**0.034**	(3,14)	0.044	(7,11)	0.037
(1,5)	0.038	(3,15)	0.049	(7,12)	0.047
(1,6)	**0.033**	(4,5)	0.049	(7,13)	0.046
(1,7)	0.045	(4,6)	0.05	(7,14)	0.061
(1,8)	**0.034**	(4,7)	0.056	(7,15)	0.056
(1,9)	0.04	(4,8)	0.044	(8,9)	0.044
(1,10)	**0.032**	(4,9)	0.052	(8,10)	**0.036**
(1,11)	0.043	(4,10)	0.046	(8,11)	0.061
(1,12)	0.052	(4,11)	0.049	(8,12)	0.059
(1,13)	0.054	(4,12)	0.055	(8,13)	0.06
(1,14)	0.044	(4,13)	0.051	(8,14)	0.043
(1,15)	0.057	(4,14)	0.045	(8,15)	0.053
(2,3)	0.051	(4,15)	**0.066**	(9,10)	0.057
(2,4)	**0.029**	(5,6)	0.047	(9,11)	0.051
(2,5)	**0.035**	(5,7)	0.055	(9,12)	0.044
(2,6)	**0.036**	(5,8)	0.055	(9,13)	0.046
(2,7)	**0.035**	(5,9)	0.046	(9,14)	0.055
(2,8)	0.042	(5,10)	0.053	(9,15)	0.054
(2,9)	0.041	(5,11)	0.042	(10,11)	0.056
(2,10)	0.049	(5,12)	0.038	(10,12)	0.046
(2,11)	0.039	(5,13)	0.047	(10,13)	0.058
(2,12)	0.055	(5,14)	0.057	(10,14)	0.054
(2,13)	0.052	(5,15)	0.047	(10,15)	0.059
(2,14)	0.053	(6,7)	0.056	(11,12)	0.058
(2,15)	0.038	(6,8)	0.06	(11,13)	0.046
(3,4)	0.037	(6,9)	**0.034**	(11,14)	0.054
(3,5)	0.038	(6,10)	0.051	(11,15)	0.055
(3,6)	0.054	(6,11)	0.041	(12,13)	0.063
(3,7)	**0.035**	(6,12)	0.06	(12,14)	0.043
(3,8)	0.04	(6,13)	0.057	(12,15)	0.043
(3,9)	**0.034**	(6,14)	0.045	(13,14)	0.056
(3,10)	0.047	(6,15)	0.041	(13,15)	**0.067**
(3,11)	0.043	(7,8)	0.042	(14,15)	0.057

Table A.9: Type I Error Study for Orthogonal Components for Zero Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.048	(3,12)	0.054	(7,9)	0.052
(1,3)	0.041	(3,13)	0.049	(7,10)	0.046
(1,4)	0.044	(3,14)	0.056	(7,11)	0.05
(1,5)	0.044	(3,15)	0.052	(7,12)	0.052
(1,6)	0.063	(4,5)	0.063	(7,13)	0.048
(1,7)	0.05	(4,6)	0.045	(7,14)	0.056
(1,8)	0.045	(4,7)	0.047	(7,15)	0.054
(1,9)	0.059	(4,8)	0.04	(8,9)	0.054
(1,10)	0.051	(4,9)	0.05	(8,10)	0.054
(1,11)	0.047	(4,10)	0.043	(8,11)	0.052
(1,12)	0.054	(4,11)	0.046	(8,12)	0.063
(1,13)	0.049	(4,12)	0.051	(8,13)	0.059
(1,14)	0.06	(4,13)	0.061	(8,14)	0.051
(1,15)	0.047	(4,14)	0.051	(8,15)	0.05
(2,3)	0.046	(4,15)	0.063	(9,10)	0.054
(2,4)	**0.033**	(5,6)	0.047	(9,11)	0.044
(2,5)	0.046	(5,7)	0.037	(9,12)	0.039
(2,6)	0.053	(5,8)	0.053	(9,13)	0.063
(2,7)	0.042	(5,9)	0.037	(9,14)	0.05
(2,8)	0.042	(5,10)	0.063	(9,15)	0.046
(2,9)	0.046	(5,11)	0.061	(10,11)	0.062
(2,10)	0.05	(5,12)	0.052	(10,12)	0.039
(2,11)	0.051	(5,13)	0.048	(10,13)	0.05
(2,12)	0.044	(5,14)	0.058	(10,14)	0.053
(2,13)	0.045	(5,15)	0.049	(10,15)	0.053
(2,14)	0.055	(6,7)	0.046	(11,12)	0.051
(2,15)	0.048	(6,8)	0.047	(11,13)	0.049
(3,4)	0.043	(6,9)	0.059	(11,14)	0.057
(3,5)	0.052	(6,10)	0.057	(11,15)	**0.066**
(3,6)	0.054	(6,11)	0.063	(12,13)	0.04
(3,7)	0.058	(6,12)	0.042	(12,14)	0.055
(3,8)	0.037	(6,13)	0.044	(12,15)	0.053
(3,9)	0.05	(6,14)	0.037	(13,14)	0.052
(3,10)	0.055	(6,15)	0.045	(13,15)	0.048
(3,11)	0.039	(7,8)	0.05	(14,15)	0.059

Table A.10: Type I Error Study for Adjusted Residuals for Zero Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.048	(3,12)	0.043	(7,9)	0.055
(1,3)	0.041	(3,13)	0.043	(7,10)	0.049
(1,4)	0.045	(3,14)	0.048	(7,11)	0.049
(1,5)	0.039	(3,15)	0.056	(7,12)	0.048
(1,6)	0.064	(4,5)	0.063	(7,13)	0.05
(1,7)	0.051	(4,6)	0.039	(7,14)	0.056
(1,8)	0.046	(4,7)	0.051	(7,15)	0.044
(1,9)	0.06	(4,8)	0.041	(8,9)	0.055
(1,10)	0.054	(4,9)	0.053	(8,10)	0.048
(1,11)	0.052	(4,10)	0.043	(8,11)	0.055
(1,12)	0.052	(4,11)	0.049	(8,12)	0.053
(1,13)	0.043	(4,12)	0.058	(8,13)	0.051
(1,14)	**0.033**	(4,13)	0.058	(8,14)	0.06
(1,15)	0.054	(4,14)	0.047	(8,15)	0.051
(2,3)	0.046	(4,15)	0.055	(9,10)	0.055
(2,4)	**0.033**	(5,6)	0.042	(9,11)	0.045
(2,5)	0.049	(5,7)	0.042	(9,12)	0.048
(2,6)	0.047	(5,8)	0.047	(9,13)	0.062
(2,7)	0.042	(5,9)	0.04	(9,14)	0.054
(2,8)	0.042	(5,10)	0.06	(9,15)	0.052
(2,9)	0.044	(5,11)	0.062	(10,11)	0.056
(2,10)	0.051	(5,12)	0.053	(10,12)	0.042
(2,11)	0.049	(5,13)	0.047	(10,13)	0.047
(2,12)	0.039	(5,14)	0.047	(10,14)	0.044
(2,13)	0.041	(5,15)	0.043	(10,15)	**0.07**
(2,14)	0.043	(6,7)	0.047	(11,12)	0.05
(2,15)	0.059	(6,8)	0.054	(11,13)	0.04
(3,4)	0.043	(6,9)	0.06	(11,14)	0.052
(3,5)	0.056	(6,10)	0.054	(11,15)	0.059
(3,6)	0.053	(6,11)	0.063	(12,13)	0.049
(3,7)	0.056	(6,12)	0.045	(12,14)	0.044
(3,8)	0.039	(6,13)	0.048	(12,15)	0.05
(3,9)	0.053	(6,14)	0.046	(13,14)	0.058
(3,10)	0.058	(6,15)	0.042	(13,15)	0.042
(3,11)	0.043	(7,8)	0.052	(14,15)	0.041

Table A.11: Type I Error Study for $\bar{\chi}^2_{ij}$ for Zero Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.048	(3,12)	0.044	(7,9)	0.055
(1,3)	0.041	(3,13)	0.042	(7,10)	0.049
(1,4)	0.045	(3,14)	0.049	(7,11)	0.047
(1,5)	0.039	(3,15)	0.056	(7,12)	0.048
(1,6)	0.064	(4,5)	0.063	(7,13)	0.05
(1,7)	0.052	(4,6)	0.047	(7,14)	0.055
(1,8)	0.046	(4,7)	0.054	(7,15)	0.045
(1,9)	0.059	(4,8)	0.042	(8,9)	0.055
(1,10)	0.054	(4,9)	0.054	(8,10)	0.048
(1,11)	0.052	(4,10)	0.044	(8,11)	0.055
(1,12)	0.055	(4,11)	0.047	(8,12)	0.055
(1,13)	0.043	(4,12)	0.059	(8,13)	0.05
(1,14)	**0.03**	(4,13)	0.061	(8,14)	0.06
(1,15)	0.055	(4,14)	0.049	(8,15)	0.052
(2,3)	0.046	(4,15)	0.057	(9,10)	0.054
(2,4)	**0.033**	(5,6)	0.046	(9,11)	0.045
(2,5)	0.048	(5,7)	0.042	(9,12)	0.049
(2,6)	0.047	(5,8)	0.05	(9,13)	0.062
(2,7)	0.041	(5,9)	0.041	(9,14)	0.058
(2,8)	0.042	(5,10)	0.061	(9,15)	0.053
(2,9)	0.044	(5,11)	0.059	(10,11)	0.056
(2,10)	0.05	(5,12)	0.052	(10,12)	0.039
(2,11)	0.05	(5,13)	0.047	(10,13)	0.044
(2,12)	0.042	(5,14)	0.046	(10,14)	0.047
(2,13)	0.039	(5,15)	0.047	(10,15)	0.07
(2,14)	0.045	(6,7)	0.05	(11,12)	0.052
(2,15)	0.059	(6,8)	0.054	(11,13)	0.039
(3,4)	0.043	(6,9)	0.058	(11,14)	0.055
(3,5)	0.057	(6,10)	0.052	(11,15)	0.058
(3,6)	0.052	(6,11)	0.064	(12,13)	0.055
(3,7)	0.056	(6,12)	0.044	(12,14)	0.043
(3,8)	0.039	(6,13)	0.048	(12,15)	0.05
(3,9)	0.053	(6,14)	0.047	(13,14)	0.054
(3,10)	0.058	(6,15)	0.043	(13,15)	0.039
(3,11)	0.044	(7,8)	0.052	(14,15)	0.046

Table A.12: Type I Error Study for Orthogonal Components for Zero Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.041	(3,12)	0.048	(7,9)	0.045
(1,3)	0.044	(3,13)	0.052	(7,10)	0.05
(1,4)	0.045	(3,14)	0.039	(7,11)	0.045
(1,5)	0.056	(3,15)	0.045	(7,12)	0.052
(1,6)	0.051	(4,5)	0.05	(7,13)	0.051
(1,7)	0.053	(4,6)	0.046	(7,14)	0.047
(1,8)	0.052	(4,7)	0.044	(7,15)	0.053
(1,9)	0.055	(4,8)	0.055	(8,9)	0.044
(1,10)	0.039	(4,9)	0.052	(8,10)	0.048
(1,11)	0.047	(4,10)	0.05	(8,11)	0.055
(1,12)	0.048	(4,11)	0.049	(8,12)	0.06
(1,13)	0.039	(4,12)	0.054	(8,13)	0.045
(1,14)	0.056	(4,13)	0.052	(8,14)	0.051
(1,15)	0.044	(4,14)	0.051	(8,15)	0.049
(2,3)	0.061	(4,15)	**0.066**	(9,10)	0.064
(2,4)	0.047	(5,6)	0.042	(9,11)	0.046
(2,5)	0.058	(5,7)	0.04	(9,12)	0.064
(2,6)	0.054	(5,8)	0.04	(9,13)	0.051
(2,7)	0.054	(5,9)	0.049	(9,14)	0.041
(2,8)	0.041	(5,10)	0.051	(9,15)	0.058
(2,9)	0.048	(5,11)	0.059	(10,11)	0.052
(2,10)	0.049	(5,12)	0.044	(10,12)	0.043
(2,11)	0.058	(5,13)	0.042	(10,13)	0.048
(2,12)	0.048	(5,14)	0.047	(10,14)	0.059
(2,13)	0.037	(5,15)	0.049	(10,15)	0.054
(2,14)	0.062	(6,7)	0.041	(11,12)	0.049
(2,15)	0.037	(6,8)	0.053	(11,13)	0.045
(3,4)	0.059	(6,9)	0.049	(11,14)	0.048
(3,5)	0.047	(6,10)	0.044	(11,15)	0.057
(3,6)	0.046	(6,11)	**0.066**	(12,13)	0.054
(3,7)	0.05	(6,12)	0.046	(12,14)	0.049
(3,8)	0.044	(6,13)	0.042	(12,15)	0.053
(3,9)	0.042	(6,14)	0.051	(13,14)	0.05
(3,10)	0.046	(6,15)	0.056	(13,15)	0.051
(3,11)	0.049	(7,8)	0.046	(14,15)	**0.067**

Table A.13: Type I Error Study for Adjusted Residuals for Zero Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.041	(3,12)	0.048	(7,9)	0.039
(1,3)	0.044	(3,13)	0.043	(7,10)	0.054
(1,4)	0.047	(3,14)	0.051	(7,11)	0.046
(1,5)	0.056	(3,15)	**0.036**	(7,12)	0.054
(1,6)	0.053	(4,5)	0.05	(7,13)	0.057
(1,7)	0.05	(4,6)	0.05	(7,14)	0.037
(1,8)	0.048	(4,7)	0.044	(7,15)	0.053
(1,9)	0.058	(4,8)	0.055	(8,9)	0.047
(1,10)	0.041	(4,9)	0.054	(8,10)	0.046
(1,11)	0.054	(4,10)	0.05	(8,11)	0.052
(1,12)	0.049	(4,11)	0.048	(8,12)	**0.07**
(1,13)	0.052	(4,12)	0.05	(8,13)	0.049
(1,14)	0.047	(4,13)	0.056	(8,14)	0.053
(1,15)	0.046	(4,14)	0.05	(8,15)	0.055
(2,3)	0.06	(4,15)	0.056	(9,10)	**0.065**
(2,4)	0.049	(5,6)	**0.036**	(9,11)	0.046
(2,5)	**0.066**	(5,7)	0.04	(9,12)	0.056
(2,6)	0.058	(5,8)	0.041	(9,13)	0.053
(2,7)	0.054	(5,9)	0.055	(9,14)	0.048
(2,8)	**0.036**	(5,10)	0.047	(9,15)	0.05
(2,9)	0.048	(5,11)	0.055	(10,11)	0.054
(2,10)	0.049	(5,12)	0.05	(10,12)	0.048
(2,11)	0.063	(5,13)	0.05	(10,13)	0.061
(2,12)	0.045	(5,14)	0.041	(10,14)	0.049
(2,13)	0.042	(5,15)	0.05	(10,15)	0.056
(2,14)	0.056	(6,7)	0.041	(11,12)	0.063
(2,15)	0.062	(6,8)	0.051	(11,13)	0.042
(3,4)	0.061	(6,9)	0.051	(11,14)	0.059
(3,5)	0.051	(6,10)	0.048	(11,15)	**0.067**
(3,6)	0.038	(6,11)	**0.07**	(12,13)	0.056
(3,7)	0.052	(6,12)	0.046	(12,14)	0.042
(3,8)	0.049	(6,13)	0.052	(12,15)	0.049
(3,9)	0.044	(6,14)	**0.035**	(13,14)	0.06
(3,10)	0.046	(6,15)	0.053	(13,15)	0.054
(3,11)	0.05	(7,8)	0.043	(14,15)	0.054

Table A.14: Type I Error Study for $\bar{\chi}^2_{ij}$ for Zero Intercept Model, q=15, n=300

Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.	Pair (i,j)	Orthgonal Comp.
(1,2)	0.041	(3,12)	0.045	(7,9)	0.039
(1,3)	0.044	(3,13)	0.046	(7,10)	0.055
(1,4)	0.047	(3,14)	0.05	(7,11)	0.046
(1,5)	0.056	(3,15)	**0.035**	(7,12)	0.052
(1,6)	0.054	(4,5)	0.048	(7,13)	0.056
(1,7)	0.049	(4,6)	0.05	(7,14)	0.039
(1,8)	0.048	(4,7)	0.043	(7,15)	0.052
(1,9)	0.058	(4,8)	0.054	(8,9)	0.047
(1,10)	0.041	(4,9)	0.053	(8,10)	0.047
(1,11)	0.053	(4,10)	0.053	(8,11)	0.052
(1,12)	0.048	(4,11)	0.048	(8,12)	**0.07**
(1,13)	0.053	(4,12)	0.051	(8,13)	0.047
(1,14)	0.046	(4,13)	0.062	(8,14)	0.053
(1,15)	0.047	(4,14)	0.051	(8,15)	0.054
(2,3)	0.06	(4,15)	0.056	(9,10)	**0.065**
(2,4)	0.046	(5,6)	0.04	(9,11)	0.045
(2,5)	**0.067**	(5,7)	0.038	(9,12)	0.057
(2,6)	0.057	(5,8)	0.039	(9,13)	0.053
(2,7)	0.054	(5,9)	0.053	(9,14)	0.047
(2,8)	**0.036**	(5,10)	0.046	(9,15)	0.05
(2,9)	0.048	(5,11)	0.055	(10,11)	0.054
(2,10)	0.05	(5,12)	0.05	(10,12)	0.052
(2,11)	0.063	(5,13)	0.054	(10,13)	0.064
(2,12)	0.044	(5,14)	0.041	(10,14)	0.049
(2,13)	0.041	(5,15)	0.05	(10,15)	0.056
(2,14)	0.056	(6,7)	0.041	(11,12)	0.062
(2,15)	0.061	(6,8)	0.049	(11,13)	0.042
(3,4)	0.062	(6,9)	0.05	(11,14)	0.059
(3,5)	0.051	(6,10)	0.051	(11,15)	**0.066**
(3,6)	0.037	(6,11)	**0.071**	(12,13)	0.06
(3,7)	0.052	(6,12)	0.049	(12,14)	0.042
(3,8)	0.049	(6,13)	0.054	(12,15)	0.051
(3,9)	0.044	(6,14)	0.037	(13,14)	0.059
(3,10)	0.045	(6,15)	0.053	(13,15)	0.052
(3,11)	0.05	(7,8)	0.042	(14,15)	0.056

Table A.15: Asymptotic and Empirical Power Comparison for Zero Intercept Model

	n=300				n=500			
Pair (i,j)	Orth. Comp.	Adj. Res.	$\bar{\chi}^2_{ij}$	Asym. Power	Orth. Comp.	Adj. Res.	$\bar{\chi}^2_{ij}$	Asym. Power
(1,2)	0.1332665	0.1412826	0.141	0.05393	0.108	0.108	0.106	0.05657
(1,3)	0.1482966	0.1533066	0.155	0.05727	0.132	0.12	0.12	0.06215
(1,4)	0.0801603	0.1082164	0.114	0.08297	0.096	0.119	0.12	0.10549
(1,5)	0.0981964	0.1042084	0.108	0.08496	0.132	0.154	0.15	0.10886
(1,6)	0.1813627	0.1042084	0.105	0.08713	0.163	0.122	0.122	0.11254
(1,7)	0.0691383	0.1342685	0.134	0.05	0.049	0.106	0.101	0.05
(1,8)	0.0851703	0.1232465	0.129	0.05001	0.059	0.096	0.095	0.05002
(2,3)	0.1603206	0.1533066	0.152	0.06543	0.126	0.108	0.107	0.07585
(2,4)	0.0971944	0.1072144	0.112	0.10329	0.151	0.14	0.141	0.13995
(2,5)	0.1072144	0.0831663	0.089	0.10693	0.154	0.118	0.117	0.14613
(2,6)	0.2154309	0.1002004	0.099	0.11096	0.215	0.14	0.139	0.15297
(2,7)	0.0711423	0.1472946	0.144	0.05	0.06	0.104	0.103	0.05
(2,8)	0.0841683	0.1392786	0.141	0.05	0.078	0.105	0.103	0.05
(3,4)	0.1543086	0.1142285	0.113	0.15727	0.23	0.138	0.141	0.23102
(3,5)	0.1853707	0.1192385	0.119	0.16609	0.254	0.142	0.139	0.24571
(3,6)	0.2885772	0.1182365	0.122	0.17602	0.352	0.143	0.15	0.26217
(3,7)	0.1052104	0.1432866	0.14	0.05	0.1	0.111	0.109	0.05
(3,8)	0.0591182	0.1332665	0.131	0.05001	0.06	0.103	0.107	0.05002
(4,5)	0.761523	0.8456914	0.843	0.93157	0.927	0.956	0.957	0.99363
(4,6)	0.7985972	0.8637275	0.864	0.9441	0.94	0.963	0.964	0.99564
(4,7)	0.0390782	0.0651303	0.065	0.05	0.043	0.096	0.098	0.05
(4,8)	0.0480962	0.0861723	0.086	0.05007	0.045	0.074	0.08	0.05011
(5,6)	0.8897796	0.8547094	0.851	0.95534	0.982	0.957	0.957	0.99714
(5,7)	0.0450902	0.0801603	0.081	0.05	0.035	0.079	0.081	0.05
(5,8)	0.0470942	0.0941884	0.093	0.05002	0.04	0.086	0.086	0.05004
(6,7)	0.0561122	0.0741483	0.078	0.05001	0.053	0.084	0.082	0.05001
(6,8)	0.0561122	0.0811623	0.079	0.05002	0.046	0.093	0.093	0.05003
(7,8)	0.0651303	0.1412826	0.139	0.05001	0.062	0.112	0.117	0.05001

* Asymptotic power was calculated only for the orthogonal components.

Table A.16: Asymptotic and Empirical Power Comparison for Asymmetric Intercept Model

	n=300				n=500			
Pair (i,j)	Orth. Comp.	Adj. Res.	$\bar{\chi}^2_{ij}$	Asym. Power	Orth. Comp.	Adj. Res.	$\bar{\chi}^2_{ij}$	Asym. Power
(1,2)	0.1183673	0.1336735	0.121	0.05388	0.197	0.197	0.198	0.05648
(1,3)	0.1081633	0.122449	0.112	0.05741	0.241	0.228	0.226	0.06238
(1,4)	0.0744898	0.1142857	0.104	0.08196	0.079	0.119	0.117	0.10378
(1,5)	0.0795918	0.0938776	0.092	0.0859	0.096	0.12	0.12	0.11045
(1,6)	0.1632653	0.1061224	0.101	0.09008	0.212	0.105	0.106	0.11754
(1,7)	0.0653061	0.1397959	0.128	0.05001	0.065	0.152	0.156	0.05002
(1,8)	0.0857143	0.1255102	0.115	0.05019	0.092	0.12	0.118	0.05032
(2,3)	0.1306122	0.15	0.138	0.06359	0.281	0.242	0.242	0.07276
(2,4)	0.094898	0.1142857	0.108	0.10114	0.127	0.151	0.152	0.1363
(2,5)	0.0867347	0.0969388	0.092	0.10792	0.13	0.133	0.131	0.1478
(2,6)	0.1867347	0.1020408	0.1	0.11529	0.293	0.108	0.108	0.1603
(2,7)	0.0785714	0.1459184	0.133	0.05001	0.056	0.178	0.178	0.05001
(2,8)	0.0785714	0.1530612	0.143	0.05	0.072	0.135	0.135	0.05
(3,4)	0.1204082	0.1173469	0.119	0.14761	0.198	0.15	0.152	0.21484
(3,5)	0.1479592	0.1122449	0.11	0.1618	0.21	0.131	0.133	0.23858
(3,6)	0.2826531	0.1153061	0.122	0.17762	0.453	0.131	0.132	0.26481
(3,7)	0.105102	0.172449	0.163	0.05007	0.093	0.213	0.213	0.05011
(3,8)	0.0795918	0.1530612	0.147	0.05001	0.067	0.175	0.172	0.05002
(4,5)	0.6173469	0.7510204	0.737	0.85718	0.697	0.8	0.798	0.97434
(4,6)	0.6826531	0.772449	0.753	0.89549	0.718	0.781	0.777	0.98582
(4,7)	0.0336735	0.1091837	0.11	0.05002	0.037	0.094	0.095	0.05003
(4,8)	0.0510204	0.1030612	0.104	0.05063	0.051	0.088	0.088	0.05105
(5,6)	0.8081633	0.7704082	0.76	0.93651	0.864	0.782	0.781	0.99447
(5,7)	0.0336735	0.0928571	0.096	0.05	0.034	0.101	0.099	0.05001
(5,8)	0.0632653	0.1142857	0.116	0.05059	0.074	0.084	0.082	0.05098
(6,7)	0.0581633	0.1040816	0.102	0.05014	0.066	0.103	0.103	0.05024
(6,8)	0.0632653	0.1142857	0.116	0.05042	0.065	0.089	0.088	0.05071
(7,8)	0.0714286	0.1897959	0.179	0.05008	0.052	0.181	0.181	0.05014

* Asymptotic power was calculated only for the orthogonal components.

Table A.17: Asymptotic and Empirical Power Comparison for Symmetric Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Residuals	$\bar{\chi}^2_{ij}$	Asymptotic power
(1,2)	0.067	0.069	0.066	0.05092
(1,3)	0.069	0.07	0.069	0.05149
(1,4)	0.05	0.052	0.051	0.0639
(1,5)	0.07	0.069	0.069	0.0664
(1,6)	0.076	0.07	0.071	0.06859
(1,7)	0.058	0.067	0.069	0.05007
(1,8)	0.055	0.065	0.067	0.05013
(1,9)	0.062	0.067	0.065	0.05167
(1,10)	0.071	0.068	0.064	0.05357
(1,11)	0.074	0.049	0.05	0.06118
(1,12)	0.062	0.075	0.074	0.05537
(1,13)	0.055	0.073	0.075	0.05632
(1,14)	0.069	0.067	0.067	0.0567
(1,15)	0.046	0.049	0.049	0.05055
(2,3)	0.071	0.071	0.072	0.0523
(2,4)	0.062	0.065	0.065	0.07208
(2,5)	0.074	0.067	0.067	0.07616
(2,6)	0.085	0.076	0.074	0.07975
(2,7)	0.062	0.079	0.08	0.05007
(2,8)	0.057	0.076	0.071	0.05014

(2,9)	0.059	0.072	0.073	0.05244
(2,10)	0.077	0.061	0.061	0.05564
(2,11)	0.086	0.046	0.049	0.07294
(2,12)	0.052	0.074	0.074	0.05524
(2,13)	0.069	0.09	0.091	0.05782
(2,14)	0.07	0.069	0.07	0.06115
(2,15)	0.055	0.041	0.04	0.05003
(3,4)	0.085	0.078	0.081	0.08574
(3,5)	0.106	0.098	0.1	0.09222
(3,6)	0.12	0.091	0.092	0.09788
(3,7)	0.067	0.073	0.073	0.0501
(3,8)	0.071	0.072	0.077	0.05018
(3,9)	0.084	0.084	0.082	0.05373
(3,10)	0.099	0.08	0.08	0.05848
(3,11)	0.1	0.065	0.068	0.08477
(3,12)	0.071	0.096	0.095	0.05886
(3,13)	0.063	0.087	0.085	0.06291
(3,14)	0.067	0.085	0.084	0.06803
(3,15)	0.047	0.032	0.033	0.05014
(4,5)	0.826	0.855	0.855	0.94927
(4,6)	0.893	0.896	0.898	0.96507
(4,7)	0.06	0.113	0.114	0.05197
(4,8)	0.073	0.111	0.11	0.05362
(4,9)	0.113	0.142	0.138	0.10065
(4,10)	0.198	0.119	0.122	0.17675
(4,11)	0.408	0.101	0.101	0.57638
(4,12)	0.335	0.628	0.627	0.33529
(4,13)	0.305	0.494	0.494	0.3553
(4,14)	0.272	0.295	0.298	0.3734
(4,15)	0.058	0.053	0.056	0.05029
(5,6)	0.933	0.923	0.922	0.98636
(5,7)	0.072	0.115	0.114	0.05312
(5,8)	0.074	0.117	0.119	0.05567
(5,9)	0.16	0.162	0.164	0.11923
(5,10)	0.243	0.132	0.13	0.22204
(5,11)	0.52	0.12	0.12	0.70376
(5,12)	0.448	0.763	0.761	0.43548
(5,13)	0.417	0.634	0.633	0.46222
(5,14)	0.373	0.438	0.439	0.4841
(5,15)	0.032	0.073	0.073	0.05009
(6,7)	0.072	0.117	0.117	0.05465
(6,8)	0.105	0.11	0.109	0.05837
(6,9)	0.175	0.153	0.153	0.1405
(6,10)	0.298	0.155	0.153	0.27253
(6,11)	0.604	0.124	0.125	0.80744
(6,12)	0.502	0.808	0.808	0.54024
(6,13)	0.514	0.703	0.703	0.57369
(6,14)	0.523	0.572	0.571	0.59835
(6,15)	0.042	0.072	0.074	0.05003
(7,8)	0.082	0.114	0.115	0.05003
(7,9)	0.094	0.123	0.127	0.05028
(7,10)	0.108	0.121	0.12	0.05051
(7,11)	0.101	0.111	0.11	0.05184
(7,12)	0.05	0.102	0.103	0.05015

(7,13)	0.043	0.082	0.08	0.05088
(7,14)	0.055	0.087	0.088	0.05209
(7,15)	0.054	0.053	0.055	0.05003
(8,9)	0.114	0.131	0.133	0.05064
(8,10)	0.1	0.123	0.125	0.05097
(8,11)	0.116	0.118	0.119	0.05313
(8,12)	0.053	0.104	0.104	0.05026
(8,13)	0.046	0.088	0.088	0.05153
(8,14)	0.068	0.083	0.083	0.05358
(8,15)	0.068	0.058	0.056	0.05
(9,10)	0.161	0.131	0.131	0.07128
(9,11)	0.202	0.132	0.131	0.0907
(9,12)	0.052	0.139	0.136	0.06431
(9,13)	0.071	0.128	0.127	0.07482
(9,14)	0.104	0.1	0.101	0.08631
(9,15)	0.056	0.084	0.083	0.05002
(10,11)	0.417	0.123	0.121	0.4527
(10,12)	0.075	0.129	0.13	0.07479
(10,13)	0.105	0.117	0.118	0.0931
(10,14)	0.153	0.094	0.094	0.11431
(10,15)	0.063	0.07	0.07	0.05015
(11,12)	0.217	0.122	0.121	0.29387
(11,13)	0.354	0.086	0.088	0.58302
(11,14)	0.575	0.1	0.101	0.96215
(11,15)	0.145	0.061	0.063	0.07986
(12,13)	0.301	0.75	0.751	0.05106
(12,14)	0.27	0.648	0.649	0.05189
(12,15)	0.048	0.048	0.048	0.06174
(13,14)	0.292	0.607	0.607	0.10883
(13,15)	0.046	0.061	0.061	0.05007
(14,15)	0.046	0.029	0.029	0.05024

* Asymptotic power was calculated only for the orthogonal components.

Table A.18: Asymptotic and Empirical Power Comparison for Zero Intercept Model, q=15, n=500

Pair (i,j)	Orthgonal Comp.	Residuals	$\bar{\chi}^2_{ij}$	Asymptotic power
(1,2)	0.11	0.108	0.109	0.05836
(1,3)	0.136	0.128	0.129	0.06075
(1,4)	0.183	0.199	0.2	0.15852
(1,5)	0.21	0.219	0.222	0.16164
(1,6)	0.201	0.189	0.19	0.16489
(1,7)	0.086	0.094	0.094	0.0506
(1,8)	0.085	0.085	0.086	0.05087
(1,9)	0.111	0.114	0.114	0.05743
(1,10)	0.144	0.119	0.123	0.06321
(1,11)	0.177	0.115	0.112	0.08004
(1,12)	0.119	0.187	0.189	0.14041
(1,13)	0.185	0.229	0.228	0.15335
(1,14)	0.214	0.193	0.191	0.16924
(1,15)	0.061	0.088	0.086	0.0508
(2,3)	0.133	0.111	0.11	0.06457

(2,4)	0.195	0.184	0.185	0.1873
(2,5)	0.207	0.18	0.179	0.19152
(2,6)	0.259	0.211	0.214	0.19594
(2,7)	0.082	0.088	0.086	0.05079
(2,8)	0.089	0.093	0.094	0.05118
(2,9)	0.108	0.1	0.096	0.06022
(2,10)	0.142	0.115	0.11	0.0691
(2,11)	0.192	0.112	0.111	0.09859
(2,12)	0.136	0.201	0.199	0.157
(2,13)	0.196	0.206	0.208	0.17666
(2,14)	0.245	0.19	0.191	0.20215
(2,15)	0.072	0.096	0.097	0.05026
(3,4)	0.258	0.202	0.204	0.23072
(3,5)	0.289	0.205	0.199	0.23628
(3,6)	0.282	0.192	0.192	0.24209
(3,7)	0.103	0.1	0.102	0.05097
(3,8)	0.105	0.105	0.104	0.05145
(3,9)	0.125	0.106	0.105	0.06362
(3,10)	0.171	0.123	0.127	0.0755
(3,11)	0.252	0.111	0.11	0.11509
(3,12)	0.168	0.2	0.198	0.19042
(3,13)	0.197	0.189	0.186	0.21653
(3,14)	0.3	0.204	0.205	0.25034
(3,15)	0.057	0.098	0.092	0.05038
(4,5)	0.941	0.968	0.967	0.99356
(4,6)	0.944	0.966	0.966	0.99471
(4,7)	0.093	0.118	0.117	0.07807
(4,8)	0.131	0.123	0.124	0.09225
(4,9)	0.26	0.2	0.203	0.23007
(4,10)	0.405	0.209	0.209	0.38159
(4,11)	0.647	0.231	0.234	0.74274
(4,12)	0.869	0.96	0.959	0.97181
(4,13)	0.904	0.966	0.968	0.98755
(4,14)	0.952	0.965	0.965	0.99587
(4,15)	0.047	0.139	0.14	0.05008
(5,6)	0.95	0.963	0.963	0.9957
(5,7)	0.121	0.122	0.126	0.07896
(5,8)	0.157	0.133	0.134	0.09359
(5,9)	0.286	0.216	0.219	0.23648
(5,10)	0.457	0.219	0.224	0.39259
(5,11)	0.678	0.211	0.214	0.75716
(5,12)	0.864	0.967	0.967	0.97549
(5,13)	0.91	0.964	0.964	0.98952
(5,14)	0.961	0.957	0.957	0.99667
(5,15)	0.05	0.136	0.138	0.05
(6,7)	0.141	0.137	0.137	0.07983
(6,8)	0.157	0.117	0.117	0.09491
(6,9)	0.304	0.201	0.205	0.24304
(6,10)	0.462	0.224	0.224	0.40376
(6,11)	0.667	0.201	0.203	0.77122
(6,12)	0.875	0.957	0.958	0.97887
(6,13)	0.918	0.964	0.963	0.99127
(6,14)	0.965	0.96	0.961	0.99735
(6,15)	0.052	0.137	0.136	0.05004

(7,8)	0.079	0.101	0.094	0.05004
(7,9)	0.084	0.091	0.092	0.05051
(7,10)	0.096	0.103	0.103	0.05095
(7,11)	0.1	0.099	0.102	0.05238
(7,12)	0.049	0.115	0.115	0.05482
(7,13)	0.071	0.12	0.121	0.05572
(7,14)	0.092	0.131	0.129	0.05691
(7,15)	0.068	0.078	0.075	0.05003
(8,9)	0.095	0.094	0.096	0.05084
(8,10)	0.102	0.111	0.113	0.05154
(8,11)	0.114	0.084	0.082	0.05383
(8,12)	0.053	0.11	0.113	0.05727
(8,13)	0.073	0.122	0.121	0.05871
(8,14)	0.133	0.125	0.129	0.06063
(8,15)	0.074	0.091	0.092	0.05003
(9,10)	0.155	0.106	0.107	0.0659
(9,11)	0.207	0.106	0.104	0.08874
(9,12)	0.103	0.196	0.199	0.1146
(9,13)	0.153	0.201	0.203	0.1291
(9,14)	0.22	0.2	0.2	0.14909
(9,15)	0.062	0.106	0.108	0.05005
(10,11)	0.401	0.103	0.104	0.20518
(10,12)	0.161	0.206	0.214	0.15524
(10,13)	0.206	0.182	0.183	0.19258
(10,14)	0.349	0.207	0.212	0.25319
(10,15)	0.068	0.095	0.095	0.05046
(11,12)	0.464	0.228	0.229	0.57338
(11,13)	0.677	0.212	0.213	0.88042
(11,14)	0.864	0.185	0.186	0.99997
(11,15)	0.301	0.096	0.099	0.52262
(12,13)	0.434	0.953	0.954	0.17059
(12,14)	0.457	0.967	0.966	0.32178
(12,15)	0.081	0.13	0.131	0.17264
(13,14)	0.654	0.952	0.952	0.82486
(13,15)	0.043	0.132	0.13	0.0501
(14,15)	0.048	0.144	0.146	0.05014

* Asymptotic power was calculated only for the orthogonal components.

APPENDIX B

FIGURES

Zero Intercept Model

Figure B.1: Orthogonal Components, n=300

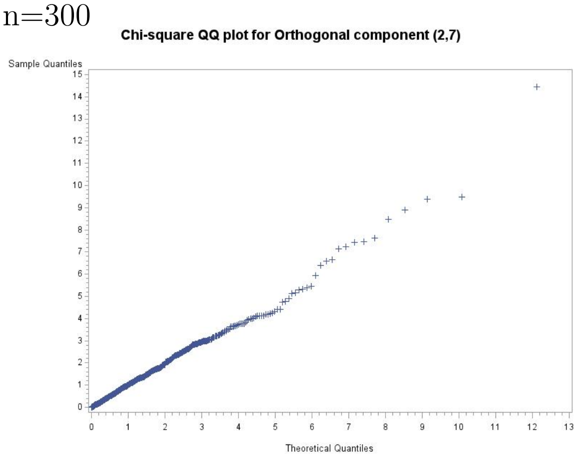


Figure B.2: Orthogonal Components, n=500

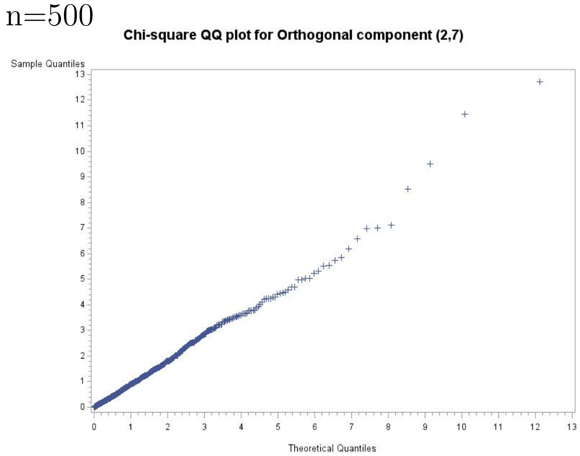


Figure B.3: Adjusted Residuals, n=300

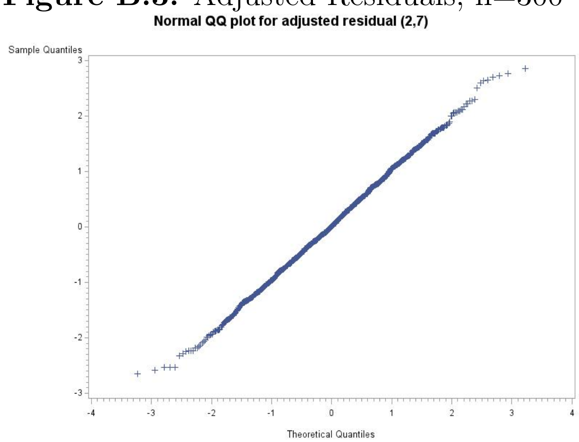


Figure B.4: Adjusted Residuals, n=500

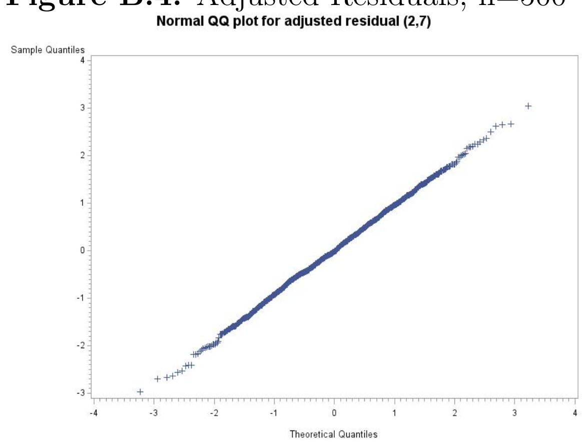


Figure B.5: $\bar{\chi}^2_{ij}$, n=300

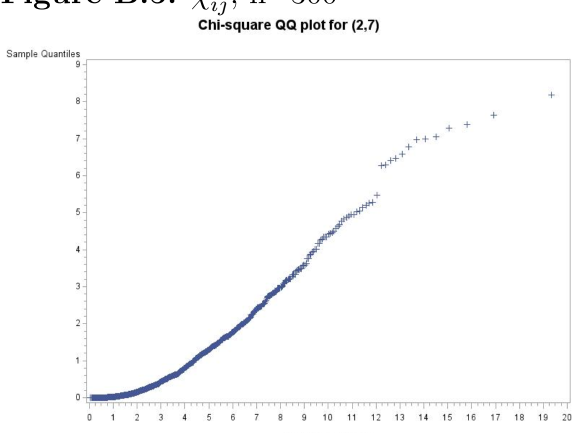
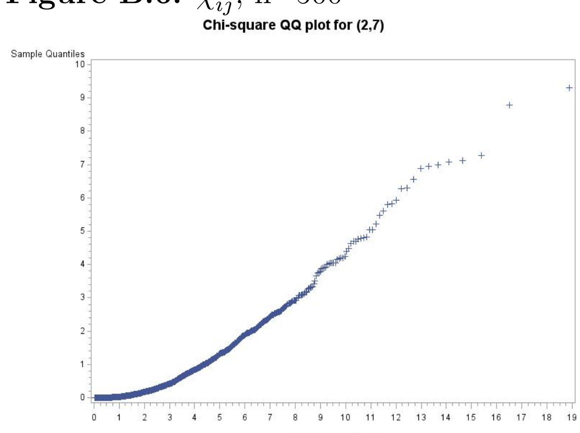


Figure B.6: $\bar{\chi}^2_{ij}$, n=500



Asymmetric Intercept Model

Figure B.7: Orthogonal Components, n=300

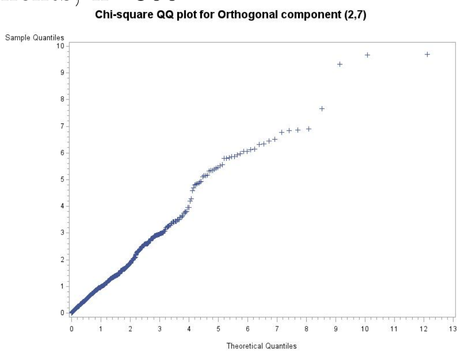


Figure B.8: Orthogonal Components, n=500

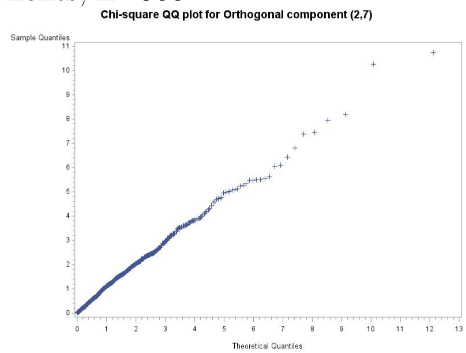


Figure B.9: Adjusted Residuals, n=300

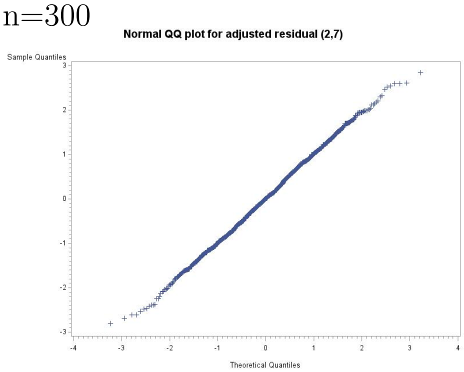


Figure B.10: Adjusted Residuals, n=500

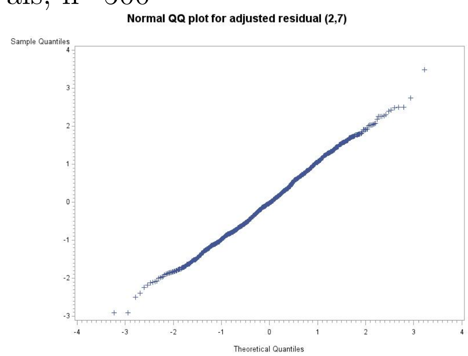


Figure B.11: $\bar{\chi}^2_{ij}$, n=300

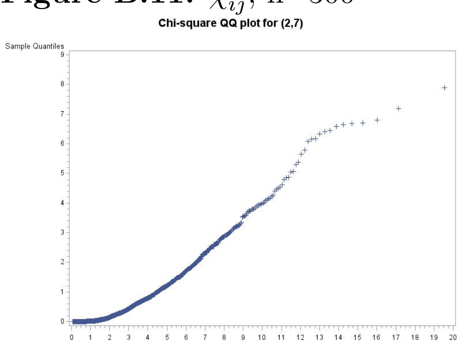


Figure B.12: $\bar{\chi}^2_{ij}$, n=500

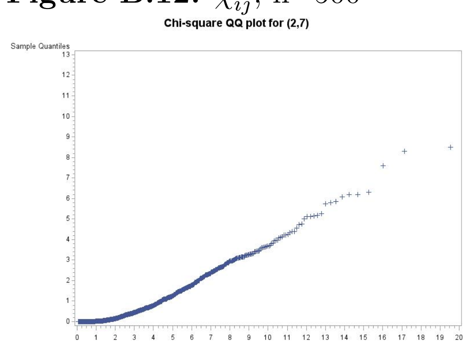


Figure B.13: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Asymmetric Intercept Model, q=15, n=500

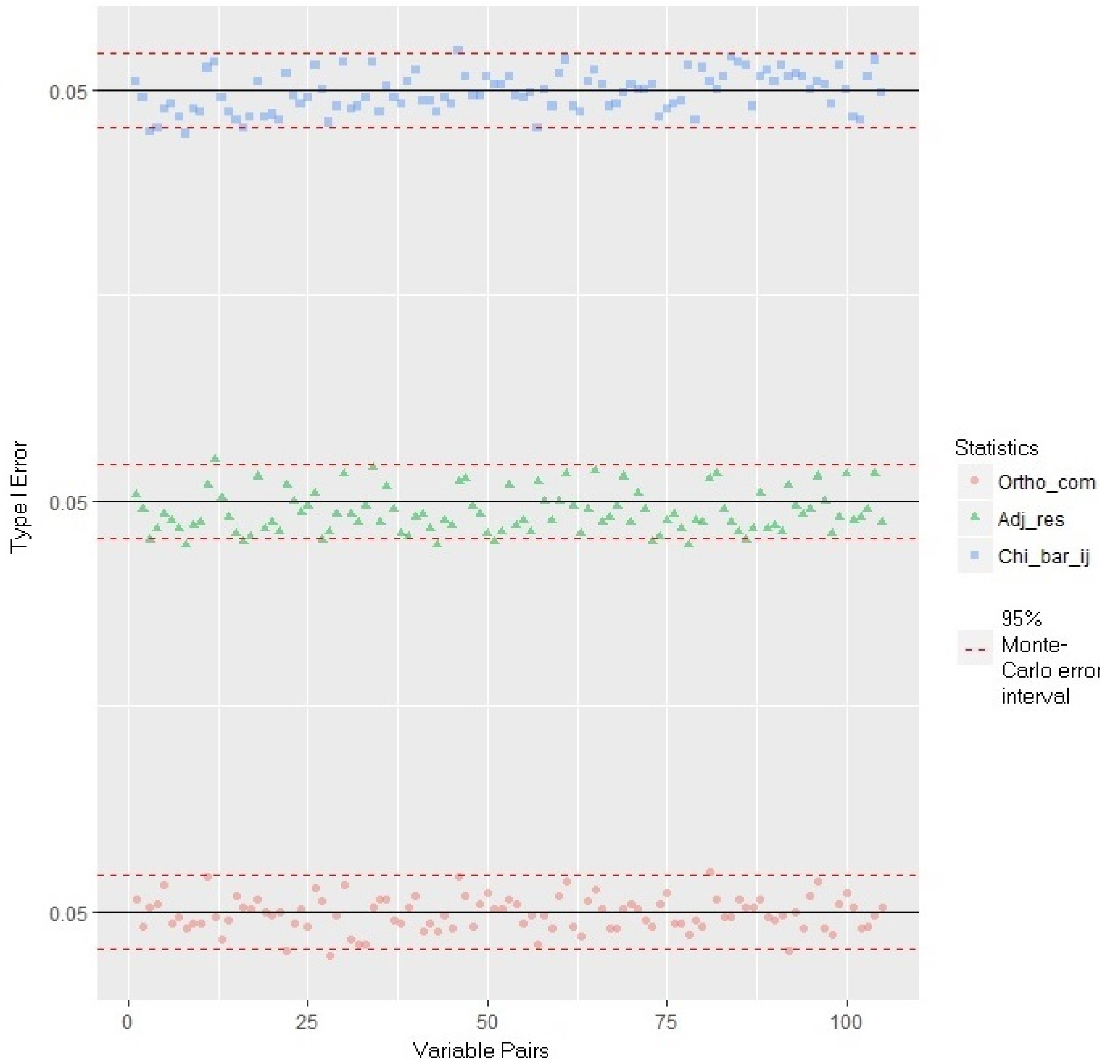


Figure B.14: Type I Error Rates for Orthogonal Components, Adjusted Residuals and $\bar{\chi}^2_{ij}$, Asymmetric Intercept Model, q=15, n=300