# Chi-Square Orthogonal Components for Assessing Goodness-of-fit of Multidimensional Multinomial Data

by

Jelena Milovanovic

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

Approved May 2011 by the
Graduate Supervisory Committee:

Dennis Young, Co-Chair
Mark Reiser, Co-Chair
Jeffrey Wilson
Randall Eubank
Yan Yang

ARIZONA STATE UNIVERSITY

August 2011

ABSTRACT

It is common in the analysis of data to provide a goodness-of-fit test to assess the performance of a model. In the analysis of contingency tables, goodness-of-fit statistics are frequently employed when modeling social science, educational or psychological data where the interest is often directed at investigating the association among multi-categorical variables. Pearson's chi-squared statistic is well-known in goodness-of-fit testing, but it is sometimes considered to produce an omnibus test as it gives little guidance to the source of poor fit once the null hypothesis is rejected. However, its components can provide powerful directional tests.

In this dissertation, orthogonal components are used to develop goodness-of-fit tests for models fit to the counts obtained from the cross-classification of multi-category dependent variables. Ordinal categories are assumed. Orthogonal components defined on marginals are obtained when analyzing multi-dimensional contingency tables through the use of the QR decomposition. A subset of these orthogonal components can be used to construct limited-information tests that allow one to identify the source of lack-of-fit and provide an increase in power compared to Pearson's test. These tests can address the adverse effects presented when data are sparse. The tests rely on the set of first- and second-order marginals jointly, the set of second-order marginals only, and the random forest method, a popular algorithm for modeling large complex data sets. The performance of these tests is compared to the likelihood ratio test as well as to tests based on orthogonal polynomial components.

The derived goodness-of-fit tests are evaluated with studies for detecting two- and three-way associations that are not accounted for by a categorical variable factor model with a single latent variable. In addition the tests are used to investigate the case when the model misspecification involves parameter constraints for large and sparse contingency tables.

The methodology proposed here is applied to data from the 38th round of the State Survey conducted by the Institute for Public Policy and Michigan State University Social Research (2005). The results illustrate the use of the proposed techniques in the context of a sparse data set.

To my family and Matori

ACKNOWLEDGMENTS

I would like to thank my two co-advisors, Dr. Mark Reiser and Dr. Dennis Young, and my two colleagues Dr. Randall Eubank and Dr. Gil.

I would also like to thank the rest of my dissertation committee for reserving some of the precious time to review this work.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

# Chapter 1: INTRODUCTION

It is common in the analysis of data to provide a goodness-of-fit test to assess the performance of a model. In the analysis of contingency tables, goodness-of-fit statistics are frequently employed when modeling social science, educational or psychological data where the interest is often directed at investigating the association among multi-categorical variables. One of the most investigated goodness-of-fit tests employed in this context is the traditional Pearson's chi-squared test. Pearson's chi-squared statistic is well-known in goodness-of-fit testing, but it is sometimes considered to produce an *omnibus* test as it gives little guidance to the source of poor fit once the null hypothesis is rejected. However, its components can provide powerful *directional* tests.

In this dissertation, limited-information goodness-of-fit tests are obtained through the use of a subset of the orthogonal components defined on marginals. Limited-information tests based on marginals are expanded to be applicable to large contingency tables formed by the cross-classification of ordinal categorical variables through the use of the QR decomposition which decomposes a matrix $A$ into an orthogonal matrix $Q$ and an upper triangular matrix $R$ i.e., $A = QR$. Orthogonal components are obtained from the decomposition of Pearson's chi-squared statistic. The QR decomposition can be applied to any matrix, unlike the Cholesky decomposition which only works well with positive definite and nonsingular diagonally dominant square matrices. In addition the random forest method, a commonly known algorithm for modeling large complex data sets, is used to develop and investigate some goodness-of-fit tests.

Rayner and Best (1989) suggested that Pearson's chi-squared test is used to test goodness-of-fit when particular alternatives are not specified; however if they were specified, more powerful directional tests could be applied. Rayner and Best emphasized that the user has the flexibility to construct test statistics that are more appropriate for the problem at hand by forming a focused test statistic that is obtained by summing a "small" number of components of Pearson's chi-squared statistic. The limited-information test is designed to either improve the chi-square approximation or the power against specified alternatives. Consequently, using subsets of these orthogonal components leads to the construction of new limited-information tests that allow one to identify the source for the lack of fit, increase the power and decrease the dilution.

These proposed tests rely on the set of first- and second-order marginals jointly, the set of second-order marginals only, or the random forest method, a popular algorithm for modeling large complex data sets.

Often, in large multi-dimensional contingency tables many response patterns which are used to identify specific cells might not occur in the sample or might have small frequencies resulting in sparse data. The size of a contingency table grows exponentially with the number of variables. For example, in a study of social life feelings, Schuessler (1982) produced a self-determination scale consisting of 14 items. An item response theory model (IRT) for 14 dichotomous items would require a table with $2^{14} = 16,384$ cells. An IRT model is a measurement model for categorical responses that provides a framework for analyzing the relationships between item responses and the latent vari-

able or variables that are not directly observable. Examples of such latent variables include intelligence, happiness and satisfaction. Sparse data implies that the $\chi^2$ approximation to the null distribution for Pearson's statistic might not be accurate (Koehler and Larantz 1980). Since marginals are essentially overlapping cells, tests that rely on them should improve the reliability of the asymptotic chi-square distribution when sparseness is present, thereby allowing for the assessment of the model fit.

However, a large multi-dimensional contingency table results in a large number of components irrespective of sparseness. The large number of components raises the issue of how to select the optimal number of components in the construction of limited-information tests. Some authors advocate a small number of components (Eubank 1997; Ledwina 1994; Rayner and Best 1990), whereas others show that a large number of components can be profitable when data are sparse (Inglot et al. 1990). Most suggestions about selecting the number of components exploit some preliminary knowledge about a possible alternative hypothesis. As proposed in this dissertation, an alternative is to use the random forest method to select the number of orthogonal components to use in the construction of a limited-information test. Such an approach is data-driven and would eliminate any disadvantage connected with fixing the number of components in advance if that number is based on inaccurate knowledge about the possible alternative hypothesis (Ledwina 1994).

The remaining chapters of this dissertation are as follows. In Chapter 2 the most commonly used goodness-of-fit statistic, namely Pearson's $\chi^2$, as well as various traditional and limited-information goodness-of-fit statistics

are discussed. Commonly used measures of sparseness in large multi-way 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 generalized linear latent variable model (GLLVM), a model for confirmatory factor analysis of ordinal categorical variables with a logistic regression function. This model is used to investigate the size and power of the proposed tests based on various components while focusing on interpretability and computational practicality. The chapter concludes with a brief review of some data-mining techniques including CART and an ensemble method, random forest (RF), (Breiman et al. 1984).

Chapter 3 describes the proposed computational methods for selecting chi-squared orthogonal components defined on marginals as a means for providing more powerful directional tests for large cross-classified tables. Type I error rates and power are also discussed in order to evaluate and compare the proposed methods to traditional methods by examining asymptotic power and using Monte Carlo simulations.

Chapter 4 discusses power comparisons for the proposed test statistics when the departure from the null hypotheses is in the form of two- or three-way association that is not accounted for by the categorical variable factor model with a single latent variable. Power comparisons for multi-category variables and large sparse contingency tables are given in Chapter 5. The model misspecification in the power comparisons involves parameter constraints using the categorical variable factor model with a single latent variable. Chapter 6 addresses the comparison of tests based on orthogonal polynomial components

to tests based on components defined on marginal frequencies. The proposed tests are applied to real-life data from the 38th round of the State Survey conducted by the Institute for Public Policy and Michigan State University Social Research (2005) in Chapter 7. 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

In this chapter the most investigated goodness-of-fit statistic, namely Pearson's $\chi^2$ statistic, as well as various traditional and limited-information goodness-of-fit statistics are discussed. Commonly used measures of sparseness in large multi-way contingency tables are reviewed, and the adverse effects of sparseness on goodness-of-fit statistics are explained. A brief review of some data-mining techniques including CART (Breiman et al. 1984) and an ensemble method, random forest (RF), is also included.

1. Quadratic Form Statistics for Multinomial Data

The two most commonly used goodness-of-fit statistics are Pearson's $\chi^2$ (PGF) and the likelihood ratio $G^2$ statistic (LR). Both of these have been found to be specific cases of a general family of goodness-of-fit statistics called the power divergence family (Cressie and Reed 1984) that is indexed by the parameter $\lambda$, where $\lambda > 0$. For a given value of $\lambda$ the corresponding test statistic is

$$CR(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{s=1}^{T} \hat{p}_s \left[ \left( \frac{\hat{p}_s}{\hat{\pi}_s} - 1 \right)^{\lambda} \right],$$

where the $\hat{p}_s$ are the observed cell proportions, $\hat{\pi}_s$ are the estimated expected cell proportions and $T$ is the number of cells. The $\chi^2$ and $G^2$ statistics are obtained with $\lambda = 1$ and, as the limiting case, with $\lambda \to 0$, respectively.

When the hypothesized model holds the statistics $\chi^2$ and $G^2$ are asymptotically equivalent in that they both have a chi-square distribution with $T - g - 1$ degrees of freedom, with $g$ being the number of estimated parameters. Investigating Pearson's $\chi^2$ is the primary objective of this research.

The discussion of the $\chi^2$ test begins in the next section with the introduction of Pearson's $\chi^2$ in 1900.

1.1. Traditional Goodness-of-fit Statistic. For a multi-way contingency table, the traditional PGF statistic is obtained by comparing observed frequencies to the expected frequencies specified under the null hypothesis as in

$$\chi^2_P = \sum_{s=1}^{T} \frac{(observed_s - expected_s)^2}{expected_s}.$$

For a simple null hypothesis where the random sample comes from a population with completely specified cumulative distribution function $F(x)$, the PGF has an approximate chi-square distribution with $T-1$ degrees of freedom in large samples and is defined as:

$$\chi^2_P = n \sum_{s=1}^{T} \frac{(\hat{p}_s - {\pi_s}^o)^2}{{\pi_s}^o},$$

where $\hat{p}_s$ is the sample proportion of the $s^{th}$ cell, ${\pi_s}^o$ is the corresponding proportion under the specified null hypothesis and $n$ is the sample size.

On the other hand, for a composite null hypothesis where the null distribution depends on a $g-$vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_g)^T$, the ${\pi_s}^o$ in $\chi^2_P$ above are replaced with corresponding estimated expected proportions $\hat{\pi}_s = \pi_s(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the vector of parameter estimates. This yields the Pearson-Fisher statistic

$$\chi^2_{PF} = n \sum_{s=1}^{T} \frac{(\hat{p}_s - \hat{\pi}_s)^2}{\hat{\pi}_s}.$$

The Pearson-Fisher statistic has an asymptotic chi-square distribution with $T-g-1$ degrees of freedom under the large sample theory conditions (Koehler

and Larantz 1980) that *i*) the null hypothesis is true, *ii*) $T$ is fixed, and *iii*) $\min_{(\,1 \le s \le T\,)}\, n\pi_s \to \infty$ for $n \to \infty$. The PGF phrase is often used to refer to both the Pearson and Pearson-Fisher $\chi^2$ statistic. However, in this research PGF will refer to the Pearson-Fisher $\chi^2$ statistic.

1.2. Decomposition of Pearson Chi-Square Statistic. A goodness-of-fit test such as the PGF is often referred to as an omnibus test. However, decomposing Pearson's chi-squared statistic into components, which has a long history as discussed in Lancaster (1969), can provide powerful directional tests.

A well known and most widely used decomposition of the components may be associated with $(T-1)$ orthonormal functions $\{g_1, g_2, ..., g_{T-1}\}$ 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 Parseval's relation)

$$\chi^2 = \sum_{j=1}^{T-1} \widehat{U^{(j)}}^2,$$

with

$$\widehat{U^{(j)}} = \sum_{s=1}^{T} g_j(x_s),$$

where $x_s$ is the observed value for the $s^{th}$ observation and therefore necessarily in $\{1, 2, ..., T\}$. These have a useful property of breaking the contributions to $\chi^2$ into component pieces that may be associated with $T-1$ orthogonal directions corresponding to the basis functions $\{g_1, g_2, ..., g_{T-1}\}$. Note that orthogonality translates into

$$\sum_{s=1}^{T} g_i(x_s) g_k(x_s) \hat{\pi}_s = \delta_{ik},$$

where $\delta_{ik}$ 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 $\widehat{U^{(j)}}$ are chosen so that they have interesting individual interpretations. Also, $\chi^2 = \sum \widehat{U^{(j)}}^2$ is invariant for any choice of the set $\{g_1, g_2, ..., g_{T-1}\}$, i.e., these can be orthonormalized indicator variables, the Walsh functions, the orthogonal polynomials on $T$ points with equal weights and so on (Lancaster 1969).

Rayner and Best (1989) also considered in detail components using the Chebyshev orthogonal polynomials. 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. Eubank (1997) also used a chi-square component decomposition

$$\chi^2 = n \sum_{s=1}^{T} ((\hat{p}_s - \pi_s^o)/\sqrt{\pi_s^o})^2 = n \sum_{s=1}^{T} \hat{f}(s)^2 = \sum_{j=1}^{T-1} n b_j^2,$$

where $b_j$ are associated (discrete) generalized Fourier coefficients

$$b_j = \sum_{s=1}^{T} \hat{f}(x_s) g_j(x_s) = \sum_{s=1}^{T} \hat{f}(s) g_j(s), \qquad j = 1, ..., T-1,$$

with $x_s = s$ for the $s^{th}$ cell and $g_j$, $j = 1, ..., T-1$, being functions on {1,...,T} satisfying certain orthogonality conditions. Note, that $\hat{f}$ is an unbiased estimator of the function

$$f(s) = (\pi_s - \pi_s^o)/\sqrt{\pi_s^o}, \qquad s = 1, ..., T.$$

The collection of test statistics

$$\chi^2_q = \sum_{j=1}^{q} nb_j^2, \qquad\qquad q = 1, ..., T-1,$$

are essentially directional tests that may outperform omnibus tests such as $\chi^2_{PF}$, provided that the order $q$ of the test is chosen optimally. It is always possible to chose $x_s$ such that $nb_j, j = 1, ..., T-1$, have interesting individual interpretations in the sense of measuring higher-frequency departures from the null as their associated indices increase (Eubank 1997). Moreover, Eubank (1997) defined the optimal value of $q$ to be the one that minimizes

$$\sum_{s=1}^{T}(f_q(s) - f(s))^2, \tag{2.1}$$

where $f_q(s) = \sum_{j=1}^{q} b_j g_j(s)$, or equivalently, maximizes

$$M(q) = -\sum_{j=1}^{q} b_j^2 + 2\sum_{j=1}^{q} \beta_j b_j, \tag{2.2}$$

with associated Fourier coefficients $\beta_j = \sum_{s=1}^{T} f(s)g_j(s),\ j = 1, ..., T-1$. Since, neither of the quantities in (2.1) and (2.2) are observable, a strategy for estimating the optimal $q$ is obtained by maximizing an unbiased estimator of $M$.

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 are that the equiprobable 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.

An alternative partition of Pearson's chi-squared statistic into independent chi-square components is discussed in Agresti (2002). This partition is not based on the orthogonal polynomial decomposition. Agresti (2002) gives the necessary conditions for determining subtables for which components are independent chi-square random variables. The sum of the chi-squared values for any separate subtables do not sum to the overall Pearson's chi-squared statistic.

Let $\mathbf{H}$ be a $q$ by $T$ matrix of constants of rank $q$, and let $\boldsymbol{\pi}_o = (\pi_1^o, \pi_2^o, ..., \pi_T^o)^T$, and set $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_T)^T$. Now consider testing a simple null hypothesis, $H_o : \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}_o$ against $H_a : \mathbf{H}\boldsymbol{\pi} \neq \mathbf{H}\boldsymbol{\pi}_o$. There are many ways to create orthogonal components via a transformation matrix $\mathbf{H}$, and Rayner and Best (1989) presented $\mathbf{H}$ as a general transformation matrix not necessarily specific to produce marginal probabilities. Under certain conditions on $\mathbf{H}$, the corresponding Pearson's chi-square test statistic can be seen as a special case of the score statistic given in Rayner and Best (1989) written as

$$\begin{aligned}
\chi^2_P &= n \left( \frac{\hat{\mathbf{p}} - \boldsymbol{\pi}_o}{\boldsymbol{\pi}_o^{\frac{1}{2}}} \right)^T \mathbf{H}^T \left( \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T \right)^{-1} \mathbf{H} \left( \frac{\hat{\mathbf{p}} - \boldsymbol{\pi}_o}{\boldsymbol{\pi}_o^{\frac{1}{2}}} \right) \\
&= n \left( \frac{\hat{\mathbf{p}} - \boldsymbol{\pi}_o}{\boldsymbol{\pi}_o^{\frac{1}{2}}} \right)^T \mathbf{H}^T \left( \mathbf{H}(\mathbf{I} - \boldsymbol{\pi}_o^{\frac{1}{2}} (\boldsymbol{\pi}_o{}^{\frac{1}{2}})^T) \mathbf{H}^T \right)^{-1} \mathbf{H} \left( \frac{\hat{\mathbf{p}} - \boldsymbol{\pi}_o}{\boldsymbol{\pi}_o^{\frac{1}{2}}} \right) \\
&= n \mathbf{z}^T \mathbf{H}^T \left( \mathbf{H}(\mathbf{I} - \boldsymbol{\pi}_o^{\frac{1}{2}} (\boldsymbol{\pi}_o{}^{\frac{1}{2}})^T) \mathbf{H}^T \right)^{-1} \mathbf{H}\mathbf{z}, \qquad (2.3)
\end{aligned}$$

where $\mathbf{z} = \left( \frac{\hat{\mathbf{p}} - \boldsymbol{\pi}_o}{\boldsymbol{\pi}_o^{\frac{1}{2}}} \right)$ is the vector of standardized residuals, $\boldsymbol{\Sigma}$ is the covariance

matrix of the standardized residuals, $\hat{\mathbf{p}}$ is the vector of observed probabilities and $\boldsymbol{\pi}_o$ is the vector of probabilities specified under the null.

If a composite null hypothesis is tested, $H_o : \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ against $H_a : \mathbf{H}\boldsymbol{\pi} \neq \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ where $\boldsymbol{\pi}(\boldsymbol{\theta})$ is the multinomial vector of cell probabilities that depend on the parameter vector $\boldsymbol{\theta}$, the PGF (according to Rayner and Best 1989) can be written as

$$\begin{aligned} \chi^2_{PF} &= n \left( \frac{\hat{\mathbf{p}} - \hat{\boldsymbol{\pi}}}{\hat{\boldsymbol{\pi}}^{\frac{1}{2}}} \right)^T \mathbf{H}^T \left( \mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}^T \right)^{-1} \mathbf{H} \left( \frac{\hat{\mathbf{p}} - \hat{\boldsymbol{\pi}}}{\hat{\boldsymbol{\pi}}^{\frac{1}{2}}} \right) \\ &= n\mathbf{z}^T\mathbf{H}^T \left( \mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}^T \right)^{-1} \mathbf{H}\mathbf{z}, \end{aligned} \tag{2.4}$$

where $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is the estimated covariance matrix of the standardized residuals evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})$ is the vector of cell probabilities evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$. Namely, $\mathbf{H}$ should be chosen such that in matrix terms $\mathbf{H}\widehat{\mathbf{D}}\mathbf{H}^T = I_{T-g-1}$, $\mathbf{H}\hat{\boldsymbol{\pi}} = \mathbf{0}$ and $\mathbf{H}\widehat{\mathbf{G}}^T = \mathbf{0}$, where $\widehat{\mathbf{D}} = \mathbf{D}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) = \text{diag}(\hat{\pi}_1, ..., \hat{\pi}_T)$, $\widehat{\mathbf{G}} = \frac{\partial \hat{\boldsymbol{\pi}}}{\partial \hat{\boldsymbol{\theta}}}$ and $g$ is the number of estimated parameters.

Since, Rayner and Best (1989) take $\mathbf{H}$ as a general matrix, not necessarily producing marginal probabilities in order to obtain orthogonal polynomial components, $\mathbf{H}$ must be chosen appropriately. Then, under certain conditions, they propose partitioning $\chi^2_{PF}$ into components defined as $\widehat{\mathbf{V}}^T\widehat{\mathbf{V}}$ such that

$$\hat{V}_j = \sum_{s=1}^{T} \frac{N_s \; g_j(x_s)}{\sqrt{n}}, \qquad\qquad j = 1, ..., T-1,$$

where $N_s$ is the frequency of the $s^{th}$ cell, $x_s$ is the observed value for the $s^{th}$ cell, $n = \sum_{s=1}^{T} N_s$ and $g_j(x)$ is a polynomial of degree $j$ in $x$ where

$g_0(x), g_1(x), ..., g_{T-1}(x)$ are orthonormal functions in that

$$\sum_{s=1}^{T} g_i(x_s)\, g_k(x_s)\hat{p}_s = \delta_{ik}, \qquad i, k = 1, ..., T-1,$$

where $\delta_{ik}$ is the Kronecker delta, $\delta_{ik} = 1$ for $i = k$, and $\delta_{ik} = 0$ for $i \neq k$. The Gram-Schmidt method may be used to construct the functions $g_j(x)$ recursively from the relation

$$g_0(x) = \left[\sum_{s=1}^{T} \hat{p}_s\right]^{-\frac{1}{2}} = 1,$$

$$q_j(x_k) = x_k^j - g_{j-1}(x_k) \sum_{s=1}^{T} \hat{p}_s\, x_s^j\, g_{j-1}(x_s) - ... - g_0(x_k) \sum_{s=1}^{T} \hat{p}_s\, x_s^j\, g_0(x_s),$$

where $k = 1, ..., T$ and the normalization

$$g_j(x_k) = \frac{q_j(x_k)}{[\sum_{s=1}^{T} \hat{p}_s\, q_j^2(x_s)]^2},$$

is performed at each step for $k = 1, ..., T$. The $\widehat{V}_j^2$ each have an asymptotic $\chi_1^2$ distribution and are jointly asymptotically independent. In the univariate case the first component is good for detecting changes in location and the second component is sensitive to scale changes. The Cholesky factorization method is, among other things, a way to implement the Gram-Schmidt method.

On the other hand, a more useful decomposition of the PGF statistic for extremely unbalanced non-equiprobable situations and for very sparse multinomials is based on orthogonal components usually defined on low-order marginals; first-order marginals are univariate distributions of variables, second-order marginals are bivariate distributions of variables, etc. The decomposition of the PGF statistic assumes that the regularity conditions (Birch 1964) for the asymptotic chi-square distribution are met. There are numerous

ways to create these orthogonal components which can be obtained via the multiplication of the multinomial vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_T)^T$ by the transformation matrix $\mathbf{H}$. Components based on these low-order marginals are most often justified as easily interpretable because they are related to the model variables and somewhat computationally practical. In general, tests based on a subset of these types of orthogonal components are essentially a low-order cell focusing test (Reiser 2008). They should have higher power than the traditional test to detect a departure from certain null hypotheses for any finite sample size, when the lack-of-fit is expected in lower-order marginals. Also, since low-order marginals are *overlapping* cells, using them to form a test statistic should improve the reliability of the asymptotic chi-square distribution when expected cell frequencies are small or when sparseness is present. Thus, when $\mathbf{H}$ produces marginal probabilities the components of PGF then become both an overlapping cells test and a focused test. Reiser (2008) proposed using (2.4) with $\mathbf{H}$ as a transformation matrix that produced all possible marginal probabilities. Reiser (2008) also partitioned PGF into orthogonal components based on lower-order marginal frequencies using the same Cholesky decomposition as Rayner and Best (1989) for a cross-classification table of $q$ dichotomous variables. He defined $\hat{\boldsymbol{\gamma}}$ as

$$\hat{\boldsymbol{\gamma}} = n^{-\frac{1}{2}} \widehat{\mathbf{H}}^* \mathbf{z}, \tag{2.5}$$

where $\widehat{\mathbf{H}}^* = \widehat{\mathbf{F}}\mathbf{H}$, $\widehat{\mathbf{F}}$ is the matrix $\mathbf{F}$ evaluated at $\hat{\boldsymbol{\theta}}$ with $\mathbf{F}$ being an upper triangular matrix such that $\mathbf{F}^T\boldsymbol{\Sigma}\mathbf{F} = \mathbf{I}$. Specifically, $\mathbf{F} = (\mathbf{C}^T)^{-1}$, where $\mathbf{C}$ is the Cholesky factor of $\boldsymbol{\Sigma}$, the asymptotic covariance matrix of the standardized

residuals. Then,

$$\chi^2_{PF} = \hat{\boldsymbol{\gamma}}^T \hat{\boldsymbol{\gamma}} = \sum_k \hat{\gamma}^2_k. \tag{2.6}$$

The orthogonal components $\hat{\gamma}^2_k$ in expression (2.6) are asymptotically independent $\chi^2_1$ random variables. Although the **H** matrix here must be full rank in order for the Cholesky decomposition algorithm to be numerically stable, Reiser (2008) demonstrated the advantage of orthogonal components in terms of power and avoiding dilution of a test, i.e., watering down the test with extra degrees of freedom.

There are many ways to create orthogonal components via the transformation matrix **H**, other than that based on marginal frequencies. Other options are Hall's (1985) approach, where **H** routinely combines pairs of adjacent cells, using Helmertian matrices where cells are combined only in the sense of contrasts, and combining each small cell probability with a high probability cell. One should keep in mind that some ingenuity may be required in interpreting the alternatives being tested with different **H** matrices (Rayner and Best 1989).

In summary, components based on orthogonal polynomials may not be applicable for large multi-way contingency tables unlike the proposed orthogonal components defined on marginals in this dissertation. The next section discusses in detail the limited-information test statistics that use information contained in the low-order marginals.

1.3. Limited-Information Goodness-of-fit Statistics. One way of remedying the problem of sparseness is to consider limited-information test statis-

tics that are based on information contained only in the low-order marginals, e.g., just the first- and second-order marginals. Any statistic formed from a sum of components, not necessarily ones based on marginal frequencies, can be considered a limited-information statistic. The idea behind this approach of summing a subset of components is that using a limited-information test statistic should further increase the power against the alternatives they can detect and decrease dilution of a test.

The limited-information approach has a long tradition in psychometrics. Christoffersson (1975) first introduced the idea of using first- and second-order marginals for a test of fit in dichotomous variable factor analysis. Muthèn (1978) improved his statistic, but both used observed proportions and neither presented their test as having higher power or as a remedy for sparse data.

Salomaa (1990) showed that lack-of-fit usually occurs in the second-order marginals for psychological data. As a result, this research proposes test statistics using components based on first- and second-order marginals and just second-order marginals. These proposed test statistics will be more focused than PGF while still providing power to detect lack-of-fit in many directions.

Reiser (1996) proposed a limited-information statistic 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 limited-information statistic for testing the fit of latent class models. Similar use of the residuals in the Generalized Linear Latent Variable Model (GLLVM) for binary data was discussed

in Bartholomew and Tzamourani (1999). GLLVM is a model for confirmatory factor analysis of ordinal categorical variables with a logistic regression function. An extension of work by Bartholomew and Tzamourani (1999) is presented by Moustaki (2007) where she proposed a goodness-of-fit test based on the residuals associated with the bivariate marginal distributions for a categorical variable factor model with ordinal variables. Her proposed goodness-of-fit test is over-parameterized compared to the most recent work by Reiser (2008) and is discussed in more detail in Chapter 3.

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. Their approach is closely related to that of Reiser (2008) but their limited-information statistic $M_2$ does not correspond to the same decomposition of the PGF. $M_2$ is discussed in more detail in Chapter 3.

Although overcoming the adverse effects of sparseness may be a reason for using orthogonal components (Reiser 2008), these effects have not been investigated. Other work based on low-order marginals includes Knott and Tzamourani (1997) who suggested that it would be informative to compare observed and fitted values for first-, second- and third-order marginal frequencies when assessing model fit.

Bartholomew and Leung (2002) developed an alternative goodness-of-fit statistic that is computationally simpler than other suggested statistics and can easily be "decomposed" into simple additive pieces to assess the contributions of individual marginals to the poor fit. The distribution of the statistic

was approximated under the simple null hypothesis. Moment adjustments were used to account for the composite null hypotheses, but they did not work very well for the extreme tail areas of the distribution.

Approaches given by Eubank (1997) and Ledwina (1994) can also be considered as limited-information. Eubank used orthogonal polynomial components and estimated the optimal value for the correct order, assuming equiprobable cells. Ledwina (1994) proposed a data-driven method consisting of using Schwatz's BIC procedure to choose the dimension of the exponential model and then using the chosen dimension as the number of components. In this way any disadvantage connected with fixing $q$ in advance that may be based on inaccurate knowledge about the possible alternative model is eliminated.

In summary, components based on first- and second-order marginals as well as only second-order marginals are useful for large multi-way contingency tables. Still, a precise method for determining which and how many of these components to use for the construction of a limited-information test statistic has not been determined.

1.4. Sparseness. The two most commonly used goodness-of-fit statistics are the Pearson's $\chi^2$ (PGF) and the likelihood ratio $G^2$ statistics (LR). In large sample sizes when the model under the null hypothesis holds, the two statistics are asymptotically equivalent and follow an approximate a chi-square distribution with $T - g - 1$ degrees of freedom, where $g$ is the number of estimated parameters. However, the problem of sparse data can arise in the presence of a large number of variables with several categories, as the num-

ber of possible response patterns can be very large. Even with a moderate sample size, many response patterns may not be realized or might have small frequencies. Sparse data have an adverse effect on goodness-of-fit tests as they may invalidate using the chi-square distribution as an approximation for the distribution of PGF and LR (Agresti and Yang 1987).

It is known that in sparse tables the empirical Type I error rates of both PGF and LR often do not match their expected rates under the chi-square approximation; in fact Cochran (1952) thought that both PGF and LR are asymptotically normally distributed under sparseness. However, according to Koehler and Larantz (1980) who examined the accuracy of the chi-squared and normal approximations for PGF and LR via a Monte Carlo study, it was found that in general the chi-squared approximation for PGF 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, LR is not well approximated by a chi-squared distribution when $n/T \leq 5$. Several other simulation studies have also confirmed that the PGF approximates a chi-squared random variable more closely than LR (Agresti and Yang 1987).

Many suggestions have been given on how to measure sparseness in multi-way contingency table. But, to date, no universal definition of sparseness has been adopted. 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 (Cochran 1954; Agresti and Yang 1987; Fisher 1941; Cramer 1946; Kendall 1952; Tate and Hyer 1973; Lancaster 1969), 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. Moreover, extreme parameters should also be considered. In existing literature, the extremity of parameters was recognized as an important factor directly contributing to the expected values. Furthermore, the extremity of parameters is also very important because the more extreme the parameters are, the more likely it is that there will be boundary parameters and estimated zeros, and thus resulting in an undefined chi-square statistic (Tollenaar and Mooijaart 2003).

Tollenaar and Mooijaart (2003) proposed a measure of sparseness $v_{eq}$, a special version of Cohen's (1988) effect size $w$, defined as

$$v_{eq} = \sqrt{\sum_{s=1}^{T} \frac{(\pi_s^M - \pi_s^{eq})^2}{\pi_s^{eq}}},$$

where $\pi_s^M$ is the probability of cell $s$ under the model and $\pi_s^{eq}$ is the probability of cell $s$ under equiprobability. Note that this measure is independent of the sample size. They showed severe sparseness present with $v_{eq} > 2$ in conjunction with $n/T$ ratios of 1 and 3.

Using limited-information statistics as a potential solution for overcoming adverse effects of sparseness has been studied by 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), Maydeu-Olivares and Joe (2006) and Moustaki (2007). When sparseness is present, overlapping cells have an advantage of improving the $\chi^2$ approximation. If a cell has a small expectation, combining cells in this manner can give a more moderate expectation improving the $\chi^2$ approximation under the null distribution. An overlapping cells approach which does not use marginal frequencies is Hall's method (Hall 1985) which involves routinely combining pairs of adjacent cells.

Other solutions include adding a small constant to the frequency of every response pattern (which could cause havoc with the distribution of the Pearson statistic), 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 $\chi^2$ and $G^2$ (Bartholomew and Tzamourani 1999 and Tollenaar and Mooijaart 2003) has become increasingly popular given today's computing power. However, this method is computationally intense since in order to obtain a stable $p$-value several hundred bootstrap resamples are needed for each model the researcher is interested in comparing (Bartholomew and Leung 2002).

On the other hand, according to Agresti and Yang (1987), a beneficial aspect of spareness is that the power of certain single-degree-of-freedom test

statistics, e.g., likelihood ratio test, tends to increase as the table becomes more sparse for a fixed sample size. The likelihood ratio statistic can be written as

$$\Lambda = \frac{\sup\{\boldsymbol{L}(\theta|\ x) : \theta \in \Theta_o\}}{\sup\{\boldsymbol{L}(\theta|\ x) : \theta \in \Theta\}},$$

where $\mathbf{L}(\theta|\ x)$ is the likelihood function, the "sup" notation refers to the Supremum function and $\Theta_o$ is a specified subset of the parameter space $\Theta$. In most cases, however, the exact distribution of the likelihood ratio corresponding to a specific hypotheses is difficult to determine. A convenient result says that as $n \to \infty$, the test statistic $-2\log(\Lambda)$ for a nested model will be asymptotically $\chi^2$ distributed with degrees of freedom equal to the difference in dimensionality of $\Theta$ and $\Theta_o$.

If a hypothesis can be expressed as the condition that some model $\mathcal{M}_1$ holds, and moreover, if it is possible to imbed that model in a slightly more complex model $\mathcal{M}_2$ that reflects the pattern of departures from the hypothesis one expects then a likelihood ratio difference statistic ($LR_{diff}$) can be computed as the difference of two likelihood ratio statistics where

$$\Lambda_{\mathcal{M}_i} = \frac{\sup\{\boldsymbol{L}(\theta_{\mathcal{M}_i}|\ x) : \theta_{\mathcal{M}_i} \in \Theta_o\}}{\sup\{\boldsymbol{L}(\theta_{\mathcal{M}_i}|\ x) : \theta_{\mathcal{M}_i} \in \Theta\}},$$

and $\theta_{\mathcal{M}_i}$ is the vector of parameters when model $\mathcal{M}_i$ holds. Even when data are sparse, the standard asymptotic approximation for the likelihood ratio difference statistic given by

$$-2\log(\Lambda_{\mathcal{M}_1}) - 2\log(\Lambda_{\mathcal{M}_2}), \tag{2.7}$$

can hold quite well. Agresti (2002) uses the expression $G^2(\mathcal{M}_1|\mathcal{M}_2)$ to refer to this statistic. In particular, for the studies presented below, the $LR_{diff}$

is computed as the difference between the likelihood ratio statistics from the constrained and unconstrained version of the categorical variable factor model which will be discussed in Section 3 of this chapter. The degrees of freedom for $LR_{diff}$ are the difference of the degrees of freedom for the two likelihood ratio statistics. Furthermore, results from Agresti and Yang (1987) showed that the likelihood ratio difference statistics may perform well when sparseness is present, so it may be a competitor to tests based on lower-order components. In terms of power, $LR_{diff}$ may outperform all other statistics as it will usually have a smaller number of degrees of freedom.

In summary, carefully defined components on overlapping cells could be useful for large multi-way contingency tables that exhibit sparseness as means for assessment of goodness-of-fit of a hypothesized model. In case of severe sparseness, one should consider that higher-order overlapping cells too could become sparse, and thus careful thought should be applied when selecting the number of components for the construction of a lower-order focused test. Also, benefits such as increased power and small number of degrees of freedom of test statistics like the likelihood ratio difference statistic should be considered as a potential remedy for sparseness.

2. Tree-Based Ensemble Methods

A recently developed popular Classification and Regression Tree (CART) (Breiman et al. 1984) algorithm ensemble method called random forest (RF) (Breiman 2001) has become a widely used method in regression and multi-class data settings. While CART and RF are primarily used for pre-

diction, they can also be used to select variables and reduce dimensionality. In many psychological and biological applications, the number of explanatory variables can be very large, into thousands. They tend to be correlated, with outliers while having only a few hundred observations. A fully parameterized hierarchal regression model in this case would not be feasible. Furthermore, a fully parameterized regression model with only main effect terms would likely yield poor estimators, as many complex interactions amongst the explanatory variables would not be included (van der Laan 2006).

A large multi-way contingency table results in a large number of components irrespective of sparseness. The large number of components raises the issue of how to select the optimal number of components. Most suggestions about selecting the number of components exploit some preliminary knowledge about a possible alternative hypothesis. An alternative proposed here is to explore using the random forest algorithm as a means for selecting the number of components of PGF to use in the construction of the limited-information test proposed in this research. CART and RF are discussed in more detail in the following section.

2.1. Random Forest. There are many statistical techniques for modeling a categorical/continuous response. However, most of them come with inherent assumptions such as linearity, additive effects, etc., which might not be part of the data structure. The Classification and Regression Tree (CART) algorithm (Breiman, Friedman, Olshen, and Stone 1984) is one technique which, unlike others, does not make any functional assumptions about the model structure of the data. In CART, a binary tree is grown using recursive

splitting based on node impurity, with a constant response value at each node. At each internal node, the values of predictors are used to determine simple binary conditions. If the condition is satisfied at the node, the left path is chosen, otherwise the right path is selected. This process continues until the terminal node is reached, and finally prediction is made. The CART algorithm builds a tree by selecting the best variables for splitting, optimizing the nodes and "pruning" in order to find the right-size tree.

The CART algorithm has some distinct advantages and disadvantages. One of CART's major advantages is that it is intuitive and simplistic. The fitted model can be easily interpreted by non-statisticians. Moreover, the algorithm can handle a large number of predictors and nonhomogeneous relationships between predictors. Also, unlike linear regression models where the model estimates unit changes in the predictor on the response, trees can identify important ranges of continuous predictors or common clusters of categorical factors. Most of all, CART models are robust with respect to outliers and misclassified observations. Every observation has weight among $N$ data points. Therefore, one essentially counts how many observations go left or right. This is similar to the robustness property of median values (Breiman et al. 1984).

On the other hand, the CART modeling process is very data dependent and therefore small changes in the data can have a dramatic impact on the final tree. This instability is attributed to the variability in the selection of the optimal variable and/or its splitting point at each internal node. Lastly, the CART algorithm defines a non-smooth prediction surface with sharp jumps.

In order to account for the lack of smoothness and instability inherent in the CART, Breiman (2001) proposed the RF algorithm, which extended the CART algorithm by adding an additional layer of randomness to bagging (Liaw and Wiener 2002). Bagging is building a set of trees using modified training sets of equal size created by bootstrapping the original training set with replacement. Consequently, slightly different trees will in turn produce different predictions. More formally, Breiman (2001) proposed a random forest construction as a process where given a specific training set $T$ with $M$ predictors and $N$ observations, one forms bootstrap samples $T_1, ..., T_K$ (with replacement) with equal size of $T$. Then using the CART algorithm, one builds tree $k$ using $T_k$ where the split at each node is determined from a randomly chosen subset of all $M$ predictors, unlike in the construction of the standard tree where each node is split using the best split amongst all predictors. The number of predictors randomly chosen is $\lfloor M/3 \rfloor$ where $\lfloor \cdot \rfloor$ is the floor function (for $M = 1$ or 2, only one predictor is randomly selected). The RF prediction is the unweighted averaged prediction across the forest. Averaging over trees in combination with the randomization used in growing them, enables random forests to approximate a rich class of functions while maintaining low generalization error. Generally, $K = 500$ unpruned trees are built to maximum depth, and because of such a large number of trees, predictions tend to be more accurate than those from a single classification tree (Breiman 2001).

Let $\widehat{f}(\mathbf{x}, \boldsymbol{\theta}_k)$ represent the prediction from the $k$th tree model in the forest where $\mathbf{x}$ is an input vector and $\boldsymbol{\theta}_k$ is a random vector: i.e., a random sample of predictors at each node. The vector $\boldsymbol{\theta}_k$ is generated independently of

the past random vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{k-1}$ but with the same distribution (Breiman 2001). We can think of the vector $\boldsymbol{\theta}_k$ as the mechanism which creates tree $k$. Given a new input vector $\mathbf{x}_0$ the RF prediction is

$$\widehat{f}_{RF}(\mathbf{x}_0, \boldsymbol{\Theta}) = K^{-1} \sum_{k=1}^{K} \widehat{f}(\mathbf{x}_0, \boldsymbol{\theta}_k), \tag{2.8}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{K}$ is the set of realized random vectors.

Great interest in the random forest method has been stimulated by its nature to easily adapt to data, fitting higher order interactions and non-linear terms, making limited model assumptions, while being robust against overfitting (Breiman 2001). In fact, as more trees are grown a limiting value of the prediction error is achieved. In addition, the RF algorithm's accuracy has been shown to be competitive with many other data mining techniques (Breiman 2001; Segal 2004) and is very user friendly in the sense that it has only two parameters (the number of predictors in the random subset at each node and the number of trees in the random forest), and is usually not very sensitive to either (Liaw and Wiener 2002).

Since the trees are built independently given the data, the RF model can be easily parallelized and is computationally faster than many ensemble methods, including Bootstrap AGGregatING or bagging (Breiman 1996a; Breiman 1996b; Dietterich 2000), output smearing (Breiman 2000) and randomizing internal decisions made by the CART algorithm (Dietterich 2000). Yet, such a method can be difficult to interpret compared to a single tree and as a result is sometimes thought as a "black box" with little to say about the relationship between the response and predictor variables.

The RF algorithm in R (The R Project for Statistical Computing 2.6.0) is able to estimate two other important properties: the variable importance scores and proximity measures (measure of the internal structure of the data). Both of these measures are calculated using the cases not selected (bagged) in the bootstrap samples, which are called *out-of-bag*, or OOB, observations. Variable importance scores are discussed in the next section.

2.2. Variable Importance Scores. Variable importance scores (VIMP) can be used on a large set of predictor variables to reduce dimensionality without any model assumptions. These scores do not depend on a specific model structure and complex interactions amongst the predictors do not need to be explicitly stated, as is the case with traditional model selection methods (Breiman 2001 and Ishwarn 2007). The VIMP method estimates the importance of a variable by looking at how much the prediction error increases when OOB data for that variable is permuted while others are left unchanged.

The prediction error can be estimated without using a separate test set. Specifically, after each tree is created, the OOB data are used to estimate the tree prediction error. The tree prediction errors are then averaged over all trees to get the RF prediction error estimator (Breiman 1996c). Wolpert and McCready (1999) compared estimating prediction error using OOB cases with estimates using cross-validation and concluded that using OOB cases produced a better estimate especially with small data sets.

Large positive values of VIMP for a variable indicate a predictive nature of that predictor, whereas zero or negative importance values identify a variable as not predictive (Ishwarn 2007). This random forest byproduct

will be used to select the number of orthogonal components of PGF for large multi-way tables, and will be described in detail in Chapter 3.

3. Generalized Linear Latent Variable Models

The categorical variable factor model will be used to investigate the size of the proposed tests and power based on various components in addition to focusing on interpretability and computational practicality. The aim of the categorical variable factor model is to describe a relationship between manifest variables and unobserved variables (latent variables) through the so-called linear predictor; that is, the link function that maps the manifest variable space to the latent one. Categories can be graded such as: (1) letter grading, (2) an attitude survey with "strongly disagree, disagree, agree and strongly agree", etc.

In psychometrics, for example, data from questionnaires and tests are used as a basis for measuring abilities, attitudes, or other variables. At its most basic level, a categorical variable factor model is based on the idea that the probability of getting an item correct is a function of a latent trait or ability. For example, a person with higher intelligence would be more likely to correctly respond to a given item on an intelligence test. The main assumption is that given the latent variables, the manifest variables are conditionally independent. In other words, the latent variables explain all the dependence structure between manifest variables. The latent variables may be assumed to have a standard multi-variate normal distribution. Note, unless otherwise stated, a categorical variable factor model in this research will have a single

latent variable. Exact tests for latent traits for an omnibus null hypothesis are still not computationally feasible as the amount of computations grows factorially (faster than exponential growth) along with the size of the contingency table (Gooijer and Yuan 2011).

If $\mathbf{y} = (y_1, y_2, ..., y_p)^T$ is a vector of $p$ ordinal observed variables, each having the same number of categories $K$, $\boldsymbol{\eta}$ is the vector of continuous latent variables and $\boldsymbol{\theta}$ is a vector of parameters, the $K$ categories have the associated probabilities $\pi_0^{(i)}(\boldsymbol{\eta};\boldsymbol{\theta}), \pi_1^{(i)}(\boldsymbol{\eta};\boldsymbol{\theta}), ..., \pi_{K-1}^{(i)}(\boldsymbol{\eta};\boldsymbol{\theta})$, $i = 1, ..., p$, for $y_i$. Moreover, the number of latent variables is essentially the number of factors in the model. Then, the probability of the $s^{th}$ response pattern $\mathbf{y}_s$ over the different response categories of the $p$ variables, is

$$\pi_s(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \pi_s(\boldsymbol{\eta};\boldsymbol{\theta}) h(\boldsymbol{\eta}) d\boldsymbol{\eta}, \tag{2.9}$$

where $h(\boldsymbol{\eta})$ is the density function of $\boldsymbol{\eta}$ that may be assumed to be multivariate normal (Bartholomew and Tzamourani 1999) and $\pi_s(\boldsymbol{\eta};\boldsymbol{\theta})$ is the conditional probability of $\mathbf{y}_s$ given $\boldsymbol{\eta}$ with the form

$$\pi_s(\boldsymbol{\eta};\boldsymbol{\theta}) = \prod_{i=1}^{p} \prod_{j=0}^{K-1} Pr(Y_i = j \mid \boldsymbol{\eta};\boldsymbol{\theta})^{y_{ij}} = \prod_{i=1}^{p} \prod_{j=0}^{K-1} (\pi_j^{(i)}(\boldsymbol{\eta};\boldsymbol{\theta}))^{y_{ij}}, \quad j = 0, ..., K-1 \tag{2.10}$$

where $y_{ij} = 1$ if the response falls in category $j$ of variable $i$ and $y_{ij} = 0$ otherwise. The conditional probabilities assuming a single latent variable in

the model are given by

$$\pi_j^{(i)}(\boldsymbol{\eta};\boldsymbol{\theta}) = Pr(Y_i = j \mid \boldsymbol{\eta};\boldsymbol{\theta}) = \begin{cases} 1 - G(\alpha_{i,1} + \beta_i\eta), & \text{if } j = 0, \\ G(\alpha_{i,j} + \beta_i\eta) - G(\alpha_{i,j+1} + \beta_i\eta), & \text{if } 0 < j < K-1, \\ G(\alpha_{i,K-1} + \beta_i\eta), & \text{if } j = K-1, \end{cases} \tag{2.11}$$

where $G(x)$ equals the standard logistic distribution function

$$G(x) = [1 + e^{-x}]^{-1}. \tag{2.12}$$

In this model, for each variable there is one slope parameter $\beta_i$ and $K-1$ intercept parameters $\alpha_{i,j}$ with $\alpha_{i,j}$ decreasing in $j$ for each $i$. The integral of expression (2.9) is evaluated using Gauss-Hermite quadrature with $r$ quadrature points and their corresponding weights ($r = 32$ in this dissertation as it was found to be adequate in many previous studies). Derivatives of expression (2.9) are also evaluated using quadrature

$$\frac{\partial \pi_s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial \pi_s(\boldsymbol{\eta};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} h(\boldsymbol{\eta}) d\boldsymbol{\eta}. \tag{2.13}$$

The quadrature method gives an approximation of a definite integral of a function, usually as a weighted sum of function values at specified points within the domain of integration. An $r$-point Gaussian-Hermite quadrature rule has the form

$$\int f(x)\, dx = \int w(x)g(x)\, dx \approx \sum_{i=1}^{r} w_i(x_i)g(x_i),$$

where $f(x) = w(x)g(x)$ and $w_i(x_i) = e^{-x_i^2}$. Gauss-Hermite calculations with a fixed number of quadrature points represent the latent variable as discrete with

weight $w_i$ at point $x_i$, $i = 1, 2, ..., r$. Note, for a fixed number of quadrature points, the accuracy of the maximum likelihood estimation may decrease as the slope parameters increase in absolute value (Maydeu-Olivares and Joe 2006). For the multidimensional latent variable case, the integral in expression (2.9) and its derivatives in expression (2.13) are evaluated as iterated integrals.

The GRM function in R fits the graded response model for ordinal polytomous data. The parameters in the GRM function in R which are estimated by maximizing the marginal log-likelihood under the conditional independence assumption, i.e., conditionally on the latent structure the items are independent Bernoulli variates under the logit link and the required integrals are approximated using the Gauss-Hermite rule. The optimization procedure used is a hybrid algorithm. The procedure initially uses a moderate number of EM iterations and then switches to quasi-Newton iterations until convergence.

A special case of the categorical variable factor model with all slopes equal to 1 is known as the unidimensional Rasch model (Rasch 1980). In reality the Rasch model is an unrealistic approach for a model because the requirement of equal slopes is too restrictive. Even so, the Rasch model is often selected to demonstrate power calculations because a log-linear version is available. Tjur (1982) and Cressie and Holland (1983) demonstrated the equivalence of the logit version of the Rasch model to a generalized log-linear version. In this generalized log-linear version of the Rasch model for 5 variables the log cell frequencies can be obtained by

$$log(m_s) = \lambda + \lambda_g^{Y_1} + \lambda_h^{Y_2} + \lambda_i^{Y_3} + \lambda_j^{Y_4} + \lambda_k^{Y_5} + \lambda_t^T, \tag{2.14}$$

where $m_s$ is the $s^{th}$ cell frequency, $\lambda_g^{Y_i}$ is the effect for level $g$ of manifest variable $i$ and $\lambda_t^T$ is an effect for respondents with the same total score, $t = 0, 1, ..., k$. Using the log-linear version of the model has the advantage that it is convenient to demonstrate the influence of higher-order interactions and to estimate the model with widely used software. Also, a generalized log-linear version of the Rasch model does not assume a specified distribution of the latent variable. However, the model given in expression (2.14) does not allow for higher-order interactions. Thus, the power calculations under the condition that the null hypothesis is false because it omits a higher-order interaction use frequencies generated from a log-linear model that includes that interaction. For 5 variables the log-linear model that generates cell frequencies with a single three-way association among variables $Y_2$, $Y_3$ and $Y_4$ can be obtained by

$$\begin{aligned} log(m_s) =& \lambda + \lambda_g^{Y_1} + \lambda_h^{Y_2} + \lambda_i^{Y_3} + \lambda_j^{Y_4} + \lambda_k^{Y_5} + \lambda_{gh}^{Y_1Y_2} + \lambda_{gi}^{Y_1Y_3} + \lambda_{gj}^{Y_1Y_4} + \lambda_{gk}^{Y_1Y_5} + \\ & \lambda_{hi}^{Y_2Y_3} + \lambda_{hj}^{Y_2Y_4} + \lambda_{hk}^{Y_2Y_5} + \lambda_{ij}^{Y_3Y_4} + \lambda_{ik}^{Y_3Y_5} + \lambda_{jk}^{Y_4Y_5} + \lambda_{hij}^{Y_2Y_3Y_4}, \end{aligned} \tag{2.15}$$

where $m_s$ is the $s^{th}$ cell expected frequency, $\lambda_g^{Y_i}$ is the effect for level $g$ of manifest variable $i$, $\lambda_{ab}^{Y_iY_j}$ is the two-way effect for a pair of manifest variables $i$ at level $a$ and $b$ at level $h$ and $\lambda_{cde}^{Y_tY_lY_k}$ is the three-way effect among manifest variables $c$ at level $t$, $d$ at level $l$ and $e$ at level $k$.

# Chapter 3: COMPUTATIONAL METHODS

## 1. Chi-Square Orthogonal Components

The primary focus of this research is calculating and selecting orthogonal components of PGF that are interpretable and computationally practical for large multi-way contingency tables. These components may serve as a means for providing more powerful directional tests and for overcoming the adverse effects of sparseness. In particular, the idea of components in this research is extended to large cross-classified tables with graded multi-category variables. Graded categories are encountered in both log-linear and logistic models.

Moses et al. (1984), for instance, reported that ordered categorical data occurred in 32 of 168 articles in volume 36 (1982) of the *New England Journal of Medicine*. Using standard log-linear models with ungraded variables for data sets where at least one variable is graded ignores important information about the data. With such large cross-classified tables it is acknowledged that there will be even more components to select from than in a binary variable model. It is also acknowledged that for some models computing components based on orthogonal polynomials may not be feasible in all applications. As a result, this research explores the full- and limited-information goodness-of-fit statistics for multidimensional multinomial data in the particular case of the categorical variable factor model for ordinal data.

The $\mathbf{H}$ transformation matrix, as discussed in Chapter 2, along with certain conditions discussed in this section, uniquely define various orthogonal

components. Thus, the orthogonal components defined on marginal frequencies for graded categorical variables can be obtained by using an expanded **H** matrix. Examples are given in the following section for ordinary marginals.

1.1. Ordinary Marginals. The focus of this section is to provide a link between the joint proportions and first-, second-order and higher-order marginal proportions as a means for constructing proposed orthogonal components.

A $p$-dimensional vector with zeros and ones is often called a response pattern and can be used to identify a specific cell from the contingency table formed by the cross-classification of $p$ response variables each at $K$ categories. A $T = K^p$-dimensional set of response patterns can thus be generated by varying the $p^{th}$ variable most rapidly in $K-1$ columns, $(p-1)^{st}$ variable next in $K-1$ columns, etc.

Define $\mathbf{V}$ as the $T$ by $p(K-1)$ matrix with response patterns as rows. For example, with 3 variables each at 2 categories (i.e., $p = 3$ and $T = 2^3 = 8$),

$$\mathbf{V}_{8\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

The first row represents the response pattern for $Y_1 = 0, Y_2 = 0$ and $Y_3 = 0$, the second row represents the response pattern for $Y_1 = 0, Y_2 = 0$ and $Y_3 = 1$,..., and the last row represents the response pattern for $Y_1 = 1, Y_2 = 1$ and $Y_3 = 1$. Similarly, for 3 variables each at 3 categories (i.e., $p = 3$ and $T = 3^3 = 27$),

$$\mathbf{V}_{27\times 6} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

The first row represents the response pattern for $Y_1 = 0, Y_2 = 0$ and $Y_3 = 0$, the second row represents the response pattern for $Y_1 = 0, Y_2 = 0$ and $Y_3 = 1$, the third row represents the response pattern for $Y_1 = 0, Y_2 = 0$ and $Y_3 = 2$,..., and the last row represents the response pattern for $Y_1 = 2, Y_2 = 2$ and $Y_3 = 2$.

Let $v_{sj}$ represent element $s$ in column vector $\mathbf{v}_j$ where $j = 1, ..., p(K-1)$. For example,

$$\mathbf{V}_{27\times 6} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 & \mathbf{v}_6 \end{pmatrix}.$$

Now let $\mathbf{Y}$ be a vector of multidimensional multinomial variables with cell probabilities $\boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi_1(\boldsymbol{\theta}) from a model of interest, \pi_2(\boldsymbol{\theta}), ..., \pi_T(\boldsymbol{\theta}))^T$ that depend on the parameter vector $\boldsymbol{\theta}$. Then, under the model, the first-order marginal proportion for category 1 of $Y_i$, can be defined as

$$P_{Y_i}(1;\boldsymbol{\theta}) = Prob(Y_i = 1 \mid \boldsymbol{\theta}) = \mathbf{v}^T_{(1+(K-1)(i-1))}\boldsymbol{\pi}(\boldsymbol{\theta}) = \sum_{s=1}^{T} v_{s,(1+(K-1)(i-1))}\ \pi_s(\boldsymbol{\theta}).$$

In general, marginal proportions are linear transformations of the cell proportions in the multinomial vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_T)^T$ with associated response patterns. They can be obtained via the multiplication of $\boldsymbol{\pi}$ by a certain transformation matrix, denoted by $\mathbf{H}$. In particular, matrix $\mathbf{H}$ can be defined from matrix $\mathbf{V}$ such that $\mathbf{H} = \mathbf{V}^T$ for first-order marginals to be obtained as $\mathbf{H}\boldsymbol{\pi}$. For example when $p = 3$ and $K = 3$, the first-order marginal proportion for category 1 of $Y_1$ is given by the first row of $\mathbf{H}\boldsymbol{\pi}$ as

$$\begin{aligned} P_{Y_1}(1;\boldsymbol{\theta}) &= Prob(Y_1 = 1 \mid \boldsymbol{\theta}) = \mathbf{v}_1^T\boldsymbol{\pi}(\boldsymbol{\theta}) \\ &= \pi_{10} + \pi_{11} + \pi_{12} + \pi_{13} + \pi_{14} + \pi_{15} + \pi_{16} + \pi_{17} + \pi_{18}. \end{aligned}$$

The first-order marginal proportion for category 2 of $Y_1$ is given by the second row of $\mathbf{H}\boldsymbol{\pi}$ as

$$\begin{aligned} P_{Y_1}(2;\boldsymbol{\theta}) &= Prob(Y_1 = 2 \mid \boldsymbol{\theta}) = \mathbf{v}_2^T\boldsymbol{\pi}(\boldsymbol{\theta}) \\ &= \pi_{19} + \pi_{20} + \pi_{21} + \pi_{22} + \pi_{23} + \pi_{24} + \pi_{25} + \pi_{26} + \pi_{27}. \end{aligned}$$

Moreover, the matrix producing a full set of marginals can be obtained by forming Hadamard products (Magnus and Neudecker 1999) either among the columns of $\mathbf{V}$ or the columns of matrix $\mathbf{H}^T$. The $k^{th}$ element in the

Hadamard product of two vectors $(\mathbf{v}_i \circ \mathbf{v}_j)$ is given by

$$(\mathbf{v}_i \circ \mathbf{v}_j)_k = \begin{cases} 1 \text{ if } v_{ki} = v_{kj} = 1 \\ \\ 0 \text{ otherwise.} \end{cases}$$

The second-order marginals can then be obtained by Hamadard products of columns of $\mathbf{V}$ as $(\mathbf{v}_i \circ \mathbf{v}_j)$, the third-order marginals $(\mathbf{v}_i \circ \mathbf{v}_j \circ \mathbf{v}_k)$, etc. For example, a matrix $\mathbf{H}^T$ that would produce marginals from first- to third-order for 3 variables each with 3 categories is given below. The first 6 rows of $\mathbf{H}\boldsymbol{\pi}$ produce all first-order marginals for $Y_1, Y_2$ and $Y_3$, the next 12 rows produce all second-order marginals and the last 8 rows produce all third-order marginals.

$$H^T_{27\times 26} = \begin{pmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0\\
1&0&1&0&0&1&1&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&0\\
1&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&1&1&0&0&1&0&0&1&0&0&0&0&0&1&0&0&0&1&0&0&0&0&0\\
1&0&0&1&0&1&0&1&0&0&0&1&0&0&0&0&0&1&0&0&0&1&0&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&0&1&0&0&0&1&0&0&0&1&0&1&0&0&0&0&0&0&1&0&0&0&0\\
0&1&1&0&0&1&0&0&1&0&0&0&0&1&0&1&0&0&0&0&0&0&1&0&0&0\\
0&1&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&0&1&1&0&0&0&0&1&0&0&1&0&0&0&1&0&0&0&0&0&0&1&0&0\\
0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&1
\end{pmatrix}$$

For example, the second-order marginal proportion for category 1 of $Y_1$

and category 1 of $Y_2$ is given by the seventh row of $\mathbf{H}\boldsymbol{\pi}$ as

$$\begin{aligned} P_{Y_1Y_2}(1,1;\boldsymbol{\theta}) = Prob(Y_1 = 1, Y_2 = 1 \mid \boldsymbol{\theta}) &= (\mathbf{v}_1 \circ \mathbf{v}_3)^T \pi(\boldsymbol{\theta}) \\ &= \sum_{s=1}^{T} v_{s1} v_{s3} \pi_s(\boldsymbol{\theta}) \\ &= \pi_{13} + \pi_{14} + \pi_{15}. \end{aligned}$$

In the same way, higher-order marginals can be defined.

The following section describes in detail methods for computing components defined on marginals.

1.2. Methods for Computing Chi-Square Orthogonal Components. A very large $\mathbf{H}$ matrix results in a high degree of collinearity among the rows. This collinearity can produce inaccuracy in the calculation of components. Because of the need for high numerical accuracy, the following procedure for calculating orthogonal components of PGF is proposed.

In what follows we use the standardized cell residual (Cochran 1954)

$$z_s = \frac{(\hat{p}_s - \hat{\pi}_s)}{\hat{\pi}_s^{\frac{1}{2}}}, \qquad s = 1, ..., T,$$

for which

$$\chi^2_{PF} = \mathbf{z}^T \mathbf{z}. \tag{3.1}$$

1.2.1. *Composite Null Hypothesis.* From Reiser (2008), PGF in terms of the standardized residuals can be obtained by

$$\chi^2_{PF} = (\mathbf{H}\widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{r})^T (\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}^T)^{-1} (\mathbf{H}\widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{r}) = (\mathbf{H}\mathbf{z})^T (\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}^T)^{-1} (\mathbf{H}\mathbf{z}) \tag{3.2}$$

where $\mathbf{H}$ may be the matrix for all possible joint marginals, $\mathbf{r}$ is the vector of raw residuals, $\mathbf{z}$ is the vector of standardized residuals, $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = n^{-1}(\mathbf{I} -$

$\hat{\boldsymbol{\pi}}^{\frac{1}{2}}(\hat{\boldsymbol{\pi}}^{\frac{1}{2}})^T - \widehat{\mathbf{A}}(\widehat{\mathbf{A}}^T\widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^T$), which is $\boldsymbol{\Sigma}$ evaluated at the maximum likelihood estimates $\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})$,

$$\widehat{\mathbf{D}} = \mathbf{D}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) = \text{diag}(\hat{\pi}_1, ..., \hat{\pi}_T),$$

and

$$\widehat{\mathbf{A}} = \widehat{\mathbf{D}}^{-\frac{1}{2}} \frac{\partial \hat{\boldsymbol{\pi}}}{\partial \hat{\boldsymbol{\theta}}}.$$

One method for obtaining orthogonal components defined on marginals is the QR decomposition. The QR decomposition (also called the QR factorization) of a matrix is a decomposition of the matrix into an orthogonal and an upper triangular matrix. Any real square matrix $\mathbf{B}$ may be decomposed as

$$\mathbf{B} = \mathbf{QR},$$

where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. If $\mathbf{B}$ is invertible, then the factorization is unique if the diagonal elements of $\mathbf{R}$ are positive. There are several methods for computing the QR decomposition, such as by means of the Gram-Schmidt process, Householder transformations, or Givens rotations. The routine in R uses the Gram-Schmidt process. The Gram-Schmidt process is applied to the columns of the full column rank matrix $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$, with inner product $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^\top \mathbf{w}$ (or $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* \mathbf{w}$ for the complex case). If the projection operator is defined by

$$\text{proj}_{\mathbf{e}}\mathbf{b} = \frac{\langle \mathbf{e}, \mathbf{b} \rangle}{\langle \mathbf{e}, \mathbf{e} \rangle} \mathbf{e}$$

then Gram-Schmidt process then works as follows

$$
\begin{aligned}
\mathbf{u}_1 &= \mathbf{b}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
\mathbf{u}_2 &= \mathbf{b}_2 - \operatorname{proj}_{\mathbf{e}_1} \mathbf{b}_2, & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
\mathbf{u}_3 &= \mathbf{b}_3 - \operatorname{proj}_{\mathbf{e}_1} \mathbf{b}_3 - \operatorname{proj}_{\mathbf{e}_2} \mathbf{b}_3, & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
&\vdots & &\vdots \\
\mathbf{u}_k &= \mathbf{b}_k - \sum_{j=1}^{k-1} \operatorname{proj}_{\mathbf{e}_j} \mathbf{b}_k, & \mathbf{e}_k &= \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}
\end{aligned}
$$

The sequence $\mathbf{u}_1, ..., \mathbf{u}_k$ is the required system of orthogonal vectors, and the normalized vectors $\mathbf{e}_1, ..., \mathbf{e}_k$ form an orthonormal set. The calculation of the sequence $\mathbf{u}_1, ..., \mathbf{u}_k$ is known as Gram-Schmidt orthogonalization, while the calculation of the sequence $\mathbf{e}_1, ..., \mathbf{e}_k$ is known as Gram-Schmidt orthonormalization as the vectors are normalized. Thus, $Q = [\mathbf{e}_1, \cdots, \mathbf{e}_n]$ and

$$
R = \begin{pmatrix}
\langle \mathbf{e}_1, \mathbf{b}_1 \rangle & \langle \mathbf{e}_1, \mathbf{b}_2 \rangle & \langle \mathbf{e}_1, \mathbf{b}_3 \rangle & \dots \\
0 & \langle \mathbf{e}_2, \mathbf{b}_2 \rangle & \langle \mathbf{e}_2, \mathbf{b}_3 \rangle & \dots \\
0 & 0 & \langle \mathbf{e}_3, \mathbf{b}_3 \rangle & \dots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
$$

If $\mathbf{z}$ is regressed on the columns of $\mathbf{H}^T$ (Reiser 2008) then

$$\mathbf{z} = \mathbf{H}^T \hat{\boldsymbol{\beta}}.$$

There is no error term when $\mathbf{H}$ contains coefficients for all possible joint mariginals and $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}^T)^{-1}\mathbf{H}\widehat{\mathbf{W}}\mathbf{z} \tag{3.3}$$

with weight matrix $\widehat{\mathbf{W}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}}$, since $\widehat{\boldsymbol{\Sigma}}$ is idempotent.

There is no intrinsic interest in $\hat{\boldsymbol{\beta}}$, since the purpose of the regression is to obtain the sequential sum of squares.

Let $\widehat{\mathbf{M}} = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^T$. Then, expression (3.3) becomes

$$\hat{\boldsymbol{\beta}} = (\widehat{\mathbf{M}}^T\widehat{\mathbf{M}})^{-1}\widehat{\mathbf{M}}^T\mathbf{z}. \tag{3.4}$$

Direct orthonormalization, via Gram-Schmidt orthonormalization or QR decomposition, can then be applied to $\widehat{\mathbf{M}}$ to obtain an orthogonal regression

$$\hat{\boldsymbol{\gamma}} = (\widehat{\mathbf{M}^*}^T\widehat{\mathbf{M}^*})^{-1}\widehat{\mathbf{M}^*}^T\mathbf{z} = \widehat{\mathbf{M}^*}^T\mathbf{z}, \tag{3.5}$$

where $\widehat{\mathbf{M}}^*$ is the orthonormalized standardized transformation matrix $\widehat{\mathbf{M}}$ and the elements of $\hat{\boldsymbol{\gamma}}$ are known as the orthogonal coefficients. The vectors $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ are related via the Cholesky factor $\mathbf{C}$ of $\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}^T$ with $\hat{\boldsymbol{\beta}} = (\mathbf{C}^T)^{-1}\hat{\boldsymbol{\gamma}}$. It follows from Reiser (2008) that $\widehat{\mathbf{M}^*}^T\mathbf{z}$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_{T-g-1}$; i.e., $\widehat{\mathbf{M}^*}^T\mathbf{z}$ has the limiting covariance matrix $\mathbf{F}^T\boldsymbol{\Sigma}\mathbf{F} = \mathbf{I}_{T-g-1}$, where $\widehat{\mathbf{F}} = (\widehat{\mathbf{C}}^T)^{-1}$ is the Cholesky factor of $\widehat{\boldsymbol{\Sigma}}$.

Using the QR decomposition does not place any constraints on the $\mathbf{H}$ matrix. In contrast, the Cholesky decomposition requires deletion of rows from $\mathbf{H}$ which have a linear dependency with $\mathbf{G} = \frac{\partial\hat{\boldsymbol{\pi}}}{\partial\hat{\boldsymbol{\theta}}}$ and which results in $\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}^T$ having deficient rank.

Since

$$\hat{\boldsymbol{\gamma}} = \widehat{\mathbf{M}^*}^T\mathbf{z}, \tag{3.6}$$

if $\text{rank}(\mathbf{H}) \geqslant T - g - 1$ and $\text{rank}(\mathbf{HG}) = g$, the PGF defined in expression (3.1) is the sum of the $\hat{\gamma}_k^2$ according to the proof given in Reiser (2008) and

$$\chi^2_{PF} = \hat{\boldsymbol{\gamma}}^T\hat{\boldsymbol{\gamma}}. \tag{3.7}$$

The chi-square statistic is obtained by summing the orthogonal components which are asymptotically independent $\chi_1^2$ random variables. Moreover, using the relationship between normal and chi-squared distributions given by Rao (1973, p. 188) implies that $\hat{\boldsymbol{\gamma}}^T\hat{\boldsymbol{\gamma}}$ is asymptotically chi-squared. Furthermore, limited-information test statistics can be obtained by summing a subset of these orthogonal components.

Alternatively, the orthogonal components of PGF can be obtained from the sequential sums of squares that result from an ordinary regression of $\mathbf{z}$ on the columns of $\mathbf{H}^T$. If $\boldsymbol{h}_l^T$ represents row $l$ of $\mathbf{H}$, and $\mathbf{h}_l$ represents column $l$ of matrix $\mathbf{H}^T$ then, using results from linear models for the regression that produces the orthogonal components, the sum of squares that constitute the first component, $\hat{\gamma}_1^2$, is given by

$$SS(\boldsymbol{h}_1) = n^{-1}\mathbf{z}^T\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_1\left(\boldsymbol{h}_1^T\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_1\right)^{-1}\boldsymbol{h}_1^T\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\mathbf{z},$$

provided $\boldsymbol{h}_1^T\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_1$ is non-zero.The orthogonal compliment of $\boldsymbol{h}_2$ to $\boldsymbol{h}_1$ is given by

$$\boldsymbol{h}_2^{\mathrm{I}} = \boldsymbol{h}_2 - \boldsymbol{h}_1\left(\boldsymbol{h}_1^T\widehat{\mathbf{W}}\boldsymbol{h}_1\right)^{-1}\boldsymbol{h}_1^T\widehat{\mathbf{W}}\boldsymbol{h}_2,$$

provided $\boldsymbol{h}_1^T\widehat{\mathbf{W}}\boldsymbol{h}_1$ is non-zero, and the sequential sum of squares that constitute the second component, $\hat{\gamma}_2^2$, is given by

$$SS(\boldsymbol{h}_2|\boldsymbol{h}_1) = n^{-1}\mathbf{z}^T\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_2^{\mathrm{I}}\left(\boldsymbol{h}_2^{\mathrm{I}T}\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_2^{\mathrm{I}}\right)^{-1}\boldsymbol{h}_2^{\mathrm{I}T}\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\mathbf{z}.$$

provided $\boldsymbol{h}_2^{\mathrm{I}T}\widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\boldsymbol{h}_2^{\mathrm{I}}$ is non-zero. Sequential sum of squares for additional orthogonal components may be obtained in a similar manner (Reiser 2008).

Unlike the QR decomposition which does not detect linear dependencies or make any adjustments, the sequential sum of squares are calculated using Goodnight's (1978) sweep operator. When a dependency is encountered for a component, the sweep operator sets that component to zero and proceeds (Goodnight 1978). The critical issue here is not the sequential sum of squares versus the QR decomposition. The issue here is writing the code for the decomposition to deal with linear dependencies as does the Goodnight's code for the sweep operator. The routines for the QR decomposition in R and in SAS IML are not written to check carefully for linear dependencies. Routines for the QR decomposition both in R and in SAS IML could be written to check for linear dependencies, and then they would be as reliable as the sweep operator in PROC REG in SAS. PROC REG in SAS is used to obtain the sequential sums of squares.

The method given here is numerically more accurate than the Cholesky factor method used by Reiser (2008) which is important when there are a large number of components. Table 1 compares the numerical accuracy of the QR decomposition versus the Cholesky factor method. A categorical variable factor model for non-sparse data was generated with intercepts with magnitude in the range (-1,1) and slopes with magnitude in the range (0,3). As the size of the contingency table increases with the number of variables the numerical accuracy of both the QR decomposition and the Cholesky factor method decrease. However, the decrease is greater for the Cholesky factor method with the exponential growth of the table size i.e., the number of orthogonal components summed for the $\chi^2$ statistic.

Table 1. Numerical accuracy for the QR decomposition and the Cholesky factor method for a categorical variable factor model for non-sparse data

| No. of variables | Degrees of freedom | $\chi^2_{PF}$ | Sum of components using the QR decomposition | Sum of components using the Cholesky factor method |
|---|---|---|---|---|
| 5 | 21 | 15.11165 | 15.11113 | 15.11613 |
| 6 | 51 | 34.12063 | 34.12100 | 34.10810 |
| 7 | 113 | 119.48340 | 119.33701 | 118.09701 |
| 8 | 239 | 226.01750 | 225.97232 | 223.47231 |
| 9 | 493 | 476.67530 | 475.39827 | 473.32277 |

An example of the calculations of orthogonal components defined on marginal frequencies is given below as a preview of a two-way association study in Chapter 4. A categorical variable factor model given in expression (2.9) for 5 variables, each at 2 categories, when $n = 1000$ was extended to include an extra two-way association between $Y_1$ and $Y_2$. In this example the model misspecification under the null hypothesis was in a two-way association not accounted for by a categorical variable factor model with a single latent variable, and as such a large second-order component between $Y_1$ and $Y_2$ was expected. The data were generated with $\boldsymbol{\alpha}_1 = (0.50, 0.25, 0.75, -0.25, -0.50)$, $\boldsymbol{\beta}_1 = (0.50, 0.50, 1.00, 0.75, 1.00)$ and an extra two-way association between $Y_1$ and $Y_2$ of 0.80 in magnitude. The model given in expression (2.9) with a single latent variable was fitted producing parameter estimates given in Table 2. $\chi^2_{32-10-1} = \chi^2_{21} = 11.41$ was observed for this data set. Orthogonal components defined on the full set of marginal frequencies, $\mathbf{H}_{31\times 32}$, were computed from

Table 2. Parameter estimates when the model under the null hypothesis a categorical variable factor model.

| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ |
|---|---|---|---|---|---|
| $\hat{\alpha}_i$ | 0.4612420 | 0.2301011 | 0.7389288 | -0.2463937 | -0.4893528 |
| $\hat{\beta}_i$ | 0.5980901 | 0.5927957 | 0.9187180 | 0.7167252 | 0.9284945 |

expression (3.5) where

$$\mathbf{z} = \begin{pmatrix} 0.001 \\ 0.001 \\ 0.001 \\ 0.001 \\ 0.003 \\ 0.004 \\ 0.004 \\ 0.006 \\ -0.002 \\ -0.002 \\ -0.002 \\ -0.001 \\ -0.005 \\ -0.003 \\ -0.003 \\ -0.001 \\ -0.002 \\ -0.002 \\ -0.002 \\ -0.001 \\ -0.005 \\ -0.003 \\ -0.004 \\ -0.001 \\ 0.009 \\ 0.001 \\ 0.002 \\ 0.000 \\ 0.005 \\ 0.001 \\ 0.001 \\ 0.000 \end{pmatrix},$$

and,

$$\widehat{\mathbf{M}}^{*}_{32\times 31} = \begin{pmatrix}
-16.072 & 9.510 & 5.446 & 1.222 & -0.847 & -0.234 & \cdots & 1.902 \\
8.891 & -5.250 & 2.307 & 1.459 & -8.715 & -0.822 & \cdots & 2.945 \\
6.366 & -5.902 & -4.842 & -0.954 & -3.012 & -2.449 & \cdots & 0.514 \\
2.843 & 2.227 & 1.823 & -3.698 & 3.635 & -4.267 & \cdots & 0.427 \\
5.056 & -10.836 & -1.507 & 5.490 & -10.210 & -3.205 & \cdots & 11.638 \\
2.671 & 4.284 & -0.843 & -6.407 & 6.279 & -9.331 & \cdots & 1.430 \\
2.881 & 3.719 & 2.401 & -5.303 & 8.153 & -8.935 & \cdots & 6.506 \\
-3.799 & 1.265 & -4.286 & 11.418 & 8.168 & -12.653 & \cdots & -2.505 \\
3.475 & -7.267 & -9.956 & -5.028 & 6.636 & 2.596 & \cdots & -0.083 \\
4.100 & 2.186 & 4.315 & -2.256 & 2.294 & 4.875 & \cdots & 0.520 \\
4.116 & 1.706 & 5.711 & 4.617 & -0.755 & 6.043 & \cdots & 1.324 \\
-3.100 & 1.873 & -3.730 & -0.427 & -2.419 & 3.822 & \cdots & -0.016 \\
3.121 & 3.420 & 6.023 & -1.839 & 6.609 & 7.572 & \cdots & 8.097 \\
-3.280 & 2.376 & -6.106 & 8.387 & -0.978 & 5.860 & \cdots & -2.730 \\
-2.754 & 3.544 & -5.335 & -0.474 & -5.364 & 5.858 & \cdots & 7.274 \\
-5.139 & -5.861 & 7.633 & -6.753 & -7.028 & 1.018 & \cdots & -5.260 \\
2.720 & -8.017 & -10.113 & -5.172 & 5.176 & 1.625 & \cdots & -6.587 \\
4.350 & 2.511 & 4.921 & -2.553 & 2.436 & 4.671 & \cdots & -0.801 \\
4.200 & 2.003 & 6.658 & 5.106 & -1.207 & 5.829 & \cdots & -0.775 \\
-3.324 & 2.039 & -4.324 & -0.241 & -2.325 & 3.675 & \cdots & 0.485 \\
3.096 & 3.952 & 6.947 & -2.120 & 7.119 & 7.015 & \cdots & 6.483 \\
-3.193 & 2.543 & -7.118 & 9.842 & 0.055 & 5.560 & \cdots & -1.460 \\
-2.755 & 3.867 & -6.211 & -0.105 & -5.353 & 5.439 & \cdots & 9.066 \\
-4.486 & -6.729 & 7.988 & -7.738 & -6.942 & 0.079 & \cdots & -1.140 \\
3.990 & 1.417 & 9.422 & 13.554 & -5.294 & -11.419 & \cdots & -2.693 \\
-2.795 & 3.336 & -3.857 & -4.196 & -5.097 & -3.445 & \cdots & -0.075 \\
-2.136 & 4.771 & -0.439 & -7.380 & -0.562 & -6.026 & \cdots & 4.287 \\
-9.086 & -6.263 & -1.685 & 3.126 & 7.266 & 0.578 & \cdots & 3.576 \\
-2.587 & 6.158 & -5.607 & -6.914 & -9.657 & -8.207 & \cdots & 9.871 \\
-4.226 & -9.126 & 5.812 & -2.881 & 3.754 & -1.027 & \cdots & -1.930 \\
-4.791 & -10.880 & 2.519 & 5.901 & 7.978 & -0.412 & \cdots & 18.051 \\
13.705 & 9.326 & -4.186 & 1.887 & 0.674 & 0.008 & \cdots & 2.762
\end{pmatrix}.$$

The QR decomposition was applied to $\widehat{\mathbf{M}}$ in expression (3.4) in order to obtain $\widehat{\mathbf{M}^{*}}$. Thus, the calculated orthogonal components, $\hat{\boldsymbol{\gamma}}^{2}$, given in expression (3.6) were

$$\hat{\boldsymbol{\gamma}}^2 = \begin{pmatrix} 0.000 \\ 0.004 \\ 0.042 \\ 0.377 \\ 0.001 \\ 10.963 \\ 0.000 \\ 0.000 \\ 0.001 \\ 0.000 \\ 0.000 \\ 0.004 \\ 0.000 \\ 0.001 \\ 0.001 \\ 0.000 \\ 0.002 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix}$$

Individual $\hat{\gamma}_k^2$'s are orthogonal components defined on marginal frequencies and are asymptotically independent $\chi_1^2$ random variables. Moreover, $\sum_{k=1}^{21} \hat{\gamma}_k^2 = 11.40$. The first 5 $\hat{\gamma}_k^2$'s are first-order marginal components, the next 10 are second-order marginal components, etc. As can be seen, $\hat{\gamma}_6^2 = 10.963$ and this component captures a two-way association between $Y_1$ and $Y_2$ not accounted for by a categorical variable factor model with a single latent variable, as expected.

Since not all components of $\mathbf{H}$ are necessary for constructing various limited-information test statistics, using the random forest method may enable a reduction in the number of components. Applying the random forest method to expression (3.4), where $\mathbf{z}$ is the response vector and $\widehat{\mathbf{M}}$ is the model matrix, components will be obtained in an exploratory mode, using the positive VIMP scores for columns of $\widehat{\mathbf{M}}$. Large positive values of VIMP for a variable indicate

the important nature of that component in terms of variable selection in the orthogonal regression (Ishwarn 2007). Note that the number of selected components is random from sample to sample. As explained in Section 2.2, VIMP scores can be used on a large set of components to reduce dimensionality of $\widehat{\mathbf{M}}$, without any model assumptions.

The notation $\chi^2_{rf}$ denotes the test statistic where the random forest method has been applied to the full set of marginals. $\chi^2_{rf_{[1:2]}}$ denotes the test statistic applied to first- and second-order marginals and $\chi^2_{rf_{[2]}}$ denotes the test statistic applied to second-order marginals only. The distribution of these statistics under the null hypothesis is a mixture of chi-squares because they are a sum of orthogonal components obtained using the random forest method, and it will be investigated in a simulation study using the moment approximation (Mathai and Provost 1992; Box 1954). This data-driven approach for selecting components of PGF should further increase the power and decrease dilution of such test, as it will not include superfluous degrees of freedom.

1.2.2. *Related Limited-information Statistics Defined on Lower-order Marginals.* Moustaki (2007) defined $GFfit^{ij}$ as part of her goodness-of-fit statistic for bivariate marginal distributions of the variables $i$ and $j$ as a remedy when sparseness is present as

$$GFfit^{ij} = n\sum_{ab} \frac{(f^{ij}_{ab} - \hat{\pi}^{ij}_{ab})^2}{\hat{\pi}^{ij}_{ab}}, \qquad i = 1, ..., p-1 \qquad j = i+1, ..., p$$

where $f^{ij}_{ab}$ is the sample proportion from the bivariate marginal distribution of the variables $i$ and $j$, each at categories $a$ and $b$ respectively and $\hat{\pi}^{ij}_{ab}$ is the corresponding estimated probability. The values $\frac{(f^{ij}_{ab} - \hat{\pi}^{ij}_{ab})^2}{\hat{\pi}^{ij}_{ab}}$ are standardized

residuals computed from the bivariate marginal distributions of the variables $i$ and $j$ and measure the discrepancies between observed and expected proportions. However, using this $GFfit^{ij}$ definition over-parameterizes the $\mathbf{H}$ matrix compared to the component approach proposed here. For example, if the $\mathbf{H}$ matrix was defined on ordinary marginals for 3 variables each at 3 categories, then the statistic $GFfit^{12}$ between variable 1 and variable 2 would imply a $\mathbf{H}_{9\times 27}$ matrix using the Moustaki (2007) approach where number of rows are given by $p^2$ , versus a $\mathbf{H}_{4\times 27}$ matrix using the components approach. Using the components approach, the $GFfit^{12}$ is computed by the sum of the components of $\mathbf{H}\boldsymbol{\pi}$ produced by rows (7, 8, 9, 10). Similarly, $GFfit^{13}$ is the sum of the components produced by rows (11, 12, 13, 14) and $GFfit^{23}$ is the sum of the components produced by rows (15, 16, 17, 18).

Maydeu-Olivares and Joe (2005) developed a family of statistics, $M_r$, that are closely related to test statistics obtained by summing components over lower-order marginals proposed here. $M_2$ and a limited-information statistic defined on first- and second-order marginals proposed in this research are not equivalent. In the quadratic form in (2.4), $M_2$ uses $\widehat{\mathbf{C}}_2$ instead of $\widehat{\boldsymbol{\Sigma}}$, where $\widehat{\mathbf{C}}_2$ is given by

$$\widehat{\mathbf{C}}_2 = (\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}^T)^{-1} - (\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}^T)^{-1}\mathbf{H}\widehat{\mathbf{G}}\left(\widehat{\mathbf{G}}^T\mathbf{H}^T(\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}^T)^{-1}\mathbf{H}\widehat{\mathbf{G}}\right)^{-1}\widehat{\mathbf{G}}^T\mathbf{H}^T(\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}^T)^{-1},$$

where $\widehat{\boldsymbol{\Gamma}} = D(\hat{\boldsymbol{\pi}}) - \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^T$ and $\mathbf{H}$ contains first- and second-order marginals (Reiser 2008). The degrees of freedom of $M_2$ are given by $\sum_{j=1}^{2}\binom{p}{j}(K-1)^j - g$.

1.2.3. *Simple Null Hypothesis.* Although the primary focus of this research is a test for a composite null hypothesis, the Pearson's chi-squared test

statistic for a simple null hypothesis $H_0 : \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}_0$ can be partitioned by using a procedure similar to the method described previously as

$$\begin{aligned}\chi^2_{T-1} &= (\mathbf{Hz})^T(\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}^T)^{-1}(\mathbf{Hz}) \\ &= n^{-1}(\mathbf{Hz})^T(\mathbf{H}((\mathbf{I} - \boldsymbol{\pi}_o^{\frac{1}{2}}(\boldsymbol{\pi}_o{}^{\frac{1}{2}})^T))\mathbf{H}^T)^{-1}(\mathbf{Hz}). \qquad (3.8)\end{aligned}$$

where $\boldsymbol{\pi}_0$ is the vector of probabilities specified under the null hypothesis and $\mathbf{H}$ may be chosen as the matrix of all possible joint marginals for all given variables.

1.2.4. *Size of the Test and Power.* Given the various $\mathbf{H}$ matrices discussed previously, comparisons were made of orthogonal components defined on marginal frequencies, for both data that are not sparse and sparse, to the traditional Pearson's chi-square test and the likelihood ratio test. These comparisons were in terms of the size of the tests and their relative power.

In the case when data are not sparse and the number of components is known, asymptotic power was computed. Asymptotic power is discussed in detail in the next section. When data are sparse, power calculations were supplemented by Monte Carlo simulations. Monte Carlo simulations were also used when the random forest method was applied as the number of selected components is unknown at the outset. Comparing proposed test statistics to orthogonal polynomials components using equal correlations among observed variables in the one parameter categorical variable factor model is also of interest in this research. Generalized categorical variable factor models for graded data will be used to generate data.

2. Performance Measures

A distinction should be drawn between omnibus (tests such as the PGF) and directional tests. Lancaster (1969) explained that omnibus tests are intended to have moderate power against all alternatives, while the directional tests are intended to detect specified alternatives well. In summary, against specified alternatives, directional tests are more powerful than the omnibus tests, while against other alternatives omnibus tests should be superior. The PGF statistic was constructed to be an omnibus test, but its components can provide powerful directional tests.

Type I error rates and power performance of the proposed test statistic based on various components was assessed. Regardless of power, if the empirical $\alpha$ is close to the nominal $\alpha$, this implies that the proportion of times for rejecting the model falsely is indeed the proportion of times it is expected to be. Of course, the test with the largest power will always be preferred, provided the size of the tests are the same.

2.1. Power. Given specific choices of sample sizes, class size, null hypothesis, and the alternative hypothesis, the power function is the most important criterion for comparing tests. Under some circumstances power can be calculated analytically; under other circumstances, Monte Carlo simulations must be used to estimate power.

When large sample assumptions are met, the asymptotic power of the Pearson's $\chi^2$ statistic for the composite null hypothesis can be considered by using a sequence of local alternatives for which the model lack-of-fit diminishes

as $n$ increases: i.e.,

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}(\boldsymbol{\theta}) + \boldsymbol{\delta}/\sqrt{n}, \tag{3.9}$$

where $\boldsymbol{\pi}_n$ is the vector of true probabilities for a sample of size $n$. The "best fit" of the model to the population gives $\pi_s(\boldsymbol{\theta})$ as the probability for cell $s$, and the true probability differs from that value by $\delta_s/\sqrt{n}$. The model lack-of-fit goes to zero at the rate $n^{-\frac{1}{2}}$ as $n$ approaches infinity. Mitra (1958) showed that under (3.9) the Pearson's $\chi^2$ statistic has a limiting noncentral chi-squared distribution, with degrees of freedom $T - g - 1$ and non-centrality parameter

$$\lambda = \boldsymbol{\delta}^T \mathbf{D}[\boldsymbol{\pi}(\boldsymbol{\theta})]^{-1}\boldsymbol{\delta}.$$

Reiser (2008) showed that the non-centrality parameter can also be expressed as

$$\lambda = \boldsymbol{\delta}^T \mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T)^{-1}\mathbf{H}\boldsymbol{\delta}.$$

Furthermore according to Reiser (2008), it is possible to decompose the non-centrality parameter into orthogonal components associated with marginals of different order, in a manner very close to the proposed decomposition of the PGF.

Using the QR decomposition of the non-centrality parameter, let

$$\boldsymbol{\zeta} = \mathbf{M}^{*T}\boldsymbol{\delta}, \tag{3.10}$$

where $\mathbf{M}^*$ has been defined in Section 2 of this chapter. Then, $\lambda = \boldsymbol{\zeta}^T\boldsymbol{\zeta}$, and the orthogonal components are $\zeta_j^2$, where $\zeta_j$ is an element of $\boldsymbol{\zeta}$. According to Agresti (2002) it is often reasonable to adopt expression (3.9) for fixed, finite $n$ in order to approximate the distribution of PGF, even though it might not be expected to hold as substantially more data are obtained.

For purposes of power calculations under fixed, finite $n$, cell proportions were generated from a known model, with parameter vector $\boldsymbol{\theta}_a$. These proportions are then multiplied by a selected initial sample size such as $n_0 = 1000$. The model of the null hypothesis was then fit using maximum likelihood on the resulting cell frequencies without any added random variability. Let $\boldsymbol{\theta}_a^*$ be the vector that maximizes the function

$$F(\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta})) = n \sum_s p_s \log(\pi_s(\boldsymbol{\theta})),$$

where $\boldsymbol{\pi}(\boldsymbol{\theta}_a)$ is the vector of multinomial proportions. The vector $\boldsymbol{\delta}^*$ is then chosen such that

$$\boldsymbol{\delta}^* = \sqrt{n}(\boldsymbol{\pi}_a - \boldsymbol{\pi}(\boldsymbol{\theta}_a^*)),$$

where $\boldsymbol{\pi}_a = \boldsymbol{\pi}(\boldsymbol{\theta}_a)$ corresponds to the known generated cell proportions. This method uses $\boldsymbol{\delta}^{*\prime}\mathbf{D}[\boldsymbol{\pi}(\boldsymbol{\theta}_a^*)]^{-1}\boldsymbol{\delta}^*$ as an approximation to $\lambda$. Assuming that $\boldsymbol{\theta}_a$ is close to the value specified by the null hypothesis, it could be expected that

$$n(\boldsymbol{\pi}_n - \boldsymbol{\pi}(\boldsymbol{\theta}_n^*))'\mathbf{D}[\boldsymbol{\pi}(\boldsymbol{\theta}_n^*)]^{-1}(\boldsymbol{\pi}_n - \boldsymbol{\pi}(\boldsymbol{\theta}_n^*)) = \lambda + o(1),$$

where $\boldsymbol{\pi}_n$ is given in expression (3.9) and $\boldsymbol{\theta}_n^*$ is the vector maximizing $\mathrm{F}(\boldsymbol{\pi}_n, \boldsymbol{\pi}(\boldsymbol{\theta}))$.

The chosen value of $\boldsymbol{\delta}^*$ can be used to approximate the non-centrality parameter for the initial sample size $n_0$. The non-centrality parameter for any other sample size, say simply $n$, can be approximated by using the expression $\lambda \approx \frac{n}{n_0}\lambda_0$. Power can then be computed for a specified $\alpha$ as

$$P(X_{(df,\,\lambda)} > \chi^2_{(df,\,\alpha)})$$

where $X$ is $\chi^2$ with $df$ degrees of freedom and non-centrality parameter $\lambda$.

When large sample theory isn't applicable, as is the case when sparseness is present, Monte Carlo simulations can be used to perform a power investigation in order to detect a false null hypothesis. Monte Carlo simulations will also be employed when computing power for test statistics constructed using the random forest method as the number of components is unknown at the outset.

2.2. Type I Error. The Type I error rate shows the percentage of rejections of the hypothesized model under the null. Monte Carlo simulations will be used to compare the empirical $\alpha$ to the nominal $\alpha$ rate. For the random forest method, in addition to power and Type I error rates, the total number of components used in the statistic as well as the individual components that are chosen in each case were also recorded.

# Chapter 4: COMPARISONS OF TEST STATISTICS WITH TWO- AND THREE-WAY ASSOCIATION EFFECTS

To investigate the performance of the proposed methods for selecting chi-squared orthogonal components as a means for providing more powerful tests for large cross-classified tables, various components are defined via the **H** matrix, as discussed previously in Chapter 3. In order to obtain test statistics based on orthogonal components of PGF, the following cases were investigated;

- Full set of marginal frequencies: These marginal frequencies are usually interpretable as they relate lack-of-fit to associations among variables. Summing the components of the full set produces $\chi^2_{PF}$ as shown in expression (3.7).

- First- and second-order marginals: First- and second-order marginals are less sparse than those of higher order, if sparseness is present. Summing components of first- and second-order marginals produces $\chi^2_{[1:2]} = \sum_{k=1}^{(c-1)p+\frac{1}{2}p(p-1)(c-1)^2} \hat{\gamma}^2_k$, where $p$ is the number of variables and $c$ is the number of categories of each variable.

- Second-order marginals only: For psychological data it has been shown that the lack-of-fit is often in the second-order marginals. Second-order marginals will enable an investigation of whether or not including first-order marginals dilutes the test. Summing components of second-order marginals produces $\chi^2_{[2]} = \sum_{k=(c-1)p+1}^{(c-1)p+\frac{1}{2}p(p-1)(c-1)^2} \hat{\gamma}^2_k$.

- Second- and third-order marginals: Summing components

of second- and third-order marginals produces $\chi^2_{[2:3]} = \sum_{k=(c-1)p+1}^{(c-1)p+\frac{1}{2}p(p-1)(c-1)^2+\frac{1}{6}p(p-1)(p-2)(c-1)^3} \hat{\gamma}^2_k$.

Section 1 describes the study of the proposed methods when the model misspecification is in a two-way association not accounted for by the categorical variable factor model with a single latent variable. Section 2 presents results of that study. Section 3 describes the study of the proposed methods when the model misspecification is in a three-way association not accounted for by the categorical variable factor model with a single latent variable. Section 4 presents results of that study.

For all studies the software used was R, and the parameters were estimated using marginal maximum likelihood estimation (MLE) with the Newton-Raphson method. All code used in this research was personally written except for the in-built GRM function in R. The code can be found at http://math.asu.edu/∼jelena. Furthermore, where appropriate, one and two sample proportion tests were performed at the 5% significance level.

1. Two-way Association Effects

The focus of Study 1 was comparing power and Type I error for $\chi^2_{PF}$, test statistics based on components defined on marginals, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, and tests based on the random forest method. The comparisons were performed using the categorical variable factor model, which by nature has large cross-classification tables that often encounter sparseness. Powers for the selected test statistics are calculated when the alternative model to $H_o$ was a two-way association not accounted for by a single latent variable model. Two models

under the null hypothesis were considered for 5 variables, each at 3 categories. In Study 1a the model under the null was the categorical variable factor model with a single latent variable given in expression (2.9) with null hypothesis

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \beta_{1,1} \\ \alpha_{2,1} & \alpha_{2,2} & \beta_{2,1} \\ \alpha_{3,1} & \alpha_{3,2} & \beta_{3,1} \\ \vdots & \vdots & \vdots \\ \alpha_{p,1} & \alpha_{p,2} & \beta_{p,1} \end{pmatrix}. \tag{4.1}$$

The Study 1b null model is a special case where all slopes are equal: i.e.,

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \beta \\ \alpha_{2,1} & \alpha_{2,2} & \beta \\ \alpha_{3,1} & \alpha_{3,2} & \beta \\ \vdots & \vdots & \vdots \\ \alpha_{p,1} & \alpha_{p,2} & \beta \end{pmatrix}. \tag{4.2}$$

The purpose of Study 1b was to investigate if components defined on first-order marginals would contain lack-of-fit information when slopes were constrained. It was suspected that with the constrained version of the model, estimation of intercepts would be affected, and therefore, components defined on first-order marginals might contribute to the power of the test.

Because Study 1 uses a composite null hypothesis, the orthogonal components were calculated from expression (3.2). Given that both models under the null hypothesis are misspecified since they omit a two-way association, the question of how well the selected statistics might perform when sample sizes

were large and not sparse ($n = 5000$) and when sample sizes were small and sparse ($n = 300, 500, 1000$) were investigated.

The alternative hypothesis for both Study 1a and Study 1b included a two-way association for a pair of variables (e.g., variables 1 and 2) not accounted for by the single latent variable. The model in expression (2.9) was extended to include two latent variables, $\boldsymbol{\eta} = (\eta_1, \eta_2)^T$. Thus, the alternative when $H_o$ is false has the parameter matrix, $\boldsymbol{\theta}$, with an additional column and is given by

$$A1_1 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \beta_{1,1} & \beta_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} & \beta_{2,1} & \beta_{2,2} \\ \alpha_{3,1} & \alpha_{3,2} & \beta_{3,1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{p,1} & \alpha_{p,2} & \beta_{p,1} & 0 \end{pmatrix}. \tag{4.3}$$

$A1_1$ denotes the $1^{st}$ alternative investigated in Study 1. In general, $Ai_j$ denotes the $i^{th}$ alternative investigated in Study $j$.

The Study 1 design was as follows:

- $H_o$ model: Categorical variable factor model
- Number of variables: $p = 5$
- Number of categories: 3
- Sample sizes: $n = 300, 500, 1000, 5000$
- Number of samples for the Monte Carlo simulations: 1000

This design produced tables with $3^5 = 243$ cells which in turn for sample sizes

$n = 300, 500, 1000$ are likely to produce cross-classified contingency tables that are sparse in nature.

The test statistics investigated were $X^2_{PF}$, $X^2_{[1:2]}$ and $X^2_{[2]}$ and tests based on the random forest method. With such an alternative hypothesis the model misfit is in the second-order associations, and it was suspected that more focused tests such as $X^2_{[1:2]}$ and $X^2_{[2]}$, along with tests based on the random forest method, would be more powerful in detecting this misfit than $X^2_{PF}$. The $X^2_{[2]}$ statistic should have higher power since $X^2_{[1:2]}$ may have superfluous degrees of freedom which dilute the test, as it includes the first-order marginals which should not detect for the lack-of-fit in the second-order associations.

In both Study 1a and Study 1b, probabilities were calculated from the categorical variable factor model with two latent variables with parameter matrix, $\boldsymbol{\theta}$, given by expression (4.3). A two-way association for variables 1 and 2, not accounted for by the single latent variable, was included. The population parameters in Study 1a for the extended model included the following vectors of $\boldsymbol{\theta}$

$\boldsymbol{\alpha}_1 = (0.75, 0.75, 0.75, 0.75, 0.75)^T$,

$\boldsymbol{\alpha}_2 = (-0.5, -0.5, -0.5, -0.5, -0.5)^T$,

$\boldsymbol{\beta}_1 = (0.5, 0.5, 1, 0.75, 1)^T$,

and

$\boldsymbol{\beta}_2 = (b, b, 0, 0, 0)^T$, where the values of $b$ investigated were $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$.

In Study 1b for the constrained version of the model

$\boldsymbol{\alpha}_1 = (0.75, 0.75, 0.75, 0.75, 0.75)^T$,

$\boldsymbol{\alpha}_2 = (-0.5, -0.5, -0.5, -0.5, -0.5)^T$,

$\boldsymbol{\beta}_1 = (1, 1, 1, 1, 1)^T$,

and

$\boldsymbol{\beta}_2 = (b, b, 0, 0, 0)^T$, where the values of $b$ investigated were $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$.

The integral of expression (2.9) was evaluated as an iterated integral when there were two latent variables, and derivatives were approximated using Gauss-Hermite quadrature with 32 quadrature points and their corresponding weights.

For evaluating the Type I error rate, a true model was fitted and Monte Carlo simulations were performed. To assess the power of a test, a false model was specified. Asymptotic power was computed for all statistics except for the random forest method. For the random forest method the number of components is unknown at the outset, so Monte Carlo simulations were performed in order to approximate power and Type I error rates. Monte Carlo simulations were also used to evaluate the accuracy of the asymptotic power calculations.

2. Results

2.1. Study 1a. Empirical Type I error rates at the nominal 5% level of significance are given in Table 3 for all statistics considered when the model under the null hypothesis is the categorical variable factor model. The proportions in Table 3 multiplied by 1000 are binomial with success proportion 0.05 and 1000 trials. If the true Type I error probability is 0.05, $\sqrt{\frac{(0.05)(0.95)}{1000}} = 0.007$ provides a standard error value that can be used when comparing the table

entries to the nominal level. In particular, it can be seen that the Type I error rates for $\chi^2_{PF}$ and $\chi^2_{[rf]}$ when $n = 300$ are inflated and significantly different from the nominal 5% level. This demonstrates the already well known adverse effects of sparseness on PGF. Power comparisons are thus not reliable for both $\chi^2_{PF}$ and $\chi^2_{[rf]}$ when $n = 300$, since inflated Type I error rates imply that power is confounded with Type I error rate. On the other hand, empirical Type I error rates for $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are not significantly different from the nominal level for all sample sizes considered.

Table 3. Empirical Type I error rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$ when the model under the null hypothesis is the categorical variable factor model. '*' denotes Type I error rates significantly different from 5% nominal level.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\chi^2_{[rf]}$ |
|---|---|---|---|---|
| 300 | 0.125* | 0.056 | 0.060 | 0.084* |
| 500 | 0.054 | 0.062 | 0.053 | 0.054 |
| 1000 | 0.062 | 0.044 | 0.037 | 0.055 |
| 5000 | 0.046 | 0.050 | 0.058 | 0.052 |

QQ-plots of the empirical quantiles for $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ versus those from the appropriate chi-squared distributions when $n = 300$ are attached in Appendix A in Figure 3 and Figure 4, respectively, along with corresponding estimated slopes and p-values. The Kolmogorov-Smirnov goodness-of-fit test statistic for the chi-squared distribution to the $\chi^2_{[1:2]}$ data was found to be $D = 0.0225$ with corresponding p-value of 0.6939. Similarly, the Kolmogorov-Smirnov statistic for $\chi^2_{[2]}$ was $D = 0.0293$ with corresponding p-value of 0.3578.

Thus, both test statistics support the asymptotic chi-square distribution for the considered chi-squared statistics when expected cell frequencies in the joint distribution are small.

QQ-plots for $\chi^2_{PF}$ when $n = 300$ and $n = 500$ are also attached in Appendix A in Figure 1 and Figure 2, respectively. In the case of $n = 300$ the QQ-plot suggests poor asymptotic chi-square approximation for this statistic. The poor asymptotic chi-square approximation was further verified by a Kolmogorov-Smirnov goodness-of-fit test that produced a test statistic value of $D = 0.0777$ having a p-value of $10^{-4}$. On the other hand, for $n = 500$ even though the QQ-plot suggests a somewhat poor fit the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0386$ with p-value of 0.1023) did not detect problems with the chi-square approximation. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are given in Table 4.

Table 4. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. '*' denotes significant p-values at the 5% significance level. '-' denotes that the test was not performed for the given sample size.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|
| 300 | $10^{-4*}$ | 0.6939 | 0.3578 |
| 500 | 0.1023 | - | - |

Table 5 shows asymptotic power rates when the model under the null hypothesis is the categorical variable factor model with no parameter constraints and the alternative of interest is $A1_1$ at the nominal 5% level for $\chi^2_{PF}$,

$\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. Asymptotic power rates of $\chi^2_{PF}$ are not comparable to those of $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $n = 300$ as discussed previously. From Table 5, it can be seen that asymptotic power increases with the increasing two-way effect size and increasing sample size. $\chi^2_{PF}$ is outperformed by both $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $n = 500, 1000, 5000$. Moreover, there is an unsubstantial difference in asymptotic power between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, so components for first-order marginals do not appear to contribute to the power of the test and appear to dilute the test to a minor degree. Also, for $\chi^2_{PF}$ power seems to increase slightly with increasing two-way effect size. Asymptotic power cannot be computed for $\chi^2_{[rf]}$. Graphs of asymptotic power versus a two-way effect size of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $n = 300, 500, 1000, 5000$ are attached in Appendix B.

Table 6 shows empirical power rates when the model under the null hypothesis is the categorical variable factor model with no parameter constraints and the alternative of interest is $A1_1$ at the nominal 5% level for $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$. Corresponding standard errors for the values in the table are given in parentheses next to each value. In finite samples, it is known that the asymptotic power rates of $\chi^2_{PF}$ are only accurate for a small number of cells when the table is not sparse (Maydeu-Olivares and Joe 2005). So, comparing asymptotic power rates in Table 5 to empirical power rates in Table 6 shows that empirical power rates are generally not significantly different from asymptotic power rates for all statistics except for $\chi^2_{PF}$ for $n = 300$. When $n = 300$, the cross-classified tables exhibited severe sparseness with $n/T$ ratio of 1.235 and all expected cell frequencies smaller or equal to 5. Table 6 reveals that $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are not affected by sparseness in this case, and there is no sig-

Table 5. Asymptotic power rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the categorical variable factor model with no parameter constraints and the alternative of interest is $A1_1$ at the nominal 5% level. '*' denotes Type I error rates significantly different from the nominal level.

| Two-way effect size (b) | Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 0.2 | 300 | 0.050* | 0.050 | 0.054 |
| | 500 | 0.050 | 0.051 | 0.051 |
| | 1000 | 0.051 | 0.051 | 0.051 |
| | 5000 | 0.053 | 0.056 | 0.057 |
| 0.4 | 300 | 0.052* | 0.055 | 0.056 |
| | 500 | 0.054 | 0.059 | 0.060 |
| | 1000 | 0.058 | 0.069 | 0.071 |
| | 5000 | 0.099 | 0.182 | 0.203 |
| 0.6 | 300 | 0.061* | 0.076 | 0.079 |
| | 500 | 0.069 | 0.096 | 0.103 |
| | 1000 | 0.092 | 0.162 | 0.180 |
| | 5000 | 0.427 | 0.840 | 0.882 |
| 0.8 | 300 | 0.082* | 0.134 | 0.147 |
| | 500 | 0.110 | 0.216 | 0.242 |
| | 1000 | 0.202 | 0.473 | 0.528 |
| | 5000 | 0.968 | 1.000 | 1.000 |
| 1.0 | 300 | 0.127* | 0.266 | 0.299 |
| | 500 | 0.205 | 0.478 | 0.532 |
| | 1000 | 0.471 | 0.878 | 0.913 |
| | 5000 | 1.000 | 1.000 | 1.000 |

Table 6. Empirical power rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$ when the model under the null hypothesis is the categorical variable factor model with no parameter constraints and the alternative of interest is $A1_1$ at the nominal 5% level. Corresponding standard errors are given in parentheses next to each value. '*' denotes Type I error rates significantly different from the nominal level.

| Two-way effect size (b) | Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\chi^2_{[rf]}$ |
|---|---|---|---|---|---|
| 0.2 | 300 | 0.127* (0.011) | 0.070 (0.008) | 0.062 (0.008) | 0.094*( 0.009) |
| | 500 | 0.054 (0.007) | 0.038 (0.006) | 0.039 (0.006) | 0.049 (0.007) |
| | 1000 | 0.064 (0.008) | 0.050 (0.007) | 0.061 (0.008) | 0.040 (0.006) |
| | 5000 | 0.054 (0.007) | 0.055 (0.007) | 0.051 (0.007) | 0.052 (0.007) |
| 0.4 | 300 | 0.140* (0.011) | 0.079 (0.009) | 0.069 (0.008) | 0.117* (0.010) |
| | 500 | 0.061 (0.008) | 0.060 (0.008) | 0.065 (0.008) | 0.065 (0.008) |
| | 1000 | 0.069 (0.008) | 0.076 (0.008) | 0.069 (0.008) | 0.059 (0.007) |
| | 5000 | 0.098 (0.009) | 0.181 (0.012) | 0.200 (0.013) | 0.095( 0.009) |
| 0.6 | 300 | 0.130* (0.011) | 0.088 (0.009) | 0.102 (0.010) | 0.125* (0.010) |
| | 500 | 0.060 (0.008) | 0.085 (0.009) | 0.088 (0.009) | 0.051 (0.007) |
| | 1000 | 0.089 (0.009) | 0.158 (0.012) | 0.180 (0.012) | 0.087 (0.009) |
| | 5000 | 0.431 (0.016) | 0.848 (0.011) | 0.883 (0.010) | 0.373 (0.015) |
| 0.8 | 300 | 0.148* (0.011) | 0.164 (0.012) | 0.161 (0.012) | 0.136* (0.011) |
| | 500 | 0.117 (0.010) | 0.207 (0.013) | 0.207 (0.013) | 0.105 (0.010) |
| | 1000 | 0.196 (0.013) | 0.473 (0.016) | 0.515 (0.016) | 0.165 (0.012) |
| | 5000 | 0.970 (0.005) | 1.000 (0.000) | 1.000 (0.000) | 0.941 (0.007) |
| 1.0 | 300 | 0.165* (0.012) | 0.249 (0.014) | 0.271 (0.014) | 0.145* (0.011) |
| | 500 | 0.183 (0.012) | 0.449 (0.016) | 0.495 (0.016) | 0.168 (0.012) |
| | 1000 | 0.432 (0.016) | 0.843 (0.012) | 0.880 (0.010) | 0.372 (0.015) |
| | 5000 | 1.000 (0.000) | 1.000 (0.000) | 1.000 (0.000) | 1.000 (0.000) |

nificant difference between $\chi^2_{[2]}$ and $\chi^2_{[1:2]}$ for all sample sizes considered. Also, $\chi^2_{PF}$ is not significantly different from $\chi^2_{[rf]}$ for $n = 500, 1000, 5000$ neither of which are competitive with $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$.

The random forest method was also applied to first- and second-order marginals, $\chi^2_{[rf_{[1:2]}]}$, and second-order marginals only, $\chi^2_{[rf_{[2]}]}$. Although both of these test statistics attained satisfactory empirical Type I error rates at the 5% level, compared to $\chi^2_{[rf]}$ they did not perform well for large sample sizes and thus no further investigation into these statistics was performed when cross-classified tables exhibited sparseness. These results are demonstrated in Table 7. Corresponding standard errors for the values in the table are given in parentheses next to each value. There is no significant difference in empirical power rates between $\chi^2_{[rf_{[1:2]}]}$ and $\chi^2_{[rf_{[2]}]}$. Furthermore, due to the overall poor performance of tests based on the random forest method the implementation of the moment approximation discussed previously in Chapter 2 was not pursued.

2.2. Study 1b. In Study 1a it was discovered that components for first-order marginals did not contribute to the power of the test. However, in Study 1b the model under the null hypothesis was the constrained version of the categorical variable factor model, i.e., all slopes equal, and estimation of intercepts under this model may be affected by constraints on slopes, which in turn might result in first-order marginal components actually contributing to the power of the test.

Empirical Type I error rates at the nominal 5% level are given in Table 8 for all statistics considered when the model under the null hypothesis is the constrained version of the categorical variable factor model. As discussed in

Table 7. Empirical power rates of $\chi^2_{[rf_{[1:2]}]}$ and $\chi^2_{[rf_{[2]}]}$ when the model under the null hypothesis is the categorical variable factor model with no parameter constraints and the alternative of interest is $A1_1$ at the nominal 5% level for $n = 5000$. Corresponding standard errors are given in parentheses next to each value.

| Two-way effect size (b) | $\chi^2_{[rf_{[1:2]}]}$ | $\chi^2_{[rf_{[2]}]}$ |
|---|---|---|
| 0.2 | 0.052 (0.007) | 0.049 (0.006) |
| 0.4 | 0.069 (0.008) | 0.072 (0.008) |
| 0.6 | 0.173 (0.012) | 0.161 (0.012) |
| 0.8 | 0.483 (0.016) | 0.507 (0.016) |
| 1.0 | 0.852 (0.010) | 0.866 (0.011) |

Study 1a, proportions in Table 8 multiplied by 1000 are binomial with success proportion 0.05 and 1000 trials. Moreover, if the true Type I error probability is 0.05, 0.007 provides the standard error value. In particular, it can be seen that the Type I error rates for all test statistics for various sample sizes are not significantly different from the nominal 5% level, which enabled reliable power comparisons for the small sample sizes considered. Even though the data in the cross-classified contingency tables when $n = 300, 500$ are sparse, $\chi^2_{PF}$ had the expected Type I error rate, which is possible if cell probabilities are fairly uniform.

QQ-plots of the empirical quantiles for $\chi^2_{PF}$ and $\chi^2_{[2]}$ when $n = 300$ are attached in Appendix C in Figure 9 and Figure 12, respectively, along with corresponding estimated slopes and p-values. The Kolmogorov-Smirnov

Table 8. Empirical Type I error rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\chi^2_{[rf]}$ |
|---|---|---|---|---|
| 300 | 0.053 | 0.049 | 0.055 | 0.046 |
| 500 | 0.058 | 0.040 | 0.049 | 0.055 |
| 1000 | 0.057 | 0.044 | 0.045 | 0.050 |
| 5000 | 0.060 | 0.050 | 0.055 | 0.052 |

goodness-of-fit test statistic for the chi-square distribution to the $\chi^2_{PF}$ data was found to be $D = 0.0186$ with corresponding p-value of 0.8793. Similarly, the Kolmogorov-Smirnov statistic for $\chi^2_{[2]}$ was $D = 0.0145$ with p-value of 0.9850. Thus, both test statistics support the asymptotic chi-square distribution for the considered chi-squared statistics when expected cell frequencies in the joint distribution are small.

QQ-plots for $\chi^2_{[1:2]}$ when $n = 300$ and $n = 500$ are also attached in Appendix C in Figure 10 and Figure 11, respectively, along with corresponding estimated slopes and p-values. In the case of $n = 300$ even though the QQ-plot does not suggest poor asymptotic chi-square approximation for this statistic the Kolmogorov-Smirnov goodness-of-fit test produced a test statistic value of $D = 0.0584$ having a p-value of 0.0022, which suggests the contrary. On the other hand, for $n = 500$ even though the QQ-plot suggests a somewhat poor fit, the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0385$ with p-value 0.1036) did not detect problems with the chi-square approximation. As discussed in Study 1a, due to the overall poor performance of tests

based on the random forest method the implementation of the moment approximation discussed previously in Chapter 2 was not pursued. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are given in Table 9.

Table 9. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. '*' denotes significant p-values at the 5% significance level. '-' denotes that the test was not performed for the given sample size.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|
| 300 | 0.7893 | $0.0022^*$ | 0.9850 |
| 500 | - | 0.1036 | - |

Table 10 shows asymptotic power rates when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ at the nominal 5% level for $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. From Table 10 it can be seen that asymptotic power rates increase with increasing two-way effect size and increasing sample size. Table 10 also suggests that there is an unsubstantial difference in asymptotic power between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, so components for first-order marginals do not appear to contribute to the power of the test. $\chi^2_{PF}$ is outperformed by both $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, for all sample sizes. For small sample sizes $n = 300, 500, 1000$ asymptotic power rates for $\chi^2_{PF}$ are generally half those of $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. Graphs of asymptotic power versus a two-way effect size of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $n = 300, 500, 1000, 5000$ are attached in Appendix D.

Table 11 shows empirical power rates when the model under the null hy-

Table 10. Asymptotic power rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ at the nominal 5% level.

| Two-way effect size (b) | Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 0.2 | 300 | 0.050 | 0.050 | 0.050 |
| | 500 | 0.050 | 0.050 | 0.050 |
| | 1000 | 0.050 | 0.051 | 0.051 |
| | 5000 | 0.052 | 0.054 | 0.055 |
| 0.4 | 300 | 0.051 | 0.053 | 0.054 |
| | 500 | 0.052 | 0.055 | 0.056 |
| | 1000 | 0.055 | 0.061 | 0.063 |
| | 5000 | 0.078 | 0.122 | 0.132 |
| 0.6 | 300 | 0.057 | 0.066 | 0.068 |
| | 500 | 0.062 | 0.078 | 0.082 |
| | 1000 | 0.075 | 0.114 | 0.124 |
| | 5000 | 0.252 | 0.584 | 0.640 |
| 0.8 | 300 | 0.070 | 0.101 | 0.109 |
| | 500 | 0.087 | 0.148 | 0.162 |
| | 1000 | 0.138 | 0.301 | 0.337 |
| | 5000 | 0.802 | 0.994 | 0.997 |
| 1.0 | 300 | 0.098 | 0.182 | 0.202 |
| | 500 | 0.144 | 0.316 | 0.353 |
| | 1000 | 0.301 | 0.676 | 0.729 |
| | 5000 | 0.999 | 1.000 | 1.000 |

pothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ at the nominal 5% level for $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$ when $n = 300, 500, 1000, 5000$. Corresponding standard errors are given in parentheses next to each value. Comparing asymptotic power rates in Table 10 to empirical power rates in Table 11, shows that empirical power rates are generally not significantly different from asymptotic power rates for all statistics considered. Empirical power rates in Table 11 for $\chi^2_{PF}$ and $\chi^2_{[rf]}$ are not

significantly different from each other, irrespective of the two-way effect size and sample size. $\chi^2_{[1:2]}$ did not result in significantly higher power compared to $\chi^2_{[2]}$ as initially suspected. The initial thought was that first-order marginals would contain useful information about lack-of-fit because of the constraint. The estimators of the intercepts may be biased, so first-order marginals may not be fit well. In such case, components for first-order marginals may contribute to the power of the test. It can be seen from Table 11, that $\chi^2_{[2]}$ is not affected by sparseness in this case. On the other hand, for $n = 300$ although the size of the test is accurate for $\chi^2_{[1:2]}$ the Kolmogorov-Smirnov goodness-of-fit test detected problems with the chi-square approximation for this sample size. Moreover, for $n = 500, 1000, 5000$ there is no significant difference between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for all two-way effect sizes considered. For scenarios with a two-way effect sizes of $0.2, 0.4, 0.6$ and small sample sizes there is no significant difference between any of the test statistics considered. However, $\chi^2_{PF}$ and $\chi^2_{rf}$ are not competitive with test statistics defined on marginal frequencies when a two-way effect is greater than 0.6 and increasing sample sizes. As in Study 1a, results of Study 1b also showed that first-order marginal components do not contribute to the power of the test.

Even though some sets of parameters produce cell frequencies where the effects of sparseness are not seen, as the case above, some parameter values on the other hand do reveal adverse effects of sparseness. Table 12 shows empirical Type I error rates for two sets of parameters when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ with two parameter value matrices

Table 11. Empirical power rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ at the nominal 5% level. Corresponding standard errors are given in parentheses next to each value.

| Two-way effect size (b) | Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\chi^2_{[rf]}$ |
|---|---|---|---|---|---|
| 0.2 | 300 | 0.044 (0.006) | 0.035 (0.006) | 0.044 (0.006) | 0.047 (0.007) |
| | 500 | 0.059 (0.007) | 0.045 (0.007) | 0.046 (0.007) | 0.045 (0.007) |
| | 1000 | 0.059 (0.007) | 0.046 (0.007) | 0.053 (0.007) | 0.048 (0.007) |
| | 5000 | 0.061 (0.008) | 0.058 (0.007) | 0.054 (0.007) | 0.054 (0.007) |
| 0.4 | 300 | 0.056 (0.007) | 0.056 (0.007) | 0.058 (0.007) | 0.058 (0.007) |
| | 500 | 0.058 (0.007) | 0.043 (0.006) | 0.045 (0.007) | 0.049 (0.007) |
| | 1000 | 0.059 (0.007) | 0.044 (0.006) | 0.057 (0.007) | 0.047 (0.007) |
| | 5000 | 0.067 (0.008) | 0.096 (0.009) | 0.113 (0.010) | 0.061 (0.008) |
| 0.6 | 300 | 0.054 (0.007) | 0.049 (0.007) | 0.062 (0.008) | 0.054 (0.007) |
| | 500 | 0.066 (0.008) | 0.080 (0.009) | 0.099 (0.009) | 0.059 (0.007) |
| | 1000 | 0.081 (0.009) | 0.102 (0.010) | 0.120 (0.010) | 0.081 (0.009) |
| | 5000 | 0.247 (0.014) | 0.547 (0.016) | 0.556 (0.016) | 0.204 (0.013) |
| 0.8 | 300 | 0.068 (0.008) | 0.085 (0.009) | 0.094 (0.009) | 0.062 (0.008) |
| | 500 | 0.103 (0.010) | 0.136 (0.011) | 0.149 (0.011) | 0.092 (0.009) |
| | 1000 | 0.145 (0.011) | 0.251 (0.014) | 0.253 (0.014) | 0.106 (0.010) |
| | 5000 | 0.807 (0.012) | 0.995 (0.002) | 0.986 (0.004) | 0.700 (0.014) |
| 1.0 | 300 | 0.108 (0.010) | 0.159 (0.012) | 0.181 (0.012) | 0.093 (0.009) |
| | 500 | 0.145 (0.011) | 0.283 (0.014) | 0.302 (0.015) | 0.115 (0.010) |
| | 1000 | 0.310 (0.015) | 0.638 (0.015) | 0.640 (0.015) | 0.244 (0.014) |
| | 5000 | 0.999 (0.001) | 1.000 (0.000) | 1.000 (0.000) | 0.996 (0.002) |

given by

$$\boldsymbol{\theta}_1 = \begin{pmatrix} 0.50 & 0.25 & 1.00 \\ 0.95 & 0.00 & 1.00 \\ 0.50 & -0.50 & 1.00 \\ 2.00 & -0.75 & 1.00 \\ 0.5 & -0.25 & 1.00 \end{pmatrix}, \tag{4.4}$$

and

$$\boldsymbol{\theta}_2 = \begin{pmatrix} 0.75 & 0.05 & 1.00 \\ 1.00 & 0.25 & 1.00 \\ 0.50 & -0.05 & 1.00 \\ 1.00 & -0.75 & 1.00 \\ 0.50 & -0.25 & 1.00 \end{pmatrix}. \tag{4.5}$$

For both sets of parameter values, Type I error rates for $\chi^2_{PF}$ are inflated at the nominal 1%, 5% and 10% levels, which demonstrates the adverse effects of sparseness for $n = 300$.

Table 12. Empirical Type I error rates of $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model with parameter value matrices given by (4.4) and (4.5) respectively, for $n = 300$.

| Parameter | $\chi^2_{PF}$ | | | $\chi^2_{[1:2]}$ | | | $\chi^2_{[2]}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| values | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $\boldsymbol{\theta}_1$ | 0.069 | 0.121 | 0.174 | 0.014 | 0.044 | 0.092 | 0.099 | 0.038 | 0.097 |
| $\boldsymbol{\theta}_2$ | 0.071 | 0.142 | 0.183 | 0.012 | 0.052 | 0.091 | 0.010 | 0.058 | 0.106 |

3. Three-way Association Effects

The primary focus of Study 2 was comparing the powers of $\chi^2_{PF}$, orthogonal components, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[2:3]}$, and $\chi^2_{[rf]}$ using the categorical variable factor model when the lack-of-fit was in a three-way association not accounted for by a single latent variable. Study 2 consists of two sub-studies, Study 2a and Study 2b. In Study 2a, the model under the null hypothesis was the categorical variable factor model given by

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta_{1,1} \\ \alpha_{2,1} & \beta_{2,1} \\ \alpha_{3,1} & \beta_{3,1} \\ \vdots & \vdots \\ \alpha_{p,1} & \beta_{p,1} \end{pmatrix}. \tag{4.6}$$

In Study 2b, the model under the null hypothesis was the constrained version of the categorical variable factor model given by

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta \\ \alpha_{2,1} & \beta \\ \alpha_{3,1} & \beta \\ \vdots & \vdots \\ \alpha_{p,1} & \beta \end{pmatrix}. \tag{4.7}$$

Because Study 2 also uses a composite null hypothesis, the orthogonal components were calculated from expression (3.2). Powers for the selected test statistics are calculated when the alternative to $H_o$ included a three-way association not accounted for by a single latent variable model. Answering the question of how well these selected test statistics might perform when sample

sizes were large and not sparse ($n = 1000$) and when sample sizes were small and sparse ($n = 100, 300, 500$) was the goal of this study.

The log-linear version of the model described in expression (2.15) has the advantage that it is convenient to demonstrate the influence of higher-order interactions (in particular a three-way association in this study) and to estimate the model with widely used software. Consequently, for the purpose of calculating cell frequencies under the alternative hypothesis, the log-linear form of the model was used. Power calculations and simulations were performed using the logistic form of the model as described previously.

In both Study 2a and Study 2b, cell frequencies were generated from the log-linear model given in expression (2.15) with

$\lambda = 0.5,$

$\lambda_1^{Y_1} = -0.15,$

$\lambda_1^{Y_2} = -0.10,$

$\lambda_1^{Y_3} = 0,$

$\lambda_1^{Y_4} = 0.10,$

$\lambda_1^{Y_5} = 0.15,$

$\lambda_{11}^{Y_iY_j} = \lambda_{00}^{Y_iY_j} = -\lambda_{01}^{Y_iY_j} = -\lambda_{10}^{Y_iY_j} = 0.2,$ for $i, j = 1, 2, 3, 4, 5,$

and,

$\lambda_{001}^{Y_2Y_3Y_4} = \lambda_{010}^{Y_2Y_3Y_4} = \lambda_{100}^{Y_2Y_3Y_4} = \lambda_{111}^{Y_2Y_3Y_4} = -\lambda_{000}^{Y_2Y_3Y_4} = -\lambda_{011}^{Y_2Y_3Y_4} = -\lambda_{101}^{Y_2Y_3Y_4} = -\lambda_{110}^{Y_2Y_3Y_4} = k,$ where the values of $k$ investigated were $\{0, 0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.175\}$.

It was suspected that both $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ would perform poorly as the lack-of-fit is in third-order marginals, and they only extend to second-order

marginals. Moreover, since the random forest method was applied to all marginals it may be able to identify this misfit in a three-way association, unlike $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. The expectation was that power calculations for $\chi^2_{[2:3]}$ and $\chi^2_{PF}$ would be fairly equal when data were not sparse for 5 variables as $\chi^2_{PF}$ has degrees of freedom given by $2^5 - 10 - 1 = 21$ and $\chi^2_{[2:3]}$ has 20 degrees of freedom from the sum of 10 second-order and 10 third-order marginals.

The Study 2 design was as follows:

- $H_o$ model: Categorical variable factor model
- Number of variables: $p = 5$
- Number of categories: 2
- Sample sizes: $n = 100, 300, 500, 1000$
- Number of samples for the Monte Carlo simulations: 1000

In this study, 5 variables, each at 2 categories resulted in tables with $2^5 = 32$ cells which in turn for sample sizes $n = 100, 300, 500$ is likely to produce cross-classified contingency tables that are sparse in nature.

For evaluating the Type I error rate, a true model was fitted and Monte Carlo simulations were performed. To assess the power of a test, a false model was specified. Asymptotic power was computed for all statistics except when the random forest method was applied since the number of components is unknown at the outset. Monte Carlo simulations were performed in order to approximate power and Type I error rates when the random forest method

was used. Monte Carlo simulations were also used to evaluate the accuracy of the asymptotic power calculations.

4. Results

4.1. Study 2a. The focus of Study 2a was comparing components defined on marginals, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[2:3]}$, to $\chi^2_{PF}$ and $\chi^2_{[rf]}$. Using the alternative described in (2.15), data were generated for 5 variables, each at 2 categories, for the large sample case of $n = 1000$. However, the investigation of components prior to performing any simulations revealed that for three-way association values components were not calculated nor ordered correctly when the QR decomposition was applied. Namely, prior to any Type I error or power calculations it was imperative that components obtained from the QR decomposition were checked against calculations of components obtained from the sequential sum of squares discussed in Chapter 3, Section 1. Since the QR decomposition is an alternative to the sequential sums of squares approach, the two methods should produce the same orthogonal components. The mismatch in component calculations between the QR decomposition and the sequential sum of squares was discovered in this study, irrespective of the parameters values selected.

To demonstrate this mismatch in component calculations, parameter values described previously were used with $k = 0.15$ in expression (2.15). The generated three-way association was between variables $Y_2, Y_3,$ and $Y_4$ and should manifest in a 'large' magnitude for component 22 when the QR decomposition was applied. Only $\chi^2_{[2:3]}$, which has 20 degrees of freedom and

is the sum of components 6 through 25 inclusive, should have included this large component. However, as can be seen from the components below, the largest component was not component 22.

$$\hat{\boldsymbol{\gamma}}^2_{(\text{QR decomposition})} = \begin{pmatrix} 3.622 \\ 1.084 \\ 0.041 \\ 0.003 \\ 0.008 \\ 0.002 \\ 0.005 \\ 0.071 \\ 0.693 \\ 0.151 \\ 0.067 \\ 0.007 \\ 0.003 \\ 1.024 \\ 0.070 \\ 0.213 \\ 1.688 \\ 0.001 \\ 2.831 \\ 0.463 \\ 5.472 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix}.$$

On the other hand, components below were obtained from the sequential sum of squares and resulted in the 'largest' value for component 22, as should have been the case.

$$\hat{\boldsymbol{\gamma}}^2_{(\text{Sequential sum of squares})} = \begin{pmatrix} 3.622 \\ 1.084 \\ 0.041 \\ 0.003 \\ - \\ 0.003 \\ 0.005 \\ 0.072 \\ 0.693 \\ 0.147 \\ 0.068 \\ 0.010 \\ 0.003 \\ 0.000 \\ - \\ 0.030 \\ 0.021 \\ 1.323 \\ 0.000 \\ 1.151 \\ 0.798 \\ 8.441 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix}.$$

The variance inflation factors, VIF, for each of the components are shown in Table 13. A '-' is recorded for components 5 and 15 since they are both linear combinations of other components. Namely, there is a linear dependency among first- as well as second-order marginal proportions such that they both sum to 1. Therefore, the number of orthogonal components for first-order marginals is one less than the number of first-order marginals. The same is true for the number of components for second-order marginals. Specifically, there are linear dependencies among the columns of $\mathbf{M}$. Large variance

inflation factors for the first 15 components suggest severe collinearity among the columns of **M**. The QR decomposition as a result fails in the calculation as well as the ordering of components. Unlike the QR decomposition which does not detect linear dependencies or make any adjustments, the sequential sum of squares are calculated using Goodnight's (1978) sweep operator. As discussed in Chapter 3, the critical issue here is not the sequential sum of squares versus the QR decomposition. The issue here is writing the code for the decomposition to deal with linear dependencies as does the Goodnight's code for the sweep operator. The routines for the QR decomposition in R and in SAS IML are not written to check carefully for linear dependencies. They both could be written to check for linear dependencies, and then they would be as reliable as the sweep operator in PROC REG in SAS which is used to obtain the sequential sums of squares.

Even though the sequential sum of squares produced orthogonal components in an accurate order which could have been used in the calculations of Type I error rates and power, since the primary focus of this research was in using the QR decomposition no further work was pursued including a three-way association. This problem was not encountered in other studies as there were no linear dependencies detected among the columns of **M**.

4.2. Study 2b. Given the outcome of Study 2a, the focus of Study 2b was to produce components accurately calculated and ordered when the model under the null hypothesis is the constrained version of the categorical factor variable model with a single latent variable. Since the constrained model exhibits less collinearity than the unconstrained version of the model the hope

Table 13. Variance inflation factors for the components obtained from the sequential sum of squares when the model under the null hypothesis is the unconstrained version of the categorical variable factor model.

| Component | VIF |
|---|---|
| 1 | 76627489032 |
| 2 | $1.052305 * 10^{12}$ |
| 3 | $1.064977 * 10^{12}$ |
| 4 | 91823087638 |
| 5 | - |
| 6 | 1387846 |
| 7 | 573399 |
| 8 | 135991 |
| 9 | 425046 |
| 10 | 727454 |
| 11 | 231749 |
| 12 | 585507 |
| 13 | 2564.886 |
| 14 | 127171 |
| 15 | - |
| 16 | 12.240 |
| 17 | 4.599 |
| 18 | 10.486 |
| 19 | 2.427 |
| 20 | 2.662 |
| 21 | 2.401 |

was that components obtained using the QR decomposition would be calculated correctly. Data were generated under the same conditions as in Study 2a. The fitted model has 25 degrees of freedom, and once again component 22 was not the largest in magnitude.

$$\hat{\gamma}^2_{(\text{QR decomposition})} = \begin{pmatrix} 3.701 \\ 1.074 \\ 0.057 \\ 0.000 \\ 0.020 \\ 0.006 \\ 0.012 \\ 0.071 \\ 0.126 \\ 0.003 \\ 0.004 \\ 0.227 \\ 0.000 \\ 0.000 \\ 0.060 \\ 0.812 \\ 0.201 \\ 1.229 \\ 1.497 \\ 8.605 \\ 0.000 \\ 0.006 \\ 0.003 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix}.$$

On the other hand, components below were obtained from the sequential sum of squares and resulted in the 'largest' value for component 22 once again.

$$\hat{\gamma}^2_{(\text{Sequential sum of squares})} = \begin{pmatrix} 3.701 \\ 1.074 \\ 0.057 \\ - \\ - \\ 0.000 \\ 0.020 \\ 0.006 \\ 0.693 \\ 0.012 \\ 0.126 \\ 0.003 \\ 0.004 \\ 0.227 \\ - \\ 0.030 \\ 0.060 \\ 0.812 \\ 0.201 \\ 1.229 \\ 1.497 \\ 8.605 \\ 0.000 \\ 0.006 \\ - \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.000 \end{pmatrix}.$$

The variance inflation factors, VIF, for each of the components are shown in Table 14. A '-' is recorded for components 4, 5, 15 and 25 since they are linear combinations of other components. The presence of the linear dependency among components was explained previously in Study 2a. Large variance inflation factors were obtained only for the first 3 components. Although severe collinearity among the components is present, components

Table 14. Variance inflation factors for components obtained from the sequential sum of squares when the model under the null hypothesis is the constrained version of the categorical variable factor model.

| Component | VIF |
|---|---|
| 1 | 236349 |
| 2 | 597210 |
| 3 | 84319 |
| 4 | - |
| 5 | - |
| 6 | 5.459 |
| 7 | 4.555 |
| 8 | 4.542 |
| 9 | 3.900 |
| 10 | 3.087 |
| 11 | 3.307 |
| 12 | 2.959 |
| 13 | 2.761 |
| 14 | 2.534 |
| 15 | - |
| 16 | 3.737 |
| 17 | 3.435 |
| 18 | 3.160 |
| 19 | 3.439 |
| 20 | 3.193 |
| 21 | 3.007 |
| 22 | 2.814 |
| 23 | 2.655 |
| 24 | 2.528 |
| 25 | - |

obtained from the QR decomposition are much 'closer' in magnitude to components obtained from the sequential sum of squares unlike the case with the unconstrained version of the model in Study 2a. Overall, using the QR decomposition in Study 2b fails in the calculation as well as the ordering of components when the model under the null hypothesis is the constrained categorical variable factor model. Even though the sequential sum of squares produced orthogonal components in an accurate order which could have been used in the calculations of Type I error rates and power, since the primary focus of this research was in using the QR decomposition no further work was pursued including a three-way association.

# Chapter 5: COMPARING NESTED MODELS

In this Chapter, Section 1 describes the study of the proposed methods when the model misspecification is in parameter constraints. The categorical variable factor model for large and sparse cross-classified contingency tables was used. Section 2 presents results of that study.

## 1. Comparing Nested Models

The focus of Study 3 was comparing power for $\chi^2_{PF}$ and $LR_{diff}$ and test statistics based on orthogonal components defined on marginals, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ for multi-category variables and large sparse contingency tables. Study 3 consists of two sub-studies, Study 3a and Study 3b. Study 3a considered two cases, 5 variables, each at 2 categories and 5 variables, each at 3 categories, when the model misspecification was in parameter constraints using the categorical variable factor model. Study 3b investigated the case of larger and sparse cross-classified contingency tables when 10 dichotomous variables were considered. Essentially, Study 3b is an extension of Study 3a with a larger number of variables, which in turn resulted in larger cross-classified tables and a large number of orthogonal components, i.e., $2^{10} = 1024$ components.

Results from Agresti and Yang (1987) showed that the likelihood ratio difference statistic performs well in sparse tables. So, it may be a competitor in terms of power to a test based on lower-order marginal components when data are sparse. As discussed in Chapter 2, the $LR_{diff}$ statistic was calculated from expression (2.7), where $\mathcal{M}_1$ is the constrained version of the categorical variable factor model and $\mathcal{M}_2$ is the unconstrained version. Agresti and Yang

(1987) also demonstrated that the likelihood ratio difference statistic behaves quite well for some sparse two-way tables.

In common log-linear models, the expected cell counts for the categorical variable factor model are functions of cell counts in the lower-dimensional marginal tables that are the minimal sufficient statistics. These tables are much less sparse than the full table (Agresti and Yang 1987).

The comparison of $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ to the closely related statistic $M_2$ is also investigated in this study. $LR_{diff}$ may have higher power followed by $\chi^2_{[2]}$ and $M_2$, which should have fairly close power rates for the scenario considered.

1.1. Multi-category Variables. Test statistics investigated in Study 3a were $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. In Study 3a the model under the null hypothesis for 5 variables, each at 2 categories was the constrained categorical variable factor model with the null hypothesis as given by expression (4.7)

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta \\ \alpha_{2,1} & \beta \\ \alpha_{3,1} & \beta \\ \vdots & \vdots \\ \alpha_{p,1} & \beta \end{pmatrix},$$

and the alternative of interest was as given by expression (4.6)

$$A1_3 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta_{1,1} \\ \alpha_{2,1} & \beta_{2,1} \\ \alpha_{3,1} & \beta_{3,1} \\ \vdots & \vdots \\ \alpha_{p,1} & \beta_{p,1} \end{pmatrix}.$$

The model under the null hypothesis for 5 variables, each at 3 categories was the constrained categorical variable factor model with the null hypothesis as given by expression (4.2)

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \beta \\ \alpha_{2,1} & \alpha_{2,2} & \beta \\ \alpha_{3,1} & \alpha_{3,2} & \beta \\ \vdots & \vdots & \vdots \\ \alpha_{p,1} & \alpha_{p,2} & \beta \end{pmatrix},$$

and the alternative of interest was as given by expression (4.1)

$$A2_3 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \beta_{1,1} \\ \alpha_{2,1} & \alpha_{2,2} & \beta_{2,1} \\ \alpha_{3,1} & \alpha_{3,2} & \beta_{3,1} \\ \vdots & \vdots & \vdots \\ \alpha_{p,1} & \alpha_{p,2} & \beta_{p,1} \end{pmatrix}.$$

This alternative departs from the null model due to the misspecified parameter constraints. It was suspected that with the constrained version of the model, estimation of intercepts would be affected and therefore, components defined on first-order marginals might contribute to the power of the test. The question of how well the selected statistics might perform when variables are multi-category in nature when sample sizes were large and not sparse and when samples sizes were small and sparse was investigated. Furthermore, $LR_{diff}$ may outperform all other statistics in terms of power as it will usually have a smaller number of degrees of freedom, as discussed previously in Chapter 2.

Study 3a design was as follows:

- $H_o$ model: Categorical variable factor model
- Number of variables: $p = 5$
- Number of categories: $2, 3$
- Sample sizes: 5 variables, each at 2 categories, $n = 100, 300, 500, 1000$

  5 variables, each at 3 categories, $n = 300, 500, 1000, 5000$
- Number of samples for the Monte Carlo simulations: 1000

In this study, 5 variables, each at 2 categories resulted in tables with $2^5 = 32$ cells which in turn for sample sizes $n = 100, 300$ are likely to produce cross-classified tables that are sparse in nature. Moreover, 5 variables, each at 3 categories resulted in $3^5 = 243$ cells which for sample sizes $n = 300, 500$ are also likely to produce cross-classified tables that are sparse in nature. Because Study 3a uses a composite null hypothesis the orthogonal components were calculated from expression (3.2).

In Study 3a probabilities were calculated for the categorical variable factor model described in expression (2.9). The parameters for the alternative model with dichotomous variables were

$\boldsymbol{\alpha}_1 = (0.75, 0.75, 0.75, 0.75, 0.75)^T$,

and

$\boldsymbol{\beta}_1 = (0.5, 0.5, 1, 0.75, 1)^T$.

For variables with 3 categories the parameters were

$\boldsymbol{\alpha}_1 = (0.75, 0.75, 0.75, 0.75, 0.75)^T$,

$\boldsymbol{\alpha}_2 = (-0.25, -0.25, -0.25, -0.25, -0.25)^T$,

and

$\boldsymbol{\beta}_1 = (0.5, 0.5, 1, 0.75, 1)^T$.

For evaluating the Type I error rate, a true model was fitted and Monte Carlo simulations were performed. To assess the power of a test, a false model was specified. Asymptotic power was computed for all test statistics using the power calculations described earlier in Section 2 of Chapter 3. Monte Carlo simulations were used to evaluate the accuracy of the asymptotic power calculations.

1.2. Large Cross-classified Contingency Tables with a Large Number of Variables. The primary interest of Study 3b was investigating how well $\chi^2_{PF}$, $LR_{diff}$, $M_2$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ performed when sample sizes were very sparse for a large number of variables. In Study 3b the model under the null hypothesis was the constrained categorical variable factor model with the null hypothesis as given by expression (4.7)

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta \\ \alpha_{2,1} & \beta \\ \alpha_{3,1} & \beta \\ \vdots & \vdots \\ \alpha_{p,1} & \beta \end{pmatrix},$$

and the alternative of interest was as given by expression (4.6)

$$A3_3 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} \alpha_{1,1} & \beta_{1,1} \\ \alpha_{2,1} & \beta_{2,1} \\ \alpha_{3,1} & \beta_{3,1} \\ \vdots & \vdots \\ \alpha_{p,1} & \beta_{p,1} \end{pmatrix}.$$

Study 3b design was as follows:

- $H_o$ model: Categorical variable factor model
- Number of variables: $p = 10$
- Number of categories: 2
- Sample sizes: $n = 300, 500, 750, 1000$
- Three levels $(0.5, 1.0, 1.5)$ for the true average slope parameter $\bar{\boldsymbol{\beta}}_1$
- Number of samples for the Monte Carlo simulations: 1000

In this study, 10 variables, each at 2 categories resulted in tables with $2^{10} = 1024$ cells which in turn for sample sizes $n = 300, 500, 750, 1000$ are likely to produce cross-classified tables that are sparse in nature. Because this study uses a composite null hypothesis the orthogonal components were calculated from expression (3.2).

For evaluating the Type I error rate, a true model was fitted and Monte Carlo simulations were performed. To assess the power of a test, a false model was specified. The population parameters for the model under the alternative

hypothesis were

$\boldsymbol{\alpha}_1 = (-2.7, -2.1, -1.5, -0.9, -0.3, 0.3, 0.9, 1.5, 2.1, 2.7)^T$,

and three levels of the true average slope parameter with $\bar{\boldsymbol{\beta}}_1 = 0.5, 1.0, 1.5$. Namely, the average of the true slope parameters of column vector

$\boldsymbol{\beta}_1 = (0.35, 0.25, 0.8, 0.5, 0.6, 0.6, 0.5, 0.8, 0.25, 0.35)^T$ is $\bar{\boldsymbol{\beta}}_1 = 0.5$,

$\boldsymbol{\beta}_1 = (0.35, 0.25, 0.8, 0.5, 0.6, 0.6, 0.5, 0.8, 0.25, 0.35)^T$ is $\bar{\boldsymbol{\beta}}_1 = 1$,

and

$\boldsymbol{\beta}_1 = (2.0, 1.35, 1.65, 1.0, 1.5, 1.5, 1.0, 1.65, 1.35, 2.0)^T$ is $\bar{\boldsymbol{\beta}}_1 = 1.5$.

Asymptotic power was computed for all test statistics using the power calculations described earlier in Chapter 3. Monte Carlo simulations were used to evaluate the accuracy of the asymptotic power calculations.

2. Results

2.1. Study 3a. The primary focus of Study 3a was comparing power for test statistics based on components defined on marginals, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, $\chi^2_{PF}$ and $LR_{diff}$ for a multi-category setting: i.e., 5 variables, each at 2 categories and 5 variables, each at 3 categories.

Empirical Type I error rates at nominal significance levels $(1\%, 5\%, 10\%)$ are given in Table 15 when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories. If the true Type I error probabilities are $0.01, 0.05$ and $0.10$, then $0.003, 0.007$ and $0.009$ provide respective standard error values that can be used when comparing the table entries to the nominal levels. In particular, it can be seen that the Type I error rates for all statistics considered

when $n = 100$ are significantly different only from the nominal 10% level. However, all power comparisons were performed at the 5% significance level which is not significantly different from its respective nominal level. On the other hand, empirical Type I error rates for all test statistics are not significantly different from their respective nominal levels for $n = 300, 500, 1000$, which enabled reliable power comparisons for the small sample sizes considered.

Table 15. Empirical Type I error rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories. '*' denotes Type I error rates significantly different from the specified nominal level.

| | $\chi^2_{PF}$ | | | $LR_{diff}$ | | |
|---|---|---|---|---|---|---|
| Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 100 | 0.008 | 0.049 | 0.081* | 0.011 | 0.062 | 0.121* |
| 300 | 0.012 | 0.045 | 0.096 | 0.012 | 0.057 | 0.106 |
| 500 | 0.010 | 0.052 | 0.098 | 0.008 | 0.049 | 0.098 |
| 1000 | 0.013 | 0.054 | 0.107 | 0.010 | 0.050 | 0.099 |
| | $\chi^2_{[1:2]}$ | | | $\chi^2_{[2]}$ | | |
| Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 100 | 0.012 | 0.036 | 0.076* | 0.012 | 0.042 | 0.081* |
| 300 | 0.013 | 0.056 | 0.089 | 0.012 | 0.060 | 0.114 |
| 500 | 0.011 | 0.053 | 0.094 | 0.011 | 0.053 | 0.098 |
| 1000 | 0.009 | 0.051 | 0.112 | 0.014 | 0.055 | 0.104 |

A QQ-plot of the empirical quantiles for $\chi^2_{[1:2]}$ when $n = 100$ is attached in Appendix E in Figure 21 along with corresponding estimated slope and p-values. The Kolmogorov-Smirnov statistic for $\chi^2_{[1:2]}$ was $D = 0.0286$ with p-value of 0.3886. Thus, the test statistic supports the asymptotic chi-square distribution for the considered chi-squared statistic when expected cell

frequencies in the joint distribution are small.

QQ-plots for $\chi^2_{PF}$ when $n = 100$ and $n = 300$ are also attached in Appendix E in Figure 17 and Figure 18, respectively, along with corresponding estimated slopes and p-values. In the case of $n = 100$ the QQ-plot and the Kolmogorov-Smirnov goodness-of-fit test that produced a test statistic value of $D = 0.0457$ having a p-value of 0.0305 suggest poor asymptotic chi-square approximation for this statistic. However, when $n = 300$ neither the QQ-plot nor the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0180$ with p-value 0.9024) detected problems with the chi-square approximation even though the contingency table was sparse in nature.

Similarly, QQ-plots for $\chi^2_{LR_{diff}}$ when $n = 100$ and $n = 300$ are also attached in Appendix E in Figure 19 and Figure 20, respectively, along with corresponding estimated slopes and p-values. In the case of $n = 100$ the QQ-plot and the Kolmogorov-Smirnov goodness-of-fit test produced a test statistic value of $D = 0.0527$ having a p-value of 0.0078, suggest poor asymptotic chi-square approximation for this statistic. However, when $n = 300$ even though the QQ-plot appears to be very poor fit for this statistic the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0197$ with p-value 0.8345) did not detect problems with the chi-square approximation.

Furthermore, QQ-plots for $\chi^2_{[2]}$ when $n = 100$ and $n = 300$ are also attached in Appendix E in Figure 22 and Figure 23, respectively, along with corresponding estimated slopes. In the case of $n = 100$ the QQ-plot and the Kolmogorov-Smirnov goodness-of-fit test that produced a test statistic value of $D = 0.4132$ having a p-value of $10^{-4}$ suggest poor asymptotic chi-square

approximation for this statistic. However, when $n = 300$ even though the QQ-plot appears to be very poor fit for this statistic the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0241$ with p-value 0.6070) did not detect problems with the chi-square approximation. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are given in Table 16.

Table 16. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. '*' denotes significant p-values at the 5% significance level. '-' denotes that the test was not performed for the given sample size.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 100 | $0.0305^*$ | $0.0078^*$ | 0.3886 | $10^{-4*}$ |
| 300 | 0.9024 | 0.8345 | - | 0.6070 |

Table 17 shows asymptotic power rates for $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ at a nominal 5% level when the model under the null is the constrained version of the categorical variable factor model and the alternative of interest is $A1_3$ for 5 variables, each at 2 categories. As seen from Table 17, $LR_{diff}$ is more powerful compared to $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for $n = 500, 1000$. $\chi^2_{[2]}$ generally outperforms both $\chi^2_{PF}$ and $\chi^2_{[1:2]}$ for $n = 300, 500, 1000$ and $\chi^2_{[1:2]}$ outperforms $\chi^2_{PF}$. $\chi^2_{[1:2]}$ is outperformed by $\chi^2_{[2]}$ which was surprising as it was suspected that when the model under the null hypothesis was the constrained version of the categorical variable factor model estimation of intercepts would be affected and, therefore, that components defined on first-order marginals might contribute to the power

Table 17. Asymptotic power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_3$ for 5 variables, each at 2 categories at a nominal 5% level.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 100 | 0.067 | 0.100 | 0.073 | 0.079 |
| 300 | 0.110 | 0.225 | 0.132 | 0.154 |
| 500 | 0.163 | 0.364 | 0.206 | 0.246 |
| 1000 | 0.332 | 0.672 | 0.424 | 0.501 |

of the test. Furthermore, slightly higher power rates of $\chi^2_{[2]}$ compared to $\chi^2_{[1:2]}$ suggest that components defined on first-order marginals do not contribute to the power of the test as initially suspected and appear to dilute the test with superfluous degrees of freedom to a minor degree.

Comparing asymptotic power rates in Table 17 to empirical power rates in Table 18 shows that empirical power rates are generally not significantly different from asymptotic power rates for all statistics. In terms of power, $LR_{diff}$ is significantly different from all other statistics for $n = 300, 500, 1000$. Moreover, no other statistic is competitive with $LR_{diff}$. There is no significant difference in empirical power between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for all sample sizes suggesting that components defined on first-order marginals do not contribute to the power of the test. Empirical power rates for $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are not significantly different for $n = 300$. However, as sample size increases difference in empirical power rates between $\chi^2_{PF}$ and both $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ becomes

Table 18. Empirical power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_3$ for 5 variables, each at 2 categories at a nominal 5% level. Corresponding standard errors are given in parentheses next to each value.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 100 | 0.057 (0.007) | 0.121 (0.010) | 0.071 (0.008) | 0.087 (0.009) |
| 300 | 0.115 (0.010) | 0.251 (0.014) | 0.125 (0.010) | 0.144 (0.011) |
| 500 | 0.154 (0.011) | 0.366 (0.015) | 0.196 (0.013) | 0.234 (0.013) |
| 1000 | 0.306 (0.015) | 0.656 (0.015) | 0.410 (0.015) | 0.436 (0.016) |

significant.

Table 19 shows empirical Type I error rates at nominal significance levels $(1\%, 5\%, 10\%)$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, now each at 3 categories. As discussed above, if the true Type I error probabilities are $0.01, 0.05$ and $0.10$, then $0.003, 0.007$ and $0.009$ provide respective standard error values that can be used when comparing the table entries to the nominal levels. In particular, it can be seen that the Type I error rates for $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when $n = 300$ are significantly different only from the nominal 10% level. However, all power comparisons were performed at the 5% significance level which is not significantly different from its respective nominal level. On the other hand, empirical Type I error rates for all test statistics are not significantly different from their respective nominal levels for $n = 300, 500, 1000$, which

enabled reliable power comparisons for the small sample sizes considered.

Table 19. Empirical Type I error rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model with 5 variables, each at 3 categories.

| | $\chi^2_{PF}$ | | | $LR_{diff}$ | | |
|---|---|---|---|---|---|---|
| Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 300 | 0.013 | 0.058 | 0.106 | 0.011 | 0.041 | 0.100 |
| 500 | 0.009 | 0.053 | 0.097 | 0.013 | 0.049 | 0.101 |
| 1000 | 0.010 | 0.045 | 0.086 | 0.007 | 0.047 | 0.094 |
| 5000 | 0.009 | 0.051 | 0.099 | 0.010 | 0.049 | 0.101 |
| | $\chi^2_{[1:2]}$ | | | $\chi^2_{[2]}$ | | |
| Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 300 | 0.005 | 0.047 | 0.080* | 0.014 | 0.066 | 0.120* |
| 500 | 0.012 | 0.049 | 0.097 | 0.019 | 0.056 | 0.105 |
| 1000 | 0.011 | 0.053 | 0.098 | 0.008 | 0.039 | 0.088 |
| 5000 | 0.010 | 0.050 | 0.104 | 0.009 | 0.051 | 0.102 |

QQ-plots of the empirical quantiles for $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when $n = 300$ are attached in Appendix F in Figure 24 to Figure 27, respectively, along with corresponding estimated slopes and p-values. The Kolmogorov-Smirnov goodness-of-fit test statistic for the chi-square distribution to the $\chi^2_{PF}$ data was found to be $D = 0.0348$ with corresponding p-value of 0.1775. Similarly, the Kolmogorov-Smirnov statistic for $LR_{diff}$ was $D = 0.0381$ with p-value of 0.1092, for $\chi^2_{[1:2]}$ was $D = 0.0202$ with p-value of 0.8082, and for $\chi^2_{[2]}$ was $D = 0.0278$ with p-value of 0.4206. Thus, all test statistics support the asymptotic chi-square distribution for the considered chi-squared statistics when expected cell frequencies in the joint distribution are small. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $LR_{diff}$ $\chi^2_{[1:2]}$

and $\chi^2_{[2]}$ are given in Table 20.

Table 20. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. '*' denotes significant p-values at the 5% significance level.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 300 | 0.1775 | 0.1092 | 0.8082 | 0.4206 |

Table 21 shows asymptotic power rates for $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ at a nominal 5% level when the model under the null is the constrained version of the categorical variable factor model and the alternative of interest is $A2_3$ for 5 variables, each at 3 categories. From Table 21 it can be seen that $LR_{diff}$ is more powerful than $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, especially for small sample sizes. $\chi^2_{[2]}$ generally outperforms both $\chi^2_{PF}$ and $\chi^2_{[1:2]}$ for all sample sizes, and $\chi^2_{[1:2]}$ outperforms $\chi^2_{PF}$. As with the 5 variable, 2 category scenario above, slightly higher power rates of $\chi^2_{[2]}$ compared to $\chi^2_{[1:2]}$ suggest that even with multiple categories components defined on first-order marginals do not contribute to the power of the test and appear to dilute the test with superfluous degrees of freedom to a minor degree.

Comparing asymptotic power rates in Table 21 to empirical power rates in Table 22, shows that empirical power rates are generally not significantly different from asymptotic power rates for all statistics. In terms of power $LR_{diff}$ is significantly different from all other statistics for all sample sizes. Moreover, no other statistic is competitive with $LR_{diff}$. There is no signifi-

Table 21. Asymptotic power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A2_3$ for 5 variables, each at 3 categories at a nominal 5% level.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 300 | 0.077 | 0.358 | 0.120 | 0.131 |
| 500 | 0.100 | 0.573 | 0.187 | 0.208 |
| 1000 | 0.176 | 0.891 | 0.403 | 0.450 |
| 5000 | 0.930 | 1.000 | 1.000 | 1.000 |

Table 22. Empirical power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A2_3$ for 5 variables, each at 3 categories. Corresponding standard errors are given in parentheses next to each value.

| Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ |
|---|---|---|---|---|
| 300 | 0.078 (0.008) | 0.371 (0.015) | 0.100 (0.009) | 0.113 (0.010) |
| 500 | 0.098 (0.009) | 0.593 (0.016) | 0.169 (0.012) | 0.192 (0.012) |
| 1000 | 0.164 (0.012) | 0.874 (0.010) | 0.398 (0.015) | 0.439 (0.016) |
| 5000 | 0.925 (0.008) | 1.000 (0.000) | 0.999 (0.000) | 0.999 (0.000) |

cant difference in empirical power between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ for all sample sizes considered. As with the 5 variables, each at 2 categories, no significant difference in empirical power between $\chi^2_{[2]}$ and $\chi^2_{[1:2]}$ suggests that components defined on first-order marginals yet again do not contribute substantially to the power of the test. Note, simulations with 1000 samples was not sufficiently large enough to detect the small difference between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ in terms of asymptotic power. $\chi^2_{PF}$ is not competitive with either $\chi^2_{[1:2]}$ or $\chi^2_{[2]}$.

2.2. Study 3b. The primary focus of Study 3b was comparing power for $\chi^2_{PF}$, $LR_{diff}$, $M_2$ and test statistics based on components defined on marginals, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$, for large and sparse cross-classification contingency tables with 10 dichotomous variables. The corresponding degrees of freedom for $\chi^2_{PF}$, $LR_{diff}$, $M_2$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ were $1003, 9, 44, 55$ and 45, respectively.

Empirical Type I error rates at nominal significance levels $(1\%, 5\%, 10\%)$ are given in Table 23 for $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model. If the true Type I error probabilities are $0.01, 0.05$ and $0.10$, then $0.003, 0.007$ and $0.009$ provide respective standard error values that can be used when comparing the table entries to the nominal level for all number of samples with converged iterations. In the simulation study depending on the various conditions considered, some iterations in the optimization procedure used in the estimation of the parameters in the categorical variable factor model did not converge. The number of samples with converged iterations were recorded for each scenario for all sample sizes considered. In particular, it can be seen that the Type I error rates for $\chi^2_{PF}$ are significantly different from the correspond-

Table 23. Empirical Type I error rates of $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model with 10 variables, each at 2 categories. '*' denotes Type I error rates significantly different from the specified nominal level.

| | | $\chi^2_{PF}$ | | | |
|---|---|---|---|---|---|
| $\bar{\beta}_1$ | Sample size | No. samples | 1% | 5% | 10% |
| 0.5 | 300 | 999 | 0.180* | 0.199* | 0.206* |
| | 500 | 1000 | 0.210* | 0.235* | 0.242* |
| | 750 | 1000 | 0.211* | 0.227* | 0.236* |
| | 1000 | 1000 | 0.214* | 0.234* | 0.247* |
| 1.0 | 300 | 996 | 0.193* | 0.202* | 0.207* |
| | 500 | 994 | 0.201* | 0.222* | 0.232* |
| | 750 | 994 | 0.209* | 0.230* | 0.238* |
| | 1000 | 995 | 0.204* | 0.233* | 0.249* |
| 1.5 | 300 | 983 | 0.192* | 0.200* | 0.208* |
| | 500 | 982 | 0.223* | 0.234* | 0.241* |
| | 750 | 981 | 0.221* | 0.240* | 0.249* |
| | 1000 | 975 | 0.230* | 0.244* | 0.252* |

ing nominal levels $(1\%, 5\%, 10\%)$ for all sample sizes and all values of the true average slope parameter values. The significant Type I error rates demonstrate the already well known adverse effects of sparseness on PGF. Thus, no reliable power comparisons of $\chi^2_{PF}$ with other statistics can be made in this study. Empirical Type I error rates at nominal significance levels $(1\%, 5\%, 10\%)$ are given in Table 24 for $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model. Convergence problems occurred as noted in Table 23. Note, convergence problems were not observed in previous studies. Empirical Type I error rates for $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ are not significantly different from their respective nominal

Table 24. Empirical Type I error rates of $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model with 10 variables, each at 2 categories. '*' denotes Type I error rates significantly different from the specified nominal level.

| | | $LR_{diff}$ | | | $\chi^2_{[1:2]}$ | | |
|---|---|---|---|---|---|---|---|
| $\bar{\beta}_1$ | Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 0.5 | 300 | 0.028* | 0.115* | 0.222* | 0.007 | 0.032* | 0.077* |
| | 500 | 0.024* | 0.086* | 0.160* | 0.020 | 0.053 | 0.090 |
| | 750 | 0.017* | 0.069* | 0.132* | 0.011 | 0.046 | 0.091 |
| | 1000 | 0.020* | 0.067* | 0.125* | 0.013 | 0.045 | 0.095 |
| 1.0 | 300 | 0.015 | 0.057 | 0.115 | 0.014 | 0.049 | 0.101 |
| | 500 | 0.013 | 0.064 | 0.115 | 0.009 | 0.046 | 0.092 |
| | 750 | 0.012 | 0.055 | 0.120 | 0.003 | 0.040 | 0.095 |
| | 1000 | 0.012 | 0.056 | 0.105 | 0.013 | 0.047 | 0.110 |
| 1.5 | 300 | 0.010 | 0.054 | 0.110 | 0.009 | 0.044 | 0.094 |
| | 500 | 0.010 | 0.051 | 0.095 | 0.012 | 0.059 | 0.098 |
| | 750 | 0.013 | 0.061 | 0.114 | 0.012 | 0.056 | 0.100 |
| | 1000 | 0.013 | 0.059 | 0.106 | 0.012 | 0.054 | 0.105 |
| | | $\chi^2_{[2]}$ | | | $M_2$ | | |
| $\bar{\beta}_1$ | Sample size | 1% | 5% | 10% | 1% | 5% | 10% |
| 0.5 | 300 | 0.007 | 0.030* | 0.067* | 0.006 | 0.030* | 0.064* |
| | 500 | 0.012 | 0.049 | 0.082 | 0.009 | 0.041 | 0.086 |
| | 750 | 0.005 | 0.048 | 0.089 | 0.008 | 0.048 | 0.093 |
| | 1000 | 0.010 | 0.051 | 0.088 | 0.011 | 0.050 | 0.098 |
| 1.0 | 300 | 0.013 | 0.056 | 0.097 | 0.013 | 0.050 | 0.092 |
| | 500 | 0.011 | 0.045 | 0.091 | 0.009 | 0.043 | 0.096 |
| | 750 | 0.007 | 0.041 | 0.097 | 0.009 | 0.043 | 0.096 |
| | 1000 | 0.010 | 0.051 | 0.112 | 0.009 | 0.051 | 0.103 |
| 1.5 | 300 | 0.012 | 0.047 | 0.100 | 0.008 | 0.048 | 0.112 |
| | 500 | 0.015 | 0.056 | 0.113 | 0.016 | 0.051 | 0.105 |
| | 750 | 0.011 | 0.048 | 0.108 | 0.011 | 0.045 | 0.114 |
| | 1000 | 0.014 | 0.052 | 0.101 | 0.007 | 0.046 | 0.112 |

levels for $n = 500, 750, 1000$. However, it can be seen that the Type I error rates for $LR_{diff}$ when the true average slope parameter $\bar{\beta}_1 = 0.5$ for all sample sizes are significantly different from their respective nominal levels. The combination of small sample size and small value of the true average slope parameter results in bias in the maximum likelihood slope estimator, which most directly affects the $LR_{diff}$ since this statistic compares the constrained model result to the unconstrained model result. In order to determine that the presence of bias caused the significant Type I error rates observed in Table 24, actual parameter values were fitted when computing the test statistics. The problem did not persist under such conditions.

In this study, the empirical mean square error, MSE, is calculated as

$$\widehat{MSE}(\theta) = \frac{1}{\text{Number of simulations}} \sum_{i=1}^{\text{Number of simulations}} (\hat{\theta}_i - \theta)^2,$$

and the empirical bias, BIAS, is calculated as

$$\widehat{BIAS}(\hat{\theta}, \theta) = \frac{1}{\text{Number of simulations}} \sum_{i=1}^{\text{Number of simulations}} (\hat{\theta}_i - \theta).$$

Table 25 shows MSE and BIAS when the model is the unconstrained version of the categorical variable factor model when the true average slope parameter $\bar{\beta}_1 = 0.5$ for $n = 300, 500, 750, 1000$. Corresponding standard errors are given in parentheses below each value. From Table 25 it can be seen that both MSE and BIAS decrease with increasing sample size when $\bar{\beta}_1 = 0.5$.

Table 26 shows MSE and BIAS for $n = 500$ and the true average slope parameters $\bar{\beta}_1 = (0.5, 1.0, 1.5)$. From Table 26 it can be seen that MSE decreases with increasing true average slope parameter for $n = 500$. The fitted unconstrained version of the model seems to produce bias in the direction that

Table 25. MSE and BIAS when the model under the null hypothesis is the unconstrained version of the categorical variable factor model for 10 dichotomous variables when $\bar{\beta}_1 = 0.5$ for $n = 300, 500, 750, 1000$. Corresponding standard errors are given in parentheses below each value.

MSE when $\bar{\beta}_1 = 0.5$

| Sample size | $\beta_{1,1}$ | $\beta_{2,1}$ | $\beta_{3,1}$ | $\beta_{4,1}$ | $\beta_{5,1}$ | $\beta_{6,1}$ | $\beta_{7,1}$ | $\beta_{8,1}$ | $\beta_{9,1}$ | $\beta_{10,1}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 300 | 4.166 | 6.792 | 4.984 | 9.431 | 5.898 | 5.251 | 3.807 | 2.357 | 4.133 | 3.492 |
| | (0.069) | (0.085) | (0.075) | (0.105) | (0.081) | (0.076) | (0.065) | (0.053) | (0.068) | (0.064) |
| 500 | 1.796 | 0.264 | 0.983 | 0.519 | 0.461 | 1.116 | 1.096 | 1.535 | 0.608 | 1.128 |
| | (0.047) | (0.023) | (0.036) | (0.029) | (0.028) | (0.038) | (0.038) | (0.043) | (0.031) | (0.038) |
| 750 | 0.100 | 0.454 | 0.101 | 0.847 | 0.229 | 0.122 | 0.100 | 0.080 | 0.070 | 0.109 |
| | (0.019) | (0.027) | (0.019) | (0.034) | (0.023) | (0.020) | (0.019) | (0.019) | (0.018) | (0.020) |
| 1000 | 0.082 | 0.060 | 0.042 | 0.066 | 0.030 | 0.155 | 0.123 | 0.039 | 0.053 | 0.079 |
| | (0.019) | (0.018) | (0.017) | (0.018) | (0.017) | (0.021) | (0.020) | (0.017) | (0.018) | (0.018) |

BIAS when $\bar{\beta}_1 = 0.5$

| Sample size | $\beta_{1,1}$ | $\beta_{2,1}$ | $\beta_{3,1}$ | $\beta_{4,1}$ | $\beta_{5,1}$ | $\beta_{6,1}$ | $\beta_{7,1}$ | $\beta_{8,1}$ | $\beta_{9,1}$ | $\beta_{10,1}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 300 | 0.423 | 0.279 | 0.332 | 0.410 | 0.370 | 0.333 | 0.218 | 0.159 | 0.300 | 0.366 |
| 500 | 0.162 | 0.029 | 0.090 | 0.048 | 0.062 | 0.105 | 0.074 | 0.096 | 0.076 | 0.095 |
| 750 | 0.020 | 0.032 | 0.013 | 0.045 | 0.033 | 0.018 | 0.013 | 0.028 | 0.009 | 0.034 |
| 1000 | 0.018 | 0.016 | 0.004 | 0.024 | 0.004 | 0.015 | 0.015 | 0.006 | 0.012 | 0.001 |

Table 26. MSE and BIAS when the model under the null hypothesis is the unconstrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ and $\bar{\beta}_1 = 0.5, 1.0, 1.5$. Corresponding standard errors are given in parentheses below each value.

MSE for $n = 500$

| $\bar{\boldsymbol{\beta}}_1$ | $\beta_{1,1}$ | $\beta_{2,1}$ | $\beta_{3,1}$ | $\beta_{4,1}$ | $\beta_{5,1}$ | $\beta_{6,1}$ | $\beta_{7,1}$ | $\beta_{8,1}$ | $\beta_{9,1}$ | $\beta_{10,1}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 1.796 | 0.264 | 0.983 | 0.519 | 0.461 | 1.116 | 1.096 | 1.535 | 0.608 | 1.128 |
| | (0.047) | (0.023) | (0.036) | (0.029) | (0.028) | (0.038) | (0.038) | (0.043) | (0.031) | (0.038) |
| 1.0 | 0.083 | 0.054 | 0.049 | 0.322 | 0.924 | 0.038 | 0.038 | 0.045 | 0.054 | 0.077 |
| | (0.034) | (0.033) | (0.033) | (0.037) | (0.045) | (0.033) | (0.033) | (0.033) | (0.033) | (0.033) |
| 1.5 | 0.085 | 0.054 | 0.048 | 0.296 | 0.906 | 0.033 | 0.035 | 0.048 | 0.054 | 0.073 |
| | (0.050) | (0.049) | (0.049) | (0.052) | (0.057) | (0.047) | (0.048) | (0.049) | (0.049) | (0.049) |

BIAS for $n = 500$

| $\bar{\boldsymbol{\beta}}_1$ | $\beta_{1,1}$ | $\beta_{2,1}$ | $\beta_{3,1}$ | $\beta_{4,1}$ | $\beta_{5,1}$ | $\beta_{6,1}$ | $\beta_{7,1}$ | $\beta_{8,1}$ | $\beta_{9,1}$ | $\beta_{10,1}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.162 | 0.029 | 0.090 | 0.048 | 0.062 | 0.105 | 0.074 | 0.096 | 0.076 | 0.095 |
| 1.0 | 0.032 | 0.009 | 0.016 | 0.031 | 0.065 | 0.015 | 0.009 | 0.025 | 0.029 | 0.013 |
| 1.5 | 0.044 | 0.003 | 0.014 | 0.032 | 0.048 | -0.008 | 0.003 | 0.023 | 0.029 | 0.012 |

the magnitude of the slopes are too large for small sample size and small true average slope parameter $\bar{\beta}_1$ values. This bias decreases with both increasing sample sizes and $\bar{\beta}_1$ values.

The constrained version of the categorical variable factor model on the other hand, does not demonstrate such severe bias in the maximum likelihood slope estimators. Table 27 shows MSE and BIAS when $\bar{\beta}_1 = 0.5$ for $n = 300, 500, 750, 1000$. As can be seen from Table 27, MSE is much smaller and more nearly constant than with the unconstrained version of the categorical

Table 27. MSE and BIAS when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables when $\bar{\beta}_1 = 0.5$ for $n = 300, 500, 750, 1000$. Corresponding standard errors are given in parentheses below each value.

MSE when $\bar{\beta}_1 = 0.5$

| Sample size | $\beta$ |
|---|---|
| 300 | 0.009 |
| | (0.003) |
| 500 | 0.005 |
| | (0.002) |
| 750 | 0.003 |
| | (0.002) |
| 1000 | 0.003 |
| | (0.002) |

BIAS when $\bar{\beta}_1 = 0.5$

| Sample size | $\beta$ |
|---|---|
| 300 | -0.003 |
| 500 | -0.001 |
| 750 | -0.001 |
| 1000 | -0.001 |

variable factor model and decreases with increasing sample sizes when $\bar{\beta}_1 = 0.5$.

Table 28 shows MSE and BIAS for $n = 500$ and $\bar{\beta}_1 = 0.5, 1.0, 1.5$. From Table 28 it can be seen that MSE decreases with increasing true slope parameter values for $n = 500$ and is smaller with the constrained version of the model than with the unconstrained version. Bias of the estimator in the fitted constrained version of the model is also significantly smaller and constant compared to the unconstrained version of the model.

Table 28. MSE and BIAS when the model under the null hypothesis the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ and $\bar{\beta}_1 = 0.5, 1.0, 1.5$. Corresponding standard errors are given in parentheses below each value.

MSE for $n = 500$

| $\bar{\boldsymbol{\beta}}_1$ | $\beta_1$ |
|---|---|
| 0.5 | 0.005 |
| | (0.002) |
| 1 | 0.004 |
| | (0.002) |
| 1.5 | 0.004 |
| | (0.002) |

BIAS for $n = 500$

| $\bar{\boldsymbol{\beta}}_1$ | $\beta_1$ |
|---|---|
| 0.5 | -0.002 |
| 1 | 0.002 |
| 1.5 | 0.002 |

QQ-plots of the empirical quantiles for $LR_{diff}$ for $n = 300$ and $\bar{\beta}_1 =$

1.0 and 1.5 and for $n = 500$ and $\bar{\beta}_1 = 1.5$ are attached in Appendix G in Figure 28 to Figure 30, respectively, along with corresponding estimated slopes and p-values. In the case of $n = 300$ and $\bar{\beta}_1 = 1.0$, the QQ-plot suggests poor asymptotic chi-square approximation for this statistic, and the Kolmogorov-Smirnov goodness-of-fit test produced a value of $D = 0.0471$ having a p-value of 0.0023, confirming poor asymptotic chi-square approximation. However, when $n = 500$ and $\bar{\beta}_1 = 1.0$ even though the QQ-plot suggest very poor fit the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0271$ with p-value 0.4543) did not detect problems with the chi-square approximation. The Kolmogorov-Smirnov statistic for $LR_{diff}$ when $n = 300$ and $\bar{\beta}_1 = 1.5$ was $D = 0.0180$ with p-value of 0.9018. Thus, the test statistics supports the asymptotic chi-square distribution for $LR_{diff}$ statistic when expected cell frequencies in the joint distribution are small ($n = 300$) and the true average slope parameter values are $\bar{\beta}_1 = 1.0$ and 1.5.

QQ-plots for $\chi^2_{[1:2]}$ for all sample sizes when $\bar{\beta}_1 = 0.5$, $n = 300$ and when $\bar{\beta}_1 = 1.0, 1.5$ are also attached in Appendix H in Figure 31 to Figure 36, respectively. In the case of $n = 300, 500, 750$ and $\bar{\beta}_1 = 0.5$, the QQ-plots suggest poor asymptotic chi-square approximation for this statistic and the Kolmogorov-Smirnov goodness-of-fit tests produced values of $D = 0.1171$ having a p-value of $10^{-4}$, $D = 0.0854$ having a p-value of $10^{-4}$, and $D = 0.0624$ having a p-value of 0.0001 confirming poor asymptotic chi-square approximation. However, when $n = 1000$ and $\bar{\beta}_1 = 0.5$ neither the QQ-plot nor the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0242$ with p-value 0.6006) detected problems with the chi-square approximation. The

Kolmogorov-Smirnov goodness-of-fit test statistic for the chi-square distribution for $\chi^2_{[1:2]}$ when $n = 300$ and $\bar{\beta}_1 = 1.0$ was found to be $D = 0.0343$ with corresponding p-value of 0.1890. Similarly, the Kolmogorov-Smirnov statistic for $\chi^2_{[1:2]}$ when $n = 300$ and $\bar{\beta}_1 = 1.5$ was $D = 0.0379$ with p-value of 0.9995. Thus, the test statistics support the asymptotic chi-square distribution for $\chi^2_{[1:2]}$ statistic when expected cell frequencies in the joint distribution are small ($n = 300$) and the true average slope parameter values are $\bar{\beta}_1 = 1.0$ and 1.5.

Similarly, QQ-plots for $\chi^2_{[2]}$ for all sample sizes when $\bar{\beta}_1 = 0.5$, $n = 300$ and when $\bar{\beta}_1 = 1.0, 1.5$ are attached in Appendix H in Figure 37 to Figure 42. In the case of $n = 300, 500, 750$ when $\bar{\beta}_1 = 0.5$ the QQ-plots suggest poor asymptotic chi-square approximation for this statistic, and the Kolmogorov-Smirnov goodness-of-fit tests produced values of $D = 0.1011$ having a p-value of $10^{-4}$, $D = 0.0761$ having a p-value of $10^{-4}$, and $D = 0.0541$ having a p-value of 0.0058 confirming poor asymptotic chi-square approximation. However, when $n = 1000$ and $\bar{\beta}_1 = 0.5$ neither the QQ-plot nor the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0203$ with p-value 0.8063) detected problems with the chi-square approximation. The Kolmogorov-Smirnov goodness-of-fit test statistic for the chi-square distribution for $\chi^2_{[2]}$ when $n = 300$ and $\bar{\beta}_1 = 1.0$ was found to be $D = 0.0133$ with corresponding p-value of 0.9946. Similarly, the Kolmogorov-Smirnov statistic for $\chi^2_{[2]}$ when $n = 300$ and $\bar{\beta}_1 = 1.5$ was $D = 0.0274$ with p-value of 0.4398. Thus, both test statistics support the asymptotic chi-square distribution for $\chi^2_{[2]}$ statistics when expected cell frequencies in the joint distribution are small and the true slope parameter is small in magnitude.

Furthermore, QQ-plots for $M_2$ for all sample sizes when $\bar{\beta}_1 = 0.5$, $n = 300$ and $\bar{\beta}_1 = 1.0, 1.5$ are also attached in Appendix H in Figure 43 to Figure 48, respectively. In the case of $n = 300, 500, 750$ and $\bar{\beta}_1 = 0.5$ the QQ-plots suggest poor asymptotic chi-square approximation for this statistic and the Kolmogorov-Smirnov goodness-of-fit tests produced test statistic values of $D = 0.1051$ having a p-value of $10^{-4}$, $D = 0.0801$ having a p-value of $10^{-4}$, and $D = 0.0647$ having a p-value of 0.0005 confirming poor asymptotic chi-square approximation. However, when $n = 1000$ and $\bar{\beta}_1 = 0.5$ neither the QQ-plot nor the Kolmogorov-Smirnov goodness-of-fit test statistic ($D = 0.0381$ with p-value 0.0968) detected problems with the chi-square approximation. The Kolmogorov-Smirnov goodness-of-fit test statistic for the chi-square distribution to the $M_2$ data for $n = 300$ when $\bar{\beta}_1 = 1.0$ was found to be $D = 0142$ with corresponding p-value of 0.9880. Similarly, the Kolmogorov-Smirnov statistic for $M_2$ for $n = 300$ when $\bar{\beta}_1 = 1.5$ was $D = 0.0425$ with p-value of 0.0537. Thus, the test statistics supports the asymptotic chi-square distribution for $M_2$ statistic when expected cell frequencies in the joint distribution are small ($n = 300$) and the true average slope parameter values are $\bar{\beta}_1 = 1.0$ and 1.5. Power comparisons are not reliable for $\chi^2_{PF}$ for all conditions considered and for $\chi^2_{LR_{diff}}$ when $\bar{\beta}_1 = 0.5$ as Type I error rates are confounded with power. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ are given in Table 29.

Table 30 shows asymptotic power rates at a nominal 5% level for $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ for $n = 300, 500, 750, 1000$ when $\bar{\boldsymbol{\beta}}_1 = 0.5, 1.0, 1.5$. However, as can be seen from Table 30 when $\bar{\beta}_1 = 1.0, 1.5$ irrespective of

Table 29. Summary of the Kolmogorov-Smirnov goodness-of-fit test p-values of $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$. '*' denotes significant p-values at the 5% significance level. '-' denotes that the test was not performed for the given sample size.

| $\beta_1$ | Sample size | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $M_2$ |
|---|---|---|---|---|---|
| 0.5 | 300 | - | $10^{-4*}$ | $10^{-4*}$ | $10^{-4*}$ |
| | 500 | - | $10^{-4*}$ | $10^{-4*}$ | $10^{-4*}$ |
| | 750 | - | $0.0001^*$ | $0.0058^*$ | $0.0005^*$ |
| | 1000 | - | 0.6006 | 0.8063 | 0.0968 |
| 1.0 | 300 | $0.0023^*$ | 0.1890 | 0.9946 | 0.9880 |
| | 500 | 0.4543 | - | - | - |
| 1.5 | 300 | 0.9018 | 0.9995 | 0.4398 | 0.0537 |

sample size $LR_{diff}$ is more powerful than all other statistics. Asymptotic power for both $\chi^2_{[2]}$ and $M_2$ is generally identical and increases with increasing true slope parameter values and increasing sample size. Similarly, power for $\chi^2_{[1:2]}$ closely follows both $\chi^2_{[2]}$ and $M_2$. Moreover, as discussed in Study 3a, slightly higher power rates of $\chi^2_{[2]}$ compared to $\chi^2_{[1:2]}$ suggest that components defined on first-order marginals do not significantly contribute to the power of the test as initially suspected and appear to dilute the test with superfluous degrees of freedom to a minor degree. Asymptotic power rates for $\chi^2_{PF}$ are not comparable to other statistics under any conditions due to extremely inflated Type I error rates discovered previously.

Table 31 shows empirical power rates at a nominal 5% level for $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ for $n = 300, 500, 750, 1000$ when $\bar{\beta}_1 = 0.5, 1.0, 1.5$ along with corresponding standard errors. Comparing empirical power rates

Table 30. Asymptotic power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A3_3$ for 10 dichotomous variables for $n = 300, 500, 750, 1000$ when $\bar{\beta}_1 = 0.5, 1.0, 1.5$ at a nominal 5% level. '*' denotes Type I error rates significantly different from the nominal level.

| $\bar{\boldsymbol{\beta}}_1$ | Sample size | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $M_2$ |
|---|---|---|---|---|---|---|
| 0.5 | 300 | 0.059* | 0.174* | 0.093* | 0.098* | 0.099* |
| | 500 | 0.066* | 0.283* | 0.130 | 0.141 | 0.142 |
| | 750 | 0.075* | 0.429* | 0.185 | 0.206 | 0.208 |
| | 1000 | 0.085* | 0.568* | 0.252 | 0.281 | 0.281 |
| 1.0 | 300 | 0.090* | 0.673 | 0.318 | 0.353 | 0.355 |
| | 500 | 0.127* | 0.910 | 0.566 | 0.619 | 0.614 |
| | 750 | 0.186* | 0.988 | 0.812 | 0.854 | 0.858 |
| | 1000 | 0.258* | 0.999 | 0.938 | 0.958 | 0.959 |
| 1.5 | 300 | 0.097* | 0.729 | 0.371 | 0.412 | 0.414 |
| | 500 | 0.143* | 0.941 | 0.651 | 0.705 | 0.703 |
| | 750 | 0.215* | 0.994 | 0.881 | 0.914 | 0.914 |
| | 1000 | 0.303* | 1.000 | 0.971 | 0.982 | 0.983 |

in Table 31 to asymptotic power rates in Table 30, it can be seen that the empirical power rates are generally not significantly different from asymptotic power rates for all statistics considered for conditions when $\bar{\boldsymbol{\beta}}_1 = 1.0, 1.5$. Empirical power of $\chi^2_{[1:2]}$ is significantly different from both $\chi^2_{[2]}$ and $M_2$ with increasing sample size and increasing true slope parameter values. Higher power of $M_2$ compared to $\chi^2_{[1:2]}$ can be attributed to the number of first- and second-order marginals used to obtain either statistic. $M_2$ uses first- and second-order marginals that result in a total of 44 degrees of freedom while

Table 31. Empirical power rates of $\chi^2_{PF}$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A3_3$ for 10 dichotomous variables for $n = 300, 500, 750, 1000$ when $\bar{\beta}_1 = 0.5, 1.0, 1.5$ at a nominal 5% level. Corresponding standard errors are given in parentheses below each value. '*' denotes Type I error rates significantly different from the nominal level.

| $\bar{\boldsymbol{\beta}}_1$ | Sample size | No. samples | $\chi^2_{PF}$ | $LR_{diff}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $M_2$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 300 | 999 | 0.244* | 0.261* | 0.095* | 0.105* | 0.112* |
| | | | (0.014) | (0.014) | (0.009) | (0.010) | (0.010) |
| | 500 | 1000 | 0.273* | 0.345* | 0.134 | 0.156 | 0.164 |
| | | | (0.014) | (0.015) | (0.011) | (0.011) | (0.012) |
| | 750 | 1000 | 0.262* | 0.472* | 0.190 | 0.209 | 0.205 |
| | | | (0.014) | (0.016) | (0.012) | (0.013) | (0.013) |
| | 1000 | 1000 | 0.309* | 0.564* | 0.201 | 0.272 | 0.286 |
| | | | (0.015) | (0.016) | (0.013) | (0.014) | (0.014) |
| 1.0 | 300 | 996 | 0.201* | 0.677 | 0.302 | 0.343 | 0.350 |
| | | | (0.013) | (0.015) | (0.015) | (0.015) | (0.015) |
| | 500 | 995 | 0.238* | 0.909 | 0.532 | 0.592 | 0.591 |
| | | | (0.014) | (0.009) | (0.016) | (0.016) | (0.016) |
| | 750 | 993 | 0.221* | 0.988 | 0.818 | 0.862 | 0.856 |
| | | | (0.013) | (0.003) | (0.012) | (0.011) | (0.011) |
| | 1000 | 997 | 0.248* | 0.999 | 0.940 | 0.968 | 0.968 |
| | | | (0.014) | (0.001) | (0.008) | (0.006) | (0.006) |
| 1.5 | 300 | 980 | 0.147* | 0.721 | 0.328 | 0.385 | 0.378 |
| | | | (0.011) | (0.014) | (0.015) | (0.016) | (0.015) |
| | 500 | 980 | 0.148* | 0.935 | 0.589 | 0.653 | 0.651 |
| | | | (0.011) | (0.008) | (0.016) | (0.015) | (0.015) |
| | 750 | 981 | 0.131* | 0.993 | 0.876 | 0.916 | 0.915 |
| | | | (0.011) | (0.003) | (0.011) | (0.009) | (0.009) |
| | 1000 | 982 | 0.134* | 0.995 | 0.970 | 0.993 | 0.992 |
| | | | (0.011) | (0.003) | (0.005) | (0.003) | (0.003) |

$\chi^2_{[1:2]}$ uses all first- and second-order marginals resulting in a total of 55 degrees of freedom. The larger number of degrees of freedom for $\chi^2_{[1:2]}$ dilute its power. There is no significant difference between $\chi^2_{[2]}$ and $M_2$ in terms of empirical power. However, the preference of $\chi^2_{[2]}$ over $M_2$ is due to $\chi^2_{[2]}$'s capability of isolating second-order marginals only, unlike $M_2$, which includes first- and second-order marginals and cannot isolate any second-order marginals, since it is not constructed out of components. When $\chi^2_{[2]}$ is used, second-order associations could be isolated from first-order associations.

# Chapter 6: COMPARING ORTHOGONAL POLYNOMIAL COMPONENTS TO ORTHOGONAL COMPONENTS DEFINED ON MARGINALS

Section 1 discusses the study of orthogonal polynomials components relative to components based on marginal frequencies. Section 2 presents results of that study.

## 1. Orthogonal Polynomial Components Compared to Orthogonal Components Defined on Marginals

The primary focus of Study 4 was on comparing orthogonal polynomial components to components defined on marginal frequencies. Included were the statistics $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\hat{V}_i^2$, $i = 4, 5, 6, 7$: i.e., orthogonal polynomial components of order 4, 5, 6 and 7, for 5 variables each at 2 categories. As discussed in Chapter 2, when the conditions for large sample theory are met, it is known that PGF tests are used to test goodness-of-fit when particular alternatives are not specified. However, if they were specified, more powerful directional tests could be applied. Consequently, tests based on orthogonal polynomial components may need fewer components and may have higher power than traditional tests (Rayner and Best 1989). Even though the lack-of-fit for a cross-classified table often occurs in lower-order marginals as discussed earlier in Chapter 2 (Salomaa 1990), $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ are still in a sense omnibus tests that are more focused than $\chi^2_{PF}$, but still probably not optimal as they may include components that are not related to the lack-of-fit. Thus, the purpose of this study was to use orthogonal polynomial components which may be able to form a limited-information statistic that is more focused on

the lack-of-fit than tests based on the marginals.

As discussed earlier in Chapter 2, orthogonal polynomial components can usually detect lack-of-fit in the first four components, which may provide a test with higher power than unfocused tests. $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ may have lower power compared to orthogonal polynomial components due to superfluous degrees of freedom, since they include all the first- and/or second-order marginals. However, orthogonal polynomial components of higher-order may be required if the cells are not ordered in a way that is associated with the lack-of-fit. Orthogonal polynomial components have been previously applied to one-dimensional tables for testing goodness-of-fit for a specified univariate distribution. The applications under consideration here include high-dimensional multi-way tables, and it may be difficult to order the cells of the multi-way table in a way that would be conducive to a test on low-order components. In general, with multi-way tables interpretations of orthogonal polynomial components may not be the same as interpretations in the univariate case where first-degree orthogonal polynomial components tend to detect change in means, second-degree orthogonal polynomial components tend to detect change in variance etc. (Eubank, LaRiccia, and Rosenstein 1987).

The comparison of orthogonal polynomial components to components defined on marginals is conveniently implemented when the simple null hypothesis is the model of equally correlated variables, i.e., all the slopes are equal to 1 in the categorical variable factor model for 5 variables each at 2

categories.. $H_o$ is given by

$$H_o : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

The model of equally correlated variables is another instance of the constrained categorical variable factor model discussed in Chapter 3. The constrained version of the model is a more realistic and useful model compared to equal probable cells, where the slopes and intercepts are all set to 0, and there is no way to order the cells, at least not with relation to the underlying latent variable. Because the study uses a simple null hypothesis, the orthogonal components defined on marginals were calculated from expression (3.8).

Since the hypothesis of equally correlated variables does not have equal probable cells, orthogonal polynomial component calculations from Emerson (1968), as used by Rayner and Best (1989), were implemented. The method for obtaining orthogonal polynomial components is given in Chapter 2 where the $x_j$ used in the orthonormal functions $g_0(x), g_1(x), ..., g_T(x)$ is a score value for $j^{th}$ cell. The score for each cell is sometimes known as a factor score. The score value was obtained from the likelihood function maximized over the latent variable for each pattern with known parameters i.e., the log likelihood function was maximized with respect to $\boldsymbol{\eta}$,

$$log(\mathbf{L}(\boldsymbol{\eta})) = log(n!) + \sum_{s=1}^{T} N_s \; log(\pi_s(\boldsymbol{\theta}|\boldsymbol{\eta})),$$

where $\mathbf{L}(\boldsymbol{\eta})$ is the likelihood function, $\pi_s(\cdot)$ denotes the probability function for the $s^{th}$ cell pattern as given in expression (2.10) and $n = \sum_{s=1}^{T} N_s$ is the sample size. The fitted model is an example of a non-linear mixed model with fixed parameters for the intercepts and slopes and random effect for the subjects which is normally distributed. As such, another way to obtain the score values is to use empirical Bayes predictors from the non-linear model (Searle, Casella, and McCulloch 1992). These score values were used to order the cells. This means that the cells were ordered from lowest to highest along the underlying factor, in an attempt to correlate the lack-of-fit with the underlying factor. There was no apparent natural way to order the frequencies for cells formed by cross-classification of a large number of variables. Cells could be ordered by the overall correct number of items scored: i.e., with possible values $0, 1, 2, 3, 4, 5$. However, this results in ties and only 6 possible outcomes limiting a polynomial to order 6.

Power calculations were performed using a large sample theory approach to power as discussed earlier in Chapter 3 in order to investigate the following five alternatives

$$A1_4 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ -0.5 & 1 \\ 0.5 & 1 \end{pmatrix},$$

$$A2_4 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0.5 & 1 \\ -0.25 & 1 \\ 0 & 1 \\ -0.25 & 1 \\ 0.5 & 1 \end{pmatrix},$$

$$A3_4 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

$$A4_4 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0 & 1.25 \\ 0 & 1 \\ 0 & 2 \\ 0 & 0.75 \\ 0 & 1.5 \end{pmatrix},$$

and

$$A5_4 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}), \boldsymbol{\theta} = \begin{pmatrix} 0 & 0.75 \\ -0.25 & 1.25 \\ 0 & 1 \\ -0.5 & 1.25 \\ 0.25 & 2 \end{pmatrix}.$$

The alternatives $A1_4$ and $A2_4$ provide an avenue for investigating the effect of fitting the null model which departs from the alternative model with respect

to intercept parameters. $A1_4$ captures a small effect size with two non-zero intercepts, since the departure of the null model from the true model is minor, and $A2_4$ captures a larger effect size as the departure of the null model from the alternative model is greater in this case. $A3_4$ and $A4_4$ give two alternatives for investigating the effect of fitting the null model which departs from the alternative model with respect to slope parameters. $A3_4$ investigates the case where two variables have a large slope, and $A4_4$ captures unevenly distributed slopes. $A5_4$ is an alternative for investigating the effect of fitting the null model to data that departs from the alternative model with respect to both intercept and slope parameters.

As discussed earlier, a test based on orthogonal polynomial components may be more focused and have higher power than tests based on orthogonal components defined on marginals. In fact, Rayner and Best (1989, p.62) stated that in their experience "lower order alternatives occur more frequently than higher order alternatives, so that unless there is some reason to expect higher order alternatives, a low-order test should be used". On the other hand, if the ordering of the cells is not related to the lack-of-fit for the model under the null hypothesis, tests based on orthogonal components for lower-order polynomials may not detect the poor fit of the model. For the alternatives given above, lack-of-fit is in the pair-wise associations among the manifest variables.

For evaluating the Type I error rate, a true model was fitted and Monte Carlo simulations were performed. To assess the power of a test for the categorical variable factor model, known cell frequencies from the models under null and alternative hypotheses were used to calculate power using the ap-

proach that relies on the large sample approximation using the chi-squared distribution. In this study, 5 variables each at 2 categories resulted in $2^5 = 32$ cells which in turn for samples sizes $n = 50, 100$ are likely to produce cross-classified tables that are sparse in nature. Overall, sample sizes considered were $n = 50, 100, 300, 500$ in order to measure the performance of selected test statistics both when sample sizes were large and not sparse and when sample sizes were small and sparse.

2. Results

2.1. Study 4. Empirical Type I error rates at nominal 5% significance level for $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$, $\hat{V}^2_4$, $\hat{V}^2_5$, $\hat{V}^2_6$ and $\hat{V}^2_7$ for 5 variables each at 2 categories are given in Table 32. As discussed previously, the proportions in Table 32 multiplied by a 1000 are binomial with success proportion 0.05 and 1000 trials. If the true Type I error probability is 0.05, 0.007 provides a standard error value that can be used when comparing the table entries to the nominal level. In particular, it can be seen that the Type I error rates for all test statistics for various sample sizes are not significantly different from the nominal 5% level, which enabled reliable power comparisons for the small sample sizes considered. It is important to note that the empirical Type I error rates given in Table 32 are not comparable to empirical Type I error rates given in Table 15 as the first uses a simple null hypothesis and the other uses a composite null hypothesis.

Tables 33 and 34 depict asymptotic results for two alternatives for investigating the effect of fitting the null model which departs from the alternative

Table 32. Empirical Type I error rates for $\chi^2_{PF}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$, $\hat{V}^2_4$, $\hat{V}^2_5$, $\hat{V}^2_6$ and $\hat{V}^2_7$ when the model under the null hypothesis depicts equal correlation among observed 5 variables each at 2 categories.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}^2_4$ | $\hat{V}^2_5$ | $\hat{V}^2_6$ | $\hat{V}^2_7$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.061 | 0.059 | 0.058 | 0.059 | 0.061 | 0.062 | 0.061 |
| 100 | 0.053 | 0.049 | 0.053 | 0.053 | 0.052 | 0.054 | 0.055 |
| 300 | 0.051 | 0.051 | 0.050 | 0.055 | 0.053 | 0.053 | 0.053 |
| 500 | 0.052 | 0.052 | 0.049 | 0.049 | 0.050 | 0.052 | 0.051 |

model with respect to intercept parameters. $A1_4$ captures a small effect size with two non-zero intercepts, since the departure of the null model from the true model is minor and $A2_4$ captures a larger effect size as the departure of the null model from the alternative model is greater in this case. Tests based on orthogonal polynomial components required higher-degree polynomials to achieve high power as the ordering of the cell frequencies was not informative about the lack-of-fit in both cases, as demonstrated in Tables 33 and 34, where the orthogonal polynomial components have lower power compared to $\chi^2_{PF}$, $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. Overall, in both Tables 33 and 34, $\chi^2_{[1:2]}$ outperforms both $\chi^2_{[2]}$ and $\chi^2_{PF}$ for all sample sizes which did not match results discovered in earlier studies, and $\chi^2_{[2]}$ slightly outperforms $\chi^2_{PF}$. $\chi^2_{[1:2]}$ has higher power than $\chi^2_{[2]}$ which indicates that some of the effect of change in the intercepts was manifested in first-order marginals. The manifestation of the lack-of-fit in first-order marginals can be attributed to parameter values being specified for both the null and alternative hypotheses. Unlike in the earlier studies, Study

4 uses the simple null hypothesis. In earlier studies slopes and intercepts were estimated due to the composite null hypothesis, and results indicated that second-order associations accounted for most of the lack-of-fit. For both, the unconstrained and constrained versions of the models under the null hypothesis in earlier studies, first-order marginals were almost perfectly fit which implied that first-order components did not contribute to identification of model misfit. This is not the case here with the simple null hypothesis, since intercepts and slopes are not fitted but are specified. Consequently, first-order marginals are not almost perfectly fit as the case with the composite null hypothesis, and as a result, contribute to the lack-of-fit.

Table 33. Asymptotic power when the alternative of interest is $A1_4$. Investigating the effect of misspecified intercept parameters.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}_4^2$ | $\hat{V}_5^2$ | $\hat{V}_6^2$ | $\hat{V}_7^2$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.164 | 0.231 | 0.196 | 0.127 | 0.118 | 0.116 | 0.123 |
| 100 | 0.337 | 0.479 | 0.394 | 0.221 | 0.202 | 0.198 | 0.216 |
| 300 | 0.906 | 0.972 | 0.920 | 0.608 | 0.570 | 0.569 | 0.624 |
| 500 | 0.996 | 1.000 | 0.996 | 0.850 | 0.822 | 0.826 | 0.873 |

Table 35 and 36 depict results for two alternatives for investigating the effect of fitting the null model which departs from the alternative model with respect to slope parameters. $A3_4$ investigates the case where two variables have a large slope, and $A4_4$ captures unevenly distributed slopes. Once again tests based on orthogonal polynomial components required high-degree polynomials to achieve high power. For both alternatives, $\chi^2_{[1:2]}$ outperforms both $\chi^2_{PF}$ and

Table 34. Asymptotic power when the alternative of interest is $A2_4$. Investigating the effect of misspecified intercept parameters.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}_4^2$ | $\hat{V}_5^2$ | $\hat{V}_6^2$ | $\hat{V}_7^2$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.198 | 0.283 | 0.240 | 0.101 | 0.104 | 0.110 | 0.112 |
| 100 | 0.422 | 0.583 | 0.488 | 0.164 | 0.169 | 0.190 | 0.191 |
| 300 | 0.963 | 0.993 | 0.969 | 0.469 | 0.472 | 0.552 | 0.554 |
| 500 | 1.000 | 1.000 | 0.999 | 0.721 | 0.724 | 0.828 | 0.829 |

$\chi^2_{[2]}$. $\chi^2_{[2]}$ performs worse in terms of power, compared to both $\chi^2_{[1:2]}$ and $\chi^2_{PF}$ for all sample sizes. $\chi^2_{[1:2]}$ has the highest power indicating that the lack-of-fit is not only manifested in second-order marginals as initially suspected but also in first-order marginals which did not match results discovered in earlier studies. The manifestation of the lack-of-fit in first-order marginals has been previously discussed above in relation to $A1_4$ and $A2_4$. Power is higher in Table 36 than in Table 35 suggesting that all the test statistics become more powerful as the slope effect increases in magnitude among the model variables. This phenomenon is expected since the $H_o$ model fits slopes equal to 1, and the greater the departure from the true model the higher the power of the test statistics.

Table 37 depicts results for investigating the effect of fitting the null model which departs from the alternative model with respect to intercept and slope parameters. In this setting, tests based on orthogonal polynomial components become competitors to the other chi-squared statistics, as the lack-of-fit seemed to be more related to the ordering of the cells. In general

Table 35. Asymptotic power when the alternative of interest is $A3_4$. Investigating the effect of misspecified slope parameters.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}^2_4$ | $\hat{V}^2_5$ | $\hat{V}^2_6$ | $\hat{V}^2_7$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.113 | 0.150 | 0.075 | 0.079 | 0.081 | 0.085 | 0.091 |
| 100 | 0.204 | 0.292 | 0.104 | 0.112 | 0.118 | 0.127 | 0.140 |
| 300 | 0.662 | 0.821 | 0.255 | 0.267 | 0.294 | 0.333 | 0.386 |
| 500 | 0.922 | 0.979 | 0.431 | 0.433 | 0.483 | 0.545 | 0.626 |

Table 36. Asymptotic power when the alternative of interest is $A4_4$. Investigating the effect of misspecified slope parameters.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}^2_4$ | $\hat{V}^2_5$ | $\hat{V}^2_6$ | $\hat{V}^2_7$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.096 | 0.122 | 0.070 | 0.078 | 0.080 | 0.083 | 0.087 |
| 100 | 0.158 | 0.221 | 0.092 | 0.110 | 0.115 | 0.116 | 0.119 |
| 300 | 0.506 | 0.674 | 0.208 | 0.260 | 0.288 | 0.291 | 0.292 |
| 500 | 0.802 | 0.919 | 0.347 | 0.423 | 0.479 | 0.481 | 0.484 |

however, it is not clear how to interpret tests based on orthogonal polynomial components in the case of multi-way contingency tables. $\chi^2_{[1:2]}$ outperformed $\chi^2_{PF}$, $\chi^2_{[2]}$ and $\hat{V}^2_4$, for all sample sizes, for the very same reason outlined above in the investigation of the effect of misspecified slope parameters. $\chi^2_{[2]}$ is slightly outperformed by $\hat{V}^2_4$ since first-order marginals do contain useful information about the lack-of-fit. $\chi^2_{[1:2]}$ does slightly better than the $\chi^2_{PF}$ since $\chi^2_{PF}$ is diluted by superfluous degrees of freedom.

Overall, an attempt to correlate the lack-of-fit with the underlying factor was unsuccessful, as tests based on orthogonal polynomial components required higher-order polynomials in order to achieve adequate power, which

Table 37. Asymptotic power when the alternative of interest is $A5_4$. Investigating the effect of misspecified intercept and slope parameters.

| Sample size | $\chi^2_{PF}$ | $\chi^2_{[1:2]}$ | $\chi^2_{[2]}$ | $\hat{V}_4^2$ | $\hat{V}_5^2$ | $\hat{V}_6^2$ | $\hat{V}_7^2$ |
|---|---|---|---|---|---|---|---|
| 50 | 0.143 | 0.194 | 0.142 | 0.069 | 0.069 | 0.070 | 0.074 |
| 100 | 0.282 | 0.398 | 0.268 | 0.080 | 0.084 | 0.085 | 0.088 |
| 300 | 0.835 | 0.935 | 0.759 | 0.187 | 0.187 | 0.189 | 0.190 |
| 500 | 0.986 | 0.998 | 0.955 | 0.297 | 0.297 | 0.299 | 0.302 |

was not desired. Consequently, empirical power was not calculated for any of the alternatives considered. Moreover, in general with multi-way contingency tables interpretations of orthogonal polynomial components may not be the same as interpretations in the univariate case. Thus, even in cases when tests based on orthogonal polynomial components became competitors to other statistics, they did not have a clear interpretation.

# Chapter 7: APPLICATIONS TO REAL LIFE DATA

Proposed limited-information test statistics based on orthogonal components defined on marginal frequencies in this research were applied to a real-life data set with sparse cell counts. The test statistic based on the random forest method was also applied here. The main focus with this data set was to assess how well the proposed statistics perform with respect to detecting the lack-of-fit when data are sparse, as is frequently the case with real-life psychological data.

The data used for this example are from the 38th round of the State Survey conducted by the Institute for Public Policy and Michigan State University Social Research (2005). The survey was administrated by telephone to 949 Michigan citizens from 28 May to July 18, 2005. The focus of the survey was on charitable giving and volunteer activities of Michigan households. Five questions measured the public's faith and trust in charity organizations. Respondents were asked to what degree they agree with the following statements.

- "Charitable organizations are more effective now in providing services than they were 5 years ago"
- "I place a low degree of trust in charitable organizations"
- "Most charitable organizations are honest and ethical in their use of donated funds"
- "Generally, charitable organizations play a major role in making our communities better places to live"

- "On the whole, charitable organizations do not do a very good job in helping those who need help"

All questions have four response categories corresponding to "strongly agree", "somewhat agree", "somewhat disagree", and "strongly disagree". For our example the responses are coded from 1 to 4, with larger scores indicating less favorable views of charities.

As discussed previously, in reality the constrained version of the categorical variable factor model with equal slopes is too restrictive. So, the unconstrained version of the model was used. This data set has an extremely large degree of sparseness with $n/T$ ratio of 0.927. Moreover, 730 cells have zero counts in this situation. In such cases the likelihood ratio statistic yields a p-value of almost 1 and PGF a p-value of $10^{-4}$, very different conclusions.

Table 38 shows goodness-of-fit results for: $\chi^2_{PF}$, $G^2$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$, $M_2$ and $\chi^2_{[rf]}$ at the nominal 5% significance level. Moreover, it lists the test

Table 38. Goodness-of-fit results of $\chi^2_{PF}$, $G^2$, $LR_{diff}$, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$, $M_2$ and $\chi^2_{[rf]}$ when the model under the null hypothesis is the categorical variable factor model at the nominal 5% level.

| Statistic | Value | df | value/df | p-value |
|---|---|---|---|---|
| $\chi^2_{PF}$ | 3220.547 | 1003 | 3 | $10^{-4}$ |
| $G^2$ | 879.800 | 1003 | 1 | 0.998 |
| $LR_{diff}$ | 59.445 | 4 | 15 | $10^{-4}$ |
| $\chi^2_{[1:2]}$ | 398.224 | 105 | 4 | $10^{-4}$ |
| $\chi^2_{[2]}$ | 308.720 | 90 | 4 | $10^{-4}$ |
| $M_2$ | 347.252 | 85 | 4 | $10^{-4}$ |
| $\chi^2_{[rf]}$ | 3218.083 | 1014 | 3 | $10^{-4}$ |

statistics and their corresponding value/df ratios to show that the order of the ratios are comparable to the ordering of power as discovered in Study 1 and Study 3. Furthermore, from the earlier research of nested models in Study 3 it was discovered that $LR_{diff}$ performs well under sparse conditions as long as the slope parameter estimates were not too small, which is the case here. Also, from earlier results it was discovered that the asymptotic p-values for $\chi^2_{[2]}$ and $M_2$ statistics are quite accurate when data are sparse. This is reflected in the equivalent value/df ratio values of 4.

Although $\chi^2_{[1:2]}$ also has value/df ratio of 4 it was discovered that this test statistic should only be included when a simple null hypothesis is used because it was discovered that first-order marginal components did not contribute to the power when a composite null hypothesis was used. Thus, the equivalent value/df ratio should not be taken to imply that $\chi^2_{[1:2]}$ is as powerful as $\chi^2_{[2]}$ and $M_2$.

Test statistics should not only be chosen based on power. $\chi^2_{[rf]}$ has a value/df ratio of 3 which might imply that this statistic could be a competitor to $\chi^2_{[2]}$ and $M_2$. However, ease of computation should also be an important factor and including a test based on the random forest method with $4^5 = 1024$ components resulted in intense and long computations.

Although in this case, the same conclusion would be reached by performing $\chi^2_{PF}$ or $\chi^2_{[2]}$ and $M_2$, earlier results show that a test based on second-order marginals has higher power when the lack-of-fit is in the second-order associations and is generally not affected by sparseness. Furthermore, since the model does not fit well, the investigation of orthogonal components of $\chi^2_{[2]}$

could reveal the source of misfit. The large components are shown in Table 39. All second-order orthogonal components in Table 39 are significantly different from 0 at the nominal 5% significance level. Unfortunately, meaningful extraction of any trends in these patterns could not be obtained.

Table 39. Second-order orthogonal components when the model under the null hypothesis is the categorical variable factor model is fit to data from the 38th round of the State Survey.

| Marginal Component | $\hat{\gamma}_k^2$ |
|---|---|
| $Y_2Y_4(1,1)$ | 25.871 |
| $Y_2Y_4(1,2)$ | 15.668 |
| $Y_2Y_5(1,1)$ | 11.360 |
| $Y_2Y_5(2,3)$ | 46.255 |
| $Y_3Y_5(1,1)$ | 14.127 |
| $Y_3Y_5(3,1)$ | 11.355 |
| $Y_4Y_5(1,2)$ | 15.219 |
| $Y_4Y_5(2,1)$ | 13.201 |

As discussed in Chapter 3, Moustaki's $GFfit^{ij}$ can be used to obtain more detailed information about the model fit. Individual $GFfit^{ij}$ identify pair-wise associations and can be obtained by summing the appropriate number of orthogonal components of PGF as demonstrated in Chapter 3. For example, the sum of 16 second-order marginal components or $GFfit^{24} = \sum_{i=1,j=1}^{4} Y_2Y_4(i,j)$. Individual $GFfit^{ij}$'s are given in Table 40.

As can be seen from Table 40 the categorical variable factor model does not fit well as 5 bivariate test statistics each with 9 degrees of freedom are significantly different from 0 at the nominal 5% level. Moreover, this

implies that these 5 pairs of variables are responsible for the poor fit of the model. Furthermore, the main contributors to the lack-of-fit as can be seen from Table 40 are variables $2, 3$ and 4. It is important to note that lack-of-fit in this example may be attributed to the fact that one or more assumptions in the model have been violated. Perhaps the number of latent variables was incorrectly chosen or the wrong model was selected. These possibilities have not been tested in this research.

Table 40. Pair-wise associations using $GFfit^{ij}$ when the model under the null hypothesis is the categorical variable factor model with a single latent variable is fit to data from the 38th round of the State Survey. '*' denotes significant bivariate test statistics at the 5% nominal level.

| $GFfit^{ij}$ | Value | $p$-value |
|---|---|---|
| (12) | 15.584 | 0.076 |
| (13) | 13.381 | 0.146 |
| (14) | 16.251 | 0.062 |
| (15) | 10.200 | 0.335 |
| (23) | 21.469 | $0.011^{*}$ |
| (24) | 60.188 | $10^{-4*}$ |
| (25) | 63.433 | $10^{-4*}$ |
| (34) | 9.327 | 0.408 |
| (35) | 38.257 | $10^{-4*}$ |
| (45) | 49.310 | $10^{-4*}$ |

# Chapter 8: CONCLUSIONS AND FURTHER RESEARCH

In this dissertation, orthogonal components of Pearson's chi-squared statistic defined on marginal frequencies were used to develop goodness-of-fit tests. A subset of these orthogonal components lead to the construction of limited-information tests that allowed one to identify the source of lack of fit and to increase the power of the tests. The derived goodness-of-fit tests ($\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $\chi^2_{[rf]}$) were evaluated in studies for detecting two-way and three-way associations that were not accounted for by a model on a multi-dimensional contingency table for a single latent variable. In addition the derived tests were also used to investigate the case when the model misspecification involved parameter constraints for large but sparse contingency tables.

In the case of detecting two-way associations that were not accounted for by the model, two versions of the model under the null hypothesis were considered: the unconstrained and the constrained versions. For the model with the unconstrained version, both PGF and $\chi^2_{[rf]}$ had inflated empirical Type I error rates when $n = 300$. So power was confounded with Type I error rate and was not comparable to power rates for other test statistics when $n = 300$. For the model under the constrained version, all statistics performed adequately when sparseness was present. With both constrained and unconstrained versions of the model under the null hypothesis, it was discovered that there was no significant difference in terms of power between $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$. Both $\chi^2_{[1:2]}$ and $\chi^2_{[2]}$ attained higher power compared to other statistics considered. However, since the power for $\chi^2_{[1:2]}$ is diluted with superfluous degrees of freedom, the recommendation is to use $\chi^2_{[2]}$. As expected, the traditional

Pearson's chi-squared test was adversely affected by sparseness and was not competitive in terms of power with the other statistics investigated in this research.

The recommendation in the case of nested models for multi-category variables and large multi-way contingency tables is that the likelihood ratio difference test should be used to select models. However, bias in the parameter estimates may cause poor results with the likelihood ratio difference test. Specifically, when the average slope parameter in a logistic regression was small in magnitude, the likelihood ratio difference test, $LR_{diff}$, resulted in inflated empirical Type I error rates for all sample sizes considered. Moreover, $LR_{diff}$ should be used especially when sparseness is present, as it resulted in higher power compared to tests defined on lower-order marginals.

For nested models with dichotomous variables in multi-dimensional contingency tables, $\chi^2_{[1:2]}$, $\chi^2_{[2]}$ and $M_2$ all had inflated empirical Type I error rates when $n = 300$, irrespective of the value of the average slope parameter. Even though the power for $\chi^2_{[2]}$ and power for $M_2$ were not different from each other, the recommendation is to use $\chi^2_{[2]}$ since $M_2$ includes first- and second-order marginals and cannot isolate components to determine source of lack of fit. On the other hand, for nested models with multi-categorical variables, all statistics performed adequately when sparseness was present. Thus, the recommendation is to use $\chi^2_{[2]}$ for reasons previously stated.

Chi-squared limited-information test statistics proposed in this dissertation are still in a sense omnibus tests but less so than PGF. Although the chi-squared test statistics were formed by adding a subset of components, a

clear method for selecting the optimal number of components using the random forest method was not achieved in this research.

The use of the random forest algorithm did not result in test statistics with higher power and fewer components. This could possibly be attributed to incorrectly specifying $\mathbf{z}$ as the response vector and/or $\widehat{\mathbf{M}}$ as the model matrix in the random forest algorithm in R. Also large positive values of VIMP for a variable might not indicate the important nature of that component in terms of variable selection in the orthogonal regression as stated by Ishwarn (2007). Thus the use of VIMP might not result in dimensionality reduction of the model matrix. Although the benefit of applying the random forest algorithm was in its lack of the need for additional assumptions, further research needs to be performed in order to more accurately relate the response vector to the model matrix in this application.

Orthogonal polynomial components have been previously applied to one-dimensional contingency tables in testing goodness of fit for a specified univariate distribution. However, in this dissertation the focus was on higher-dimensional multi-way contingency tables. In this setting it was difficult to order the cells of the multi-way table in a way that would be conducive to a test on low-order polynomial components. Also, with multi-way tables interpretations of the orthogonal polynomial components may not be the same as interpretations with the univariate case, where first-degree orthogonal polynomial components tend to detect a change in mean, second-degree orthogonal polynomial components tend to detect change in variance, etc. Overall, an attempt to correlate the lack of fit with the underlying factor was unsuccessful,

as tests based on orthogonal polynomial components required a larger number of higher-order components than desired in order to achieve adequate power.

Among the conclusions it is important to note that for tests based on marginals, inclusion of first-order components should be considered when testing a simple null hypothesis, as higher power was obtained when they were included. However, when a composite null hypothesis was tested, first-order components did not contribute to increased power.

Further, it was found that when linear dependencies among the components were present, components obtained from the QR decomposition were usually not calculated accurately nor were they ordered correctly. The routines for the QR decomposition in R and in SAS IML are not written to check carefully for linear dependencies. If the routines for the QR decomposition were written to check for linear dependencies, then they would be more reliable. Goodnight's sweep operator in SAS PROC REG produced components from sequential sum of squares and performed reliably even when linear dependencies were encountered.

Overall, this research has confirmed that tests based on first-order marginals do not contribute to the power of the test for a composite null hypothesis. Test statistics defined on just second-order marginals can serve as a remedy for sparseness and are competitive with other limited-information tests such as $M_2$. However, the preference of using a test statistic defined on components over $M_2$ is due to the capability of isolating lack of fit.

Research in this dissertation was directed toward the specific case of multi-dimensional contingency tables when a model using a single latent vari-

able was employed. Only settings with 5 and 10 variables with 2 and 3 categories for each variable were considered. Further research needs to be explored to examine the cases of large numbers of variables i.e., $p = 15, 20$ with varying degrees of sparseness. Based on the cases studied here, it can be concluded that tests formed on orthogonal components defined on marginal frequencies ($\chi^2_{[2]}$) can provide more powerful directional tests both when data are sparse or not sparse. The applications of the proposed limited-information tests are especially important in applications which exhibit severe sparseness as does the real life data set examined in Chapter 7.

# APPENDIX A

Figure 1. QQ-plot for $\chi^2_{PF}$ for $n = 300$ when the model under the null hypothesis is the categorical variable factor model for 5 variables, each at 3 categories. The estimated slope in the QQ-plot is 1.022 and the corresponding p-value is $10^{-4}$.

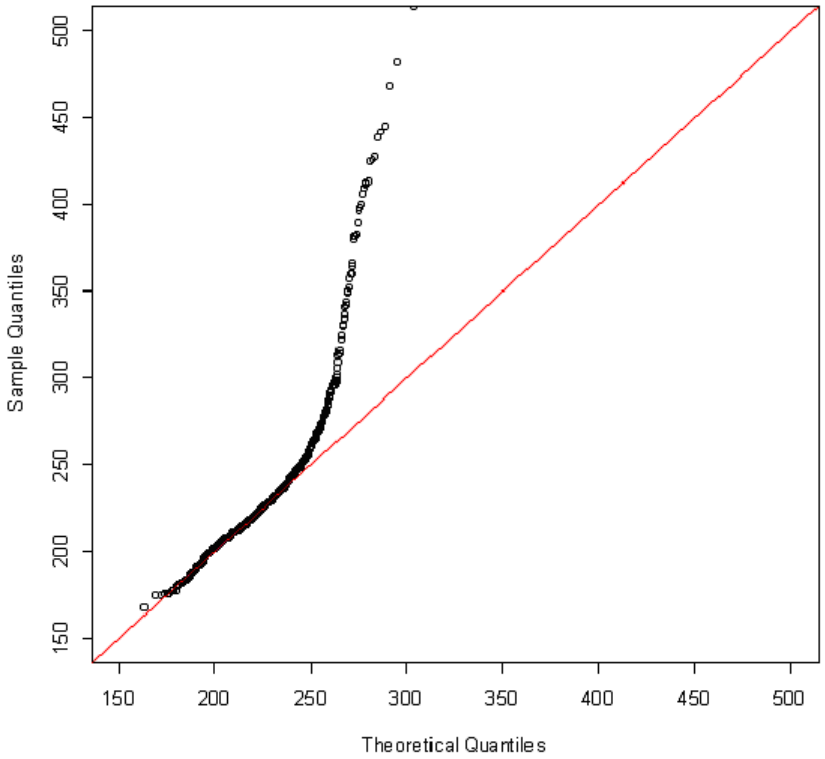


Figure 2. QQ-plot for $\chi^2_{PF}$ for $n = 500$ when the model under the null hypothesis is the categorical variable factor model for 5 variables, each at 3 categories. The estimated slope in the QQ-plot is 0.997 and the corresponding p-value is 0.1023.

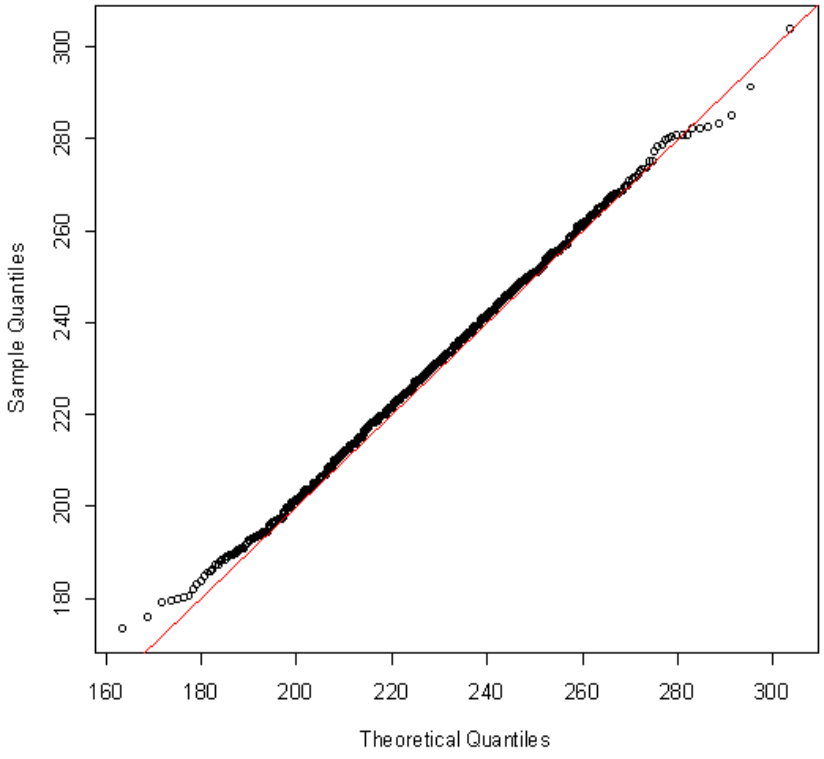


Figure 3. QQ-plot for $\chi^2_{[1:2]}$ for $n = 300$ when the model under the null hypothesis is the categorical variable factor model for 5 variables, each at 3 categories. The estimated slope in the QQ-plot is 0.965 and the corresponding p-value is 0.6939.

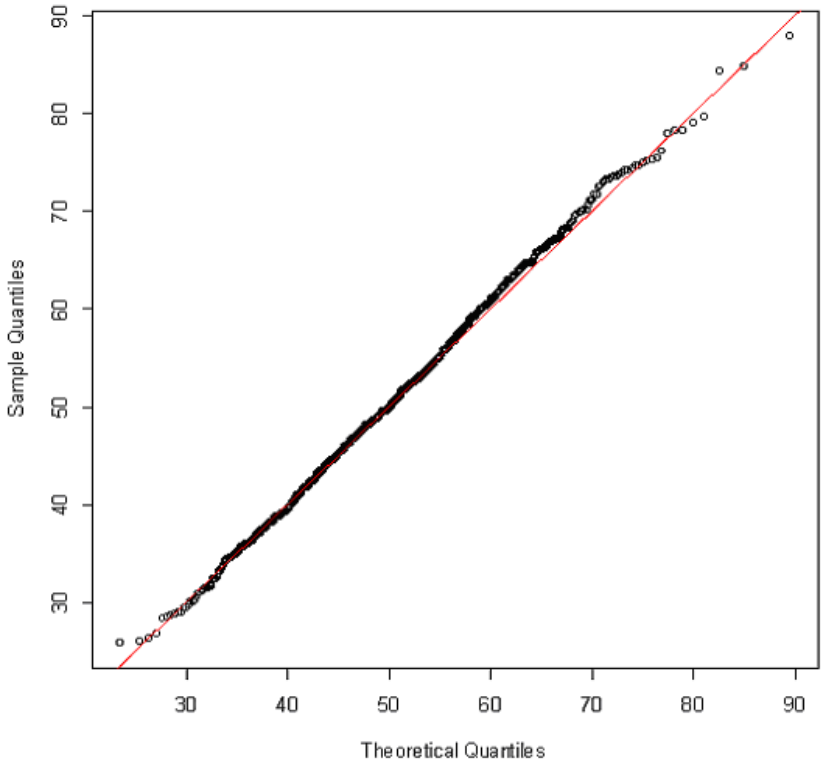
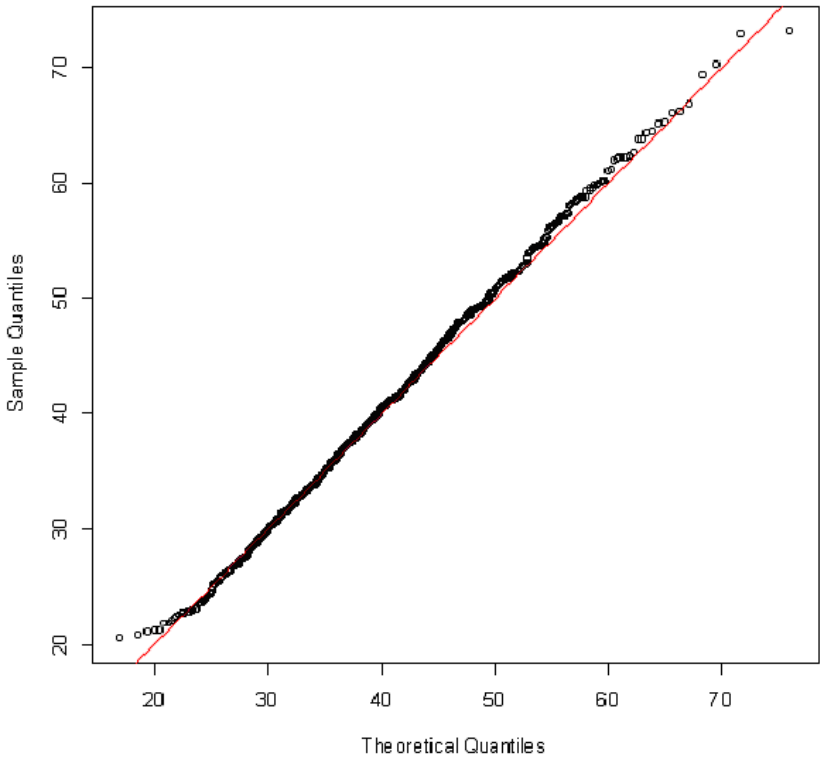
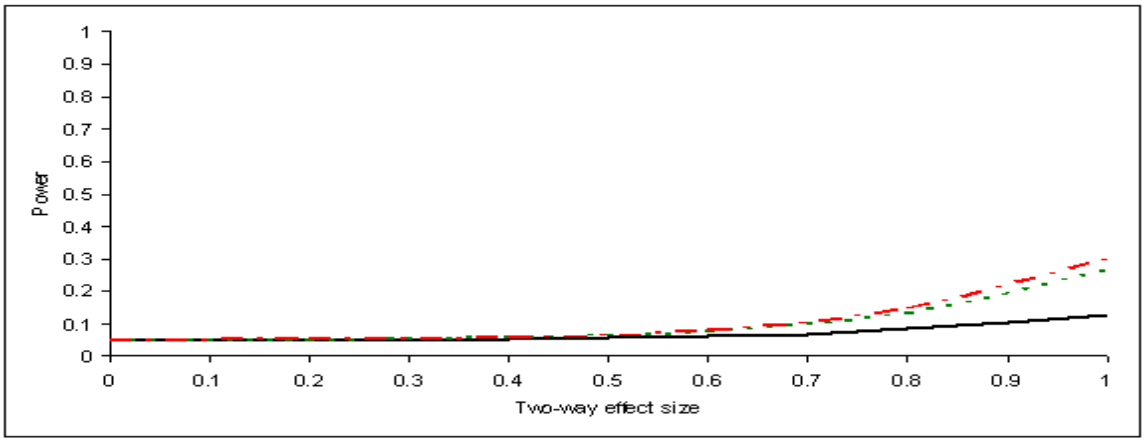


Figure 4. QQ-plot for $\chi^2_{[2]}$ for $n = 300$ when the model under the null hypothesis is the categorical variable factor model for 5 variables, each at 3 categories. The estimated slope in the QQ-plot is 0.957 and the corresponding p-value is 0.3578.



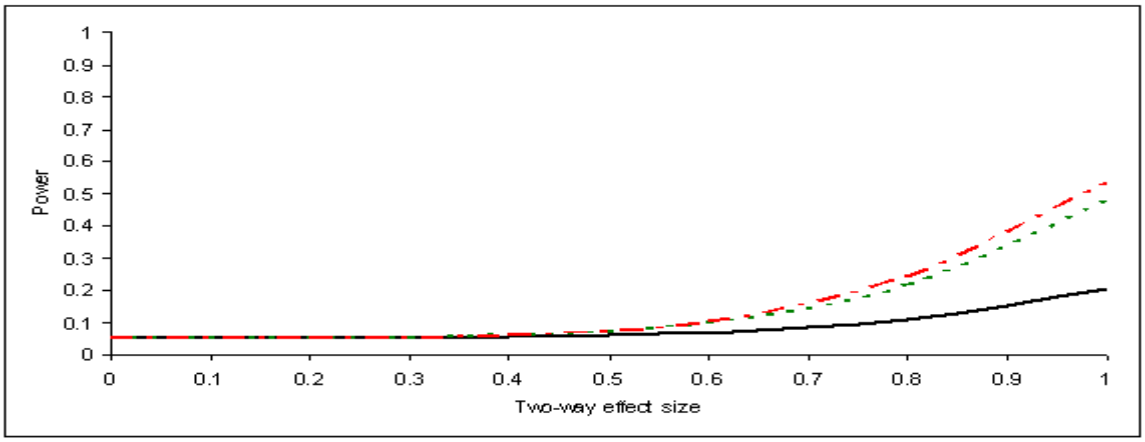
# APPENDIX B

Figure 5. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 300$.



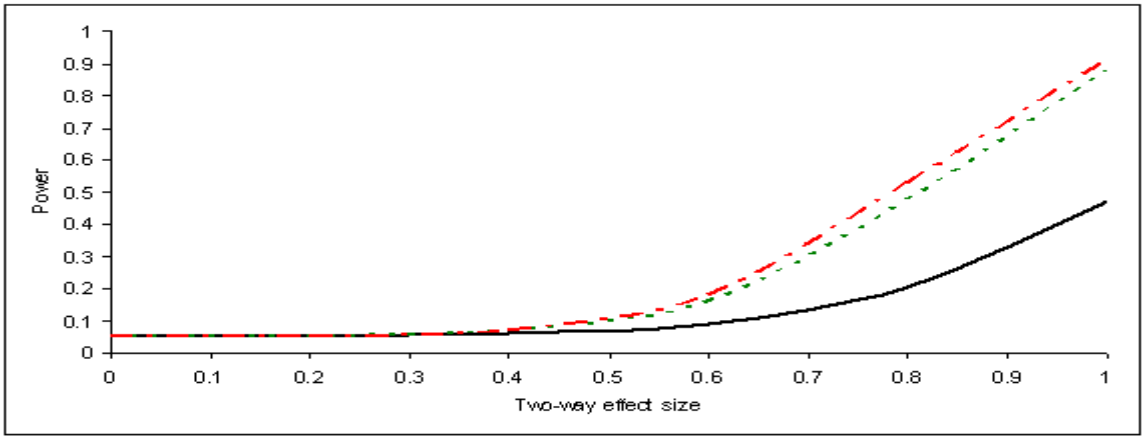

Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, – · – $\chi^2_{[2]}$

Figure 6. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 500$.



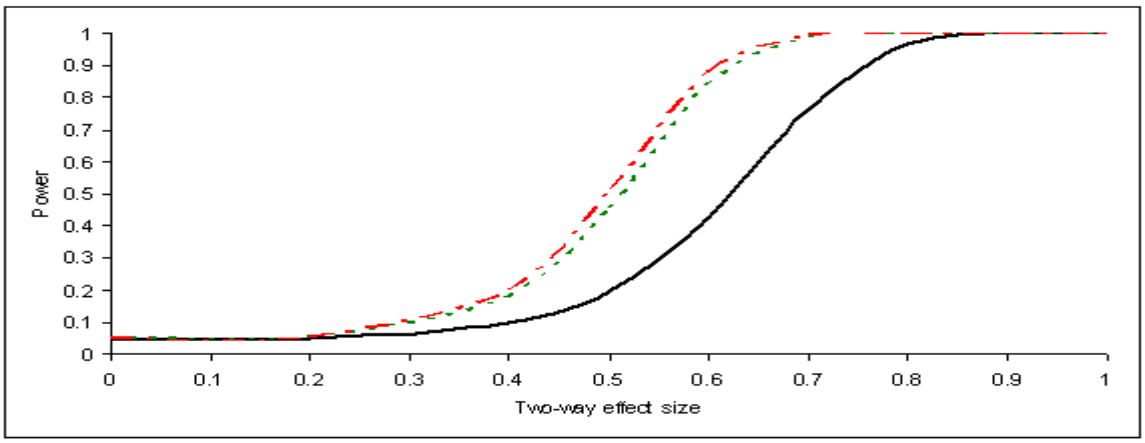

Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, – · – $\chi^2_{[2]}$

Figure 7. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 1000$.




Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, – · – $\chi^2_{[2]}$

Figure 8. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 5000$.




Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, $- \cdot -$ $\chi^2_{[2]}$

# APPENDIX C

Figure 9. QQ-plot for $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.991 and the corresponding p-value is 0.8793.

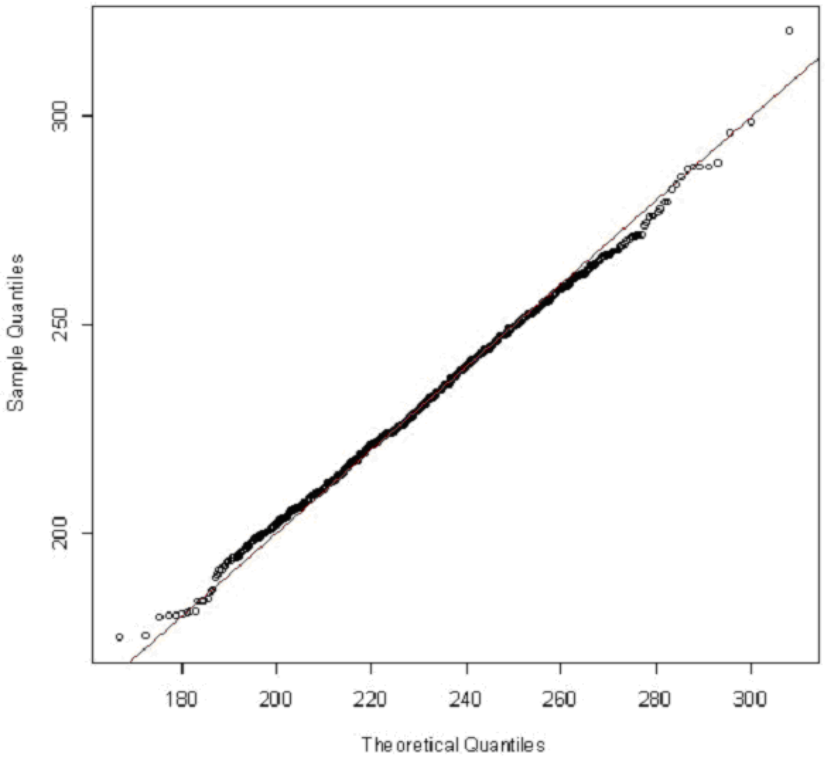


Figure 10. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.951 and the corresponding p-value is 0.0022.

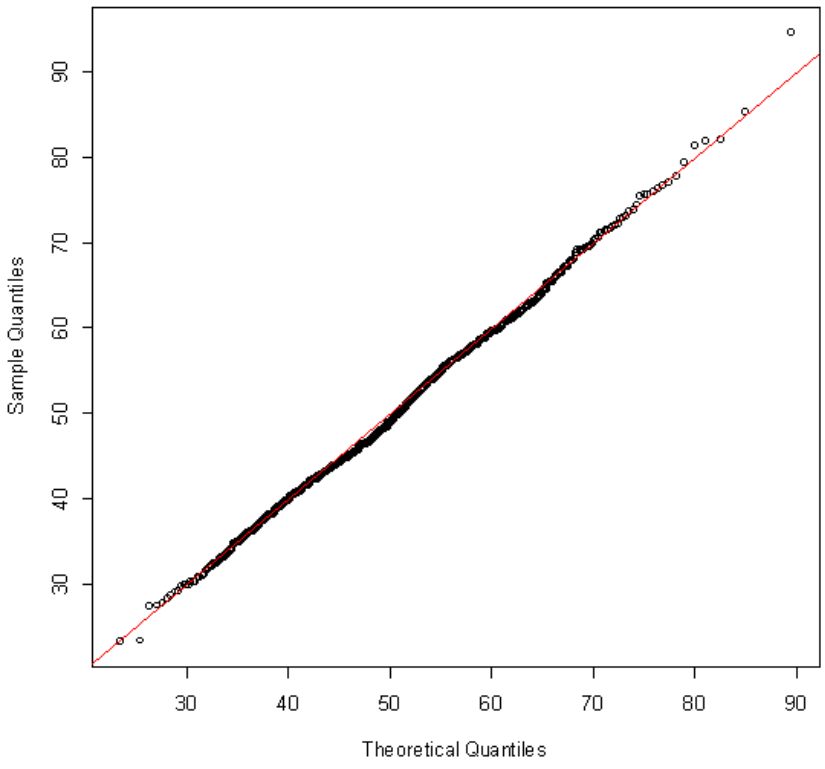


Figure 11. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 500$. The estimated slope in the QQ-plot is 0.948 and the corresponding p-value is 0.1036.

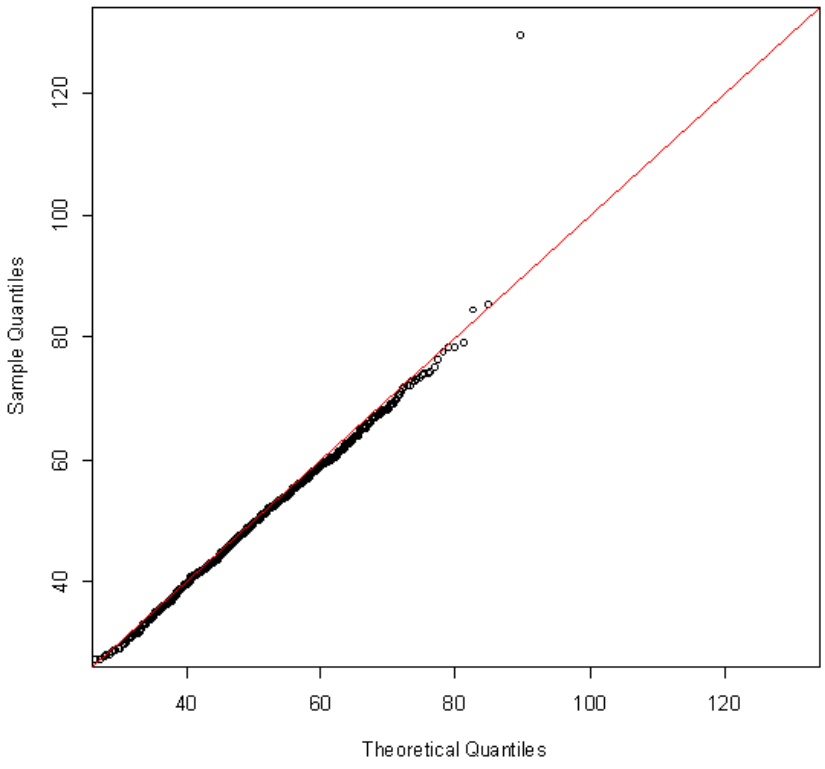
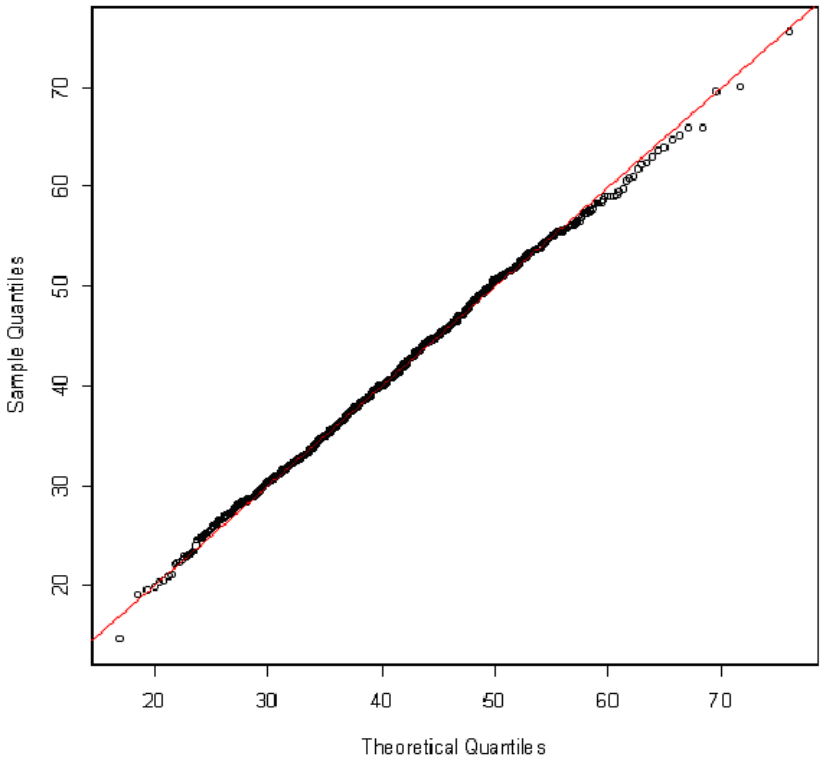
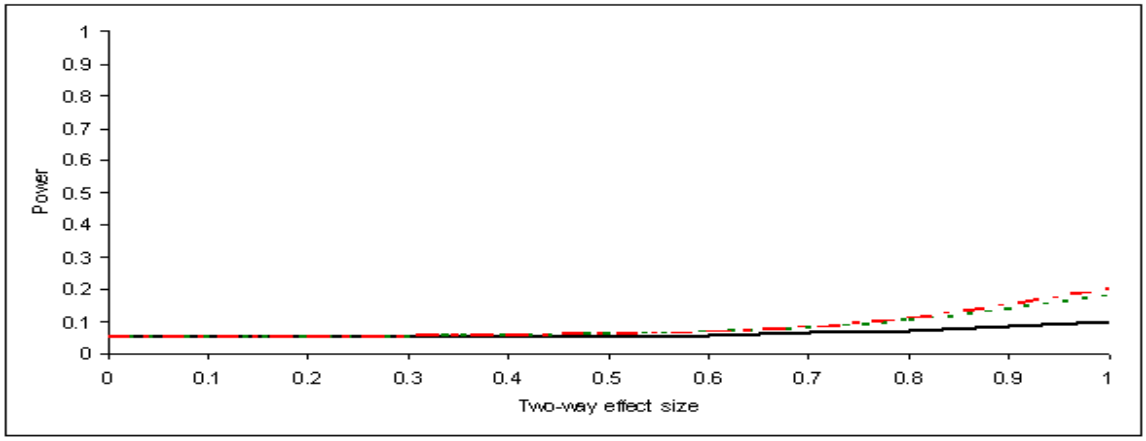


Figure 12. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.952 and the corresponding p-value is 0.9850.



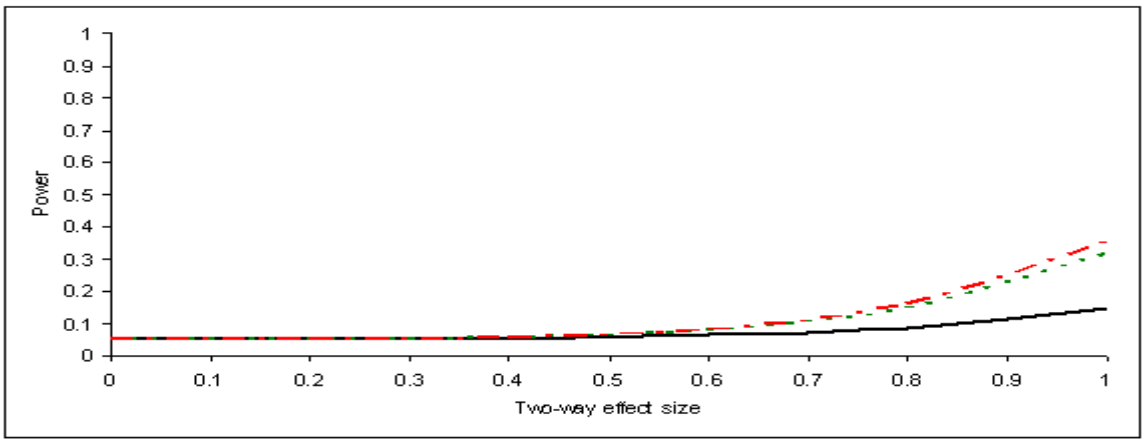
# APPENDIX D

Figure 13. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 300$.



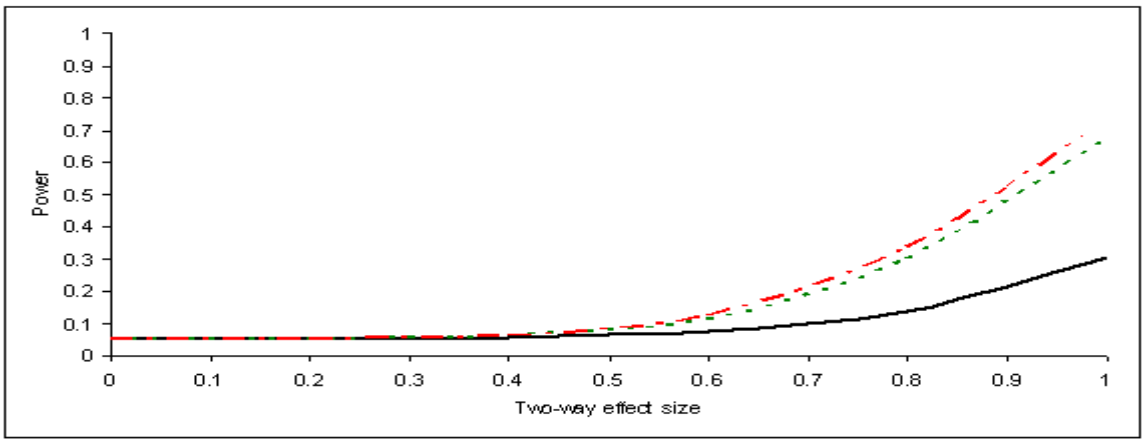

Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, $- \cdot -$ $\chi^2_{[2]}$

Figure 14. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 500$.



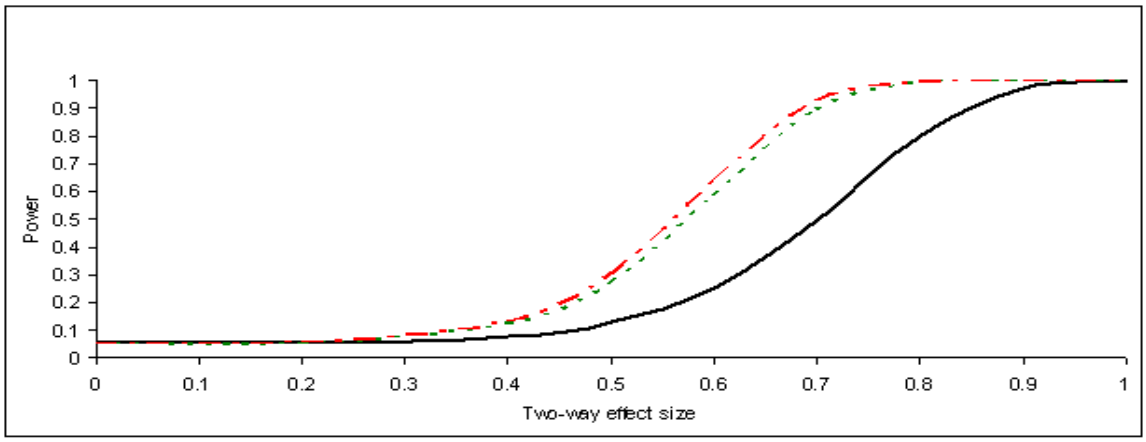

Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, – · – $\chi^2_{[2]}$

Figure 15. Asymptotic power Vs Two-way effect size when the model under the null hypothesis is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 1000$.




Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, – · – $\chi^2_{[2]}$

Figure 16. Asymptotic power Vs Two-way effect size when the true model is the constrained version of the categorical variable factor model and the alternative of interest is $A1_1$ for 5 variables, each at 3 categories for $n = 5000$.




Key: — $\chi^2_{PF}$ , $\cdots$ $\chi^2_{[1:2]}$, $- \cdot -$ $\chi^2_{[2]}$

# APPENDIX E

Figure 17. QQ-plot for $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 100$. The estimated slope in the QQ-plot is 0.908 and the corresponding p-value is 0.0305.

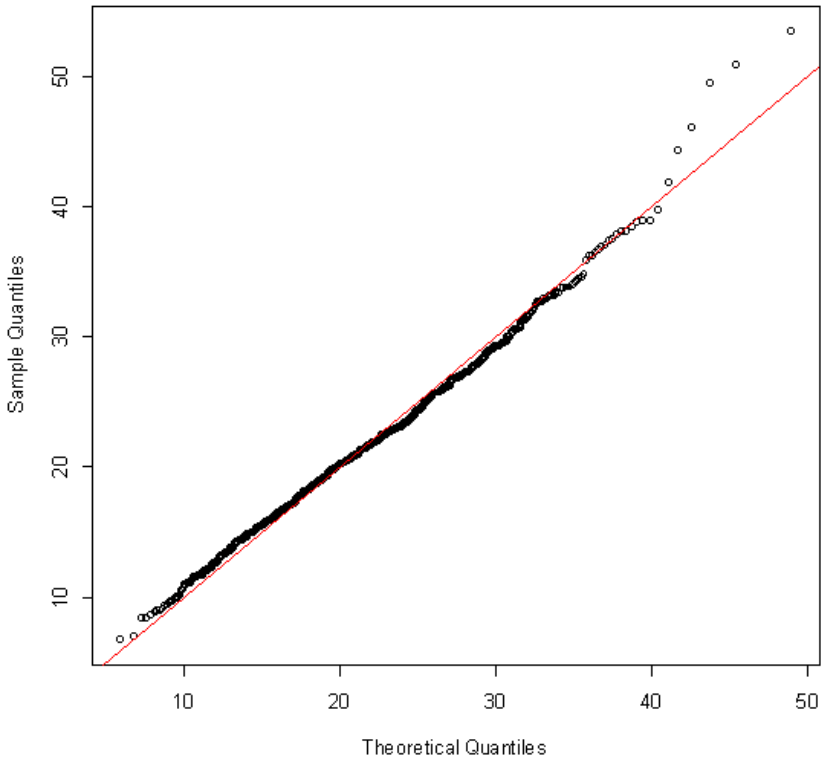


Figure 18. QQ-plot for $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 300$. The estimated slope in the QQ-plot is 0.907 and the corresponding p-value is 0.9024.

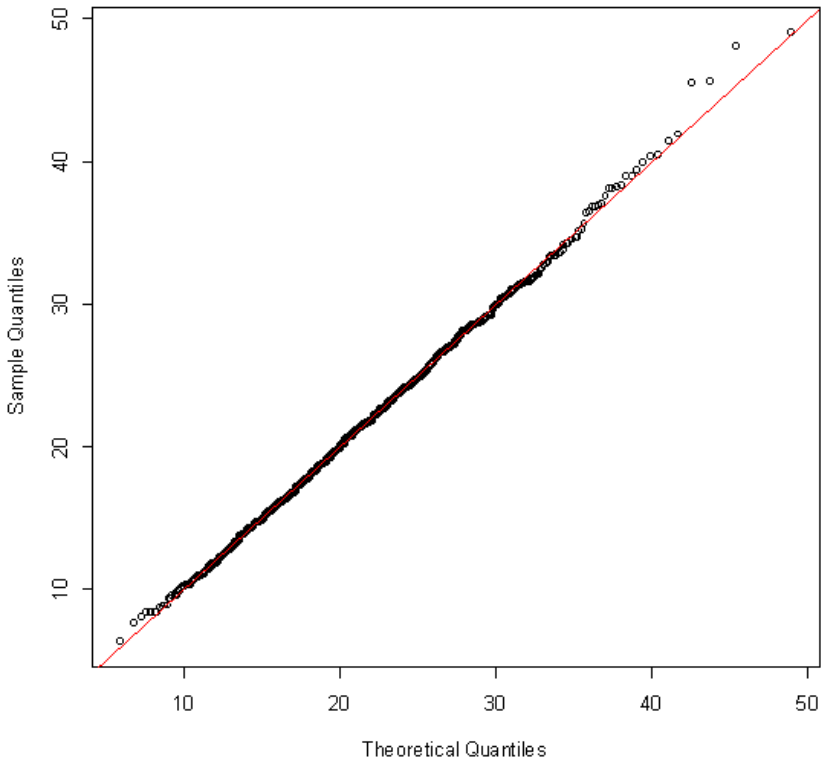


Figure 19. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 100$. The estimated slope in the QQ-plot is 0.702 and the corresponding p-value is 0.0078.

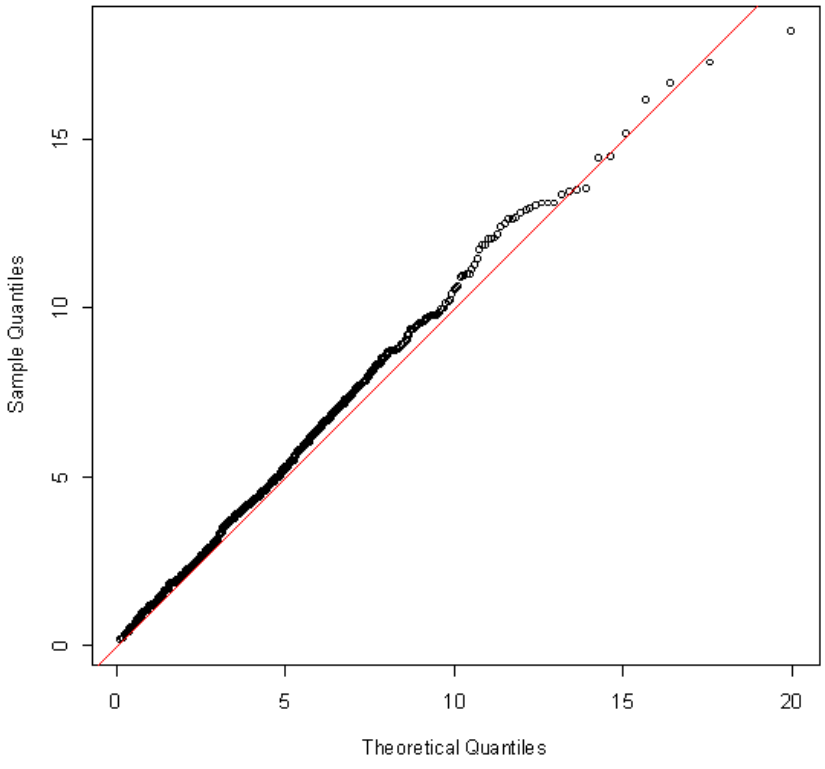


Figure 20. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 300$. The estimated slope in the QQ-plot is 0.686 and the corresponding p-value is 0.8345.

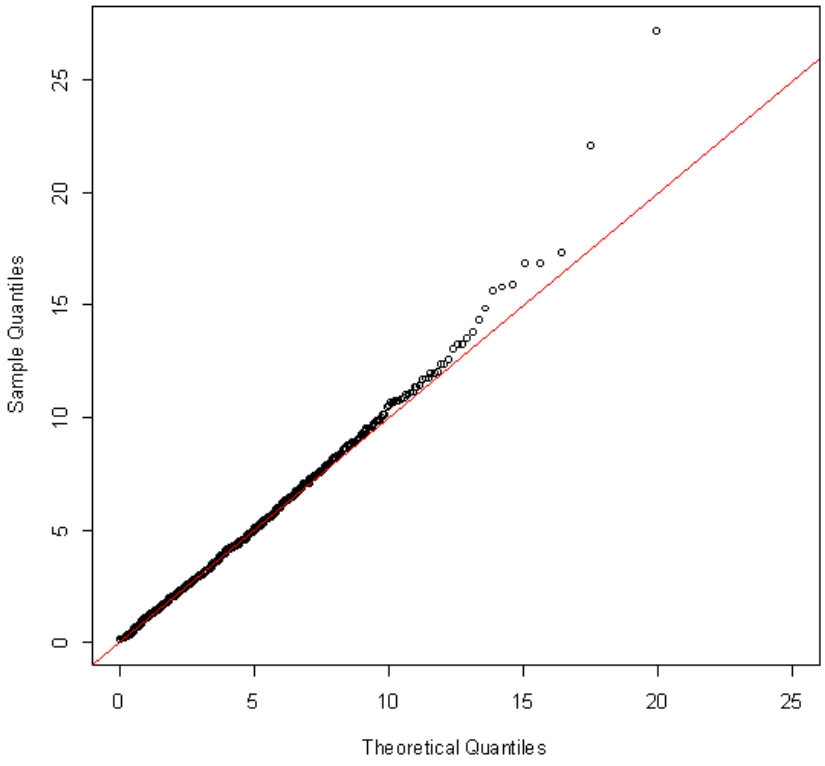


Figure 21. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 100$. The estimated slope in the QQ-plot is 0.871 and the corresponding p-value is 0.3886.

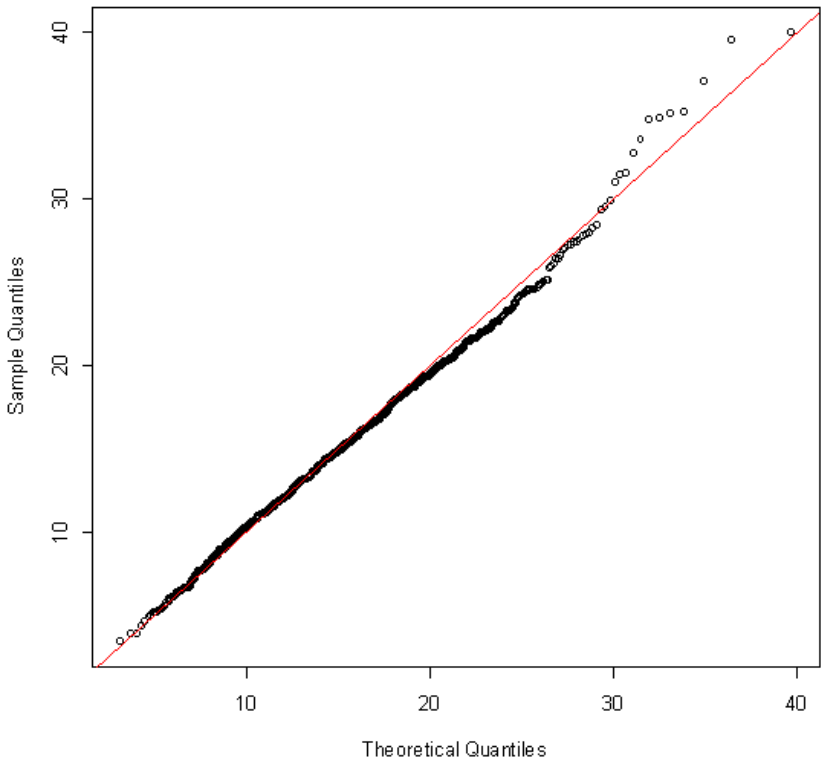


Figure 22. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 100$. The estimated slope in the QQ-plot is 0.817 and the corresponding p-value is $10^{-4}$.

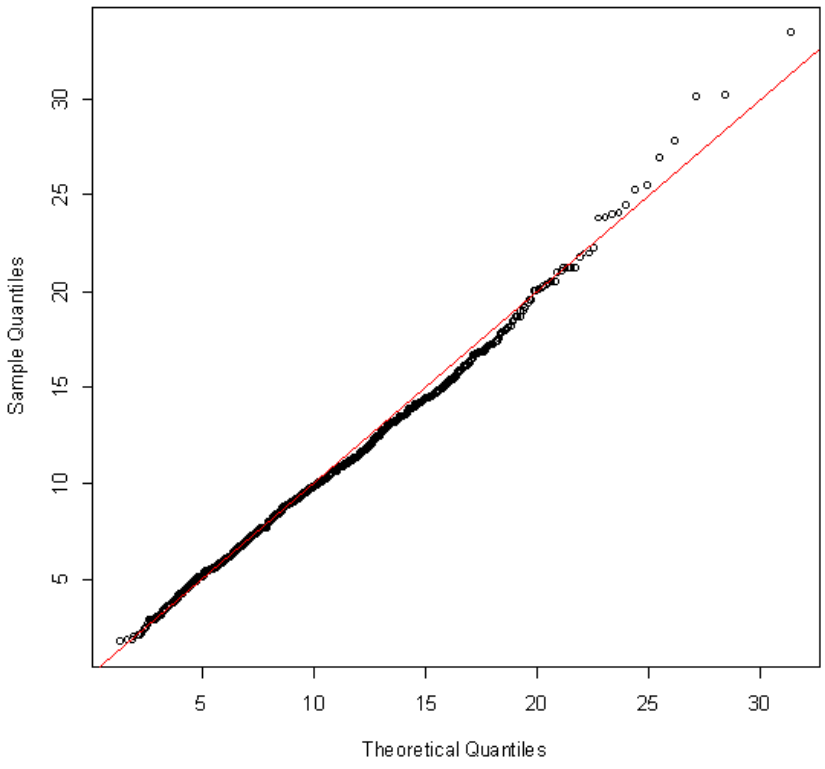
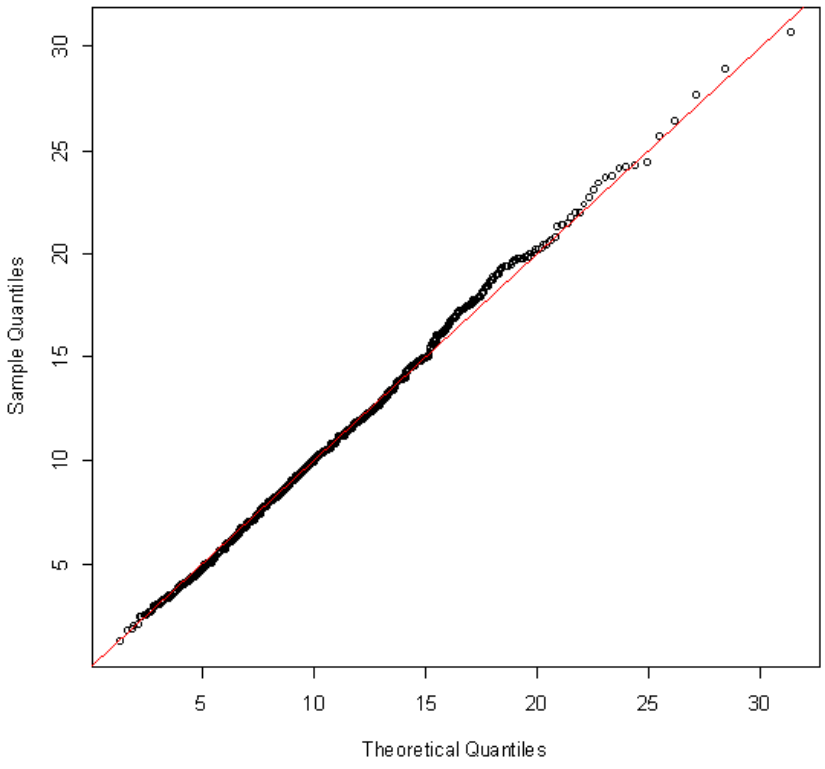


Figure 23. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 2 categories for $n = 300$. The estimated slope in the QQ-plot is 0.824 and the corresponding p-value is 0.6070.



# APPENDIX F

Figure 24. QQ-plot for $\chi^2_{PF}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.993 and the corresponding p-value is 0.1775.

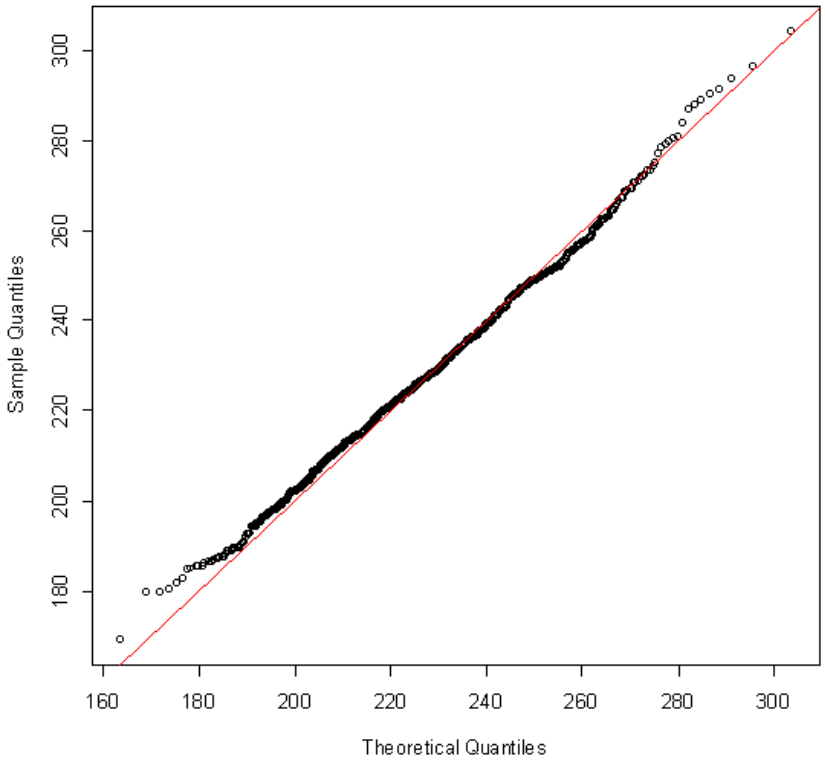


Figure 25. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 1.020 and the corresponding p-value is 0.1092.

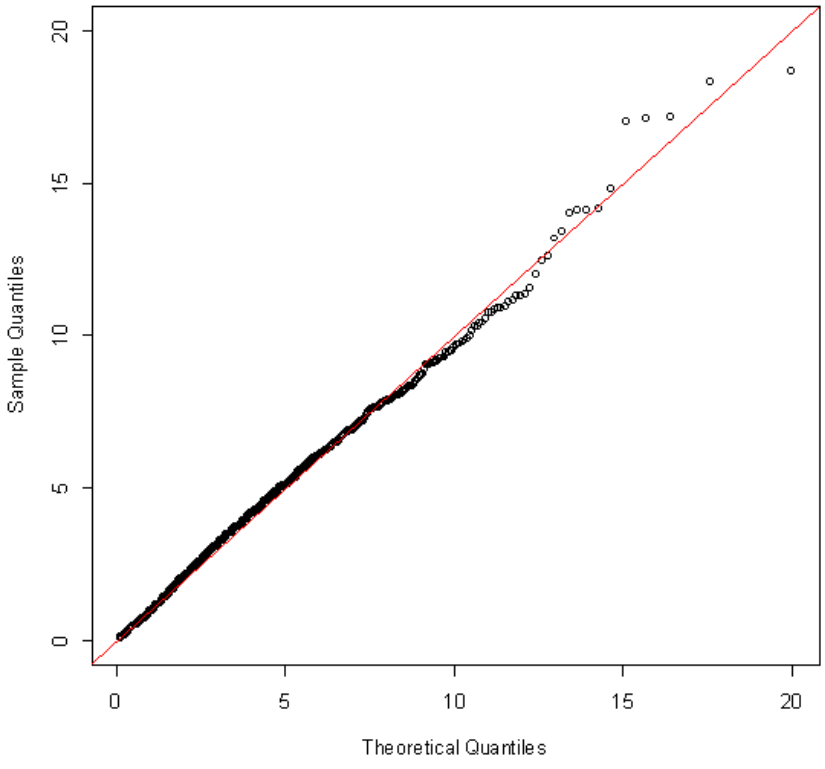


Figure 26. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.961 and the corresponding p-value is 0.8082.

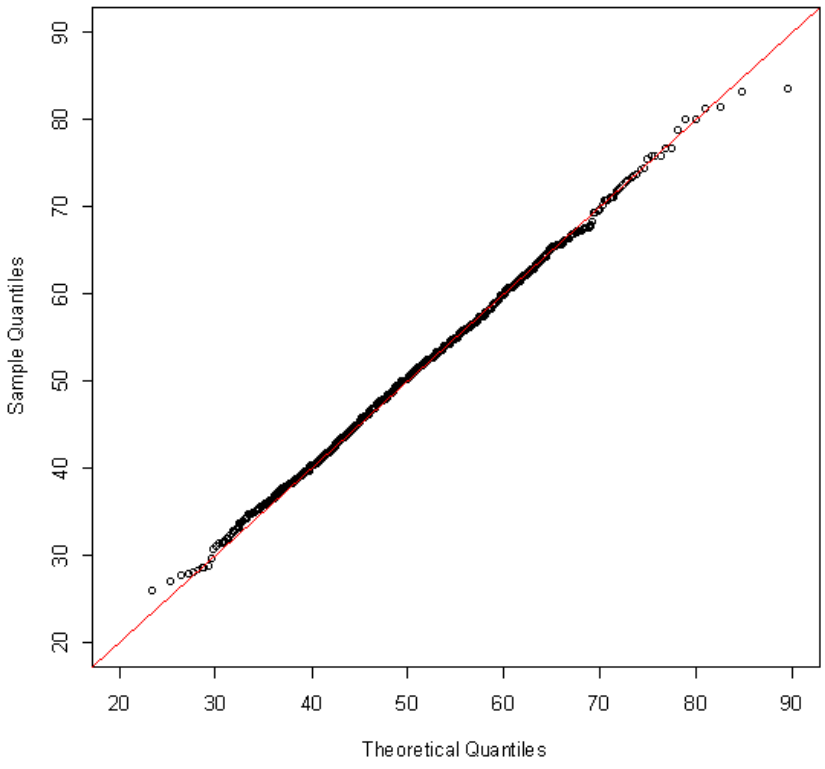
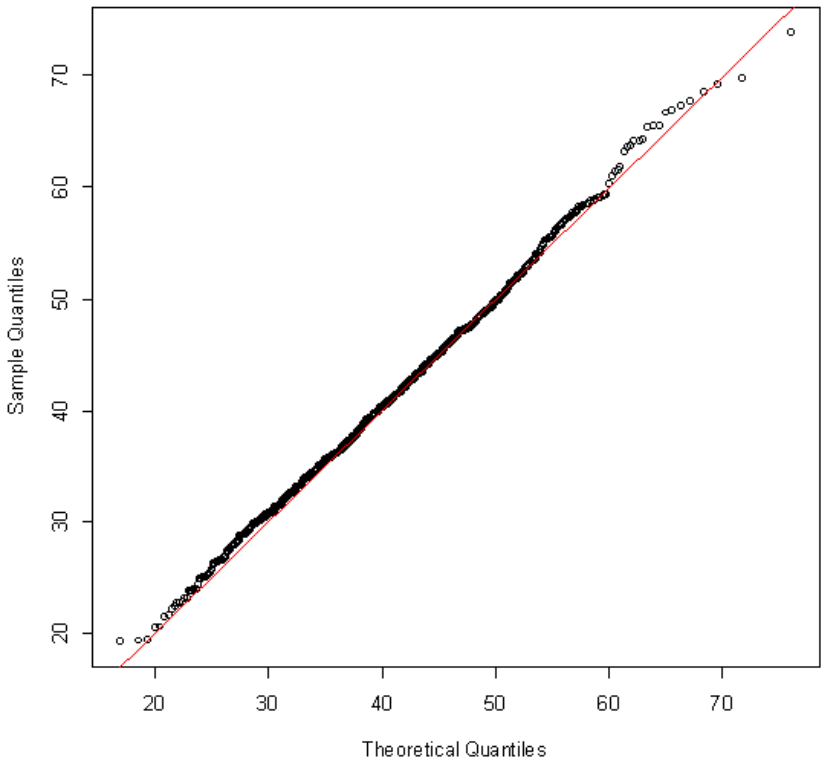


Figure 27. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 5 variables, each at 3 categories for $n = 300$. The estimated slope in the QQ-plot is 0.956 and the corresponding p-value is 0.4206.



# APPENDIX G

Figure 28. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.0$. The estimated slope in the QQ-plot is 1.086 and the corresponding p-value is 0.0023.

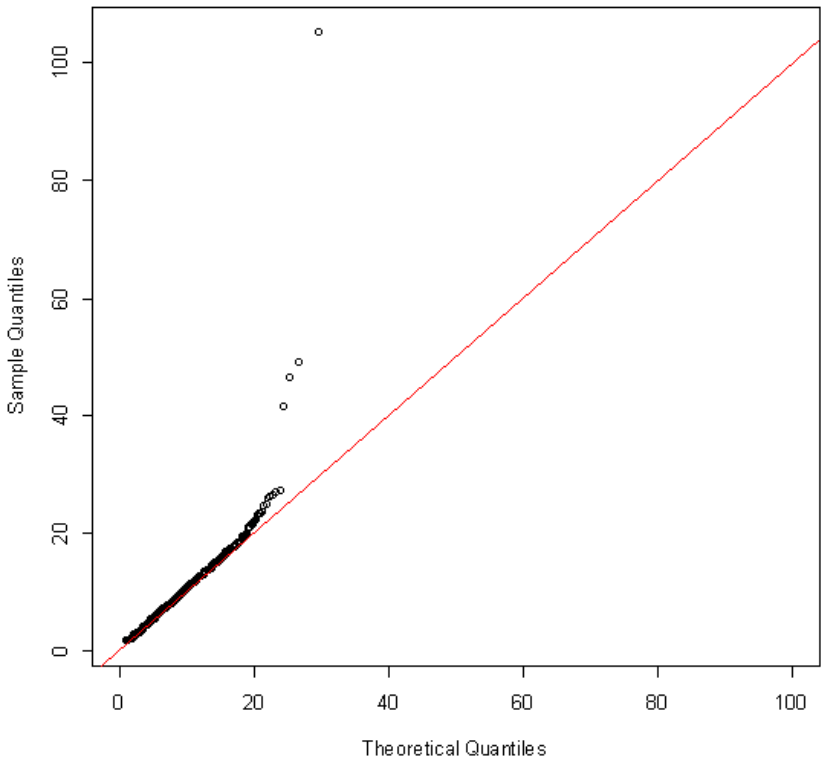


Figure 29. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ when $\bar{\beta}_1 = 1.0$. The estimated slope in the QQ-plot is 1.031 and the corresponding p-value is 0.4543.

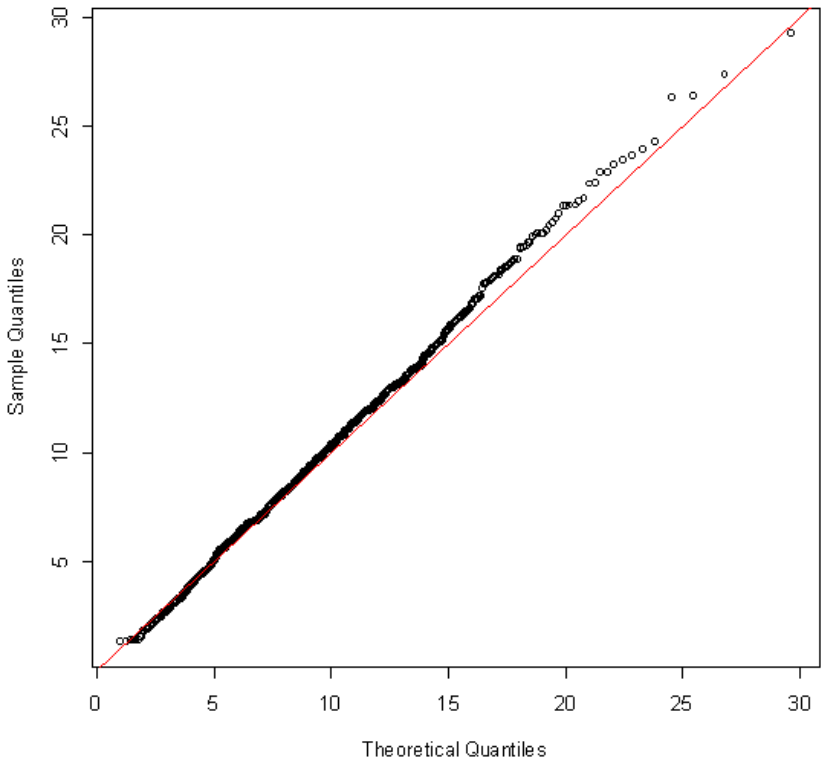
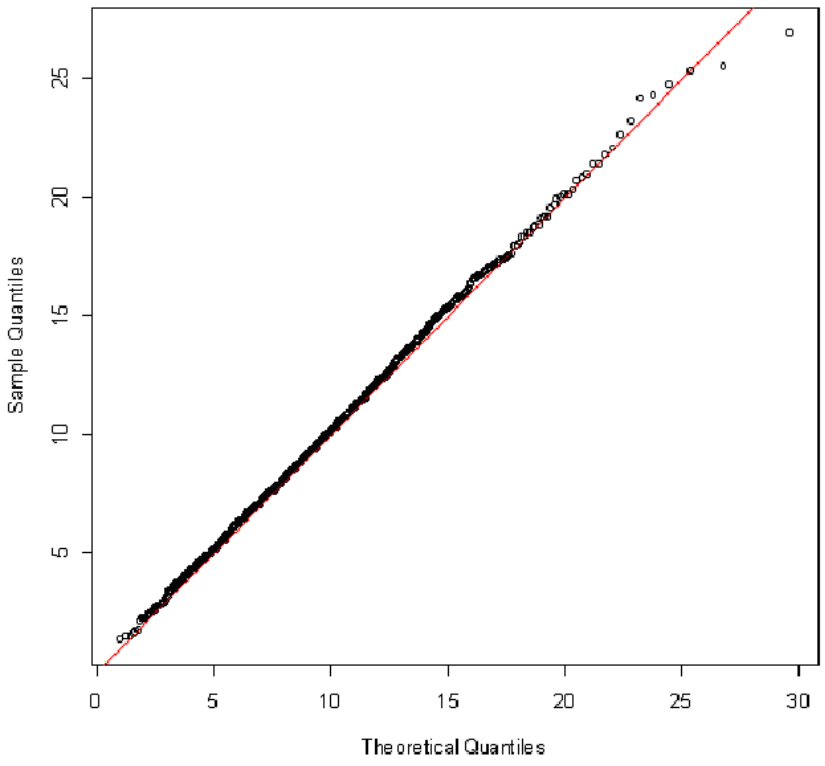


Figure 30. QQ-plot for $LR_{diff}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.5$. The estimated slope in the QQ-plot is 1.003 and the corresponding p-value is 0.9018.



# APPENDIX H

Figure 31. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.919 and the corresponding p-value is $10^{-4}$.

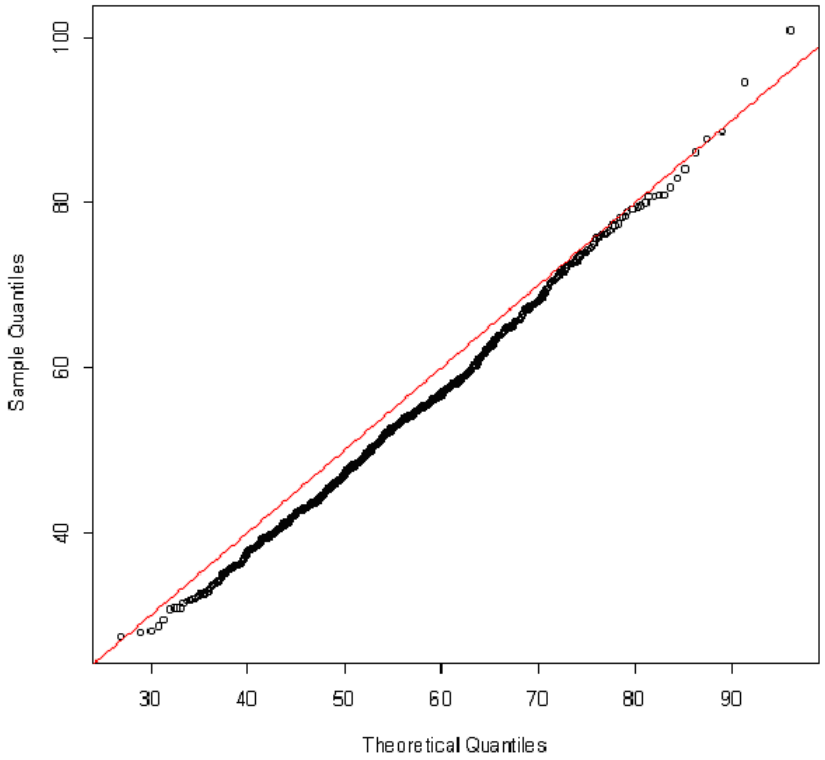


Figure 32. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.946 and the corresponding p-value is $10^{-4}$.

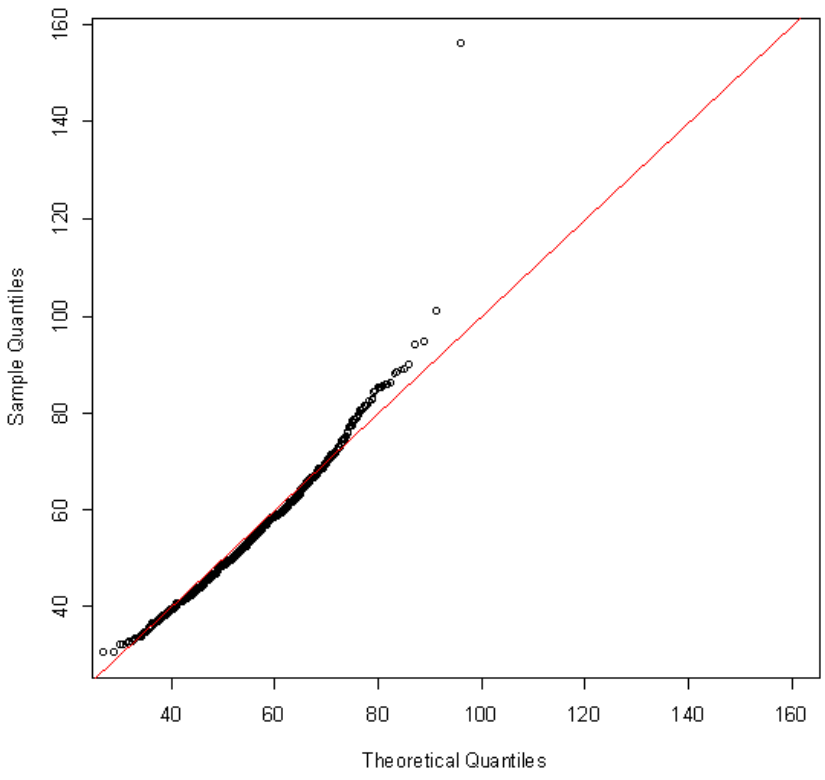


Figure 33. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 750$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.949 and the corresponding p-value is 0.0001.

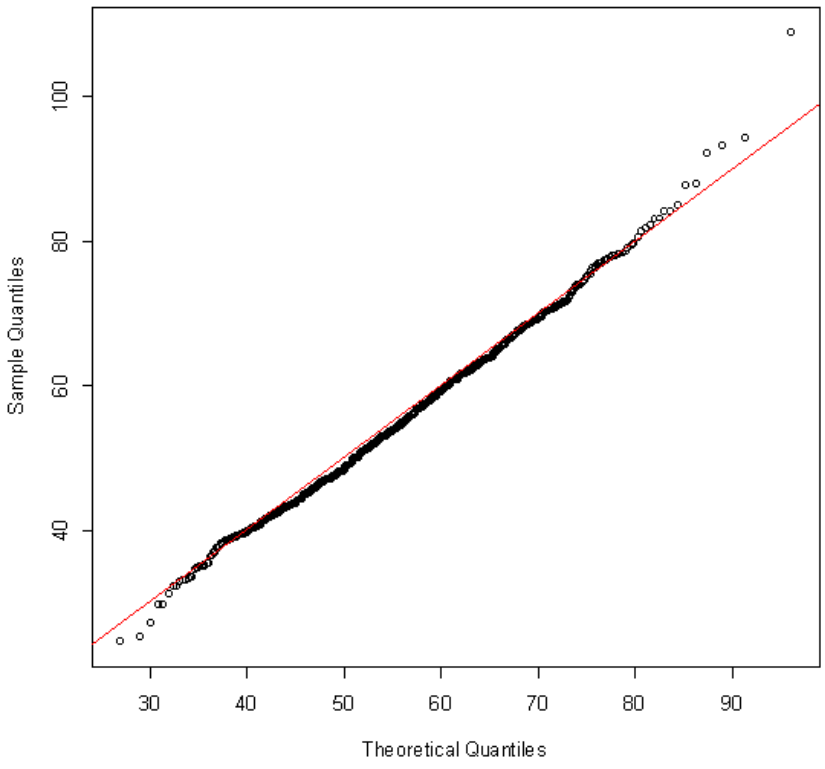


Figure 34. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 1000$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.961 and the corresponding p-value is 0.6006.

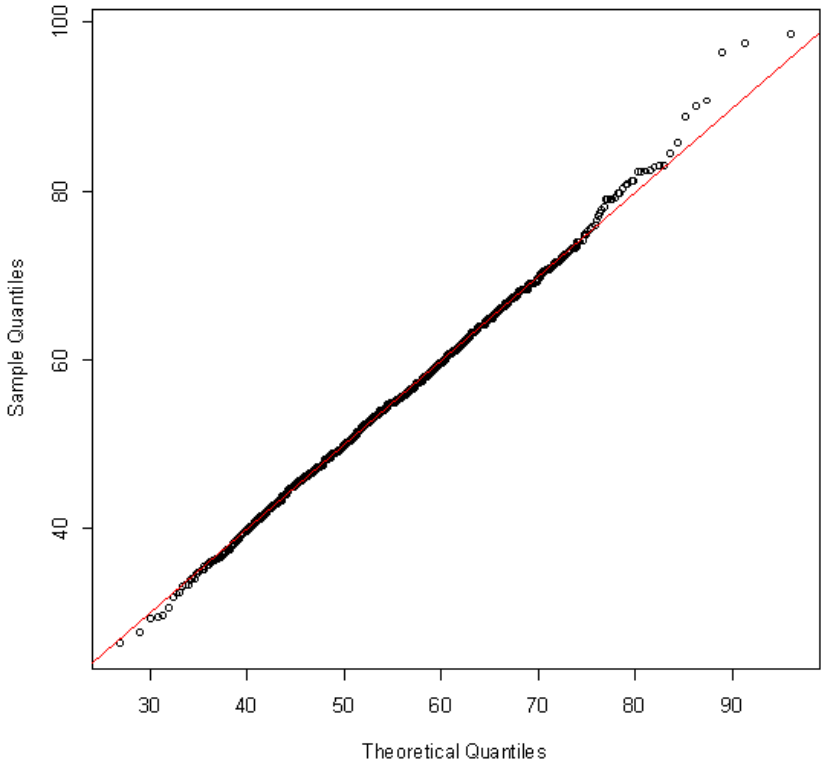


Figure 35. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.0$. The estimated slope in the QQ-plot is 0.962 and the corresponding p-value is 0.1890.

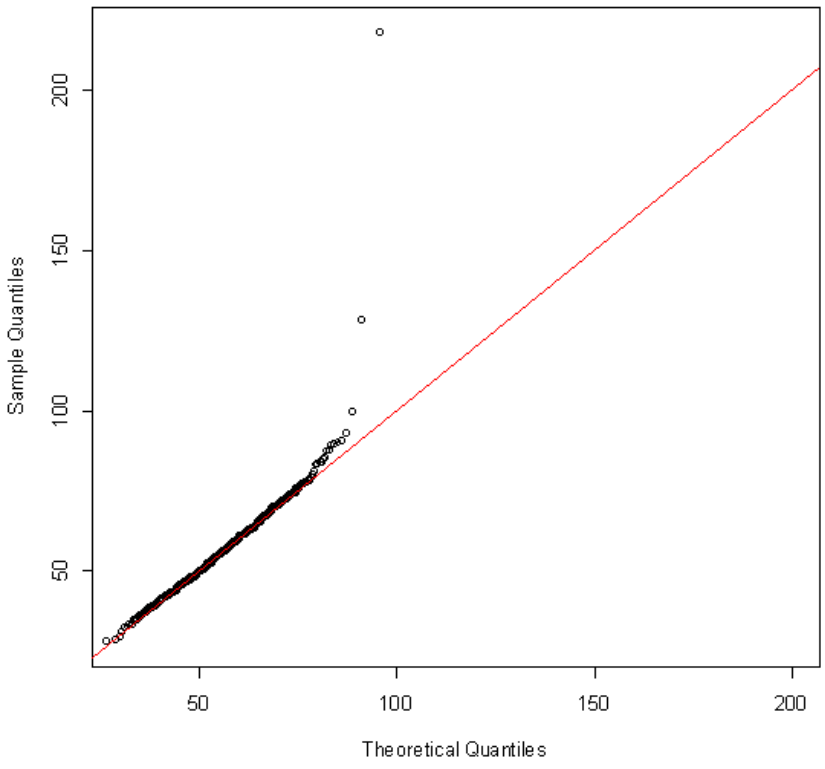


Figure 36. QQ-plot for $\chi^2_{[1:2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.5$. The estimated slope in the QQ-plot is 1.000 and the corresponding p-value is 0.9995.

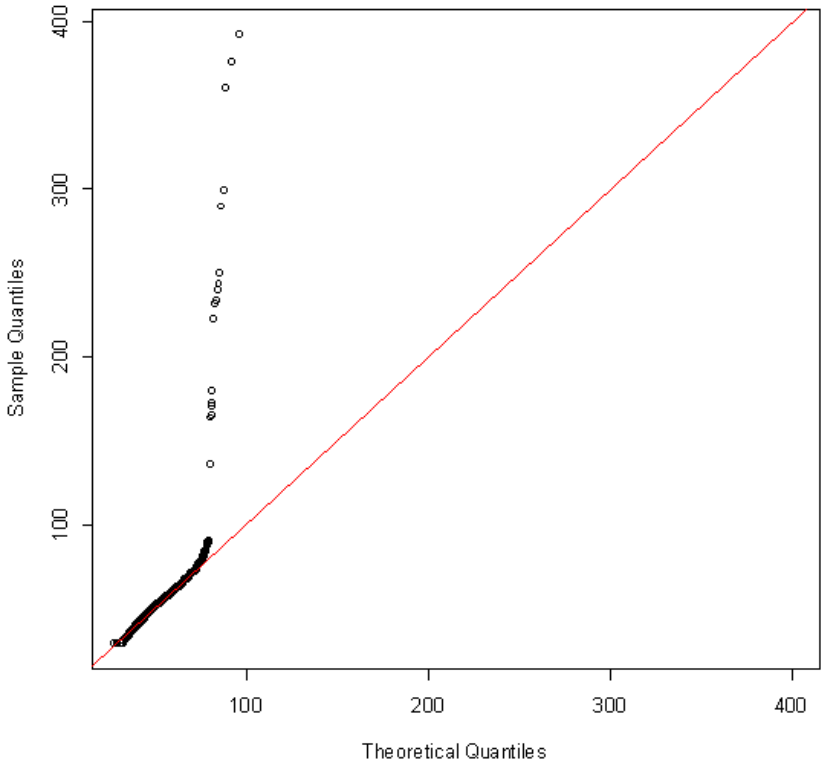


Figure 37. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.916 and the corresponding p-value is $10^{-4}$.

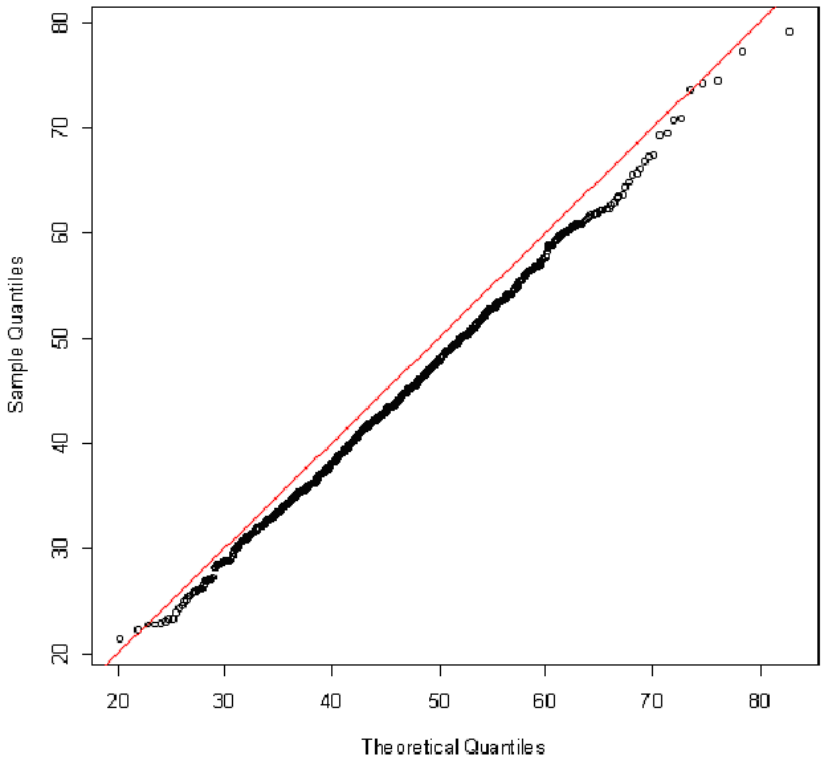


Figure 38. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.936 and the corresponding p-value is $10^{-4}$.

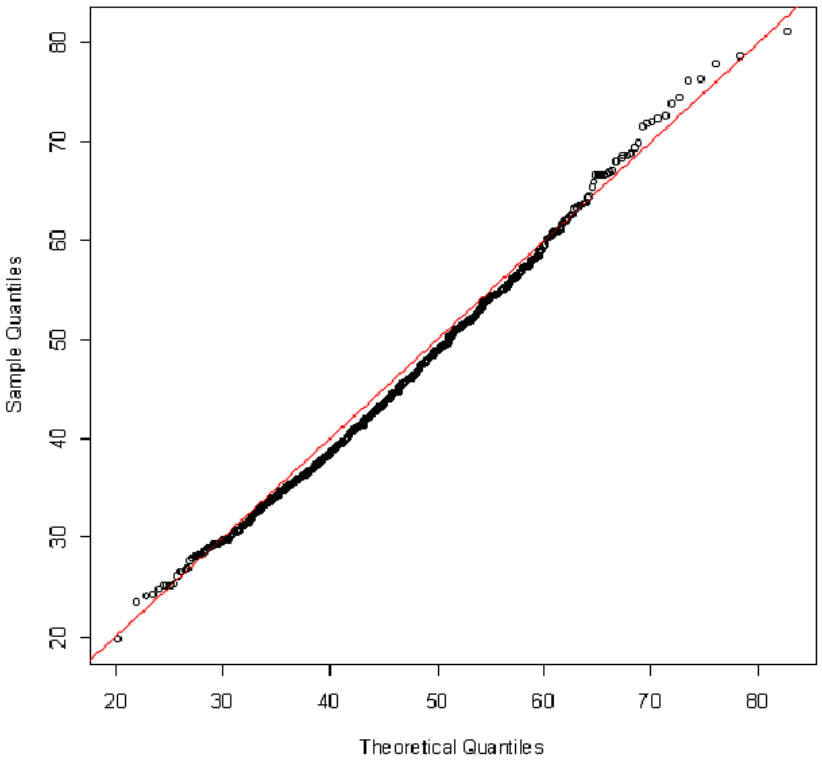


Figure 39. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 750$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.940 and the corresponding p-value is 0.0058.

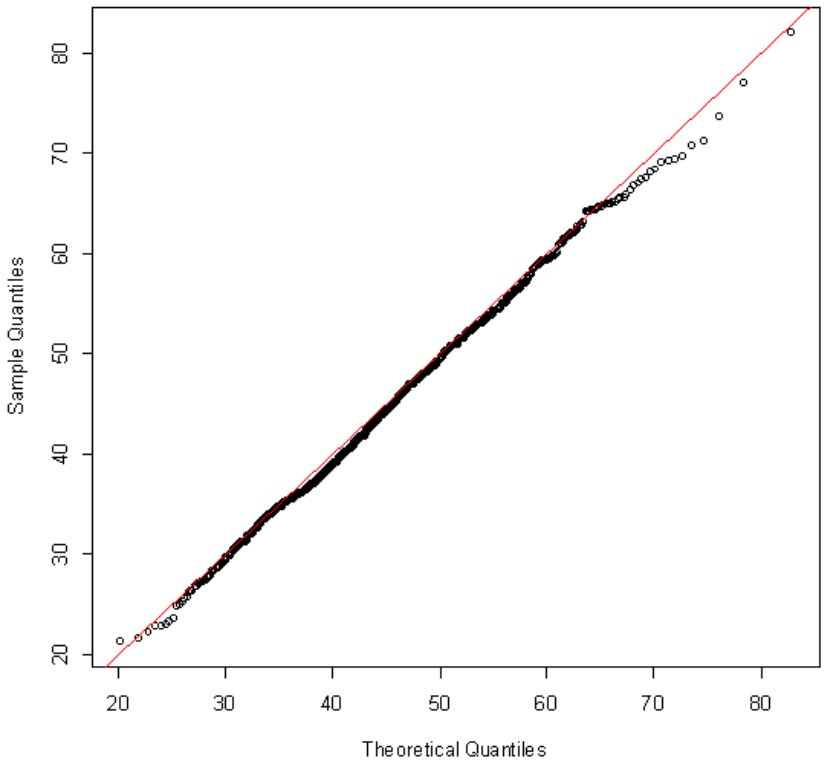


Figure 40. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 1000$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.953 and the corresponding p-value is 0.8063.

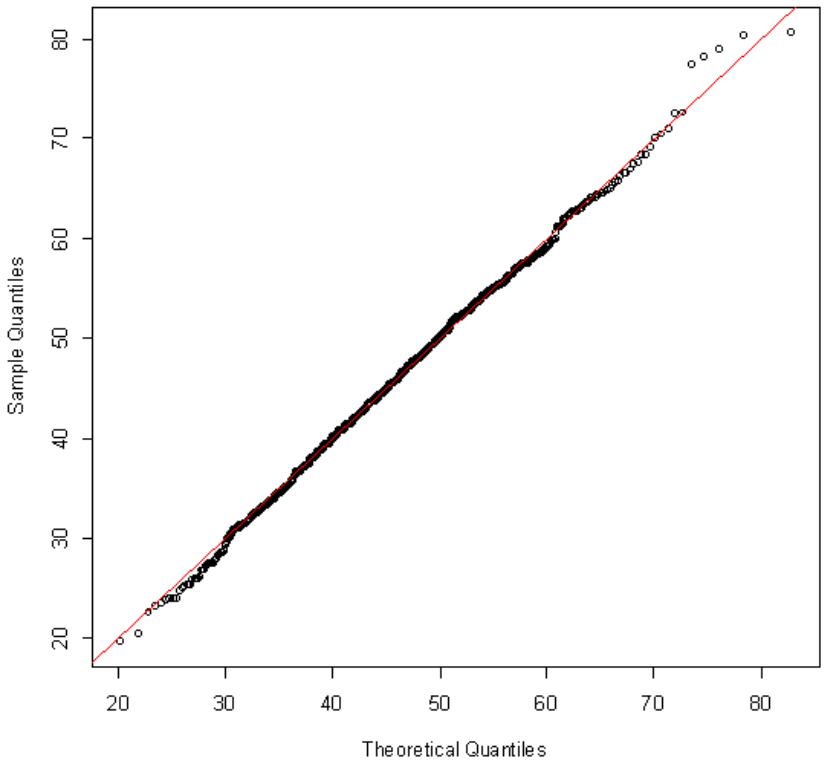


Figure 41. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.0$. The estimated slope in the QQ-plot is 0.958 and the corresponding p-value is 0.9946.

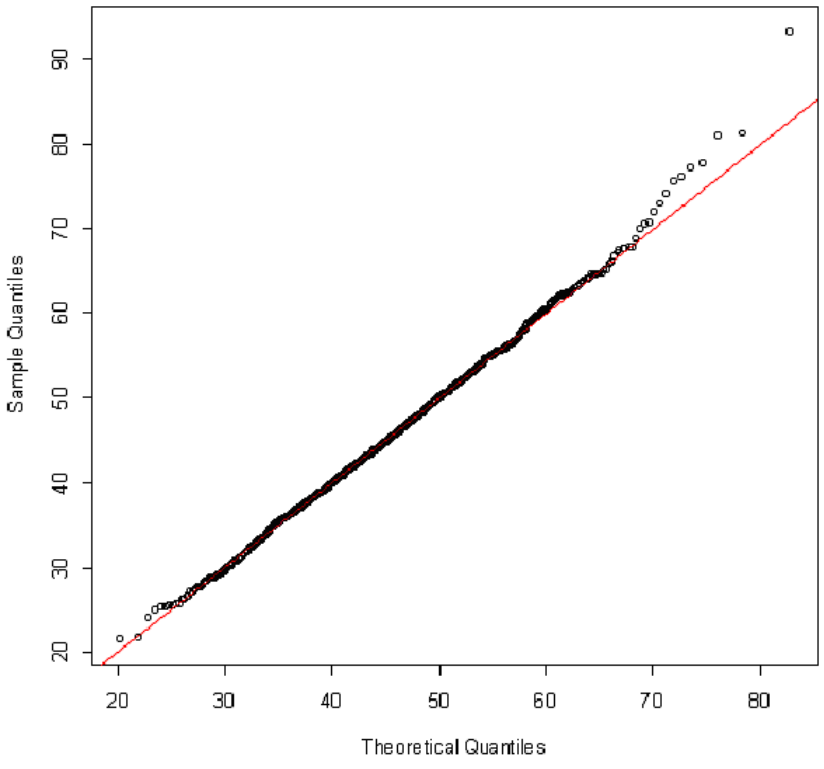


Figure 42. QQ-plot for $\chi^2_{[2]}$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.5$. The estimated slope in the QQ-plot is 0.978 and the corresponding p-value is 0.4398.

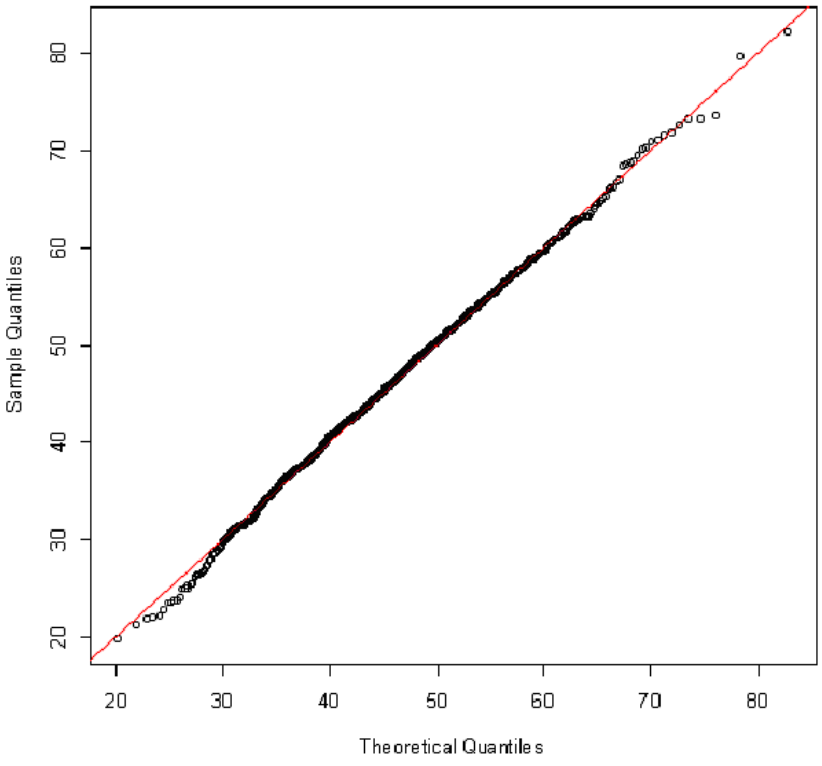


Figure 43. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.901 and the corresponding p-value is $10^{-4}$.

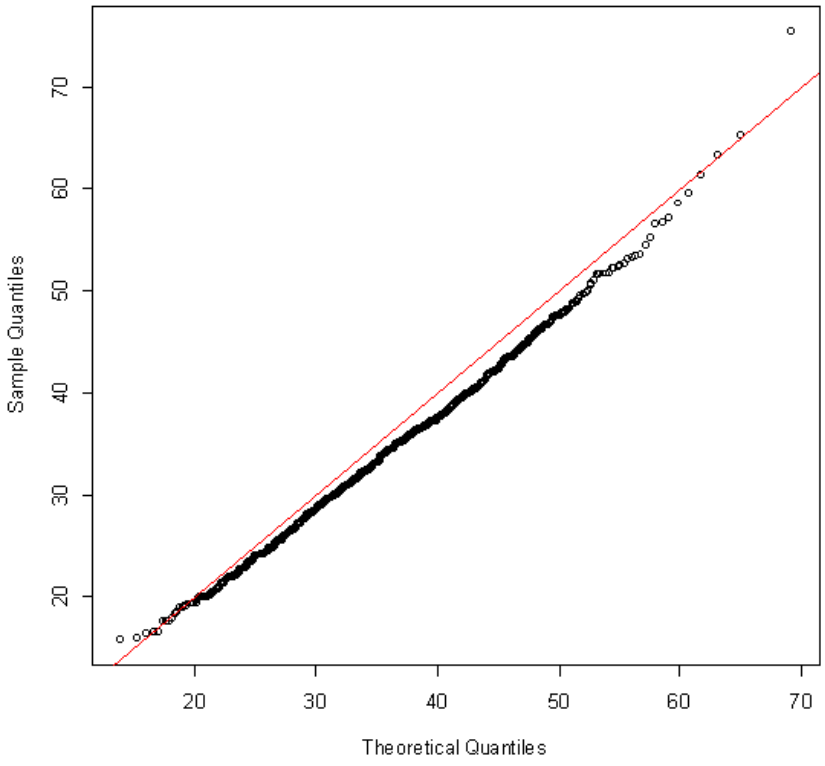


Figure 44. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 500$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.916 and the corresponding p-value is $10^{-4}$.

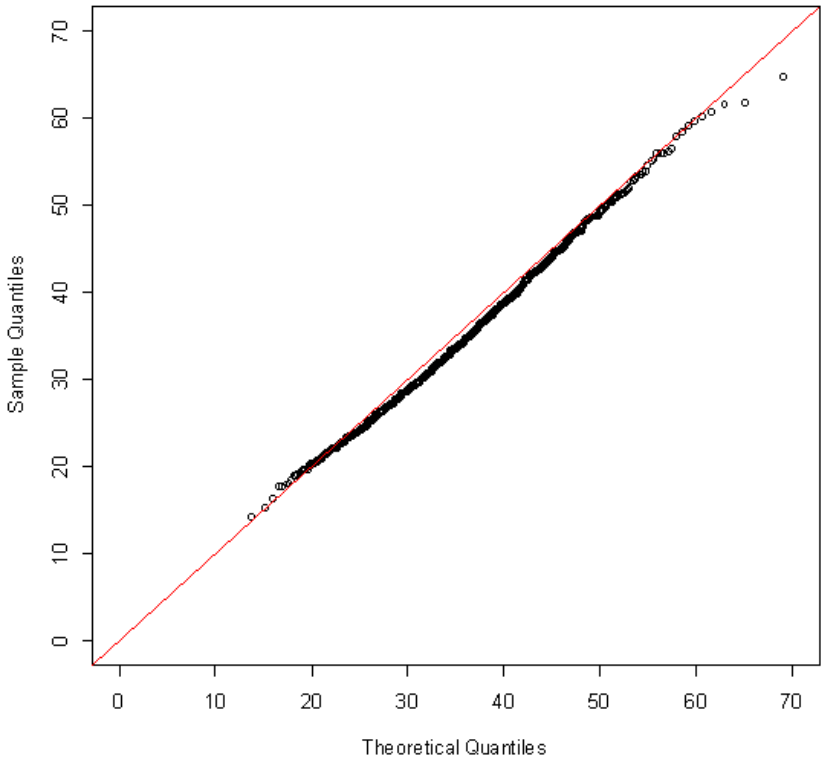


Figure 45. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 750$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.922 and the corresponding p-value is 0.0005.

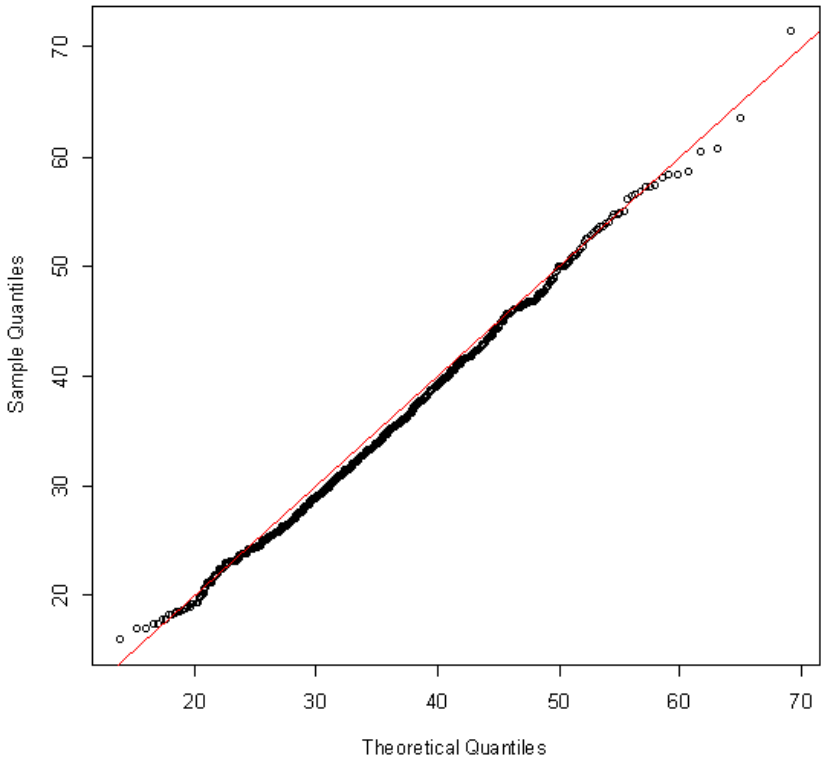


Figure 46. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 1000$ when $\bar{\beta}_1 = 0.5$. The estimated slope in the QQ-plot is 0.933 and the corresponding p-value is 0.0968.

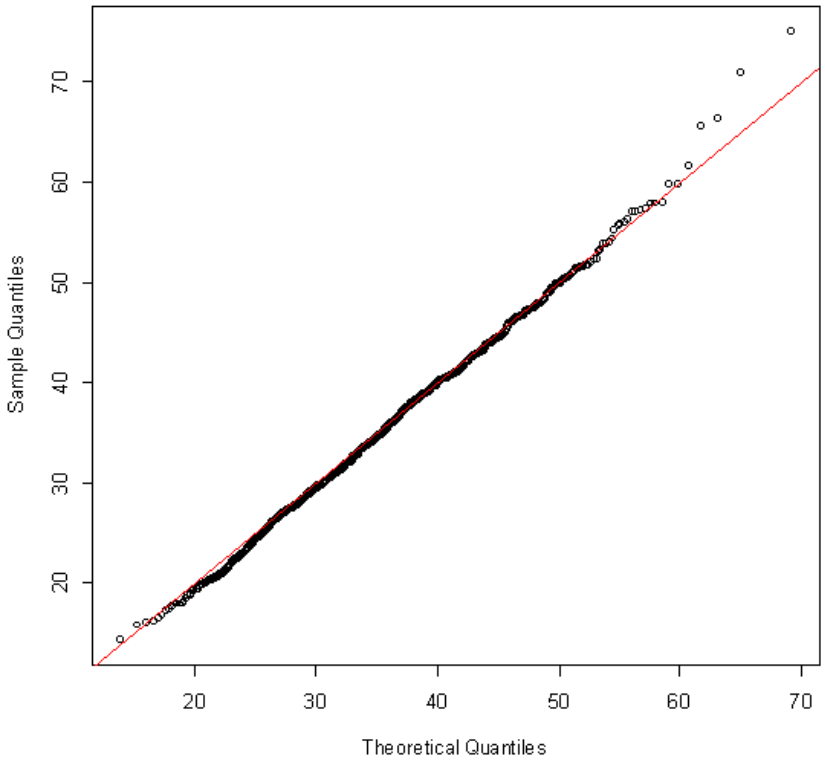


Figure 47. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.0$. The estimated slope in the QQ-plot is 0.945 and the corresponding p-value is 0.9880.

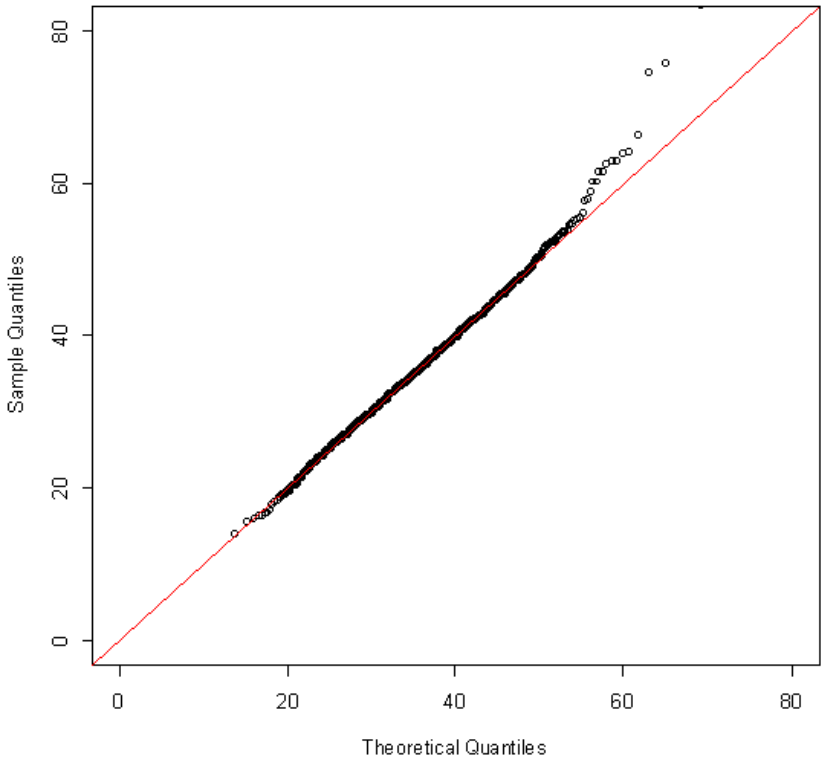
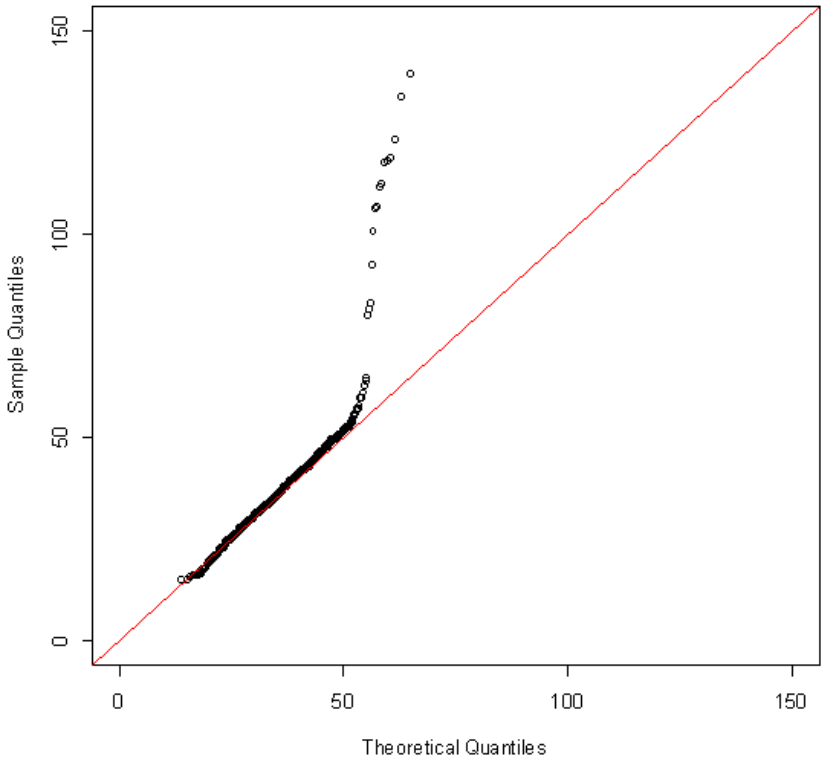


Figure 48. QQ-plot for $M_2$ when the model under the null hypothesis is the constrained version of the categorical variable factor model for 10 dichotomous variables for $n = 300$ when $\bar{\beta}_1 = 1.5$. The estimated slope in the QQ-plot is 0.970 and the corresponding p-value is 0.0537.



References


Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.

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

Bartholomew, D. J. and S. O. Leung (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. and P. Tzamourani (1999). A goodness of fit of a latent trait models in attitude measurement. *Sociological Methods and Research 27*, 525–546.

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

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics 25*, 290–302.

Breiman, L. (1996a). Bagging predictors. *Machine Learning 24*(2), 123–140.

Breiman, L. (1996b). Heuristics of instability and stabilization in model selection. *The Annals of Statistics 24*(6), 2350–2383.

Breiman, L. (1996c). Out-of-bag estimation. `ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps.Z`.

Breiman, L. (2000). Randomizing outputs to increase prediction accuracy. *Machine Learning 40*(3), 229–242.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.

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

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

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

Cramer, H. (1946). *Mathemartical Methods of Statistics*. Princeton, N. J: Princeton University Press.

Cressie, N. and P. W. Holland (1983). Characterizing the manifest probabilities of latent trait models. *Psychopmetrika 48*, 129–141.

Cressie, N. and T. Reed (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B 46*, 440–464.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning 40*(2), 139–157.

Emerson, P. L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics 24*, 695–701.

Eubank, R. L. (1997). Testing goodness of fit with multinomial data. *Journal of the American Statistical Association 92*, 1084–1093.

Eubank, R. L., V. N. LaRiccia, and R. B. Rosenstein (1987). Test statistics derived as components of pearson's phi-squared distance measure. *Journal of the American Statistical Association 82*, 816–825.

Fisher, R. A. (1941). *Statistical Methods for Researxh Workers, 8th edition*. Edinburgh: Oliver and Boyd, Ltd.

Goodnight, J. H. (1978). The sweep operator: Its importance in statistical computing. Technical report, Cary, NC.

Gooijer, J. G. D. and A. Yuan (2011). Exact tests for some latent traits. *Computational Statistics and Data Analysis 38*, 34–44.

Hall, P. (1985). Tailor-made tests of goodness of fit. *Journal of the Royal Statistical Society 109*, 125–131.

Inglot, T., T. Jurlewicz, and T. Ledwina (1990). On Neyman-type smooth tests of fit. *Statistics 21*, 549–568.

Institute for Public Policy and Michigan State University Social Research (2005). State of the state survey-38. spring 2005. http://www.ippsr.msu.edu/SOSS.

Ishwarn, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics 1*, 529–537.

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

Kendall, M. G. (1952). *The Advanced Theory of Statistics, Vol. 1 (5th edition)*. London: Griffin.

Knott, M. and P. Tzamourani (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*, pp. 224–252.

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

Lancaster, H. O. (1969). *The Chi-Squared Distribution*. New York: Wiley.

Ledwina, T. (1994). Data-driven version of neyman's smooth test of fit. *American Statistical Association 89*, 1000–1005.

Liaw, A. and M. Wiener (2002). Classification and regression by random forest. *R News 2*(3), 18–22.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus*. New York: Wiley.

Mathai, A. and S. Provost (1992). *Quadratic Forms in Random Variables: Theory and Applications*. New York: M. Dekker.

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

Maydeu-Olivares, A. and H. Joe (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.

Moses, E. L., J. D. Emerson, and H. Hosseini (1984). Analyzing data from ordered categories. *New England Journal of Medicine 311*, 442–448.

Moustaki, I. (2007). Assessing the goodness of fit of a latent variable model for ordinal data. Unpublished manuscript.

Rasch, G. (1960/1980). *Probabilistic Models for Some Intellegence and Attainment Tests.* (Conephagen, Danish Institute for Educational Research), exapanded edition (1980) with forward and afterword by B. D. Wright: Chicago: University of Chicago Press.

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

Rayner, J. C. and D. J. Best (1990). Smooth tests of goodness-of-fit: An overview. *International Statistical Review 58*, 9–17.

Reiser, M. (1996). Analysis of residulas 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. *Psychometrika*, 224–252.

Reiser, M. and G. Lin (1999). A goodnes-of-fit test for the latent class model when frequencies are small. *Sociological Methodology 39*, 81–111.

Salomaa, H. (1990). *Factor Analysis of Dichotomous Data.* Helsinki, Finland: Statistical Society.

Schuessler, K. F. (1982). *Measuring Social Life Feelings.* San Franciso: Jossey-Bass.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components.* New York: Wiley.

Segal, M. R. (2004). Machine learning benchmarks and random forest regression. `http://repositories.cdlib.org/cbmb/bench_rf_regn`.

Tate, M. W. and L. A. Hyer (1973). Inaccuracy of the chi-squared test of goodness of fit when expected freqencies are small. *Journal of the American Statistical Association 68*, 836–841.

Tjur, T. (1982). Type i error and power of the parametric bootstrap goodness-of-fit test: Full and limited-information. *British Journal of Mathematical and Statistical Psychology 56*, 271–288.

Tollenaar, N. and A. Mooijaart (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.

van der Laan, M. J. (2006). Statistical inference for variable importance. *International Journal for Biostatistics 2*(1). Article 2.

Wolpert, D. H. and W. G. McCready (1999). An efficient method to estimate bagging's generalization error. *Machine Learning 35*(1), 41–55.