July 2021

Model-Free Descriptive Modeling for Multivariate Categorical Data with An Ordinal Dependent Variable

Li Wang

University of Massachusetts Amherst

Model-Free Descriptive Modeling for Multivariate Categorical Data with An Ordinal Dependent Variable

A Dissertation Presented

by

LI WANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2021

Mathematics and Statistics

Model-Free Descriptive Modeling for Multivariate Categorical Data with An Ordinal Dependent Variable

A Dissertation Presented

by

LI WANG

Approved as to style and content by:

____

Daeyoung Kim, Chair

____

Krista Gile, Member

____

Anna Liu, Member

____

Jing Qian, Member

____

Nathaniel Whitaker, Department Head
Mathematics and Statistics

DEDICATION

This dissertation is dedicated to Professor Daeyoung Kim and my parents, Jinqiu Wang and Peimeng Tan, who supported me all the time in pursuit of statistics.

ACKNOWLEDGMENTS

First, my sincere gratitude goes to Professor Daeyoung Kim who taught me everything about research and guided me through the journey towards the completion of this dissertation. I am always your humble student.

Secondly, my deepest appreciation goes to my parents, Jinqiu Wang and Peimeng Tan, who tolerated the separation for more than 10 years since the first day I came to the U.S. as an undergraduate student and encouraged me to pursuit a Ph.D. degree with their unconditional and unreserved love. I hope you can always be proud of me.

Finally, I would like to give my thanks to Professor Krista Gile, Anna Liu and Jing Qian for their valuable comments and suggestions about my current and future work related to this dissertation.

ABSTRACT

Model-Free Descriptive Modeling for Multivariate Categorical Data with An Ordinal Dependent Variable

MAY 2021

Li WANG, B.S., IOWA STATE UNIVERSITY
M.S., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Daeyoung Kim

In the process of statistical modeling, the descriptive modeling plays an essential role in accelerating the formulation of plausible hypotheses in the subsequent explanatory modeling and facilitating the selection of potential variables in the subsequent predictive modeling. Especially, for multivariate categorical data analysis, it is desirable to use the descriptive modeling methods for uncovering and summarizing the potential association structure among multiple categorical variables in a compact manner. However, many classical methods in this case either rely on strong assumptions for parametric models or become infeasible when the data dimension is higher. To this end, we propose a model-free method for the descriptive modeling to delineate and quantify the association structure between an ordinal dependent variable and a set of categorical independent variables in a

multi-dimensional contingency table.

The proposed method consists of four components: subcopula score, subcopula regression, subcopula regression based association measure and its (sequential/non-sequential) decompositions. The subcopula score is a data-dependent scoring method for an ordinal variable reflecting the ordered nature of its categories. The subcopula regression leverages the subcopula scores to identify the association structure between the ordinal dependent variable and a set of categorical independent variables. The subcopula regression based association measure exploits the subcopula regression to quantify the strength of the association structure in a model-free manner. The sequential and non-sequential decompositions of the proposed association measure evaluate the contribution of the subsets of independent variables to the overall association in various forms such as marginal, conditional, interactive and correlative association.

We first study the theoretical properties of the subcopula score, subcopula regression, subcopula regression based association measure and its (sequential/non-sequential) decompositions. Next we develop the statistical inference for the proposed method including point estimation, (asymptotic/bootstrap) confidence intervals and permutation based hypothesis testing. Then we examine the finite-sample properties of the proposed overall, marginal and conditional association measures in multi-dimensional contingency tables. Finally, we demonstrate the potential use of the proposed method in real-world applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Ordinal data has been playing a fundamental role in many application areas. It has been widely adopted in social sciences for measuring attitudes and opinions. For example, each individual in a social science survey could be questioned for the opinion on whether tech companies respect or disrespect customers' privacy with the ordered categories such as the *Likert scale* (e.g. "strongly disagree", "disagree", "neural", "agree" and "strongly agree") (Likert, 1932). Ordinal data is also commonly found in the study of medicine and public health. For example, each subject in a medical test could be asked to report the severity of migraine before and after taking some drug with the categories such as "none", "mild", "moderate" and "severe". More scientific studies that heavily rely on ordinal data can be found in the areas including biology, physiology, psychology, to name a few.

In practice, one of the main interests of analyzing ordinal data is to identify the patterns of association between a set of dependent and independent variables. The observations on the ordinal dependent variables are usually collected with the observations on the categorical independent variables and a contingency table can be used to display the frequency (i.e the number of observations) for each combination of the categories of the variables. Various methods have been developed as the tools of statistical modeling for describing and measuring the association structure in the contingency tables with ordinal dependent variables and some well-known references include Agresti (2002), Agresti (2010), Liu and Agresti (2005), Beh and Lombardo (2014), Tutz (2011), Kateri (2014) and Harrell (2015).

Generally, the entire process of statistical modeling starts with purpose establishment, study design and data collection. Then it proceeds with data preparation, exploratory data analysis, variable selection, methodology determination, evaluation and validation, and model selection. Lastly, the whole process of statistical modeling ends with model application and report. Each of the existing methods for contingency tables with ordinal dependent variables introduced in the references above may be considered as one of the following three types of statistical modeling tools, depending on the research purpose: descriptive, explanatory and predictive modeling (Shmueli et al., 2010).

If the goal of the research is to explore and reveal potential association structures among the variables of interest in a compact manner, then one should perform the descriptive modeling to accurately capture the patterns of association in the data with minimal assumptions. If the purpose of the research is to develop and test causal theories, then one should conduct the explanatory modeling to justify how the factors measured by the set of independent variables are able to result in the outcomes measured in the dependent variables. Finally, if the intent of the research is to forecast the events of interest in the future, then one should adopt the predictive modeling to learn the underlying association structure of the data at hand and infer the values of the dependent variables based on a new set of values for the independent variables.

Although the (explanatory/predictive) modeling is the prime tool for the major steps (e.g. methodology determination, evaluation and validation) in the process of statistical modeling, its significance largely depends on the success of upstream steps especially exploratory data analysis. The descriptive modeling lies at the heart of exploratory data analysis consisting of various methods for visualization, dimension reduction and pattern identification. It plays a crucial role in facilitating the formulation of plausible hypotheses in the subsequent explanatory modeling or the selection of important variables in the subsequent predictive modeling. An exercise of the techniques mentioned above has been essential to many statistical learning problems nowadays because the number of variables

in the data sets for statistical learning is often as large as hundreds or thousands.

Several descriptive modeling techniques are available for multivariate continuous data, e.g. Pearson's correlation coefficient, principal component analysis, kernel density estimation, smoothing splines and so on. For the contingency tables with ordinal variables, the descriptive modeling can be the first and vital step to identify and quantify potential association structures among the variables of interest before model-based methods are employed in the formal modeling step. For two- and three-dimensional contingency tables with ordinal variables, a few methods are available, e.g. *Kendall's tau* (Kendall, 1938), *Spearman's rho* (Spearman, 1904) and *Goodman-Kruskal's tau* (Goodman and Kruskal, 1954), *Gray-Williams' index* (Gray and Williams, 1981) and *Marcotorchino's index* (Marcotorchino, 1984*a*,*b*,*c*).

To detect the potential association pattern for a contingency table, one may consider using a model-based approach incorporating reasonable hypotheses of (in)dependence and the predetermined scores for the ordinal variables. However, unlike a two-dimensional table, there are more than one possible hypotheses of (in)dependence for a multi-dimensional table. For example, for three, four and five-dimensional tables, there are 8, 113 and thousands of hierarchical log-linear models with all the main effects included, respectively, aside from non-hierarchical log-linear models (Guo and Thompson, 1989; Fienberg, 2007). Note that the number of possible models of independence (both mutual and conditional independence) in a 10-dimensional table is 3,475,978, which is smaller than the total number of different hierarchical log-linear models (Good, 1975). Given a large number of possible (hierarchical/non-hierarchical) parametric models for a multi-dimensional table, it would be very challenging to find a few subsets of parametric models that can be reasonably fit to the data.

Given the importance of the explanatory data analysis to the statistical modeling process, it is desirable to perform the descriptive modeling by making as few assumptions as possible. To our best knowledge, however, there is a lack of a purely data-driven and

non-model based method for the descriptive modeling in the analysis of multi-dimensional contingency tables. As a result, we propose a novel model-free method to identify and quantify the interesting patterns of association between an ordinal dependent variable and a set of categorical independent variables in a multi-dimensional contingency table. Here we mean “interesting patterns of association” by the non-identically distributed patterns associated with the dependent variable across different combinations of the categories of the independent variables. The proposed method can also be used for the prediction of the categories of the dependent variable and the variable selection for independent variables. Note that the proposed method is completely model-free in the sense that it doesn’t require any parametric assumptions or tuning parameters for studying the association structure in the multi-dimensional contingency table with an ordinal dependent variable. Such a model-free approach indeed empowers the method for efficiently and effectively discovering the complex pattern of association in the multi-dimensional contingency table, which is desirable for the descriptive modeling.

The rest of this dissertation will consist of eight chapters organized as follows. Chapter 2 will review existing scoring methods for ordinal variables and non-model based measures for the association between the (nominal/ordinal) categorical variables in two- and three-dimensional contingency tables. We will then discuss the limitations of existing association measures which will motivate the development of the proposed methodology in the dissertation.

Chapter 3 and 4 will propose the model-free descriptive modeling methodology for studying the association between the ordinal dependent variable and a set of categorical independent variables in a multi-dimensional contingency table. The proposed methodology will consist of three core components: subcopula scores, subcopula regression and subcopula regression based association measure.

Chapter 3 will formally define the subcopula scores and subcopula regression function, and investigate their theoretical properties. A subcopula score is a type of numerical

representation of each category in an ordinal variable. It is the marginal cumulative probability evaluated at each category of a variable, which reflects the natural order of the ordered categorical scale. Moreover, unlike the existing scoring methods, the subcopula scores provide a basis for coupling the joint distribution of the ordinal variables with their respective marginal distributions. The subcopula regression function leverages the subcopula scores to examine and capture the association structure between the ordinal dependent variable and the set of categorical independent variables in a model-free manner. It calculates the expectation of the subcopula scores of the dependent variable with respect to its conditional distribution given the set of independent variables and predicts the category of the dependent variable for a given combination of the categories of the independent variables. The subcopula scores and subcopula regression function will be illustrated by a running example.

Chapter 4 will define the overall subcopula regression based association measure for the full-dimensional contingency table and its sequential and non-sequential decompositions. The subcopula regression based association measure, extended from the work of Wei and Kim (2017) for two-dimensional contingency tables, is designed to represent the average proportion of variance for the subcopula scores of the ordinal dependent variable with respect to its marginal distribution explained by the categorical independent variables in the subcopula regression. Therefore it is able to quantify the association between the dependent variable and a set of the independent variables captured by the subcopula regression. In this chapter, we will study the theoretical properties of the proposed overall association measure. Note that it is applicable to not only ordinal independent variables but also nominal ones.

We will also propose the sequential and non-sequential decompositions of the overall subcopula regression based association measure to uncover the different types of contribution of each or a subset of independent variables of interest to the overall subcopula regression association measure based on the entire set of independent variables. To this

end, we will define the marginal, conditional, interactive and correlative association measures to quantify the explanatory power of a subset of independent variables of interest in a model-free manner. In addition, the theoretical properties of both decompositions will be investigated for three-dimensional contingency tables first and then they will be extended to multi-dimensional cases. They will also be illustrated by the same running example used in Chapter 3.

Chapter 5 will develop the statistical inference of the model-free descriptive modeling methodology proposed in Chapter 3 and 4. We will first present the estimation for the proposed subcopula score, subcopula regression function and its application to the prediction of the categories of the dependent variable. Then both of the point and interval estimation will be provided for the overall subcopula regression based association measure, and the marginal, conditional, interactive and correlative association measures appearing in the sequential and non-sequential decompositions of the overall association measure proposed in Chapter 4. Furthermore, we will propose using the permutation test to assess the statistical significance of the estimates of the overall association measure and its decompositions of interest.

Chapter 6 will conduct simulation studies to examine the finite-sample performance of the proposed overall association measure and its decompositions under various conditions of multi-dimensional ordinal contingency table. We will also compare the performance of the overall association measure with an existing non-model based association measure in terms of the sensitivity to the number of categories in the dependent variable. It will be shown that the proposed association measure is insensitive to the increase of the number of categories in the dependent variables.

Chapter 7 will demonstrate the applicability of the proposed model-free methods to real data sets with different focuses. First, the intent of the use of the two-dimensional *ice cream* data set (*The Ice Cream Study at Penn State*, 2012) is to demonstrate the capability of the proposed method to identify and quantify the non-monotone (i.e. quadratic)

association between two ordinal variables. Next, the purpose of the adoption of the three-dimensional *acute migraine* data set (Vandenhende and Lambert, 2000) is to evaluate the ability of the proposed method to measure the strength of the overall, marginal, conditional, interactive and correlative association between the dependent variable and two independent variables through the proposed association measures. Similarly, the goal of the analysis of the *nuclear accident* data set (Fienberg et al., 1985) is to further assess the ability of the proposed method to identify and quantify a potential time-dependent association structure in a five-dimensional contingency table. Such capabilities will be confirmed respectively by comparing the results discovered by the proposed model-free method to those found by the existing parametric models designed exclusively for the *acute migraine* and *nuclear accident* data sets. Finally, the purpose of applying the proposed method to the nine-dimensional *post-operative patients* data set (Budihardjo et al., 1991) is to display the viability of the proposed method on the variable selection problem for a multi-dimensional contingency table which is regarded critical in statistical learning.

Finally, Chapter 8 will summarize the contribution of this dissertation and discuss future research directions.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we will first present some commonly used scoring methods to numerically represent the ordered categories of ordinal variables for subsequent statistical analysis. Then we will review several classical non-model based association measures for two- and three-dimensional contingency tables, and discuss their limitations.

2.1 Score assignment for ordinal variable

In the analysis of ordinal data, it is well known that utilizing the ordering information in the ordinal variables results in more powerful inferences than ignoring it. Thus, one of the primary steps towards the analysis of ordinal data is the conversion of the categories of an ordinal variable into numerical scores that reflect the natural ordering of the categories. Note that a scoring method is considered to be a basic tool to describe marginal and conditional distributions of ordinal variables.

Many scoring methods have been proposed to systematically assign numerical scores to the categories of an ordinal variable. In particular, they can be classified into three types: preassigned scores, distribution-based scores, and optimal scores (Thomas and Kiwanga, 1993).

Methods of preassigned scores are simple but often involve some degree of arbitrariness or subjectiveness in score assignment. One of the commonly used preassigned scoring method is the *equal-space* score assigning $\{1, \ldots, I\}$ to the I categories of an ordinal vari-

able and it is considered to be an adequate system of preassigned scores for the purpose of data exploration (Box and Jones, 1986). However, if the inter-category distances vary throughout all the categories, additional cares must be taken in formal statistical modeling.

Methods of distribution-based scores rely on the proportions of an ordinal variable or its pre-assumed underlying continuous distribution. They are usually preferred to the methods of preassigned scores when the inter-category distances are different. Bross (1958) introduced a data-driven scoring method by averaging the cumulative proportions of two adjacent categories of an ordinal variable, which is called the *Ridit scores*. To be specific, let X be an ordinal variable with I categories. The i-th category of X is denoted by x_i with $p_i = P(X = x_i)$ for $1 \leq i \leq I$. Then the *Ridit score* for the i-th category of X is defined to be

$$s_i^R = \sum_{k=1}^{i-1} p_k + \frac{1}{2} p_i = \frac{F_{i-1} + F_i}{2},$$

where $F_0 = 0$, $F_I = 1$ and $F_i = P(X \leq x_i)$ is the cumulative proportion at x_i. Note that the *Ridit scores* have the same ordering as the original categories $x_1, \ldots, x_I$ and its expectation with respect to $p_1, \ldots, p_I$ is $\sum_{i=1}^{I} s_i^R p_i = 0.5$. In addition, there is a linear relationship between the *Ridit score* and another widely used scoring method called *midrank* (Agresti, 2010, p. 11):

$$s_i^M = n s_i^R + 0.5,$$

where n is the total number of observations and s_i^M is the *midrank* for the i-th category of X given as

$$s_i^M = \frac{(\sum_{k=1}^{i-1} n p_k + 1) + \sum_{k=1}^{i} n p_k}{2}.$$

Note that, if two adjacent categories x_{i-1} and x_i are combined into a new category x_{i^*},

then the *Ridit score* and *midrank*, $s_{i^*}^R$ and $s_{i^*}^M$, for the new category are equal to $s_i^R - \frac{1}{2}p_{i-1}$ and $s_i^M - \frac{n}{2}p_{i-1}$ while the *Ridit scores* and *midranks* for the other categories of X remain unchanged. In addition, if the ordering of the categories of X is reversed, then the *Ridit score* for the i-th category of X becomes $1 - s_i^R$.

In spite of the comprehensible definitions and properties of the *Ridit score* and *midrank*, Mantel (1963) and Chen and Wang (2014) argued that when the adjacent categories of an ordinal variable have low probabilities of occurrence, they will receive the similar *Ridit score* or *midrank* even if they may represent distinct states. In particular, Chen and Wang (2014) pointed out that *Ridit score* or *midrank* may degrade further statistical analysis if the ordinal variable has a skewed and/or imbalanced distribution (i.e. the observations in some categories greatly outnumber those in the others). Specifically, the *Ridit scores* or *midranks* for the categories with few observations tend to be indistinguishable.

One way to relieve this issue is to assume a pre-determined latent continuous distribution underlying the ordinal variable and then assign scores to the categories through the inverse of the distribution function. As a consequence, the scores for the categories with few observations in a (skewed/imbalanced) distribution can be more separated. For example, suppose that $\Phi(\cdot)$ is the cumulative distribution function (c.d.f.) of the standard normal distribution underlying an ordinal variable X and let s_i^A denote the score for the i-th category of X obtained from the assumed latent standard normal distribution. Then $s_i^A = \Phi^{-1}(s_i^R)$ or $s_i^A = \Phi^{-1}(s_i^M)$ is derived from the *Ridit score* or *midrank* for the i-th category of X (Agresti, 2010, p.11).

To generalize the idea of using the underlying continuous distribution in the score assignment for ordinal variables, Fielding (1993) and Chen and Wang (2014) invented similar scoring methods based on the conditional mean of a category of the ordinal variable X by assuming it is generated from some continuous distribution whose parameters are known. Let G denote the latent c.d.f. of X and g is the corresponding probability density function (p.d.f.). Then the respective scores for the i-th category of X proposed in Fielding

(1993) and Chen and Wang (2014), denoted by s_i^F and s_i^C, are defined to be

$$s_i^F = s_i^C = E(X^*|x_{i-1} < X^* \le x_i) = \frac{\int_{G^{-1}(F_{i-1})}^{G^{-1}(F_i)} wg(w)dw}{p_i} = \frac{\int_{F_{i-1}}^{F_i} G^{-1}(w)dw}{p_i}.$$

where X^* is the continuous latent variable underlying the ordinal variable X with the finite $E(X^*)$. Note that, if X^* is the continuous uniform random variable over $[0,1]$, then $s_i^F = s_i^C = s_i^R$ for the i-th category, x_i, of X.

Methods of optimal scores determine the numerical value assigned to each category of an ordinal variable by minimizing the difference between the proportions of an ordinal variable and its pre-specified latent distribution. Note that they can incorporate the other two types of scoring methods. For example, Brockett (1981) designed a method named *conditional median score*. Let G and H denote the latent cumulative distribution function with known parameters and the empirical cumulative distribution function of X such that $H(s_i^B) = F_i$, respectively. Then the score for the i-th category of X, $s_i^B = G^{-1}(s_i^R)$ minimizes the distance $d(H,G) = \max_i |H(s_i^B) - G(s_i^B)|$, where s_i^R is the *Ridit score* for the i-th category of X.

The existing scoring methods reviewed above share a common characteristic that they determine the scores for an ordinal variable only in terms of its observable proportions and the assumed latent marginal distribution without considering its relationship with the other variables. In contrast, the score assignment method, proposed and discussed in Chapter 3, will take into account the joint relationship among multiple categorical variables and hence produce the scores of ordinal variables which will be a set of basis for uniquely connecting the multivariate joint distribution of multiple ordinal variables and their marginal distributions.

2.2 Non-model based measures of association

For the purpose of the descriptive analysis, it is essential to explore and identify the association structure among the categorical variables using non-model based summary measures. In this section, several well-known non-model based association measures will be reviewed for two- and three-dimensional contingency tables with symmetrical and non-symmetrical association structures, and the limitations of those measures will be discussed. Note that by non-symmetrical association we mean that one variable is considered as the dependent variable and the other(s) are treated as the independent variable(s), and by symmetrical association we mean the mutually dependent relationship, not the dependent-independent one.

2.2.1 Measures of association for two-dimensional contingency tables

In literature, several well-known non-model based measures were proposed to identify and measure the symmetrical and non-symmetrical association structure among the categorical variables in a two-dimensional contingency table, including *Pearson's chi-square statistic* (Pearson, 1900), *Goodman-Kruskal's gamma* (Goodman and Kruskal, 1954), *Kendall's tau-a* (Kendall, 1938), *tau-b* (Kendall, 1945), *Spearman's rho* (Spearman, 1904), Goodman-Kruskal's tau (Goodman and Kruskal, 1954), *lambda* (Goodman and Kruskal, 1954), *Theil's uncertainty coefficient* (Theil, 1970) and *Somers' D* (Somers, 1962).

To clearly present these measures in the following subsections, we define a two-dimensional contingency table as follows. Suppose that $X_1 = \{x_{1_1}, \ldots, x_{1_I}\}$ and $X_2 = \{x_{2_1}, \ldots, x_{2_J}\}$ are two categorical variables in a two-dimensional contingency table with I and J categories, respectively. Let $p_{ij} = P(X_1 = x_{1_i}, X_2 = x_{2_j})$ and n_{ij} denote the joint probability mass function (p.m.f.) and the frequency associated with the i-th and j-th category of X_1 and X_2, (x_{1_i}, x_{2_j}), where $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

2.2.1.1 Measures of symmetrical association

Pearson (1900) proposed the famous *Pearson's chi-square statistic* for identifying the existence of symmetrical association among the variables in a two-dimensional contingency table:

$$\chi^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - p_i p_j)^2}{p_i p_j}.$$

where $p_i = \sum_{j=1}^{J} p_{ij} = P(X_1 = x_{1_i})$, $p_j = \sum_{i=1}^{I} p_{ij} = P(X_2 = x_{2_j})$ and $n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$. Under the assumption that X_1 and X_2 are independent, i.e. $p_{ij} = p_i p_j$ for every pair of (x_{1_i}, x_{2_j}), χ^2 follows asymptotically the chi-square distribution with $(I-1)(J-1)$ degrees of freedom as $n \to \infty$. In addition, χ^2 ranges from 0 to $n[\min(I, J) - 1]$ (Liebetrau, 1983, p.13), and it is equal to 0 if and only if X_1 and X_2 are independent.

Besides detecting the existence of association among categorical variables, *Pearson's chi-square statistic* also provides a critical basis for the development of other association measures to quantify the deviation of the dependence from the independence, such as *the log-likelihood ratio statistic* G^2 (Wilks, 1935), *Freeman-Tukey statistics* T^2 (Freeman and Tukey, 1950), *Neyman's modified chi-square statistic* N^2 (Neyman, 1949) and *the modified log-likelihood ratio statistic* M^2 (Kullback, 1997):

$$G^2 = 2n \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \ln\left(\frac{p_{ij}}{p_i p_j}\right), \quad T^2 = 4n \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\sqrt{p_{ij}} - \sqrt{p_i p_j}\right)^2,$$
$$N^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - p_i p_j)^2}{p_{ij}}, \quad M^2 = 2n \sum_{i=1}^{I} \sum_{j=1}^{J} p_i p_j \ln\left(\frac{p_i p_j}{p_{ij}}\right).$$

Like *Pearson's chi-square statistic*, each of the statistics above follow asymptotically the chi-square distribution with $(I-1)(J-1)$ degrees of freedom, under the assumption of independence (Beh and Lombardo, 2014, p.48). In addition, each statistic above is equal to 0 if and only if X_1 and X_2 are independent. Furthermore, Cressie and Read (1984)

proposed the *power divergence statistic* which unifies all the association measures above:

$$CR(\lambda) = \frac{2n}{\lambda(\lambda+1)} \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \left[\left(\frac{p_{ij}}{p_i p_j} \right)^{\lambda} - 1 \right].$$

It is easy to see that $CR(1) = \chi^2$, $CR(0) = G^2$, $CR(-0.5) = T^2$, $CR(-1) = M^2$ and $CR(-2) = N^2$. Note that $CR(\lambda)$ also follows asymptotically the same chi-square distribution with $(I-1)(J-1)$ degrees of freedom, regardless of λ (García-Pérez and Núñez-Antón, 2009). Cressie and Read (1984) suggested $\lambda \in [0, 3/2]$ for lack of knowledge of the true association structure among the categorical variables.

There are other well-known measures for symmetrical associations including *Goodman Kruskal's gamma* (Goodman and Kruskal, 1954), *Kendall's tau-a* (Kendall, 1938), *tau-b* (Kendall, 1945) and *Spearman's rho* (Spearman, 1904). They are different from *Pearson's chi-square statistic* and related statistics reviewed above in that they are designed to measure the association between two ordinal variables and use either the probabilities of *concordance* and *discordance* between them or the scores for the categories in each ordinal variable.

Consider that X_1 and X_2 are the ordinal variables in a two-dimensional contingency table with the joint p.m.f. p_{ij} and the ordered categories for X_1 and X_2, $x_{1_1} < \cdots < x_{1_I}$ and $x_{2_1} < \cdots < x_{2_J}$. For two pairs (x_{1_i}, x_{2_j}) and $(x_{1_{i^*}}, x_{2_{j^*}})$, they are called *concordant* if $x_{1_i} < x_{1_{i^*}}$ $(x_{1_i} > x_{1_{i^*}})$ and $x_{2_j} < x_{2_{j^*}}$ $(x_{2_j} > x_{2_{j^*}})$ or *discordant* if $x_{1_i} < x_{1_{i^*}}$ $(x_{1_i} > x_{1_{i^*}})$ and $x_{2_j} > x_{2_{j^*}}$ $(x_{2_j} < x_{2_{j^*}})$. Note that they are called *tied* on X_1 only if $x_{1_i} = x_{1_{i^*}}$, on X_2 only if $x_{2_j} = x_{2_{j^*}}$, or on both X_1 and X_2 if $x_{1_i} = x_{1_{i^*}}$ and $x_{2_j} = x_{2_{j^*}}$. Then *Goodman Kruskal's gamma*, *Kendall's tau-a* and *tau-b* are defined by

$$\gamma^{GK} = \frac{q_c - q_d}{q_c + q_d}, \tau_a^K = \frac{q_c - q_d}{q_c + q_d + t_{x_1} + t_{x_2} + t_{x_1,x_2}}, \tau_b^K = \frac{q_c - q_d}{\sqrt{(q_c + q_d + t_{x_1})(q_c + q_d + t_{x_2})}}$$

where

$$q_c = \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} p_{ij}\left(\sum_{r=i+1}^{I}\sum_{t=j+1}^{J} p_{rt}\right), \quad q_d = \sum_{i=1}^{I-1}\sum_{j=2}^{J} p_{ij}\left(\sum_{r=i+1}^{I}\sum_{t=1}^{j-1} p_{rt}\right),$$
$$t_{x_1} = \sum_{i=1}^{I}\sum_{j=1}^{J-1} p_{ij}\left(\sum_{t=j+1}^{J} p_{it}\right), \quad t_{x_2} = \sum_{i=1}^{I-1}\sum_{j=1}^{J} p_{ij}\left(\sum_{r=i+1}^{I} p_{rj}\right),$$
$$t_{x_1,x_2} = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}\left(p_{ij} - \frac{1}{n}\right)$$

are the probabilities of *concordance*, *discordance*, *ties* on X_1 only, *ties* on X_2 only and *ties* on both X_1 and X_2, respectively. Hence, γ^{GK} can be interpreted to measure the probabilistic excess of *concordance* over *discordance* among the total of them (Somers, 1962). Note that τ_a^K and τ_b^K are different from γ^{GK} by adjusting the total of *concordance* and *discordance* with the *ties* on X_1, X_2 or both. Furthermore, the association measures γ^{GK}, τ_a^K and τ_b^K range from -1 to 1. *Goodman Kruskal's gamma* and *Kendall's tau-b* are equal to -1 or 1 when X_2 and X_1 has a perfectly negative (i.e. $q_c = t_{x_1} = t_{x_2} = 0$) or perfectly positive relationship (i.e. $q_d = t_{x_1} = t_{x_2} = 0$), respectively. However, *Kendall's tau-a* cannot attain -1 or 1, because it is not adjusted for the ties on both X_1 and X_2 (Agresti, 2010, p. 189). Finally, when X_2 and X_1 are independent (and so $q_c = q_d$), they are all equal to 0.

Spearman (1904) proposed *Spearman's rho*, an analogy of *Pearson's r* (Pearson, 1895), based on the *rank scores* for two ordinal variables:

$$\rho^S = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(s_{x_{1_i}}^K - \sum_{i=1}^{I} s_{x_{1_i}}^K p_i\right)\left(s_{x_{2_j}}^K - \sum_{i=j}^{J} s_{x_{2_j}}^K p_j\right) p_{ij}}{\sqrt{\left[\sum_{i=1}^{I}\left(s_{x_{1_i}}^K - \sum_{i=1}^{I} s_{x_{1_i}}^K p_i\right)^2 p_i\right]\left[\sum_{j=1}^{J}\left(s_{x_{2_j}}^K - \sum_{j=1}^{J} s_{x_{2_j}}^K p_j\right)^2 p_j\right]}},$$

where $s_{x_{1_i}}^K$ and $s_{x_{2_j}}^K$ are the respective *rank scores* for the i-th and j-th category of X_1 and X_2. Kendall (1948) proposed an analogy of *Spearman's rho* based on the *Ridit scores*:

$$\rho^K = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(s_{x_{1_i}}^R - 0.5\right)\left(s_{x_{2_j}}^R - 0.5\right) p_{ij}}{\sqrt{\left[\sum_{i=1}^{I}\left(s_{x_{1_i}}^R - 0.5\right)^2 p_i\right]\left[\sum_{j=1}^{J}\left(s_{x_{2_j}}^R - 0.5\right)^2 p_j\right]}},$$

where $s^R_{x_{1_i}}$ and $s^R_{x_{2_j}}$ are the respective *Ridit scores* for the i-th and j-th category of X_1 and X_2. Both association measures range from -1 to 1, and they are equal to -1 or 1 when X_2 and X_1 has a perfectly negative or positive monotone relationship, respectively. In addition, when X_2 and X_1 are uncorrelated, they are both equal to 0.

2.2.1.2 Measures of non-symmetrical association

Suppose one is interested in the dependent-independent association between two categorical variables X_2 and X_1 in a two-dimensional contingency table with the joint p.m.f. p_{ij}, where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Goodman and Kruskal (1954) proposed *Goodman-Kruskal's tau* to measure the relative (increase/decrease) in the proportion of (correct/incorrect) predictions in X_2 given X_1:

$$\tau^{GK}_{X_2} = \frac{(1 - \sum_{j=1}^J p_j^2) - (1 - \sum_{i=1}^I \sum_{j=1}^J p_{ij}^2/p_i)}{1 - \sum_{j=1}^J p_j^2} = \frac{\sum_{i=1}^I \sum_{j=1}^J p_i(p_{ij}/p_i - p_j)^2}{1 - \sum_{i=1}^I p_j^2}.$$

Here $1 - \sum_{j=1}^J p_j^2$ is the probability of incorrectly predicting the categories of X_2, and $1 - \sum_{i=1}^I \sum_{j=1}^J p_{ij}^2/p_i$ is the probability of incorrectly predicting the categories of X_2 given those of X_1. Note that, under the assumption that X_2 and X_1 are independent, $(n-1)(J-1)\tau^{GK}_{X_2}$ follows asymptotically the chi-square distribution with $(I-1)(J-1)$ degrees of freedom (Light and Margolin, 1971). Moreover, $\tau^{GK}_{X_2}$ ranges from 0 to 1. When X_2 and X_1 are independent, then $\tau^{GK}_{X_2} = 0$, and when X_2 is a function of X_1, e.g. $p_{ij} = p_i$ for every (x_{1_i}, x_{2_j}), then $\tau^{GK}_{X_2} = 1$.

Goodman-Kruskal's lambda (Goodman and Kruskal, 1954) is another non-symmetrical association measure:

$$\lambda^{GK}_{X_2} = \frac{\Big(1 - \max_j p_j\Big) - \Big(1 - \sum_{i=1}^I \max_j p_{ij}\Big)}{1 - \max_j p_j} = \frac{\sum_{i=1}^I \max_j p_{ij} - \max_j p_j}{1 - \max_j p_j}.$$

where $1 - max_j p_j$ and $1 - \sum_{i=1}^I max_j p_{ij}$ represent the probabilities of making mistakes in guessing the category of X_2 without and with knowing X_1. The range of $\lambda^{GK}_{X_2}$ is between 0

and 1. Furthermore, if X_2 and X_1 are independent, i.e. $p_{ij} = p_i p_j$ for every (x_{1_i}, y_{2_j}), then $\lambda_{X_2}^{GK} = 0$, and if X_2 is a function of X_1, e.g. $p_{ij} = p_i$ for every $(x_{1_i} y_{2_j})$, then $\lambda_{X_2}^{GK} = 1$.

Rather than quantifying the relative reduction in the errors of predicting the dependent variable, Theil (1970) proposed *Theil's uncertainty coefficient* to measure the proportional reduction in the entropy of a dependent variable:

$$UC_{X_2} = -\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \log(p_{ij}/p_i p_j)}{\sum_{j=1}^{J} p_j \log p_j},$$

where $\sum_{j=1}^{J} p_j \log p_j$ is the entropy of X_2 measuring the level of uncertainty in X_2 itself. *Theil's uncertainty coefficient* also ranges from 0 to 1. If X_2 and X_1 are independent, $UC_{X_2} = 0$, and if X_2 is a function of X_1, i.e. $p_{ij} = p_i$ for every (x_{1_i}, x_{2_j}), then $UC_{X_2} = 1$.

Although *Goodman-Kruskal's tau*, *lambda* and *Theil's uncertainty coefficient* can be used to assess the non-symmetrical association structure with ordinal variables, they do not explicitly incorporate the fact that the categories of the variables are ordered. To take into account the ordering information, Somers (1962) invented *Somers' D* for a two-dimensional contingency table with the ordinal dependent and independent variable, X_2 and X_1, respectively:

$$d_{X_2} = \frac{q_c - q_d}{q_c + q_d + t_{x_2}},$$

where q_c, q_d and t_{x_2} are the probabilities of *concordance*, *discordance*, and *ties* on the dependent variable X_2. Like *Kendall's tau-b*, *Somers' D* can also be interpreted to measure the probabilistic excess of *concordance* over *discordance* among the total of them. However, it becomes non-symmetrical by adjusting the total of *concordance* and *discordance* with the *ties* on X_2 only. Moreover, d_{X_2} also ranges from -1 to 1, and it is equal to -1 or 1 when X_2 and X_1 has a perfectly negative (i.e. $q_c = t_{x_2} = 0$) or perfectly positive relationship (i.e. $q_d = t_{x_2} = 0$), respectively. When X_2 and X_1 are independent (i.e. $q_c = q_d$), it is equal to 0.

2.2.2 Measures of association for three-dimensional contingency tables

For a three-dimensional contingency table, three notable non-model based measures were proposed to capture the symmetrical and non-symmetrical association structure among the categorical variables including *Pearson's mean square contingency coefficient* (Carlier and Kroonenberg, 1996), *Gray-Williams' index* (Gray and Williams, 1981) and *Marcotorchino's index* (Marcotorchino, 1984*a*,*b*,*c*).

For a clear presentation of these measures in the subsections, we define a three-dimensional contingency table as follows. Let $X_1 = \{x_{1_1}, \ldots, x_{1_I}\}$, $X_2 = \{x_{2_1}, \ldots, x_{2_J}\}$ and $X_3 = \{x_{3_1}, \ldots, x_{3_K}\}$ be the three categorical variables in a three-dimensional contingency table with the joint p.m.f. $P(X_1 = x_{1_i}, X_2 = x_{2_j}, X_3 = x_{3_k}) = p_{ijk}$, where $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$.

2.2.2.1 Measures of symmetrical association

For the symmetrical association among X_1, X_2 and X_3, *Pearson's mean square contingency coefficient* (Carlier and Kroonenberg, 1996) is considered to be the extension of *Pearson's chi-square statistic* for two-dimensional contingency tables:

$$\Phi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(p_{ijk} - p_i p_j p_k)^2}{p_i p_j p_k},$$

where $p_i = \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk}$, $p_j = \sum_{i=1}^{I} \sum_{k=1}^{K} p_{ijk}$ and $p_k = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk}$ are the marginal p.m.f.s of X_1, X_2 and X_3, respectively. Under the assumption that X_1, X_2 and X_3 are independent, i.e. $p_{ijk} = p_i p_j p_k$ for every $(x_{1_i}, x_{2_j}, x_{3_k})$, $n\Phi^2$ follows asymptotically the chi-square distribution with $IJK - I - J - K + 2$ degrees of freedom (Roy and Mitra, 1956; Carlier and Kroonenberg, 1996). Note that when X_1, X_2 and X_3 are mutually independent, i.e. $p_{ijk} = p_i p_j p_k$ for every $(x_{1_i}, x_{2_j}, x_{3_k})$, $\Phi^2 = 0$.

2.2.2.2 Measures of non-symmetrical association

For a three-dimensional contingency table with the dependent variable X_2 and the two independent variables (X_1, X_3), *Gray-Williams' index* (Gray and Williams, 1981) and *Marcotorchino's index* (Marcotorchino, 1984*a*,*b*,*c*) are defined as:

$$\tau_{X_2}^{GW} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} p_{ijk}^2/p_{ik} - \sum_{j=1}^{J} p_j^2}{1 - \sum_{j=1}^{J} p_j^2},$$
$$\tau_{X_2}^{M} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} p_{ijk}^2/p_j p_k - \sum_{j=1}^{J} p_j^2}{1 - \sum_{j=1}^{J} p_j^2}.$$

where $p_{ik} = \sum_{j=1}^{J} p_{ijk}$ is the marginal p.m.f. of (X_1, X_3). Note that $\tau_{X_2}^{GW}$ and $\tau_{X_2}^{M}$ are considered to be the generalizations of $\tau_{X_2}^{GK}$ designed for two-dimensional contingency tables, and $\tau_{X_2}^{M}$ is a special case of $\tau_{X_2}^{GW}$ by assuming that X_1 and X_3 are independent.

It is easy to see that *Gray-Williams' index* and *Marcotorchino's index* quantify the relative (increase/decrease) in the proportion of (correct/incorrect) predictions in X_2 given X_1 and X_3. Under the assumption that X_2 is independent of (X_1, X_3), i.e. $p_{ijk} = p_{ik}p_j$ for every $(x_{1_i}, x_{2_j}, x_{3_k})$, $(n-1)(J-1)\tau_{X_2}^{GW}$ follows asymptotically the chi-square distribution with $IJK - I - JK + 1$ degrees of freedom, as $n \to \infty$ (Lombardo and Camminatiello, 2010). Under the assumption of mutually independence, $(n-1)(J-1)\tau_{X_2}^{M}$ follows asymptotically the chi-square distribution with $IJK - I - J - K + 2$ degrees of freedom, as $n \to \infty$ (Beh et al., 2007). Moreover, $\tau_{X_2}^{GW}$ ranges from 0 to 1, and it is equal to 0 or 1 when X_2 is independent of (X_1, X_3), or is a function of (X_1, X_3), e.g. $p_{ijk} = p_{ik}$ for every $(x_{1_i}, x_{2_j}, x_{3_k})$, respectively. $\tau_{X_2}^{M}$ also ranges from 0 to 1, and it is equal to 0 or 1 when X_1, X_2 and X_3 are mutually independent, or is a function of X_1 and X_3, e.g. $p_{ijk} = p_i p_k$ for every $(x_{1_i}, x_{2_j}, x_{3_k})$, respectively.

2.2.3 Demand for novel model-free association measures for an ordinal dependent variable

Two major issues with *Goodman-Kruskal's tau*, *Theil's uncertainty coefficient*, as well as *Gray-Williams' index* and *Marcotorchino's index*, were reported in the literature. The first one, as mentioned in Section 2.2.1.2, is that they cannot explicitly take into account the ordinality in the categories of the ordinal variables, especially for the dependent variable. The second one is that they are sensitive to the number of categories in the dependent variable (Agresti, 2002; Beh and Lombardo, 2014, p. 182, 183, 461; Wei and Kim, 2017). That is, their values tend to be smaller as the number of categories of the dependent variable increases, regardless of the strength of association between the dependent variable and the independent ones. Thus, small values of those association measures do not necessarily indicate weak association given a large number of categories in the dependent variable. To examine the significance of the non-symmetrical association identified by those measures, one needs to conduct the asymptotic test of independence associated with the association measure (Light and Margolin, 1971; Beh et al., 2007). Note that the strength of the non-symmetrical association should be analyzed further when the association measure is small but statistically significant (Beh and Lombardo, 2014, p. 185).

To exploit the ordering of the categories of the ordinal dependent variable and overcome the sensitivity to the number of categories in the dependent variable as described above, Wei and Kim (2017) proposed a non-symmetrical association measure for two-dimensional contingency tables with an ordinal dependent variable X_2 and independent variable X_1:

$$\rho^2_{(X_1 \to X_2)} = \frac{\sum_{i=1}^{I} \left(\sum_{j=1}^{J} v_{2_j} p_{j|i} - \sum_{j=1}^{j} v_{2_j} p_j \right)^2 p_i}{\sum_{j=1}^{J} \left(v_{2_j} - \sum_{j=1}^{J} v_{2_j} p_j \right)^2 p_j},$$

where $v_{2_j} = \sum_{t=1}^{j} p_t$ and $p_{j|i} = p_{ij}/p_i$ is the conditional p.m.f. of X_2 given X_1. Here v_{2_j},

the cumulative proportion at the j-th category of X_2, can be viewed as the score assigned to the category x_{2_j}, and $\sum_{j=1}^{J} v_{2_j} p_{j|i}$ is the mean score for X_2 with respect to its conditional p.m.f. of X_2 given X_1.

Anderson and Landis (1982) proposed a similar association measure for two- and three-dimensional contingency tables with an ordinal dependent variable, X_2:

$$R^2_{(X_1 \to X_2)} = \frac{\sum_{i=1}^{I} \left(\sum_{j=1}^{J} a_{2_j} p_{j|i} - \sum_{j=1}^{j} a_{2_j} p_j \right)^2 p_i}{\sum_{j=1}^{J} \left(a_{2_j} - \sum_{j=1}^{J} a_{2_j} p_j \right)^2 p_j},$$

$$R^2_{(X_1, X_3 \to X_2)} = \frac{\sum_{i=1}^{I} \sum_{k=1}^{K} \left(\sum_{j=1}^{J} a_{2_j} p_{j|ik} - \sum_{j=1}^{j} a_{2_j} p_j \right)^2 p_{ik}}{\sum_{j=1}^{J} \left(a_{2_j} - \sum_{j=1}^{J} a_{2_j} p_j \right)^2 p_j},$$

where a_{2_j} is some score assigned to the j-th category of X_2.

Although the definitions of the two proposed measures, $\rho^2_{(X_1 \to X_2)}$ and $R^2_{(X_1 \to X_2)}$, appear to be similar, the motivation and formulation are completely different. First, $\rho^2_{(X_1 \to X_2)}$ aimed to address the sensitivity problem of existing association measures to the number of categories in the dependent variable, while quantifying the non-symmetrical association in a two-dimensional contingency table with an ordinal dependent variable. On the other hand, $R^2_{(X_1 \to X_2)}$ was originally proposed as a test statistic for testing hypotheses in the analysis of variance for ordinal data, in the same way as the analysis of variance in the regression model for continuous data. Secondly, the assignment of the scores v_{2_j} for X_2 was a part of the construction of $\rho^2_{(X_1 \to X_2)}$ and the effect of the selected scores of X_2 was taken into account in the examination of the finite-sample and asymptotic properties of $\rho^2_{(X_1 \to X_2)}$ (Wei and Kim, 2017). However, $R^2_{(X_1 \to X_2)}$ assumed that the scores for X_2 was specified by the user in advance (e.g. the equal-spaced integer scores, *Ridit scores*, *midranks*), independent of the construction of $R^2_{(X_1 \to X_2)}$. Note that the properties and the performance of $R^2_{(X_1 \to X_2)}$ with different types of scores were not investigated and hence remain unknown.

Despite the contribution of $\rho^2_{(X_1 \to X_2)}$ proposed by Wei and Kim (2017), there are some limitations to be addressed. First, $\rho^2_{(X_1 \to X_2)}$ is limited to the case of two-dimensional

contingency tables. For the multi-dimensional categorical data, it is critical to efficiently and effectively delineate the complex association structure between the ordinal dependent variable and a set of the independent variables in a model-free manner. Therefore, it is urgently needed to develop a novel model-free descriptive modeling method to accelerate the formulation of plausible hypotheses in the subsequent explanatory modeling and facilitate the selection of important variables in the subsequent predictive modeling in the multi-dimensional categorical data with the ordinal dependent variable.

Secondly, the scoring method for the dependent variable X_2 employed by Wei and Kim (2017) was not theoretically well justified and its properties are not well studied. As will be shown in Chapter 3, the scoring method for the ordinal dependent variable plays a crucial role in the novel data-dependent and model-free method, because it provides a basis for coupling the joint distribution of the ordinal variables in a multi-dimensional contingency table with their respective marginal distributions. Therefore, a complete study of the theoretical properties of the scoring method is essential for understanding the theoretical properties of the novel method.

Thirdly, the properties of the mean score for the dependent variable X_2 with respect to its conditional p.m.f. given the independent variable X_1, $\sum_{j=1}^{J} v_{2_j} p_{j|i}$, are not well investigated. As will also be shown in Chapter 3, the mean score for the ordinal dependent variable leverages the scoring method above to identify and quantify the association between the ordinal dependent variable and independent variables in a multi-dimensional contingency table. Thus, a comprehensive investigation of the theoretical properties of the mean score of the dependent variable conditional on the independent variables is vital for acknowledging the theoretical properties of the novel method.

Finally, the statistical inferences for $\rho^2_{(X_1 \to X_2)}$, especially interval estimation and hypothesis testing, were not fully developed. As will be shown in Chapter 7, an extensive development of statistical inference methods for the association measure is critical for assessing the performance of the novel method in its applications to real data sets.

CHAPTER 3

SUBCOPULA SCORES AND SUBCOPULA REGRESSION

In this chapter, we first introduce the subcopula that can link the joint probability mass function of ordinal variables in an arbitrary d-dimensional contingency table and their marginal probability mass functions. We then construct the subcopula scores to numerically represent the categories in the ordinal variables. Finally, we propose the subcopula regression to capture the dependence structure among the variables in the contingency table. Note that, starting from this chapter, we use an artificial running example to illustrate the proposed methods and related theoretical properties in each chapter.

3.1 A d-dimensional contingency table with ordinal variables

Let $X_1, X_2, \ldots, X_d$ be the d ordinal variables with $M_1, M_2, \ldots, M_d$ ordered categories, respectively. The respective supports for $X_1, X_2, \ldots, X_d$ are $\mathbf{S}_{x_1} = \{x_{1_1} < \cdots < x_{1_{M_1}}\}, \mathbf{S}_{x_2} = \{x_{2_1} < \cdots < x_{2_{M_2}}\}, \ldots, \mathbf{S}_{x_d} = \{x_{d_1} < \cdots < x_{d_{M_d}}\}$, where the sign "$<$" highlights the ordering nature of categories of X_i for $i = 1, 2, \ldots, d$. Let $\mathbf{X}_d = \{X_1, \ldots, X_d\}$ and $\mathbf{M}_d = \{M_1, \ldots, M_d\}$. Consider a $M_1 \times M_2 \times \cdots \times M_d$ contingency table of $\mathbf{X}_d$ with the joint probability mass function (p.m.f.)

$$p_{\mathbf{m}_d} = p_{m_1, m_2, \ldots, m_d} = P(\mathbf{X}_d = \mathbf{x}_{\mathbf{m}_d}) = P(X_1 = x_{1_{m_1}}, \ldots, X_d = x_{d_{m_d}})$$

where $\mathbf{x}_{\mathbf{m}_d} = \{x_{1_{m_1}}, \ldots, x_{d_{m_d}}\}$ is a vector of the categories of $\mathbf{X}_d$, $\mathbf{m}_d = \{m_1, \ldots, m_d\}$ is the category index for $\mathbf{X}_d$, $\mathbf{1}_d$ is a vector of all ones with size d, and $\sum_{\mathbf{m}_d = \mathbf{1}_d}^{\mathbf{M}_d} p_{\mathbf{m}_d} =$

$\sum_{m_1=1}^{M_1} \cdots \sum_{m_d=1}^{M_d} p_{m_1,m_2,\ldots,m_d} = 1$. The corresponding one-dimensional marginal p.m.f. of X_i is then given by

$$p_{m_i} = \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} p_{\mathbf{m}_d} = \sum_{m_1=1}^{M_1} \cdots \sum_{m_{i-1}=1}^{M_{i-1}} \sum_{m_{i+1}=1}^{M_{i+1}} \cdots \sum_{m_d=1}^{M_d} p_{m_i,m_2,\ldots,m_d},$$

where $\mathbf{M}_{-i} = \{M_1, \ldots, M_{i-1}, M_{i+1}, \ldots, M_d\}$, $\mathbf{m}_{-i} = \{m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_d\}$ and $\mathbf{1}_{-i}$ is a vector of all ones with size $d-1$. Let $\mathbf{X}_{d^*}$ be a subset with size d^* of $\mathbf{X}_d$ where $1 \leq d^* \leq d$. Then, the d^*-dimensional marginal p.m.f. of $\mathbf{X}_{d^*}$ is then given by

$$p_{\mathbf{m}_{d^*}} = \sum_{\mathbf{m}_{d-d^*}=\mathbf{1}_{d-d^*}}^{\mathbf{M}_{d-d^*}} p_{\mathbf{m}_d} = \sum_{\substack{m_i=1,\\ m_i \in \mathbf{m}_{d-d^*}}}^{M_i \in \mathbf{M}_{d-d^*}} p_{m_1,m_2,\ldots,m_d},$$

where $\mathbf{M}_{d^*} \subseteq \mathbf{M}_d$, $\mathbf{m}_{d^*} \subseteq \mathbf{m}_d$, $\mathbf{M}_{d-d^*} = \mathbf{M}_d - \mathbf{M}_{d^*}$, $\mathbf{m}_{d-d^*} = \mathbf{m}_d - \mathbf{m}_{d^*}$, and $\mathbf{1}_{d-d^*}$ is a vector of all ones with size $d - d^*$.

3.2 Subcopula

Given a $M_1 \times M_2 \times \cdots \times M_d$ contingency table of ordinal variables $\mathbf{X}_d$ and the joint p.m.f. $p_{\mathbf{m}_d}$, let the joint and d one-dimensional marginal cumulative distribution functions (c.d.f.) be $F_{\mathbf{X}_d}(\mathbf{x}_{\mathbf{m}_d}) = F_{X_1,\ldots,X_d}(x_{1_{m_1}}, \ldots, x_{d_{m_d}})$ and $F_{X_1}(x_{1_{m_1}}), \ldots, F_{X_d}(x_{d_{m_d}})$, respectively. Let $U_1 = F_{X_1}(X_1), \ldots, U_d = F_{X_d}(X_d)$. Then $U_1, \ldots, U_d$ are the random variables with the supports

$$\mathbf{D}_{u_1} = \{u_{1_0} = 0 < \cdots < u_{1_{m_1}} < \cdots < u_{1_{M_1}} = 1\},$$
$$\mathbf{D}_{u_2} = \{u_{2_0} = 0 < \cdots < u_{2_{m_2}} < \cdots < u_{2_{M_2}} = 1\},$$
$$\cdots$$
$$\mathbf{D}_{u_d} = \{u_{d_0} = 0 < \cdots < u_{d_{m_d}} < \cdots < u_{d_{M_d}} = 1\},$$

where $u_{i_0} = 0$, $u_{i_{M_i}} = 1$ and

$$u_{i_{m_i}} = \sum_{r_i=1}^{m_i} p_{r_i}. \tag{3.1}$$

Note that $U_1, \ldots, U_d$ do not have a continuous standard uniform distribution.

By Sklar's theorem (Sklar, 1959; Nelsen, 2007), there exists a unique d-dimensional function $C(u_{1_{m_1}}, \ldots, u_{d_{m_d}}) = P(U_1 \leq u_{1_{m_1}}, \ldots, U_d \leq u_{d_{m_d}})$, called subcopula, only on the set $\mathbf{D}_{\mathbf{u}_d} = \mathbf{D}_{u_1} \times \ldots \times \mathbf{D}_{u_d}$, such that

$$F_{\mathbf{X}_d}(\mathbf{x}_{\mathbf{m}_d}) = \sum_{\mathbf{r}_d=\mathbf{1}_d}^{\mathbf{m}_d} p_{\mathbf{r}_d} = \sum_{r_1=1}^{m_1} \cdots \sum_{r_d=1}^{m_d} p_{r_1,r_2,\ldots,r_d} = C(\mathbf{u}_{\mathbf{m}_d}) = C(u_{1_{m_1}}, \ldots, u_{d_{m_d}}), \tag{3.2}$$

where $\mathbf{r}_d = \{r_1, r_2, \ldots, r_d\}$. Note that C satisfies the three properties as follows:

1. $C(u_{1_{m_1}}, \ldots, u_{d_{m_d}}) = 0$ if at least one $u_{i_{m_i}} = 0$.
2. $C(1, \ldots, u_{i_{m_i}}, \ldots, 1) = u_{i_{m_i}}$ for every $u_{i_{m_i}} \in \mathbf{D}_{u_i}$.
3. Given $\mathbf{D}_{\mathbf{v}_d} = [v_1, v_1^*] \times \cdots \times [v_d, v_d^*] \subseteq \mathbf{D}_{\mathbf{u}_d}$ where $[v_i, v_i^*] \subseteq \mathbf{D}_{u_i}$, then

$$0 \leq \sum_{\mathbf{u}_d \in \mathbf{D}_{\mathbf{v}_d}} (-1)^{\left[\sum_{i=1}^d I(u_{i_{m_i}}=v_i)\right]} C(\mathbf{u}_{\mathbf{m}_d}) \leq 1,$$

   where $I(\cdot)$ is the indicator function.

All the three properties ensure C is a valid multivariate cumulative function of $U_1, \ldots, U_d$. The first one corresponds to the fact that $0 \leq P(U_1 \leq u_{1_{m_1}}, \ldots, U_i \leq 0, \ldots, U_d \leq u_{d_{m_d}}) \leq P(U_i \leq 0) = 0$ for $1 \leq i \leq d$. The second one shows that $P(U_1 \leq 1, \ldots, U_i \leq u_{i_{m_i}}, \ldots, U_d \leq 1) = P(U_i \leq u_{i_{m_i}}) = u_{i_{m_i}}$. The third one indicates that the probability of C over any subset of the unit hypercube is non-negative. Furthermore, from Eq. (3.2), we obtain the joint p.m.f. of $C(\mathbf{u}_{\mathbf{m}_d})$:

$$c(\mathbf{u}_{\mathbf{m}_d}) = c(u_{1_{m_1}}, \ldots, u_{d_{m_d}}) = p_{\mathbf{m}_d}, \tag{3.3}$$

and then derive the conditional p.m.f. of U_i given $\mathbf{U}_{-i} = \{U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_d\}$:

$$p_{m_i|\mathbf{m}_{-i}} = c(u_{i_{m_i}}|\mathbf{u}_{\mathbf{m}_{-i}}) = \frac{c(\mathbf{u}_{\mathbf{m}_d})}{c(\mathbf{u}_{\mathbf{m}_{-i}})} = \frac{p_{\mathbf{m}_d}}{p_{\mathbf{m}_{-i}}} = \frac{p_{\mathbf{m}_d}}{\sum_{m_i=1}^{M_i} p_{\mathbf{m}_d}}, \tag{3.4}$$

where $\mathbf{u}_{\mathbf{m}_{-i}} = \{u_{1_{m_1}}, \ldots, u_{i-1_{m_{i-1}}}, u_{i+1_{m_{i+1}}}, \ldots, u_{d_{m_d}}\}$ and $p_{\mathbf{m}_{-i}} = p_{m_1,\ldots,m_{i-1},m_{i+1},\ldots,m_d}$.

Given that $U_i = F_{X_i}(X_i)$ for each X_i, some important properties of $\mathbf{U}_d = \{U_1, \ldots, U_d\}$ are given below:

1. $\mathbf{U}_d$ is a d-dimensional random vector associated with $\mathbf{X}_d$ such that the support of $\mathbf{U}_d$ is $\mathbf{D}_{u_1} \times, \ldots, \times \mathbf{D}_{u_d}$ and the joint p.m.f. of $\mathbf{U}_d$ is the same as the joint p.m.f. of $\mathbf{X}_d = \{X_1, \ldots, X_d\}$.

2. Unlike X_i, the support of U_i defined in Eq. (3.1), $\mathbf{D}_{u_i}$, is numerical.

3. $\mathbf{D}_{u_i}$ maintains the natural ordering of categories of X_i in that each category of X_i corresponds to one element of $\mathbf{D}_{u_i}$, i.e. $U_i = F_{X_i}(X_i) : \{x_{i_1} < \cdots < x_{i_{m_i}} < \cdots < x_{i_{M_i}}\} \mapsto \{u_{i_1} < \cdots < u_{i_{m_i}} < \cdots < u_{i_{M_i}}\}$.

4. $\mathbf{D}_{u_1}, \mathbf{D}_{u_2}, \ldots, \mathbf{D}_{u_d}$ contain the elements of sets over which the joint distribution of $\mathbf{X}_d$ is uniquely linked to their marginal distributions. Such a distributional linkage is known to be a subcopula function C in Eq. (3.2).

3.3 Subcopula scores

In order to exploit the ordering information contained in the categories of the ordinal variable of interest, we desire some naturally ordered scores for the categories of the variable in the methods for analyzing ordinal data. To this end, we call $\mathbf{D}_{u_i}$ in Eq. (3.1) the *the set of subcopula scores* for the ordinal variable X_i. Then $\mathbf{U}_d$ consists of d new random variables, each of which takes on the values of its own subcopula scores and the joint p.m.f. of $\mathbf{U}_d$ is the same as that of $\mathbf{X}_d$.

The subcopula scores for X_i is a rank-type score such as the *Ridit score* and *midrank* reviewed in Section 2.1. Proposition 3.1 below shows the relationship between the subcopula scores and the (*Ridit scores/midranks*).

PROPOSITION 3.1 *Suppose that $s^R_{i_{m_i}}$ and $s^M_{i_{m_i}}$ are the respective Ridit score and midrank for the m_i-th category of X_i in a d-dimensional contingency table. Then,*

$$s^R_{i_{m_i}} = \frac{u_{i_{m_i-1}} + u_{i_{m_i}}}{2}, \quad s^M_{i_{m_i}} = \frac{n}{2}(u_{i_{m_i-1}} + u_{i_{m_i}}) + 0.5,$$

where $u_{i_{m_i}}$ is the subcopula score for the m_i-th category of X_i defined in Eq. (3.1) and n is the sample size of the d-dimensional contingency table.

PROOF *See Appendix A.*

We see that both the *Ridit scores* (*midranks*) and the subcopula scores are constructed based on the marginal distribution of each ordinal variable. Like the *Ridit scores* (*midranks*), the proposed subcopula scores are data-dependent, which can be estimated by the relative frequencies in the contingency table, as will be shown in Chapter 6. Note that the sets of the subcopula scores for the ordinal variables $X_1, \ldots, X_d$ provide a basis for the unique link between their joint distribution function and their marginal distributions.

Proposition 3.2 below gives the basic properties of the proposed subcopula scores.

PROPOSITION 3.2 *Consider $U_i = F_{X_i}(X_i)$ and the corresponding set of subcopula scores $\mathbf{D}_{u_i}$.*

(a) If s_i adjacent categories $\{x_{m_i+1}, \ldots, x_{m_i+s_i}\}$ of X_i are combined into a new category $x_{m_i^*}$ where $1 < s_i \leq M_i - m_i$ and $m_i + 1 \leq m_i^* \leq m_i + s_i$, then the subcopula score $u_{m_i^*}$ for the new category is equal to $u_{m_i+s_i}$ while the subcopula scores for the other categories of X_i remain unchanged.

(b) If the ordering of the categories in X_i is reversed, then the subcopula score for x_{m_i} becomes $1 - u_{m_i-1}$.

(c) The mean of U_i is given by

$$E(U_i) = \frac{1}{2} + \frac{1}{2}\sum_{m_i=1}^{M_i} p_{m_i}^2 \tag{3.5}$$

and the variance of U_i is given by

$$\begin{aligned} Var(U_i) = &\left(\frac{1}{2}\sum_{m_i=1}^{M_i} p_{m_i}^2 - \frac{1}{4}\left[\sum_{m_i=1}^{M_i} p_{m_i}^2\right]^2 - \frac{1}{4}\right) + \left(\sum\sum_{m_i<m_i^*} p_{m_i} p_{m_i^*}^2\right) + \\ &\left[2I(M_i > 2)\sum\sum\sum_{m_i<m_i^*<m_i^{**}} p_{m_i} p_{m_i^*} p_{m_i^{**}}\right], \end{aligned} \tag{3.6}$$

where $I(\cdot)$ is an indicator function. Typically, if X_i follows a discrete uniform distribution with the support $\mathbf{S}_{x_i} = \{x_{i_1}, \ldots, x_{i_{M_i}}\}$ such that each $x_{m_i} \in \mathbf{S}_{x_i}$ occurs with equal probability $1/M_i$, then $p_{m_i} = 1/M_i$ for every m_i and hence the mean and variance of U_i in Eq. (3.5) and (3.6) become

$$E(U_i) = \frac{1}{2} + \frac{1}{2M_i}$$

and

$$Var(U_i) = \frac{(M_i - 1)(M_i + 3)}{4M_i^2} + 2I(M_i > 2)\left[\frac{(M_i + 1)(4M_i - 1)}{12M_i^2}\right],$$

respectively.

PROOF *See Appendix B.*

To illustrate the construction of subcopula scores for ordinal variables, an artificial three-dimensional contingency table is given below as a running example. We also compute the mean and variance of the subcopula scores for each ordinal variable using Eq. (3.5) and (3.6), respectively.

EXAMPLE 3.1 *Table 3.1 below is an artificial $3 \times 3 \times 2$ contingency table for ordinal variables X_1, X_2 and X_3. The supports of X_1, X_2 and X_3 are $\mathbf{S}_{x_1} = \{x_{1_1} < x_{1_2} < x_{1_3}\}$,*

$\mathbf{S}_{x_2} = \{x_{2_1} < x_{2_2} < x_{2_3}\}$ *and* $\mathbf{S}_{x_3} = \{x_{3_1} < x_{3_2}\}$*, respectively. The joint p.m.f. of* X_1*,* X_2 *and* X_3 *is* $p_{\mathbf{m}_3} = p_{m_1,m_2,m_3} = P(X_1 = x_{1_{m_1}}, X_2 = x_{2_{m_2}}, X_3 = x_{3_{m_3}})$*, where* $m_1 = m_2 = 1, 2, 3$ *and* $m_3 = 1, 2$*. The marginal p.m.f.s of* X_1*,* X_2 *and* X_3 *are* $p_{m_1} = P(X_1 = x_{1_{m_1}}) = \{24/54, 13/54, 17/54\}$*,* $p_{m_2} = P(X_2 = x_{2_{m_2}}) = \{24/54, 27/54, 3/54\}$ *and* $p_{m_3} = P(X_3 = x_{3_{m_3}}) = \{40/54, 14/54\}$*, respectively. Note that we design the table for the purpose of establishing a functional relationship among the three variables. That is,* X_3 *is a function of* X_1 *and* X_2 *in the sense that only one category of* X_3 *has a non-zero probability mass for every combination of the categories of* X_1 *and* X_2*.*

To understand the dependence structure between X_3 *and* (X_1, X_2) *in Table 3.1, we plot the categories of* X_3 *against the combinations of the categories of* (X_1, X_2) *in Figure 3.1. Note that the size of each circle in both plots is proportional to the probability of the category of* X_3 *at each combination of the categories of* (X_1, X_2) *in the contingency table. We also plot the categories of* X_3 *versus the combinations of the categories of* (X_2, X_1) *in Figure 3.2. It turns out that Figure 3.1 shows a monotone (increasing/decreasing) pattern at a fixed level of* X_1 *and an overall wavy pattern as the level of* X_1 *increases, while Figure 3.2 follows a quadratic (up-down/down-up) pattern at a fixed level of* X_2 *and an overall zig-zag pattern as the level of* X_2 *increases.*

X_3	x_{3_1}			x_{3_2}		
X_1 \ X_2	x_{2_1}	x_{2_2}	x_{2_3}	x_{2_1}	x_{2_2}	x_{2_3}
x_{1_1}	12/54	11/54	0	0	0	1/54
x_{1_2}	0	0	1/54	5/54	7/54	0
x_{1_3}	7/54	9/54	0	0	0	1/54

Table 3.1: The artificial $3 \times 3 \times 2$ contingency table for ordinal variables X_1, X_2 and X_3

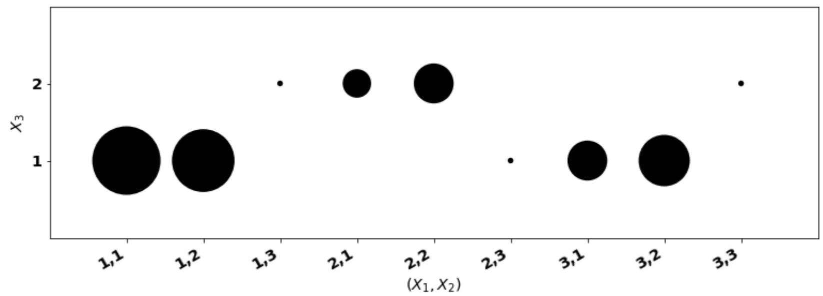


Figure 3.1: Categories of X_3 against the combinations of the categories of (X_1, X_2)

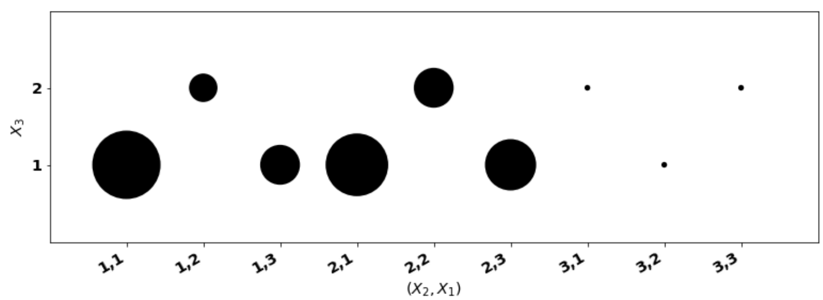


Figure 3.2: Categories of X_3 against the combinations of the categories of (X_2, X_1)

According to Eq.(3.1), the sets of subcopula scores for X_1, X_2 and X_3 in Table 3.1 are $\mathbf{D}_{u_1} = \{0, 24/54, 37/54, 1\}$, $\mathbf{D}_{u_2} = \{0, 24/54, 51/54, 1\}$ and $\mathbf{D}_{u_3} = \{0, 40/54, 1\}$, respectively. Consider that X_3 is the dependent variable, X_1 and X_2 are the independent variables. Table 3.2 below shows the means and variances for U_1, U_2 and U_3 using Eq. (3.5) and (3.6). Table 3.3 below gives the joint p.m.f. of U_1 and U_2, which is denoted by $p_{\mathbf{m}_{-3}} = p_{m_1,m_2}$. Table 3.4 below provides the joint p.m.f. of the subcopula C, which is $c(\mathbf{u}_{\mathbf{m}_3}) = p_{\mathbf{m}_3}$ defined in Eq. (3.3) and the conditional p.m.f. of U_3 given (U_1, U_2), which is denoted by $p_{m_3|\mathbf{m}_{-3}} = c(u_{3_{m_3}}|u_{1_{m_1}}, u_{2_{m_2}})$ in Eq. (3.4).

	$E(U_i)$	$Var(U_i)$
U_1	0.808	0.013
U_2	0.725	0.063
U_3	0.677	0.057

Table 3.2: The mean and variance (rounded to 3 decimals) for U_1, U_2 and U_3, respectively

$\mathbf{D}_{u_1}$ \ $\mathbf{D}_{u_2}$	24/54	51/54	1
24/54	12/54	11/54	1/54
37/54	5/54	7/54	1/54
1	7/54	9/54	1/54

Table 3.3: The sets of subcopula scores $\mathbf{D}_{u_1}$ and $\mathbf{D}_{u_2}$ and the joint p.m.f. $p_{\mathbf{m}_{-3}}$

$\mathbf{D}_{u_3}$	40/54			1		
$\mathbf{D}_{u_1}$ \ $\mathbf{D}_{u_2}$	24/54	51/54	1	24/54	51/54	1
24/54	12/54 (1)	11/54 (1)	0 (0)	0 (0)	0 (0)	1/54 (1)
37/54	0 (0)	0 (0)	1/54 (1)	5/54 (1)	7/54 (1)	0 (0)
1	7/54 (1)	9/54 (1)	0 (0)	0 (0)	0 (0)	1/54 (1)

Table 3.4: The joint p.m.f. $c(\mathbf{u}_{\mathbf{m}_3}) = p_{\mathbf{m}_3}$ and the conditional p.m.f. $p_{m_3|\mathbf{m}_{-3}}$ (i.e. the numbers in parentheses)

3.4 Subcopula regression

To identify the association among the ordinal variables $\mathbf{X}_d$ in a model-free manner, we propose the subcopula regression function using the subcopula scores and corresponding conditional p.m.f. for X_i defined in Section 3.1–3.3.

DEFINITION 3.1 *Consider a $M_1 \times \cdots \times M_d$ contingency table of ordinal variables $\mathbf{X}_d$ with the joint p.m.f. $p_{\mathbf{m}_d}$ and its associated subcopula $C(\mathbf{u}_{\mathbf{m}_d})$. Then the subcopula regression function of X_i on $\mathbf{X}_{-i} = \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_d\}$ is then defined to be*

$$\begin{aligned} r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}}) = E_C(U_i|\mathbf{U}_{-i} = \mathbf{u}_{\mathbf{m}_{-i}}) &= \sum_{u_{im_i} \in \mathbf{D}_{u_i}} u_{im_i} c_{U_i|\mathbf{U}_{-i}}(u_{im_i}|\mathbf{u}_{\mathbf{m}_{-i}}) \\ &= \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i|\mathbf{m}_{-i}}, \end{aligned} \tag{3.7}$$

where $U_i = F_{X_i}(X_i)$.

Note that $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$, the mean of U_i with respect to the conditional p.m.f. of the subcopula $c_{U_i|\mathbf{U}_{-i}}$, can be viewed as the mean subcopula score for X_i with respect to the conditional distribution of X_i given $\mathbf{X}_{-i}$, and one may identify a pattern using Eq. (3.7) over the combinations of the categories of $\mathbf{X}_{-i}$.

Proposition 3.3 below gives the range, mean and variance of the subcopula regression function in Eq. (3.7).

PROPOSITION 3.3 *Consider the subcopula regression function $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$ in Eq. (3.7).*

(a) The range of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$ is $[0, 1]$.

(b) The mean and variance of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$ are given by

$$E\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})\right] = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i} = E(U_i),$$
$$Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})\right] = \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left[\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right]^2 p_{\mathbf{m}_{-i}}.$$

PROOF *See Appendix C.*

We also show the ability of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$ to capture the association structure by predicting the categories of X_i for a combination of categories of $\mathbf{X}_{-i}$. First, we obtain the subcopula scores $\mathbf{u}_{\mathbf{m}_{-i}}$ for the corresponding $\mathbf{U}_{-i}$ given a vector of categories of $\mathbf{X}_{-i}$, $\mathbf{x}_{\mathbf{m}_{-i}} = \{x_{1_{m_1}}, \ldots, x_{i-1_{m_{i-1}}}, x_{i+1_{m_{i+1}}}, \ldots, x_{d_{m_d}}\}$. Next, we calculate a predicted value of $u_{i_{m_i}}$ denoted by $u^*_{i_{m_i}}$ through $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$. Then we locate the index m_i^* of $\mathbf{D}_{u_i}$ (the set of subcopula scores for X_i) such that $u_{i_{m_i^*-1}} < u^*_{i_{m_i}} \leq u_{i_{m_i^*}}$. Finally, the $x_{i_{m_i^*}}$ from $\mathbf{S}_{x_i}$ (the support of X_i) corresponding to $u_{i_{m_i^*}}$ is considered as the predicted category.

As will be shown in the running example below and the real data examples in Chapter 7, the plot of the predicted categories of X_i across the combinations of the categories of $\mathbf{X}_{-i}$ will reveal the pattern of association between X_i and $\mathbf{X}_{-i}$.

We resume Example 3.1 below to illustrate the proposed subcopula regression and the subsequent prediction method.

EXAMPLE 3.2 *(Example 3.1 Continued) Following Eq. (3.7), we first compute the subcopula regression of* (X_1, X_2) *on* X_3 *below, using the set of subcopula scores* $\mathbf{D}_{u_1}$*,* $\mathbf{D}_{u_2}$*,* $\mathbf{D}_{u_3}$ *and the conditional p.m.f. in Table 3.4:*

$$r^C_{U_3|\mathbf{U}_{-3}}(\mathbf{u}_{\mathbf{m}_{-3}}) = r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}}) = \begin{cases} 40/54, & (u_{1_1}, u_{2_1}) = (24/54, 24/54) \\ 40/54, & (u_{1_1}, u_{2_2}) = (24/54, 51/54) \\ 1, & (u_{1_1}, u_{2_3}) = (24/54, 1) \\ 1, & (u_{1_2}, u_{2_1}) = (37/54, 24/54) \\ 1, & (u_{1_2}, u_{2_2}) = (37/54, 51/54) \\ 40/54, & (u_{1_2}, u_{2_3}) = (37/54, 1) \\ 40/54, & (u_{1_3}, u_{2_1}) = (1, 24/54) \\ 40/54, & (u_{1_3}, u_{2_2}) = (1, 51/54) \\ 1, & (u_{1_3}, u_{2_3}) = (1, 1) \end{cases} . \tag{3.8}$$

Then, according to Proposition 3.3, we find the mean and variance of $r^C_{U_3|\mathbf{U}_{-3}}(\mathbf{u}_{\mathbf{m}_{-3}})$ *are 0.808 and 0.013, respectively.*

To illustrate how well the proposed subcopula regression function in Eq. (3.7) can capture the dependence structure shown in Figure 3.1 and 3.2, we plot the subcopula regression function $r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})$ *and* $r^C_{U_3|U_2,U_2}(u_{2_{m_2}}, u_{1_{m_1}})$ *against the combinations of the subcopula scores* $(U_1, U_2) = (u_{1_{m_1}}, u_{2_{m_2}})$ *and* $(U_2, U_1) = (u_{2_{m_2}}, u_{1_{m_1}})$ *in Figure 3.3 and 3.4, respectively. We can see that Figure 3.3 follows exactly the same wavy pattern as in Figure 3.1, and Figure 3.4 follows exactly the same zig-zag pattern as in Figure 3.2.*

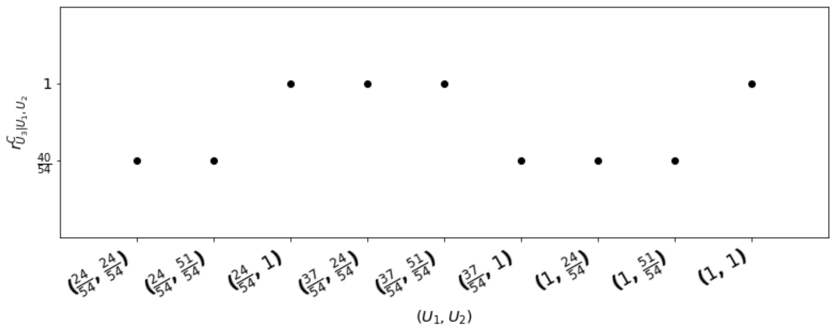


Figure 3.3: $r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})$ against the combinations of the subcopula scores $(U_1, U_2) = (u_{1_{m_1}}, u_{2_{m_2}})$

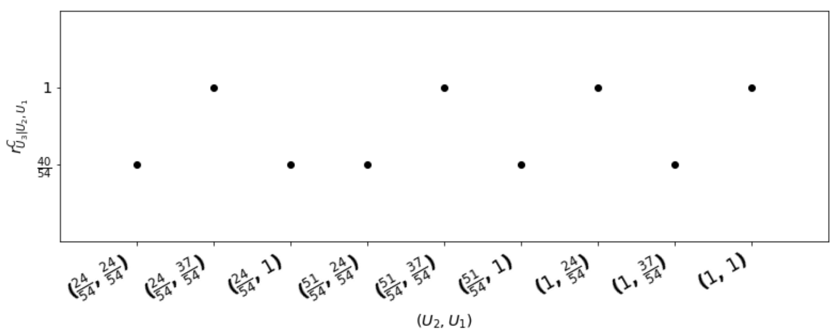


Figure 3.4: $r^C_{U_3|U_2,U_1}(u_{2_{m_2}}, u_{1_{m_1}})$ against the combinations of the subcopula scores $(U_2, U_1) = (u_{2_{m_2}}, u_{1_{m_1}})$

Now assume that the category of X_1 and X_2 are given by x_{1_1} and x_{2_2}, respectively: $(X_1, X_2) = (x_{1_1}, x_{2_2})$. Then, from Eq. (3.8), we find the corresponding subcopula scores (u_{1_1}, u_{2_2}) are $(24/54, 51/54)$ and the predicted value of the subcopula regression $r^C_{U_3|U_1,U_2}(24/54, 51/54)$ is $40/54$. Since $u_{3_1} = 40/54$, we conclude that the predicted category of X_3 given $X_1 = x_{1_1}$ and $X_2 = x_{2_2}$ is x_{3_1}. The predicted category of X_3 at every combination of the categories of (X_1, X_2) are provided in Table 3.5 and visualized in Figure 3.5 and 3.6 for each combination of the categories of (X_1, X_2) and (X_2, X_1), respectively. We can see that the predicted categories of X_3 by (X_1, X_2) are the same as the categories of X_3 whose probabilities are non-zero as shown in Table 3.1

$X_1 \backslash X_2$	x_{2_1}	x_{2_2}	x_{2_3}
x_{1_1}	x_{3_1} (40/54)	x_{3_1} (40/54)	x_{3_2} (1)
x_{1_2}	x_{3_2} (1)	x_{3_2} (1)	x_{3_1} (40/54)
x_{1_3}	x_{3_1} (40/54)	x_{3_1} (40/54)	x_{3_2} (1)

Table 3.5: The subcopula regression of (X_1, X_2) on X_3 (the numbers in parentheses) and the corresponding predicted category of X_3 for every combination of the categories of X_1 and X_2

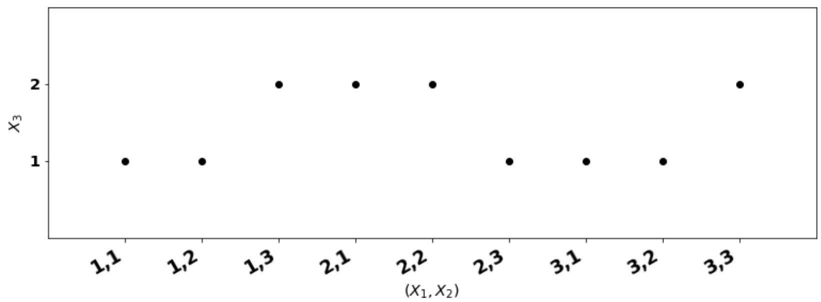


Figure 3.5: Predicted categories of X_3 against the combinations of the categories of (X_1, X_2)

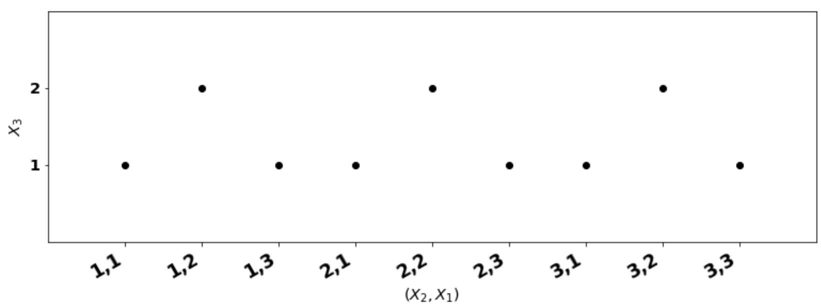


Figure 3.6: Predicted categories of X_3 against the combinations of the categories of (X_2, X_1)

Proposition 3.4 below shows that the predicted categories of the dependent variable X_i using the proposed subcopula regression function $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$ is invariant under the permutation on the categories of the independent variables $\mathbf{X}_{-i}$.

PROPOSITION 3.4 *Considering a d-dimensional contingency table of ordinal variables $\mathbf{X}_d$, let $f_{X_i|\mathbf{X}_{-i}}$ be the predicted category of X_i for a given combination of categories of $\mathbf{X}_{-i}$ using the subcopula regression function in Eq. (3.7). Then the predicted category of X_i is invariant with respect to the element-wise injective transformation on $\mathbf{X}_{-i}$.*

That is, if $\tilde{\mathbf{X}}_{-i} = \{g_1(X_1), \ldots, g_{i-1}(X_{i-1}), g_{i+1}(X_{i+1}), \ldots, g_d(X_d)\} = g_{-i}(\mathbf{X}_{-i})$*, where* $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_d$ *are injective functions, then*

$$f_{X_i|\tilde{\mathbf{X}}_{-i}}(\tilde{\mathbf{x}}_{\tilde{\mathbf{m}}_{-i}}) = f_{X_i|\mathbf{X}_{-i}}(\mathbf{x}_{\mathbf{m}_{-i}}),$$

for a given vector of categories of $\tilde{X}_{-i}$*,* $\tilde{\mathbf{x}}_{\tilde{\mathbf{m}}_{-i}} = \{\tilde{x}_{1_{\tilde{m}_1}}, \ldots, \tilde{x}_{i-1_{\tilde{m}_{i-1}}}, \tilde{x}_{i+1_{\tilde{m}_{i+1}}}, \ldots, \tilde{x}_{d_{\tilde{m}_d}}\}$*, and* $\tilde{\mathbf{m}}_{-i} = \{\tilde{m}_1, \ldots, \tilde{m}_{i-1}, \tilde{m}_{i+1}, \ldots, \tilde{m}_d\}$.

PROOF *See Appendix D.*

Note that Proposition 3.4 guarantees that the proposed subcopula regression based prediction method is valid with not only ordinal independent variables but also nominal ones. This is because the computation of the subcopula regression function of X_i for a given combination of categories of $\mathbf{X}_{-i}$ in Eq. (3.7) only requires the subcopula scores of X_i and the corresponding conditional distribution of X_i given a combination of categories of $\mathbf{X}_{-i}$. We resume Example 3.1 to show the invariance in predicting the categories of the dependent variable with respect to a permutation of the categories of independent variables.

EXAMPLE 3.3 *(Example 3.1 Continued) Without loss of generality, we only apply a permutation to* X_2*, where we switch the second category of* X_2 *with the third one. Let* $\tilde{X}_2$ *denote the permuted* X_2*, which is* $\{x_{2_1}, x_{2_3}, x_{2_2}\}$*. The contingency table of* X_1*,* $\tilde{X}_2$ *and* X_3 *is shown in Table 3.6. We again want to predict the category of* X_3 *on* $(X_1 = x_{1_1}, \tilde{X}_2 = x_{2_2})$*. Consider* $\tilde{U}_2 = F_{\tilde{X}_2}(\tilde{X}_2)$ *with the corresponding set of subcopula scores* $\mathbf{D}_{\tilde{u}_2} = \{0, 24/54, 27/54, 1\}$*. Then we compute the predicted value of the subcopula regression* $r^C_{U_3|U_1,\tilde{U}_2}(24/54, 1) = 40/54$ *and hence conclude the predicated category of* X_3 *is still* x_{3_1} *as above. Table 3.7 shows the predicted category of* X_3 *for every combination of the categories of* X_1 *and* $\tilde{X}_2$ *which is the same as those in Table 3.5.*

X_3	x_{3_1}			x_{3_2}		
X_1 \ X_2	x_{2_1}	x_{2_3}	x_{2_2}	x_{2_1}	x_{2_3}	x_{2_2}
x_{1_1}	12/54	0	11/54	0	1/54	0
x_{1_2}	0	1/54	0	5/54	0	7/54
x_{1_3}	7/54	0	9/54	0	1/54	0

Table 3.6: The artificial $3 \times 3 \times 2$ contingency table for ordinal variables X_1, $\tilde{X}_2$ and X_3

X_1 \ X_2	x_{2_1}	x_{2_3}	x_{2_2}
x_{1_1}	1 (40/54)	2 (1)	1 (40/54)
x_{1_2}	2 (1)	1 (40/54)	2 (1)
x_{1_3}	1 (40/54)	2 (1)	1 (40/54)

Table 3.7: The subcopula regression of X_3 on $\{X_1, \tilde{X}_2\}$ (the numbers in parentheses) and the corresponding predicted category of X_3 for every combination of the categories of X_1 and $\tilde{X}_2$

3.5 Use of the subcopula regression for descriptive modeling

The applications of the subcopula regression function proposed in Chapter 3.4 are two-fold. First, it can be employed to explore the potential association structures between the dependent variable and a set of independent variables in a model-free manner. That is, we can tabulate or plot the predicted category of the dependent variable for each combination of the categories of the independent variables. Then we can partition all combinations of the categories of the independent variables into a few subsets depending on the resulting category of the dependent variable to identify the patterns of change in the predicted category of the dependent variable across the combinations of categories of the independent variables. Secondly, the subcopula regression can be adopted to identify potentially important independent variables. That is, we can compare the results of the subcopula regression based on different subsets of independent variables so that we may be able to distinguish the independent variables that may influence the predicted categories of the dependent variable from those that may not.

We will first illustrate in Chapter 7.1 the application of the subcopula regression function and its prediction to identify the potential non-monotone (i.e. quadratic) association structure in a two-dimensional real data set called *ice cream* (*The Ice Cream Study at*

Penn State, 2012). Then we will use the subcopula regression function in Chapter 7.2 to study more complicated patterns of association in a three-dimensional real data set called *acute migraine* (Vandenhende and Lambert, 2000), where the goal is to compare the pain scores of patients at eight occasions across different treatment groups. In particular, we will list and plot the predicted pain score against each combination of the occasions and treatment groups to capture a relationship between the pain scores and occasion for each treatment group.

CHAPTER 4

SUBCOPULA REGRESSION BASED ASSOCIATION MEASURE AND ITS DECOMPOSITIONS

In this chapter, we propose the subcopula regression based association measure to quantify the overall dependence among $\mathbf{X}_d$ in the d-dimensional contingency table, and its decompositions to study the contribution of each or a subset of independent variables considered in the subcopula regression to the overall association. In addition, we investigate the theoretical properties for both the overall association measure and its decompositions.

4.1 Subcopula regression based association measure

We first give the definition of the subcopula regression based association measure below for the d-dimensional contingency table with one ordinal dependent variable.

DEFINITION 4.1 *Consider a d-dimensional contingency table of ordinal variables $\mathbf{X}_d$ with the joint p.m.f. $p_{\mathbf{m}_d}$ and its associated subcopula $C(\mathbf{u}_{\mathbf{m}_d})$. Then the subcopula regression-based association measure of X_i on $\mathbf{X}_{-i}$ is defined to be*

$$\begin{aligned}
\rho^2_{(\mathbf{X}_{-i}\to X_i)} &\equiv \frac{Var\left(r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})\right)}{Var(U_i)} \\
&= \frac{\displaystyle\sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left(\sum_{m_i=1}^{M_i} u_{i m_i} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{i m_i} p_{m_i}\right)^2 p_{\mathbf{m}_{-i}}}{\displaystyle\sum_{m_i=1}^{M_i} \left(u_{i m_i} - \sum_{m_i=1}^{M_i} u_{i m_i} p_{m_i}\right)^2 p_{m_i}},
\end{aligned} \tag{4.1}$$

where $U_i = F_{X_i}(X_i)$.

Proposition 4.1 below provides the theoretical properties of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$.

PROPOSITION 4.1

(a) $0 \le \rho^2_{(\mathbf{X}_{-i}\to X_i)} \le 1$.

(b) If X_i *and* $\mathbf{X}_{-i}$ *are independent, then* $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$.

(c) If $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$*, then the mean subcopula score for the categories of* X_i *with respect to the conditional distribution of* X_i *given* $\mathbf{X}_{-i}$ *is constant over every combination of categories in* $\mathbf{X}_{-i}$*. Furthermore,* $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ *implies that* U_i *is uncorrelated with every subset of* $\mathbf{U}_{-i}$.

(d) $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 1$ *if and only if* $X_i = g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d) = g(\mathbf{X}_{-i})$ *almost surely for some measurable function* g.

(e) If $U_i = g(\mathbf{U}_{-i}) + \epsilon$ *where* ϵ*, being independent of* $\mathbf{U}_{-i}$*, is a random variable with finite second moment and* g *is a measurable function, then*

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \frac{Var(g(\mathbf{U}_{-i}))}{Var(g(\mathbf{U}_{-i})) + Var(\epsilon)}$$

(f) $\rho_{(\mathbf{X}_{-i}\to X_i)} = corr(U_i, r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}}))$.

(g) If $\tilde{\mathbf{X}}_{-i} = g_{-i}(\mathbf{X}_{-i})$ *defined in Proposition 3.2, then* $\rho^2_{(\tilde{\mathbf{X}}_{-i}\to X_i)} = \rho^2_{(\mathbf{X}_{-i}\to X_i)}$*. Especially,* $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ *is invariant over the permutation of the categories of every* $X_j \in \mathbf{X}_{-i}$.

(h) $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ *is invariant with respect to the permutation on the categories of* X_i *when* X_i *is a binary variable.*

PROOF *See Appendix E.*

Proposition 4.1 (a) shows that the range of the proposed subcopula regression based association measure $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is between 0 and 1. (b), (c) and (d) present the sufficient and necessary conditions for $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 1$, respectively. In particular, (c) indicates that when $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is zero, $\mathbf{X}_{-i}$ and its subsets have no influence on the mean subcopula score of X_i. Moreover, (d) and (e) imply that $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is able to measure not only monotone linear but also non-monotone nonlinear relationship between the dependent variable and independent variables.

By Proportion 4.1 (e), $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ represents the average proportion of variance for the subcopula scores of the dependent variable X_i with respect to the marginal p.m.f. of X_i explained by all the independent variables $\mathbf{X}_{-i}$ in the subcopula regression. (f) reveals that the positive square root of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is equal to the correlation between the subcopula scores of X_i and its mean subcopula scores $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{\mathbf{m}_{-i}})$. Hence $\rho_{(\mathbf{X}_{-i}\to X_i)}$ can measure the explanatory power of $\mathbf{X}_{-i}$. Lastly, (g) and (h) guarantee that $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is still valid even when the predictors $\mathbf{X}_{-i}$ is nominal and/or the dependent variable X_i is binary.

We use Example 3.1 below to illustrate the computation of the proposed subcopula regression based association measure in Eq. (4.1).

EXAMPLE 4.1 *(Example 3.1 Continued) Given* $\mathbf{D}_{u_3} = \{0, 40/54, 1\}$ *and the marginal p.m.f.* $p_{m_3} = \{40/54, 14/54\}$, *we first calculate* $E_C\left[r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})\right] = E_C(U_3) = 0.677$ *given in Table 3.2. Then, using the joint p.m.f.* $p_{12_{m_1,m_2}}$ *in Table 3.3 and the values of the subcopula regression* $r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})$ *in Table 3.5, we obtain* $\rho^2_{(X_1,X_2\to X_3)} = 1$. *Similarly, we also find* $\rho^2_{(X_1,X_3\to X_2)} = 0.078$ *and* $\rho^2_{(X_2,X_3\to X_1)} = 0.008$. *The results for* $\rho^2_{(X_1,X_2\to X_3)}$, $\rho^2_{(X_1,X_3\to X_2)}$ *and* $\rho^2_{(X_2,X_3\to X_1)}$ *support Proposition 4.1 (d) because the true relationship shown in Table 3.1 is that* X_3 *is a function of* X_1 *and* X_2. *For comparison, we also compute the Gray-Williams' index reviewed in Section 2.2.2.2 for this example:* $\tau^{GW}_{X_3} = 1$, $\tau^{GW}_{X_2} = 0.151$ *and* $\tau^{GW}_{X_1} = 0.434$.

4.2 Decompositions of the subcopula regression based association measure

To quantify the contribution of a set of independent variables of interest to the overall association in the full-dimensional contingency table, we first introduce two different ways to decompose the subcopula regression based association measure proposed in Eq. (4.1) for three-dimensional contingency tables and investigate their properties. Then we show the properties of the decompositions under four types of independence: mutual, joint, marginal and conditional. Lastly, we extend the decompositions and corresponding properties for three-dimensional contingency tables to d-dimensional ones.

4.2.1 Decompositions for three-dimensional contingency tables

Let X_1, X_2 and X_3 be the three ordinal variables in a contingency table. Without loss of generality, we consider that X_3 is the dependent variable, and X_1 and X_2 are the independent variables. For $i = 1, 2, 3$, we define U_i, u_i, M_i and m_i for X_i following the definitions given in Chapter 3.

According to Definition 3.1, we have the subcopula regressions of X_3 on X_1, X_2 and (X_1, X_2), denoted by $r^C_{U_3|U_1}(u_{1_{m_1}})$, $r^C_{U_3|U_2}(u_{2_{m_2}})$ and $r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})$, respectively:

$$r^C_{U_3|U_1}(u_{1_{m_1}}) = \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_1}, \quad r^C_{U_3|U_2}(u_{2_{m_2}}) = \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_2},$$
$$r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}}) = \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_1,m_2}.$$

From Definition 3.2, we also have the subcopula regression-based association measures of

X_3 on X_1, X_2 and (X_1, X_2), denoted by $\rho^2_{(X_1 \to X_3)}$, $\rho^2_{(X_2 \to X_3)}$ and $\rho^2_{(X_1, X_2 \to X_3)}$, respectively:

$$\rho^2_{(X_1 \to X_3)} = \frac{Var\left[r^C_{U_3|U_1}(u_{1_{m_1}})\right]}{Var(U_3)} = \frac{\sum_{m_1=1}^{M_1}\left(\sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_1} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_1}}{\sum_{m_3=1}^{M_3}\left(u_{3_{m_3}} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_3}},$$

$$\rho^2_{(X_2 \to X_3)} = \frac{Var\left[r^C_{U_3|U_2}(u_{2_{m_2}})\right]}{Var(U_3)} = \frac{\sum_{m_2=1}^{M_2}\left(\sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_2} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_2}}{\sum_{m_3=1}^{M_3}\left(u_{3_{m_3}} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_3}},$$

$$\begin{aligned}\rho^2_{(X_1, X_2 \to X_3)} &= \frac{Var\left[r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})\right]}{Var(U_3)} \\ &= \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\left(\sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3|m_1,m_2} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}\left(u_{3_{m_3}} - \sum_{m_3=1}^{M_3} u_{3_{m_3}} p_{m_3}\right)^2 p_{m_3}}.\end{aligned}$$

In three-dimensional contingency tables, we call $\rho^2_{(X_1, X_2 \to X_3)}$ *the overall subcopula regression based association measure* and call $\rho^2_{(X_1 \to X_3)}$ and $\rho^2_{(X_2 \to X_3)}$ *the marginal subcopula regression based association measures* of X_1 and X_2, respectively.

For the purpose of clear illustration, let $r_{u_{1_{m_1}}} = r^C_{U_3|U_1}(u_{1_{m_1}})$, $r_{u_{2_{m_2}}} = r^C_{U_3|U_2}(u_{2_{m_2}})$ and $r_{u_{1_{m_1}}, u_{2_{m_2}}} = r^C_{U_3|U_1,U_2}(u_{1_{m_1}}, u_{2_{m_2}})$. We also let $r = E_C(U_3)$ denote the expectation of subcopula score of the dependent variable X_3. Note that we have $r = E_C(r_{u_{1_{m_1}}}) = E_C(r_{u_{2_{m_2}}}) = E_C(r_{u_{1_{m_1}}, u_{2_{m_2}}})$.

Theorem 4.1 below defines the decompositions of *the overall subcopula regression based association measure*.

THEOREM 4.1 *Consider a three-dimensional contingency table of ordinal variables X_1, X_2 and X_3, with the joint p.m.f. p_{m_1,m_2,m_3}. The decomposition of the subcopula regression*

based measure $\rho^2_{(X_1,X_2\to X_3)}$ is given by

$$\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_1\to X_3)} + \rho^2_{(X_2\to X_3|X_1)}, \tag{4.2}$$

$$= \rho^2_{(X_2\to X_3)} + \rho^2_{(X_1\to X_3|X_2)}, \tag{4.3}$$

$$= \rho^2_{(X_1\to X_3)} + \rho^2_{(X_2\to X_3)} + \rho^2_{(X_1X_2\to X_3)} - 2\gamma_{(X_1\to X_3,X_2\to X_3)}, \tag{4.4}$$

where

$$\rho^2_{(X_2\to X_3|X_1)} = \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}(r_{u_{1m_1},u_{2m_2}} - r_{u_{1m_1}})^2 p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}(u_{3m_3} - r)^2 p_{3m_3}},$$

$$\rho^2_{(X_1\to X_3|X_2)} = \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}(r_{u_{1m_1},u_{2m_2}} - r_{u_{2m_2}})^2 p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}(u_{3m_3} - r)^2 p_{3m_3}},$$

$$\rho^2_{(X_1X_2\to X_3)} = \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}(r_{u_{1m_1},u_{2m_2}} - r_{u_{1m_1}} - r_{u_{2m_2}} + r)^2 p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}(u_{3m_3} - r)^2 p_{3m_3}},$$

$$\gamma_{(X_1\to X_3,X_2\to X_3)} = \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}(r_{u_{1m_1}} - r)(r_{u_{2m_2}} - r) p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}(u_{3m_3} - r)^2 p_{3m_3}}.$$

PROOF *See Appendix F.*

From the decompositions in Eq. (4.2) and (4.3), we first notice that $\rho^2_{(X_2\to X_3|X_1)}$ and $\rho^2_{(X_1\to X_3|X_2)}$ are non-negative. Thus, we conclude that $\rho^2_{(X_1,X2\to X_3)}$ is non-decreasing when another independent variable is added to the subcopula regression, as given in Proposition 4.2 below:

PROPOSITION 4.2 *Consider a three-dimensional contingency table of ordinal variables*

X_1, X_2 and X_3 with the joint p.m.f. p_{m_1,m_2,m_3}. Suppose $\rho^2_{(X_1,X_2\to X_3)}$ is decomposed as in Eq. (4.2) and (4.3). Then we have

$$\rho^2_{(X_1,X_2\to X_3)} \geq \rho^2_{(X_1\to X_3)}, \quad \rho^2_{(X_1,X_2\to X_3)} \geq \rho^2_{(X_2\to X_3)}.$$

Note that the equality holds if and only if $r_{u_{1m_1},u_{2m_2}} = r_{u_{1m_1}}$ and $r_{u_{1m_1},u_{2m_2}} = r_{u_{2m_2}}$ for every (m_1, m_2). This suggests that the mean subcopula score for X_3 with respect to the conditional distribution of X_3 given (X_1, X_2) is the same as that with respect to the conditional distribution of X_3 given X_1 or X_2 only, independent of X_2 or X_1, respectively. This may further indicate that $\rho^2_{(X_2\to X_3|X_1)}$ ($\rho^2_{(X_1\to X_3|X_2)}$) quantifies the increment in the overall subcopula regression based association measure when $X_2(X_1)$ is added to the subcopula regression that $X_1(X_2)$ is already in. Thus, we call $\rho^2_{(X_2\to X_3|X_1)}$ in Eq. (4.2) and $\rho^2_{(X_1\to X_3|X_2)}$ in Eq. (4.3) *the conditional subcopula regression based association measures* of X_2 on X_3 given X_1 and of X_1 on X_3 given X_2, respectively.

Regarding the decomposition given in Eq. (4.4), we make the following observations. First, we interpret that $\rho^2_{(X_1X_2\to X_3)}$ may measure the contribution of the mean subcopula score for X_3 predicted by the interaction between X_1 and X_2 in the three-dimensional contingency table for (X_1, X_2, X_3). We call it *the interactive subcopula regression based association measure* of (X_1, X_2) on X_3 given X_1 and X_2. Secondly, we claim that $\gamma_{(X_1\to X_3,X_2\to X_3)}$ may measure the contribution of the (unnormalized) correlation between two mean subcopula scores for X_3 predicted by X_1 and X_2 in the two-dimensional marginal contingency tables for (X_1, X_3) and (X_2, X_3), respectively. Hence we call it *the correlative subcopula regression based association measure* of (X_1, X_2) on X_3. Thirdly, it can be easily shown that

$$\rho^2_{(X_1\to X_3)} + \rho^2_{(X_1\to X_3)} - 2\gamma_{(X_1\to X_3,X_2\to X_3)} = \frac{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}(r_{u_{1m_1}} - r_{u_{2m_2}})^2 p_{m_1,m_2}}{\sum_{m_3=1}^{M_3}(u_{3m_3} - r)^2 p_{3m_3}}. \tag{4.5}$$

where Eq. (4.5) is equal to 0 if and only if $r_{u_{1m_1}} = r_{u_{2m_2}}$ for every (m_1, m_2), i.e. X_1 and X_2 equipped with the corresponding subcopula scores and the joint p.m.f. provide the same information in predicting the mean subcopula score of X_3. According to Eq. (4.5), we find that the correlative association measure $\gamma_{(X_1\to X_3, X_2\to X_3)}$ tends to play a role in taking the (redundant/beneficial) explanatory power of X_1 and X_2 into account when the overall association measure $\rho^2_{(X_1,X_2\to X_3)}$ is computed. That is, when X_1 and X_2 provide the similar (different) information in predicting the mean subcopula score of X_3 in the sense that the deviation between two mean subcopula scores of X_3 predicted by X_1 and X_2, measured by $(r_{u_{1m_1}} - r_{u_{2m_2}})^2$ is small for all pairs of categories of X_1 and X_2 (large for at least one pair of categories of X_1 and X_2), the information from both of X_1 and X_2 (at least one of X_1 and X_2) tends to be redundant (beneficial) in understanding the association structure in the three-dimensional contingency table of (X_1, X_2, X_3).

We continue Example 3.1 below to illustrate the decompositions in Theorem 4.1 and the properties shown in Proposition 4.2.

EXAMPLE 4.2 *(Example 3.1 continued) Given the dependent variable X_3 and independent variables X_1, X_2 from Table 3.1, we compute each term in Eq. (4.2), (4.3) and (4.4). Recall that $\rho^2_{(X_1,X_2\to X_3)} = 1.000$. From Table 4.1, we first confirm the non-decreasing property of the proposed overall association measure with respect to the marginal association measures: $\rho^2_{(X_1,X_2\to X_3)} = 1 > \rho^2_{(X_1\to X_3)} = 0.728$ and $\rho^2_{(X_1,X_2\to X_3)} = 1 > \rho^2_{(X_2\to X_3)} = 0.054$. We also see that the (marginal/conditional) contribution of X_1 to the overall association measure is much larger than that of X_2: $\rho^2_{(X_1\to X_3)} = 0.728 > \rho^2_{(X_2\to X_3)} = 0.054$ and $\rho^2(X_1 \to X_3|X_2) = 0.946 > \rho^2(X_2 \to X_3|X_1) = 0.272$. We further notice that the interactive association measure $\rho^2_{(X_1X_2\to X_3)} = 0.244$ is much larger than $\rho^2_{(X_2\to X_3)} = 0.054$. This implies that the contribution of the interaction between X_1 and X_2 on X_3 is non-negligible. Finally, we observe that $\gamma_{(X_1\to X_3,X_2\to X_3)} = 0.013$ and $\rho^2_{(X_1\to X_3)} + \rho^2_{(X_2\to X_3)} - 2\gamma_{(X_1\to X_3,X_2\to X_3)} = 0.728 + 0.054 - 2 * 0.013 = 0.756$. These indicate that X_1 provides essential information for the overall association structure while*

Decomposition	$\rho^2_{(X_1\to X_3)}$	$\rho^2_{(X_2\to X_3)}$	$\rho^2_{(X_2\to X_3\mid X_1)}$	$\rho^2_{(X_1\to X_3\mid X_2)}$	$\rho^2_{(X_1X_2\to X_3\mid X_1,X_2)}$	$\gamma_{(X_1\to X_3,X_2\to X_3)}$
Eq. (4.2)	0.728	–	0.272	–	–	–
Eq. (4.3)	–	0.054	–	0.946	–	–
Eq. (4.4)	0.728	0.054	–	–	0.244	0.013

Table 4.1: Decompositions of $\rho^2_{(X_1,X_2\to X_3)}$

X_2 provides little but still non-negligible information.

To further understand how the types of association between the dependent variable and independent variables alters the relationship between the overall and (marginal/conditional/interactive/correlative) association measures in a systematic way, we investigate the properties of the decompositions given in Theorem 4.1 under four types of independence available in the three-dimensional contingency table: mutual, joint, marginal and conditional independence.

PROPOSITION 4.3

(a) *If X_3 is jointly independent of X_1 and X_2 (i.e. $p_{m_1,m_2,m_3} = p_{m_1,m_2}p_{m_3}$ for all the combinations of m_1, m_2 and m_3), then $\rho^2_{(X_1,X_2\to X_3)} = 0$. Hence, $\rho^2_{(X_1\to X_3)} = 0$ and $\rho^2_{(X_2\to X_3)} = 0$.*

(b) *If X_1, X_2 and X_3 are mutually (or complete) independent (i.e. $p_{m_1,m_2,m_3} = p_{m_1}p_{m_2}p_{m_3}$ for all the combinations of m_1, m_2 and m_3), then $\rho^2_{(X_1,X_2\to X_3)} = 0$. Hence, $\rho^2_{(X_1\to X_3)} = 0$ and $\rho^2_{(X_2\to X_3)} = 0$. Note that the mutual independence is a special case of joint independence.*

(c) *If X_3 is marginally independent of X_1 in the marginal contingency table for (X_1, X_3) (i.e. $p_{m_1,m_3} = p_{m_1}p_{m_3}$ for all the combinations of m_1 and m_3), then $\rho^2_{(X_1\to X_3)} = 0$ and $\gamma_{(X_1\to X_3,X_2\to X_3)} = 0$, and hence Eq. (4.4) becomes*

$$\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_2\to X_3)} + \rho^2_{(X_1X_2\to X_3)}$$

(d) If X_3 is marginally independent of X_2 in the marginal contingency table for (X_2, X_3) (i.e. $p_{m_2,m_3} = p_{m_2}p_{m_3}$ for all the combinations of m_2 and m_3), then $\rho^2_{(X_2\to X_3)} = 0$ and $\gamma_{(X_1\to X_3, X_2\to X_3)} = 0$, and hence Eq. (4.4) becomes

$$\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_1\to X_3)} + \rho^2_{(X_1X_2\to X_3)}.$$

(e) If X_1 is marginally independent of X_2 in the marginal contingency table for (X_1, X_2) (i.e. $p_{m_1,m_2} = p_{m_1}p_{m_2}$ for all the combinations of m_1 and m_2), then $\gamma_{(X_1\to X_3, X_2\to X_3)}$ $= 0$ and hence Eq. (4.4) becomes

$$\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_1\to X_3)} + \rho^2_{(X_2\to X_3)} + \rho^2_{(X_1X_2\to X_3)} \tag{4.6}$$

(f) If X_3 is conditionally independent of X_1 given X_2 (i.e. $p_{m_3|m_1,m_2} = p_{m_3|m_2}$ for all the combinations of m_1, m_2 and m_3), then $\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_2\to X_3)}$.

(g) If X_3 is conditionally independent of X_2 given X_1 (i.e. $p_{m_3|m_1,m_2} = p_{m_3|m_1}$ for all the combinations of m_1, m_2 and m_3), then $\rho^2_{(X_1,X_2\to X_3)} = \rho^2_{(X_1\to X_3)}$.

PROOF *See Appendix G.*

Proposition 4.3 (a) and (b) show that the proposed overall and marginal association measures are all zero when X_3 is jointly independent of X_1 and X_2. (c) implies that, if X_3 is marginally independent of X_1, the marginal association between X_3 and X_1 is zero and the mean subcopula score for X_3 predicted by X_1 is equal to the unconditional mean subcopula score of X_3. Note that the contribution of the mean subcopula score for X_3 predicted by the interaction between X_1 and X_2 still remains. Similarly, we can interpret (d) by considering the marginal independence between X_3 and X_2.

Next, (e) indicates that when two independent variables are marginally independent, the correlative association measure in Eq. (4.4) becomes zero because the (unnormalized)

correlation between two mean subcopula scores $r_{u_{1m_1}}$ and $r_{u_{2m_2}}$ is zero for the marginal independence of X_1 and X_2. Thus, $\rho^2_{(X_1 \to X_3)} + \rho^2_{(X_2 \to X_3)} \leq \rho^2_{(X_1, X_2 \to X_3)}$. The equality holds if and only if $r_{u_{1m_1}, u_{2m_2}} - r_{u_{1m_1}} - r_{u_{2m_2}} + r = 0$ for each (m_1, m_2), i.e. $r_{u_{1m_1}, u_{2m_2}} - r_{u_{1m_1}} = r_{u_{2m_2}} - r$ and $r_{u_{1m_1}, u_{2m_2}} - r_{u_{2m_2}} = r_{u_{1m_1}} - r$, which means no contribution of the interaction between X_1 and X_2 in predicting the mean subcopula score of X_3.

Finally, (f) means that, if X_3 is conditionally independent of X_1 given X_2, then X_1 will make no contribution to the overall subcopula regression based association measure. We can apply a similar interpretation for (g).

Proposition 4.4 below investigates the implications of the converse of Proposition 4.3 (a), (b), (f) and (g).

PROPOSITION 4.4

(a) If $\rho^2_{(X_1, X_2 \to X_3)} = 0$, then $r_{u_{1m_1}, u_{2m_2}}$ for every (m_1, m_2) is a constant and equal to r. If $\rho^2_{(X_1 \to X_3)} = 0$ or $\rho^2_{(X_2 \to X_3)} = 0$, then $r_{u_{1m_1}} = r$ for every m_1 or $r_{u_{2m_2}} = r$ for every m_2.

(b) If $\rho^2_{(X_1, X_2 \to X_3)} = \rho^2_{(X_1 \to X_3)}$, then $r_{u_{1m_1}, u_{2m_2}} = r_{u_{1m_1}}$ for every (m_1, m_2).

(c) If $\rho^2_{(X_1, X_2 \to X_3)} = \rho^2_{(X_2 \to X_3)}$, then $r_{u_{1m_1}, u_{2m_2}} = r_{u_{2m_2}}$ for every (m_1, m_2).

(d) If $\rho^2_{(X_1, X_2 \to X_3)} = \rho^2_{(X_1 \to X_3)} + \rho^2_{(X_2 \to X_3)}$, then $r_{u_{1m_1}, u_{2m_2}} - r_{u_{1m_1}} - r_{u_{2m_2}} + r = 0$ for every (m_1, m_2), and $r_{u_{1m_1}}$ and $r_{u_{2m_2}}$ are uncorrelated.

PROOF *See Appendix H.*

In Proposition 4.4, (a) implies that the mean subcopula score for X_3 with respect to the conditional distribution of X_3 given (X_1, X_2) is constant over every combination of the categories of X_1 and X_2. It also implies that the mean subcopula score for X_3 with respect to the conditional distribution of X_3 given X_1 or X_2 is also constant over each category of X_1 or X_2, respectively. (b) and (c) indicate that no additional contribution of

X_2 (X_1) to the prediction of the mean subcopula score of X_3 is made for every combination of the categories of X_1 and X_2. (d) suggests that there is no additional contribution of the interaction and correlation between X_1 and X_2 to the prediction of the mean subcopula score of X_3 for every combination of the categories of X_1 and X_2, and two mean scores for X_3 predicted by X_1 and X_2 are not linearly correlated.

4.2.2 Decompositions for d-dimensional contingency tables

We consider the decompositions of the overall subcopula regression based association measure $\rho^2_{(\mathbf{X}_{-i} \to X_i)}$ in the d-dimensional contingency table with respect to a sequence of variables $\mathbf{X}_{(d-1)} = \{X_{(1)}, \ldots, X_{(d-1)}\}$ which reflects the order of the independent variables in $\mathbf{X}_{-i} = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{d-1}\}$ entering the subcopula regression for X_i. That is, $X_{(1)}$ and $X_{(d-1)}$ are the first and last independent variable entering the subcopula regression for X_i, respectively. We first let $\mathbf{X}^P_{(1)}, \ldots, \mathbf{X}^P_{(J)}$ form a partition P of $\mathbf{X}_{(d-1)}$, where $1 \le J \le d-1$. Then we let $\mathbf{X}^\Omega_{(j)} = \mathbf{X}^P_{(1)} \cup \ldots \cup \mathbf{X}^P_{(j)}$ denote the union of the first j sets in the partition, where $1 \le j \le J$. Finally, according to Section 3.1 and 3.2, we define $\mathbf{M}^P_{(j)}$, $\mathbf{m}^P_{(j)}$, $\mathbf{U}^P_{(j)}$ and $\mathbf{u}^P_{(j)}$ for each $\mathbf{X}^P_{(j)}$, and $\mathbf{M}^\Omega_{(j)}$, $\mathbf{m}^\Omega_{(j)}$, $\mathbf{U}^\Omega_{(j)}$ and $\mathbf{u}^\Omega_{(j)}$ for each $\mathbf{X}^\Omega_{(j)}$.

Additionally, we let $r_{\mathbf{m}^\Omega_{(j)}} = r^C_{U_i|\mathbf{U}^\Omega_{(j)}}(\mathbf{u}^\Omega_{(j)_{\mathbf{m}^\Omega_{(j)}}})$, $r_{\mathbf{m}^P_{(j+1)}} = r^C_{U_i|\mathbf{U}^P_{(j+1)}}(\mathbf{u}^P_{(j+1)_{\mathbf{m}^P_{(j+1)}}})$ and $r_{\mathbf{m}^\Omega_{(j)}, \mathbf{m}^P_{(j+1)}} = r^C_{U_i|\mathbf{U}^\Omega_{(j)}, \mathbf{U}^P_{(j+1)}}(\mathbf{u}^\Omega_{(j)_{\mathbf{m}^\Omega_{(j)}}}, \mathbf{u}^P_{(j+1)_{\mathbf{m}^P_{(j+1)}}})$. Here $r_{\mathbf{m}^\Omega_{(j)}}$, $r_{\mathbf{m}^P_{(j+1)}}$ and $r_{\mathbf{m}^\Omega_{(j)}, \mathbf{m}^P_{(j+1)}}$ are the subcopula regressions of $\mathbf{X}^\Omega_{(j)}$, $\mathbf{X}^P_{(j+1)}$ and $\mathbf{X}^\Omega_{(j+1)} = \mathbf{X}^\Omega_{(j)} \cup \mathbf{X}^P_{(j+1)}$ on X_i, respectively. Furthermore, let $r = E_C(U_i)$. Note that we have $r = E_C(r_{\mathbf{m}^\Omega_{(j)}}) = E_C(r_{\mathbf{m}^P_{(j+1)}}) = E_C(r_{\mathbf{m}^\Omega_{(j)}, \mathbf{m}^P_{(j+1)}})$.

For the purpose of clear illustration of the decompositions, we define *the conditional, interactive and correlative subcopula regression based association measures* for d-dimensional contingency tables below.

DEFINITION 4.2 *The conditional subcopula regression based association measure of*

$\mathbf{X}^{P}_{(j+1)}$ *on* X_i *given* $\mathbf{X}^{\Omega}_{(j)}$ *is defined by*

$$\rho^2_{(\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)})} = \frac{Var\left(r_{\mathbf{m}^{\Omega}_{(j)}, \mathbf{m}^{P}_{(j+1)}} - r_{\mathbf{m}^{\Omega}_{(j)}}\right)}{Var(U_i)}$$

$$= \frac{\sum_{\mathbf{m}^{\Omega}_{(j)} = \mathbf{1}^{\Omega}_{(j)}}^{\mathbf{M}^{\Omega}_{(j)}} \sum_{\mathbf{m}^{P}_{(j+1)} = \mathbf{1}^{P}_{(j+1)}}^{\mathbf{M}^{P}_{(j+1)}} \left(r_{\mathbf{m}^{\Omega}_{(j)}, \mathbf{m}^{P}_{(j+1)}} - r_{\mathbf{m}^{\Omega}_{(j)}}\right)^2 p_{\mathbf{m}^{\Omega}_{(j)}, \mathbf{m}^{P}_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}},$$

where $1 \leq j \leq J$.

The non-negative conditional association measure in Definition 4.2 quantifies the contribution of $\mathbf{X}^{P}_{(j+1)}$ to the overall association measure given that $\mathbf{X}^{\Omega}_{(j)}$ is already in the subcopula regression function. One important property of the conditional association measure is additivity, as shown in Proposition 4.5 below:

PROPOSITION 4.5 *Consider a* d*-dimensional contingency table of ordinal variables* $\mathbf{X}_d$ *with the joint p.m.f.* $p_{\mathbf{m}_d}$*. For a new partition* P^* *of* $\mathbf{X}_{(d-1)}$ *where* $\mathbf{X}^{P^*} = \mathbf{X}^{P}_{(j+1)} \cup \mathbf{X}^{P}_{(j+2)}$*, we have*

$$\rho^2_{(\mathbf{X}^{P^*} \to X_i | \mathbf{X}^{\Omega}_{(j)})} = \rho^2_{(\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)})} + \rho^2_{(\mathbf{X}^{P}_{(j+2)} \to X_i | \mathbf{X}^{\Omega}_{(j+1)})},$$

where $\mathbf{X}^{\Omega}_{(j+1)} = \mathbf{X}^{P}_{(j+1)} \cup \mathbf{X}^{\Omega}_{j}$.

PROOF *See Appendix F.*

DEFINITION 4.3 *The interactive subcopula regression based association measure of* $(\mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})$ *on* X_i *given* $\mathbf{X}^{\Omega}_{(j)}$ *and* $\mathbf{X}^{P}_{(j+1)}$ *is defined by*

$$\rho^2_{(\mathbf{X}^{\Omega}_{(j)} \mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = \frac{Var\left(r_{\mathbf{m}^{\Omega}_{(j)}, \mathbf{m}^{P}_{(j+1)}} - r_{\mathbf{m}^{\Omega}_{(j)}} - r_{\mathbf{m}^{P}_{(j+1)}}\right)}{Var(U_i)}$$

$$=\frac{\sum_{\mathbf{m}_{(j)}^{\Omega}=\mathbf{1}_{(j)}^{\Omega}}^{\mathbf{M}_{(j)}^{\Omega}}\sum_{\mathbf{m}_{(j+1)}^{P}=\mathbf{1}_{(j+1)}^{P}}^{\mathbf{M}_{(j+1)}^{P}}\left(r_{\mathbf{m}_{(j)}^{\Omega},\mathbf{m}_{(j+1)}^{P}}-r_{\mathbf{m}_{(j)}^{\Omega}}-r_{\mathbf{m}_{(j+1)}^{P}}+r\right)^{2}p_{\mathbf{m}_{(j)}^{\Omega},\mathbf{m}_{(j+1)}^{P}}}{\sum_{m_i=1}^{M_i}\left(u_{i_{m_i}}-r\right)^2 p_{m_i}}.$$

The non-negative interactive association measure in Definition 4.3 explains the contribution of the mean subcopula score for X_i predicted by the interaction between $\mathbf{X}_{(j)}^{\Omega}$ and $\mathbf{X}_{(j+1)}^{P}$ to the overall association measure.

DEFINITION 4.4 *The correlative subcopula regression based association measure of* $(\mathbf{X}_{(j)}^{\Omega}, \mathbf{X}_{(j+1)}^{P})$ *on* X_i *is defined by*

$$\begin{aligned}\gamma_{(\mathbf{X}_{(j)}^{\Omega}\to X_i,\mathbf{X}_{(j+1)}^{P}\to X_i)} &= \frac{Cov\left(r_{\mathbf{m}_{(j)}^{\Omega}}-r, r_{\mathbf{m}_{(j+1)}^{P}}-r\right)}{Var(U_i)}\\ &= \frac{\sum_{\mathbf{m}_{(j)}^{\Omega}=\mathbf{1}_{(j)}^{\Omega}}^{\mathbf{M}_{(j)}^{\Omega}}\sum_{\mathbf{m}_{(j+1)}^{P}=\mathbf{1}_{(j+1)}^{P}}^{\mathbf{M}_{(j+1)}^{P}}\left(r_{\mathbf{m}_{(j)}^{\Omega}}-r\right)\left(r_{\mathbf{m}_{(j+1)}^{P}}-r\right)p_{\mathbf{m}_{(j)}^{\Omega},\mathbf{m}_{(j+1)}^{P}}}{\sum_{m_i=1}^{M_i}\left(u_{i_{m_i}}-r\right)^2 p_{m_i}}.\end{aligned}$$

The correlative association measure in Definition 4.4 represents the contribution of the (unnormalized) correlation between the two mean subcopula scores for X_i predicted by $\mathbf{X}_{(j)}^{\Omega}$ and $\mathbf{X}_{(j+1)}^{P}$ separately to the overall association measure.

We also notice that the conditional, interactive and correlative association measures are invariant with respect to the permutation on the categories of the independent variables and binary dependent variable, as shown in Proposition 4.6 below.

PROPOSITION 4.6 *Consider a* d*-dimensional contingency table of ordinal variables* $\mathbf{X}_d$ *with the joint p.m.f.* $p_{\mathbf{m}_d}$*. Let* $\tilde{\mathbf{X}}_{-i}$ *denote the permuted* $\mathbf{X}_{-i}$ *defined in Proposition 3.4 and* $\tilde{X}_i$ *denote the permuted binary* X_i*.*

(a) The conditional subcopula regression based association measure is invariant with

respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$. That is

$$\rho^2_{(\tilde{\mathbf{X}}^P_{(j+1)}\to X_i|\tilde{\mathbf{X}}^\Omega_{(j)})} = \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})}, \quad \rho^2_{(\mathbf{X}^P_{(j+1)}\to \tilde{X}_i|\mathbf{X}^\Omega_{(j)})} = \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})}.$$

(b) *The interactive subcopula regression based association measure is invariant with respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$. That is*

$$\rho^2_{(\tilde{\mathbf{X}}^\Omega_{(j)}\tilde{\mathbf{X}}^P_{(j+1)}\to X_i|\tilde{\mathbf{X}}^\Omega_{(j)},\tilde{\mathbf{X}}^P_{(j+1)})} = \rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})},$$
$$\rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to \tilde{X}_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} = \rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to \tilde{X}_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})}.$$

(c) *The correlative subcopula regression based association measure is invariant with respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$. That is*

$$\gamma_{(\tilde{\mathbf{X}}^\Omega_{(j)}\to X_i,\tilde{\mathbf{X}}^P_{(j+1)}\to X_i)} = \gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)},$$
$$\gamma_{(\mathbf{X}^\Omega_{(j)}\to \tilde{X}_i,\mathbf{X}^P_{(j+1)}\to \tilde{X}_i)} = \gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)}.$$

PROOF *See Appendix F.*

Now we extend the decompositions of $\rho^2_{(X_1,X_2\to X_3)}$ for three-dimensional tables in Theorem 4.1 to a general case of the decompositions of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ for d-dimensional tables below.

THEOREM 4.2 *Consider a d-dimensional contingency table of ordinal variables $\mathbf{X}_d$ with the joint p.m.f. $p_{\mathbf{m}_d}$.*

(a) *The sequential decomposition of the overall subcopula regression based measure $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is given by*

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)}\to X_i)} + \sum_{j=1}^{J-1} \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})}, \tag{4.7}$$

(b) The non-sequential decomposition of the overall subcopula regression based measure $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is given by

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \sum_{j=1}^{J} \rho^2_{(\mathbf{X}^P_{(j)}\to X_i)} + \sum_{j=1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} - 2\sum_{j=1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)}, \tag{4.8}$$

where

$$\rho^2_{(\mathbf{X}^P_{(j)}\to X_i)} = \frac{\sum\limits_{\mathbf{m}^P_{(j)}=\mathbf{1}^P_{(j)}}^{\mathbf{M}^P_{(j)}} \left(r_{\mathbf{m}^P_{(j)}} - r\right)^2 p_{\mathbf{m}^P_{(j)}}}{\sum\limits_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}}.$$

(c) The sequential and non-sequential decomposition of the overall subcopula regression based measure in Eq. (4.7) and (4.8) are invariant with respect to the permutation on the categories of every independent variable $X_j \in \mathbf{X}_{-i}$ and the binary dependent variable X_i.

PROOF *See Appendix F.*

Note that Eq. (4.7) is the extension of Eq. (4.2) and (4.3) in Theorem 4.1, while Eq. (4.8) is the extension of Eq. (4.4) in Theorem 4.1. Moreover, the sequential and non-sequential decompositions are still valid even when the independent variables are nominal and the dependent variable is binary, according to Theorem 4.2 (c).

We call Eq. (4.7) the *sequential decomposition* because it sums up the conditional association measures, each of which quantifies the contribution of a newly added set of independent variables $\mathbf{X}^P_{(j+1)}$ to the overall subcopula association measure $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ given that all the previous independent variables $\mathbf{X}^\Omega_{(j)}$ are already in the subcopula regression for X_i. For example, consider that $X_1, \ldots, X_5$ are the ordinal variables in a five-dimensional contingency table where X_5 is the dependent variable and $X_1, \ldots, X_4$ are

the independent variables. Suppose that the order of $X_1, \dots, X_4$ entering the subcopula regression for X_5 is captured by $\mathbf{X}_{(4)} = \{X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}\} = \{X_4, X_2, X_1, X_3\}$. We partition $\mathbf{X}_{(4)}$ by $\mathbf{X}^P_{(1)} = \{X_{(1)}\} = \{X_4\}$, $\mathbf{X}^P_{(2)} = \{X_{(2)}, X_{(3)}\} = \{X_2, X_1\}$ and $\mathbf{X}^P_{(3)} = \{X_{(4)}\} = \{X_3\}$. Then we obtain the unions $\mathbf{X}^\Omega_{(1)} = \mathbf{X}^P_{(1)} = \{X_{(1)}\} = \{X_4\}$ and $\mathbf{X}^\Omega_{(2)} = \mathbf{X}^P_{(1)} \cup \mathbf{X}^P_{(2)} = \{X_{(1)}, X_{(2)}, X_{(3)}\} = \{X_4, X_2, X_1\}$. Therefore, according to Eq. (4.7), the sequential decomposition of $\rho^2_{(X_1,X_2,X_3,X_4 \to X_5)}$ is given by

$$\begin{aligned}\rho^2_{(X_1,X_2,X_3,X_4 \to X_5)} &= \rho^2_{(\mathbf{X}^P_{(1)} \to X_5)} + \rho^2_{(\mathbf{X}^P_{(2)} \to X_5 | \mathbf{X}^\Omega_{(1)})} + \rho^2_{(\mathbf{X}^P_{(3)} \to X_5 | \mathbf{X}^\Omega_{(2)})} \\ &= \rho^2_{(X_4 \to X_5)} + \rho^2_{(X_2,X_1 \to X_5 | X_4)} + \rho^2_{(X_3 \to X_5 | X_4,X_2,X_1)}.\end{aligned}$$

In contrast, we call Eq. (4.8) the *non-sequential decomposition* because it contains no conditional association measures but the marginal ones for the J sets of independent variables, which doesn't explicitly reflect the order of the independent variables entering the subcopula regression for X_i. For example, consider the same five-dimensional contingency table used for illustrating the sequential decomposition above. For the same partition $\mathbf{X}^P_{(1)}, \mathbf{X}^P_{(2)}, \mathbf{X}^P_{(3)}$ and unions $\mathbf{X}^\Omega_{(1)}, \mathbf{X}^\Omega_{(2)}$ above, the non-sequential decomposition of $\rho^2_{(X_1,X_2,X_3,X_4 \to X_5)}$ based on Eq. (4.8) is then given by

$$\begin{aligned}\rho^2_{(X_1,X_2,X_3,X_4 \to X_5)} =& \rho^2_{(\mathbf{X}^P_{(1)} \to X_5)} + \rho^2_{(\mathbf{X}^P_{(2)} \to X_5)} + \rho^2_{(\mathbf{X}^P_{(3)} \to X_5)} + \rho^2_{(\mathbf{X}^\Omega_{(1)} \mathbf{X}^P_{(2)} \to X_5 | \mathbf{X}^\Omega_{(1)}, \mathbf{X}^P_{(2)})} \\ &+ \rho^2_{(\mathbf{X}^\Omega_{(2)} \mathbf{X}^P_{(3)} \to X_5 | \mathbf{X}^\Omega_{(2)}, \mathbf{X}^P_{(3)})} - 2\rho^2_{(\mathbf{X}^\Omega_{(1)} \to X_5, \mathbf{X}^P_{(2)} \to X_5)} - 2\rho^2_{(\mathbf{X}^\Omega_{(2)} \to X_5, \mathbf{X}^P_{(3)} \to X_5)} \\ =& \rho^2_{(X_4 \to X_5)} + \rho^2_{(X_2,X_1 \to X_5)} + \rho^2_{(X_3 \to X_5)} + \rho^2_{(X_4 X_2 X_1 \to X_5 | X_4,X_2,X_1)} \\ &+ \rho^2_{(X_4 X_2 X_1 X_3 \to X_5 | X_4,X_2,X_1,X_3)} - 2\gamma_{(X_4 \to X_5, X_2,X_1 \to X_5)} \\ &- 2\gamma_{(X_4,X_2,X_1 \to X_5, X_3 \to X_5)}.\end{aligned}$$

In Proposition 4.7 below, we extend the properties of the decompositions of the subcopula regression based association measure under four types of independence given in Proposition 4.3 for d-dimensional contingency tables. Recall the four types of independence are: joint, mutual, marginal and conditional independence.

PROPOSITION 4.7

(a) *If X_i is jointly independent of $\mathbf{X}_{-i}$, then $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and hence $\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} = 0$ for $1 \le k \le d-1$.*

(b) *If $\mathbf{X}_d = \{X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_d\}$ are mutually (or complete) independent, then $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and hence $\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} = 0$ for $1 \le k \le d-1$. Note that the mutual independence is a special case of joint independence.*

(c) *Suppose that X_i is marginally independent of $\mathbf{X}^\Omega_{(j)}$ in the marginal contingency table of $(\mathbf{X}^\Omega_{(j)}, X_i)$. For the sequential decomposition, we then have $\rho^2_{(\mathbf{X}^P_{(1)}\to X_i)} = 0$ and $\rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})} = 0$ for $1 < k \le j-1$. Thus Eq. (4.7) becomes*

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i)} + \sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})}.$$

For the non-sequential decomposition, we then have $\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} = 0$ and $r_{\mathbf{m}^\Omega_{(k)}} = r$ for all $\mathbf{m}^\Omega_{(k)}$, where $k = 1, \ldots, j$. Thus Eq. (4.8) becomes

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i}\to X_i)} = &\sum_{k=j+1}^{J} \rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i|\mathbf{X}^P_{(j+1)})} + \\ &\sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)},\mathbf{X}^P_{(k+1)})} - 2\sum_{k=j+1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)}.\end{aligned}$$

(d) *Suppose that $\mathbf{X}^P_{(j+1)}$ is marginally independent of $\mathbf{X}^\Omega_{(j)}$ in the marginal contingency table of $\mathbf{X}^\Omega_{(j+1)}$ for $1 \le j \le J-1$. For the non-sequential decomposition, we then have $\gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)} = 0$. Thus Eq. (4.8) becomes*

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i}\to X_i)} = &\sum_{k=1}^{J} \rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \sum_{k=1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)},\mathbf{X}^P_{(k+1)})} \\ &- 2\sum_{k=1}^{j-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)} - 2\sum_{k=j+1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)}.\end{aligned}$$

(e) *Suppose that X_i and $\mathbf{X}^P_{(j+1)}$ are conditionally independent given $\mathbf{X}^\Omega_{(j)}$. For the sequential decomposition, we then have $\rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ and hence Eq. (4.7) becomes*

$$\rho^2_{(\mathbf{X}_{-i} \to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)} \to X_i)} + \sum_{k=1}^{j-1} \rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})} + \sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})}.$$

For the non-sequential decomposition, we then have $r_{\mathbf{m}^\Omega_{(j)}, \mathbf{m}^P_{(j+1)}} = r_{\mathbf{m}^\Omega_{(j)}}$. Thus Eq. (4.8) becomes

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i} \to X_i)} =& \sum_{k=1}^{j} \rho^2_{(\mathbf{X}^P_{(k)} \to X_i)} + \sum_{k=j+2}^{J-1} \rho^2_{(\mathbf{X}^P_{(k)} \to X_i)} + \sum_{k=1}^{j-1} \rho^2_{(\mathbf{X}^\Omega_{(k)} \mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)}, \mathbf{X}^P_{(k+1)})} \\ &+ \sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(k)} \mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)}, \mathbf{X}^P_{(k+1)})} - 2 \sum_{k=1}^{j-1} \gamma_{(\mathbf{X}^\Omega_{(k)} \to X_i, \mathbf{X}^P_{(k+1)} \to X_i)} \\ &- 2 \sum_{k=j+1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(k)} \to X_i, \mathbf{X}^P_{(k+1)} \to X_i)}.\end{aligned}$$

PROOF *See Appendix G.*

Similar to Proposition 4.4, we finally explore the property of the converse of Proposition 4.7 (a), (b) and (e) for d-dimensional tables.

PROPOSITION 4.8

(a) *For $1 \le j \le J$, if $\rho^2_{(\mathbf{X}^P_{(j)} \to X_i)} = 0$, then $r_{\mathbf{m}^P_{(j)}}$ for every $\mathbf{m}^P_{(j)}$ is a constant and equal to r.*

(b) *For $1 \le j \le J$, if*

$$\rho^2_{(\mathbf{X}_{-i} \to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)} \to X_i)} + \sum_{k=1}^{j-1} \rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})},$$

then we have $\rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})} = 0$ and hence $r_{\mathbf{m}^\Omega_{(k)}, \mathbf{m}^P_{(k+1)}} = r_{\mathbf{m}^\Omega_{(k)}}$ for every $\mathbf{m}^\Omega_{(k)}$ and $\mathbf{m}^P_{(k+1)}$, where $k = j, \ldots, J-1$.

PROOF *See Appendix H.*

4.3 Use of the overall association measure and its decompositions for descriptive modeling

For the purpose of descriptive modeling, the overall subcopula regression association measure and its decompositions proposed in Chapter 4.1 and 4.2 can be applied in the following ways.

First, the sequential and non-sequential decomposition can be utilized to deepen the understanding of the association structure in a model-free manner. This is because they can provide the qualification of (marginal/conditional/interactive/correlative) contribution for a set of independent variables of interest to the overall association measure for the full-dimensional contingency table.

Secondly, the overall subcopula regression based association measure can be adopted to quantify the association structure between the dependent variable and every subset of the independent variables, namely *all-possible-subset subcopula regressions*. That is, we can calculate the association measures for all possible subsets of the independent variables, identify one or a few subsets of the independent variables with the largest or larger value of the association measure per subset size and use them for further statistical inference. Here the subset size refers to the number of independent variables included in the subcopula regression.

Thirdly, one can utilize the sequential decomposition of the overall association measure for the full-dimensional contingency table to obtain the conditional contribution of every independent variable to the overall association and build up the best subset of independent variables. To be specific, one can add a new independent variable with the largest conditional association measure to the existing subset of independent variables already in the subcopula regression until a newly added variable has a negligible value of conditional association measure.

We will illustrate in Chapter 7.2 and 7.3 the application of the sequential and non-

sequential decomposition of the overall association measure using the real data sets *acute migraine* (Vandenhende and Lambert, 2000) and *nuclear accident* (Fienberg et al., 1985), respectively. For the *acute migraine* data, we will use the (sequential/non-sequential) method of decomposition to quantify different types of contribution of the independent variables to the overall association measure. For the analysis of *nuclear accident*, we will show the use of the marginal and conditional association measures from the sequential decomposition of the overall association measure to capture the time-dependent association structure.

We will also illustrate the process of using *all-possible-subset subcopula regressions* to build up the best subset of independent variables in Chapter 7.4, with a real data set called *post-operative patients* (Budihardjo et al., 1991) whose main goal is to determine the place that a patient in the post-operative recovery area should be sent to next, which is related to a classification problem in statistical learning.

CHAPTER 5

STATISTICAL INFERENCE FOR THE SUBCOPULA REGRESSION AND ITS ASSOCIATION MEASURES

In this chapter, we study the point estimation of the proposed subcopula scores, subcopula regression function, subcopula regression based association measure and the decompositions (i.e. the sequential and non-sequential decomposition) for d-dimensional tables. We also investigate the interval estimation of the proposed association measures to quantify the uncertainty of their point estimators. Furthermore, we perform the permutation test to confirm the statistical significance of the observed patterns of association between the ordinal dependent variable and a set of independent variables.

First, we follow the plug-in principle using the relative frequencies in a d-dimensional contingency table to calculate the point estimation of the proposed subcopula scores, subcopula regression, its overall association measure and the corresponding decompositions.

Then we consider the interval estimation of the proposed measures using two approaches. The first approach is to use the asymptotic distribution of the point estimator of the overall subcopula regression based association measure to construct the Wald-type asymptotic confidence interval. The second approach is to employ the bias-corrected and accelerated (BCa) bootstrap confidence interval (Efron and Tibshirani, 1994; Davison and Hinkley, 1997). Note that we consider only the bootstrap confidence intervals for the prediction of the category of the dependent variable and the decompositions of the overall subcopula regression based association measure.

At last, we design the permutation tests for the hypotheses of the proposed overall association measure, $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$, as well as its marginal and conditional association measures obtained from the sequential decomposition, in a d-dimensional contingency table to assess the statistical significance of the (joint/marginal/conditional) contribution of the independent variables.

5.1 Point estimation

Let $n_{\mathbf{m}_d} = n_{m_1,m_2,\ldots,m_d}$ denote the cell count in a d-dimensional contingency table corresponding to the m_1-th, m_2-th, $\ldots$, m_d-th category of $X_1, X_2, \ldots, X_d$, respectively. Then, the marginal count for $X_i = x_{i_{m_i}}$ is denoted by

$$n_{m_i} = \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} n_{\mathbf{m}_d} = \sum_{m_1=1}^{M_1} \cdots \sum_{m_{i-1}=1}^{M_{i-1}} \sum_{m_{i+1}=1}^{M_{i+1}} \cdots \sum_{m_d=1}^{M_d} n_{m_1,m_2,\ldots,m_d},$$

the count for $\mathbf{X}_{-i} = \mathbf{x}_{\mathbf{m}_{-i}}$ is denoted by

$$n_{\mathbf{m}_{-i}} = \sum_{m_i=1}^{M_i} n_{\mathbf{m}_d} = \sum_{m_i=1}^{M_i} n_{m_1,m_2,\ldots,m_d}$$

and the total count for $\mathbf{X}_d = x_{d_{\mathbf{m}_d}}$ is denoted by

$$n = \sum_{\mathbf{m}_d=\mathbf{1}_d}^{\mathbf{M}_d} n_{\mathbf{m}_d} = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_d=1}^{M_d} n_{m_1,m_2,\ldots,m_d}.$$

We define the estimator for the marginal p.m.f. p_{m_i} of X_i by

$$\hat{p}_{i_{m_i}} = \frac{n_{m_i}}{n},$$

and the estimator for the conditional p.m.f. $p_{m_i|\mathbf{m}_{-i}}$ of X_i given $\mathbf{X}_{-i}$ by

$$\hat{p}_{m_i|\mathbf{m}_{-i}} = \frac{n_{\mathbf{m}_d}}{n_{\mathbf{m}_{-i}}}.$$

The set of subcopula scores of X_i is estimated by

$$\hat{\mathbf{D}}_{u_i} = \{\hat{u}_{i_0} = 0 < \cdots < \hat{u}_{i_{m_i}} < \cdots < \hat{u}_{i_{M_i}} = 1\},$$

where

$$\hat{u}_{i_{m_i}} = \sum_{r_i=1}^{m_i} \hat{p}_{r_i}.$$

Now the estimator for the subcopula regression function of X_i on $\mathbf{X}_{-i}$ in Eq. (3.7) is obtained by

$$\hat{r}^C_{U_i|\mathbf{U}_{-i}}(\hat{\mathbf{u}}_{\mathbf{m}_{-i}}) = \sum_{m_i=1}^{M_i} \hat{u}_{i_{m_i}} \hat{p}_{m_i|\mathbf{m}_{-i}}.$$

Following the prediction process introduced in Section 3.4, we can estimate the predicted category of X_i for a combination of categories of $\mathbf{X}_{-i}$ via the estimated subcopula regression above. Given the observed categories $\mathbf{x}_{\mathbf{m}_{-i}}$ for $\mathbf{X}_{-i}$ again, we first calculate the corresponding $\hat{\mathbf{u}}_{\mathbf{m}_{-i}}$ using the estimated marginal p.m.f. $\hat{p}_{\mathbf{m}_{-i}}$. Next, we compute the predicted value of $\hat{u}_i$ via the estimated subcopula regression, i.e. $\hat{u}_i^* = \hat{r}^C_{U_i|\mathbf{U}_{-i}}(\hat{\mathbf{u}}_{\mathbf{m}_{-i}})$. Then, from the estimated set of the subcopula scores of X_i, $\hat{\mathbf{D}}_{u_i}$, we locate the index $\hat{m}_i^*$ such that $\hat{u}_{i_{\hat{m}_i^*-1}} < \hat{u}_i^* \leq \hat{u}_{i_{\hat{m}_i^*}}$. Finally, we obtain the estimate of the predicted category of X_i, i.e. $x_{i_{\hat{m}_i^*}}$.

Next, the estimator for the subcopula regression based association measure in Eq.

(4.1) is given by

$$\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)} == \frac{\sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left(\sum_{m_i=1}^{M_i} \hat{u}_{i_{m_i}} \hat{p}_{m_i|\mathbf{m}_{-i}} - \hat{r} \right)^2 \hat{p}_{\mathbf{m}_{-i}}}{\sum_{m_i=1}^{M_i} \left(\hat{u}_{i_{m_i}} - \hat{r} \right)^2 \hat{p}_{m_i}}, \tag{5.1}$$

where $\hat{r} = \sum_{m_i=1}^{M_i} \hat{u}_{i_{m_i}} \hat{p}_{i_{m_i}}$

We can also obtain the estimator for the sequential decomposition of the subcopula regression based association measure in Eq. (4.7) as follows:

$$\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)} = \hat{\rho}^2_{(\mathbf{X}^P_{(1)}\to X_i)} + \sum_{j=1}^{J-1} \hat{\rho}^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})},$$

where

$$\hat{\rho}^2_{(\mathbf{X}^P_{(j)}\to X_i)} = \frac{\sum_{\mathbf{m}^P_{(j)}=\mathbf{1}^P_{(j)}}^{\mathbf{M}^P_{(j)}} \left(\hat{r}_{\mathbf{m}^P_{(j)}} - \hat{r} \right)^2 \hat{p}_{\mathbf{m}^P_{(j)}}}{\sum_{m_i=1}^{M_i} \left(\hat{u}_{i_{m_i}} - \hat{r} \right)^2 \hat{p}_{m_i}},$$

$$\hat{\rho}^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})} = \frac{\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(\hat{r}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} - \hat{r}_{\mathbf{m}^\Omega_{(j)}} \right)^2 \hat{p}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(\hat{u}_{i_{m_i}} - \hat{r} \right)^2 \hat{p}_{m_i}},$$

where $\hat{r}_{\mathbf{m}^P_{(j)}}$, $\hat{r}_{\mathbf{m}^\Omega_{(j)}}$ and $\hat{r}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}$ are the estimators for the subcopula regression functions $r^C_{U_i|\mathbf{U}^P_{(j)}}(\mathbf{u}^P_{(j)_{\mathbf{m}^P_{(j)}}})$, $r^C_{U_i|\mathbf{U}^\Omega_{(j)}}(\mathbf{u}^\Omega_{(j)_{\mathbf{m}^\Omega_{(j)}}})$ and $r^C_{U_i|\mathbf{U}^\Omega_{(j)},\mathbf{U}^P_{(j+1)}}(\mathbf{u}^\Omega_{(j)_{\mathbf{m}^\Omega_{(j)}}}, \mathbf{u}^P_{(j+1)_{\mathbf{m}^P_{(j+1)}}})$, respectively.

Lastly, the estimator for the non-sequential decomposition of the subcopula regres-

sion based association measure in Eq. (4.8) is given by

$$\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)} = \sum_{j=1}^{J} \hat{\rho}^2_{(\mathbf{X}^P_{(j)}\to X_i)} + \sum_{j=1}^{J-1} \hat{\rho}^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} - 2\sum_{j=1}^{J-1} \hat{\gamma}_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)},$$

where

$$\hat{\rho}^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} = \frac{\displaystyle\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(\hat{r}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} - \hat{r}_{\mathbf{m}^\Omega_{(j)}} - \hat{r}_{\mathbf{m}^P_{(j+1)}} + \hat{r}\right)^2 \hat{p}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\displaystyle\sum_{m_i=1}^{M_i} \left(\hat{u}_{i_{m_i}} - \hat{r}\right)^2 \hat{p}_{m_i}}.$$

and

$$\hat{\gamma}_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)} = \frac{\displaystyle\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(\hat{r}_{\mathbf{m}^\Omega_{(j)}} - \hat{r}\right)\left(\hat{r}_{\mathbf{m}^P_{(j+1)}} - \hat{r}\right) \hat{p}_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\displaystyle\sum_{m_i=1}^{M_i} \left(\hat{u}_{i_{m_i}} - \hat{r}\right)^2 \hat{p}_{m_i}}.$$

5.2 Interval estimation

For the practical significance of the estimates of the proposed association measures, it is important to estimate the size of any effect, e.g. by constructing the confidence intervals for them. In the following we first provide the asymptotic distribution of the point estimator of the overall subcopula regression based association measure $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ for constructing the asymptotic confidence interval. Then we introduce different bootstrap resampling approaches for constructing the bootstrap confidence intervals for all the proposed association measures. Finally, we describe how to utilize the bootstrap method to quantify the uncertainty in the predicted categories of the response variable obtained from the estimated subcopula regression.

5.2.1 Asymptotic confidence interval

We present the asymptotic distribution of the point estimator $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ in Theorem 5.1 below:

THEOREM 5.1 *Let $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ be the overall subcopula regression based association measure of X_i on $\mathbf{X}_{-i}$ defined in Eq. (4.1) and $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ be the corresponding estimator in Eq. (5.1). Suppose that $0 < \rho^2_{(\mathbf{X}_{-i}\to X_i)} < 1$. Then,*

$$\sqrt{n}\left(\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)} - \rho^2_{(\mathbf{X}_{-i}\to X_i)}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbf{\Sigma}\right),$$

where $\mathbf{p} = (p_{1,1,\ldots,1}, p_{2,1,\ldots,1}, \ldots, p_{M_1,1,\ldots,1}, \ldots, p_{M_1,M_2,\ldots,M_d})^T$,

$$\nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p}) = \left[\frac{\partial \rho^2_{(\mathbf{X}_{-i}\to X_i)}}{\partial p_{1,1,\ldots,1}}, \ldots, \frac{\partial \rho^2_{(\mathbf{X}_{-i}\to X_i)}}{\partial p_{M_1,M_2,\ldots,M_d}}\right]^T,$$
$$\mathbf{\Sigma} = \nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p})^T(diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^T)\nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p}),$$

$\frac{\partial \rho^2_{(\mathbf{X}_{-i}\to X_i)}}{\partial p_{m_i,\mathbf{m}_{-i}}}$ is given in Appendix J, $\mathcal{N}(\cdot)$ is a univariate normal distribution, and $diag(\mathbf{p})$ is a matrix with diagonal values equal to $\mathbf{p}$ and 0 elsewhere.

PROOF *See Appendix I.*

Theorem 5.1 allows us to obtain an estimator for the asymptotic variance of $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ by the plug-in principle and then construct the asymptotic Wald-type confidence interval.

5.2.2 Bootstrap confidence intervals

The asymptotic confidence interval obtained from Theorem 5.1 works well when the sample size is sufficiently large and the true value of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ of interest is away from the boundary of the parameter space (0 or 1). For small and moderate sized data sets, the bootstrap confidence interval is often adopted to take into account the finite sample

variation in terms of parameter estimation. Moreover, the bootstrap approach can preserve the range of the proposed subcopula regression based association measure which is between 0 and 1. For the bootstrap resampling from a multi-dimensional contingency table with the ordinal dependent variable and a set of independent variables, we consider three types of bootstrap implementation: parametric bootstrap, unconditional paired nonparametric bootstrap and conditional paired nonparametric bootstrap (Efron, 1981; Fox and Weisberg, 2018).

Parametric bootstrap starts with fitting the saturated log-linear model to the contingency table and then uses the estimated parameters in the saturated log-linear model to generate bootstrap samples.

Unconditional paired nonparametric bootstrap converts the full-dimensional contingency table into a case-form data set, applies the nonparametric bootstrap and converts back to a table-form data set. For example, consider a $3 \times 3 \times 2$ contingency table of ordinal variables X_1, X_2 and X_3. Suppose that X_1 has categories $\{x_{1_1}, x_{1_2}, x_{1_3}\}$, X_2 has categories $\{x_{2_1}, x_{2_2}, x_{2_3}\}$ and X_3 has categories $\{x_{3_1}, x_{3_2}\}$ as in Example 3.1. First, we convert the three-dimensional contingency table into a case-form data set, where the cases are defined by the combinations of the categories $(x_{1_1}, x_{2_1}, x_{3_1}), (x_{1_1}, x_{2_1}, x_{3_2}), \ldots, (x_{1_3}, x_{2_3}, x_{3_2})$. Note that the number of duplicates for each case is determined by its corresponding count in the original contingency table. After obtaining a bootstrap sample based on the case-form data set using the nonparametric bootstrap, we then convert it back to a three-dimensional contingency table that has the same layout as the original table.

Conditional paired nonparametric bootstrap first fixes the counts in the marginal contingency table of the independent variables and converts the marginal contingency table of the dependent variable into a case-form data set with respect to each combination of the categories of the independent variables. It then applies the non-parametric bootstrap to each case-form data set and integrates them back to the table form. For instance, suppose that X_3 is the dependent variable in the three-dimensional table as in Example 3.1. First,

we fix the counts in the marginal contingency table of (X_1, X_2). Next, given a combination of the categories of (X_1, X_2), we convert the marginal table of X_3 into a case-form data set, where the cases are defined by the categories of X_3. Note that the number of duplicates for each case is determined by its corresponding count in the original contingency table. After obtaining a bootstrap sample based on the case-form data set using the nonparametric bootstrap for each combination of the categories of (X_1, X_2), we then integrate them back to a three-dimensional contingency table that has the same layout as the original table.

For the construction of confidence intervals using bootstrap samples, there exist a few available methods, including the standard, the percentile, and the bias-corrected and accelerated (BCa) method (Efron, 1981, 1987; Davison and Hinkley, 1997). The BCa method, an improved version of the percentile method, has been proved to be more reliable than other methods especially when the sample size is small (Efron, 1987; Davison and Hinkley, 1997). It introduces the bias-correction constant for reducing the median bias of the bootstrap point estimates measured by the average deviation between the median of bootstrap point estimates and the original point estimate (due to the skewness). It also considers the acceleration constant for controlling the rate of change in the standard error of the original point estimate with respect to the true parameter value. We will employ the BCa method to construct the confidence intervals for the overall association measure $\rho^2_{(\mathbf{X}_{-i} \to X_i)}$ as well as the conditional, interactive and correlative association measures obtained from the sequential and non-sequential decomposition of $\rho^2_{(\mathbf{X}_{-i} \to X_i)}$.

Using the bootstrap method, we will also quantify the uncertainty of the predicted category of X_i for a given combination of categories of $\mathbf{X}_{-i}$. Specifically, we will first generate B bootstrap contingency tables using one of the three implementations described above. Then we will predict the category of X_i for each combination of the categories of $\mathbf{X}_{-i}$ in a bootstrap contingency table by following the prediction method given in Section 3.4. At last, we will compute the proportion that each category of X_i is predicted for each combination of the categories of $\mathbf{X}_{-i}$ over all the bootstrap contingency tables.

5.3 Permutation test

Compared to the interval estimation in Section 5.2 that offers practical significance for the estimates of the proposed association measures, hypothesis testing will provide the statistical significance of the hypothesis of no association between the ordinal dependent variable and a set of categorical independent variables, which is equally important in the inference for the proposed methods. To this end, one can consider two approaches: the Neyman-Pearson (NP) population model (Neyman and Pearson, 1928*a*,*b*) and the Fisher-Pitman (FP) permutation model (Fisher, 1992; Pitman, 1937*a*,*b*, 1938). To obtain the level of statistical significance, the NP population model first calculates the test statistic that exactly or asymptotically follows the reference distribution under the null hypothesis for the observed random sample which is assumed to be generated from a pre-specified distribution for the population. Then it estimates the frequency with which the null hypothesis would be rejected in the repeated random samples with the same sample size as that of the observed sample from the same population. On the other hand, the FP permutation model computes the same test statistics for the observed random sample as well as all or a large number (e.g. $\geq 1,000,000$) of possible permutations of the observed sample under the null hypothesis of interest. Then it estimates the frequency with which the values of test statistics for the permuted samples are larger than or equal to that for the observed sample.

Given that the proposed subcopula regression based association measure, $\rho^2_{(\mathbf{X}_{-i} \to X_i)}$, is non-negative, the asymptotic distribution of Theorem 5.1 is not applicable to the testing of $H_0 : \rho^2_{(\mathbf{X}_{-i} \to X_i)} = 0$ because the asymptotic theory we proved requires the assumption that the overall association is positive and the value of $\rho^2_{(\mathbf{X}_{-i} \to X_i)}$ under H_0 is on the boundary of its parameter space. This means that one may consider deriving the asymptotic distribution of the estimator $\hat{\rho}^2_{(\mathbf{X}_{-i} \to X_i)}$ under H_0. However, it is well known that the large-sample based testing procedure under the (NP) population model could be problematic when the sample size is not sufficiently large compared to the size of contingency

table (determined by the number of variables and the number of categories in each variable), e.g. when the contingency table is sparse (Guo and Thompson, 1989; Agresti, 2002, p.392–397).

In this section, we adopt the permutation test under the FP permutation model to assess the statistical significance of the null hypotheses concerning three types of proposed association measures: overall, marginal and conditional. For hypothesis testing, it is well-known that the FP permutation model has several advantages over the NP population model (Berry et al., 2011, 2018, 2019). First, the permutation test gives or estimates the exact probability corresponding to the permutation distribution of equally-likely test statistic values. Secondly, the permutation test is data-dependent in the sense that all the required information for testing is from the observed data. Thirdly, the permutation test is distribution-free as it does not assume a specific distribution for the population from which the sample was drawn. Finally, the permutation test is robust for small samples and sparse contingency tables (Guo and Thompson, 1989).

For the rest of this section, we will first explain the respective permutation strategies for testing the proposed overall, marginal and conditional association measures, describe the estimation of p-values in the permutation tests, and illustrate the relationships among the tests for the three association measures.

5.3.1 Permutation strategy for the hypothesis of "no overall association"

According to Proposition 4.1 (b) and (c), we know that $p_{m_i,\mathbf{m}_{-i}} = p_{m_i}p_{\mathbf{m}_{-i}}$ for every $(m_i, \mathbf{m}_{-i})$ (i.e. $\mathbf{X}_{-i}$ is independent of X_i) implies $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$, which further suggests $r_{\mathbf{m}_{-i}} = r$ for every $\mathbf{m}_{-i}$. Although $r_{\mathbf{m}_{-i}} = r$ for every $\mathbf{m}_{-i}$ does not imply $p_{m_i,\mathbf{m}_{-i}} = p_{m_i}p_{\mathbf{m}_{-i}}$ for every $(m_i, \mathbf{m}_{-i})$, we expect that if we managed to draw a sample from the factorized joint distribution $p_{m_i}p_{\mathbf{m}_{-i}}$, the estimated $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ would be almost zero and hence we could use the generated sample to test the null hypothesis $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ against

the observed sample from $p_{m_i,\mathbf{m}_{-i}}$. However, the factorized joint distribution $p_{m_i}p_{\mathbf{m}_{-i}}$ is unknown and hence we have no way to generate an i.i.d. (independent and identically distributed) sample from it. Instead, when the observed sample is i.i.d., we can permute X_i and leave $\mathbf{X}_{-i}$ invariant to approximately simulate an i.i.d. sample from the factorized joint distribution $p_{m_i}p_{\mathbf{m}_{-i}}$ (Doran et al., 2014). Note that the permutations of the independent variables $\mathbf{X}_{-i}$ lead to the same value of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ according to Proposition 4.1 (g).

To generate such a permutation of the observed d-dimensional contingency table in this case, we first convert it into a case-form data set. Note that the number of duplicates for each case is determined by its corresponding count in the original contingency table. Then we permute the categories in the column of X_i only in the case-form data set. Finally, we convert the permuted case-form data set back to a d-dimensional contingency table.

5.3.2 Permutation strategy for the hypothesis of "no marginal association"

The same permutation strategy used in Section 5.3.1 can be applied to the test of a marginal association measure with the null hypothesis $H_0 : \rho^2_{(\mathbf{X}^P_{(j)}\to X_i)} = 0$, where $\mathbf{X}^P_{(j)}$ is the j-th partition of $\mathbf{X}_{-i}$ defined in Section 4.2.2. By Definition 4.1, we know that if $p_{m_i,\mathbf{m}^P_{(j)}} = p_{m_i}p_{\mathbf{m}^P_{(j)}}$ for every $(m_i, \mathbf{m}^P_{(j)})$, then $\rho^2_{(\mathbf{X}^P_{(j)}\to X_i)} = 0$. Then we can permute X_i and leave $\mathbf{X}^P_{(j)}$ invariant in the marginal table of $(\mathbf{X}^P_{(j)}, X_i)$ to approximately simulate i.i.d. samples from $p_{m_i}p_{\mathbf{m}^P_{(j)}}$ (Doran et al., 2014).

To generate such a permutation of the marginal table of $(\mathbf{X}^P_{(j)}, X_i)$ in this case, we first marginalize the full table along each independent variable in $\mathbf{X}_{-i} - \mathbf{X}^P_{(j)}$ and convert the resulting marginal table of $(\mathbf{X}^P_{(j)}, X_i)$ into a case-form data set. Note that the number of duplicates for each case is determined by its corresponding count in the marginal table. Then we permute the categories in the column of X_i only in the case-form data set. Finally, we convert the permuted case-form data set back to the contingency table with the same size as the observed marginal table of $(\mathbf{X}^P_{(j)}, X_i)$.

5.3.3 Permutation strategy for the hypothesis of "no conditional association"

The permutation strategy for the overall and marginal association measures in Section 5.3.1 and 5.3.2 can be further extended for the test of a conditional association measure with the null hypothesis $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$, where $\mathbf{X}^\Omega_{(j)}$ is the union of the first j partitions of $\mathbf{X}_{-i}$ defined in Section 4.2.2.

By Proposition 4.7 (e) and Proposition 4.8 (b), we know that $p_{m_i, \mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)}} = p_{m_i | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^P_{(j+1)} | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^\Omega_{(j)}}$ for every $(m_i, \mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)})$ (i.e. $\mathbf{X}^P_{(j+1)}$ is independent of X_i given $\mathbf{X}^\Omega_{(j)}$) implies $\rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$, which further suggests $r_{\mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)}} = r_{\mathbf{m}^\Omega_{(j)}}$ for every $(\mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)})$. Although $r_{\mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)}} = r_{\mathbf{m}^\Omega_{(j)}}$ for every $(\mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)})$ does not imply $p_{m_i, \mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)}} = p_{m_i | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^P_{(j+1)} | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^\Omega_{(j)}}$ for every $(m_i, \mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)})$, we expect that if we managed to draw a sample from the factorized joint distribution $p_{m_i | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^P_{(j+1)} | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^\Omega_{(j)}}$, the estimated $\rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})}$ would be almost zero and hence we could use it to test the null hypothesis $\rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ against the observed sample from $p_{m_i, \mathbf{m}^P_{(j+1)}, \mathbf{m}^\Omega_{(j)}}$. However, the factorized joint distribution $p_{m_i | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^P_{(j+1)} | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^\Omega_{(j)}}$ is unknown and hence we have no way to generate an i.i.d. sample from it. Instead, we can permute X_i only and leave $\mathbf{X}^P_{(j+1)}$ invariant for every combination of categories of $\mathbf{X}^\Omega_{(j)}$ to approximately simulate an i.i.d. sample from the factorized joint distribution $p_{m_i | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^P_{(j+1)} | \mathbf{m}^\Omega_{(j)}} p_{\mathbf{m}^\Omega_{(j)}}$ (Tsamardinos and Borboudakis, 2010; Doran et al., 2014).

To generate such a permutation of the marginal table of $(\mathbf{X}^P_{(j+1)}, \mathbf{X}^\Omega_{(j)}, X_i)$ in this case, we first marginalize the full table along each independent variable in $\mathbf{X}_{-i} - \mathbf{X}^P_{(j+1)} - \mathbf{X}^\Omega_{(j)}$ and convert the resulting marginal table of $(\mathbf{X}^P_{(j+1)}, \mathbf{X}^\Omega_{(j)}, X_i)$ into a case-form data set. Note that the number of duplicates for each case is determined by its corresponding count in the marginal table. Then, we permute the categories in the column of X_i only in the case-form data set for every combination of categories of $\mathbf{X}^\Omega_{(j)}$. Finally, we convert the permuted case-form data set back to the contingency table with the same size as the

observed marginal table of $(\mathbf{X}^P_{(j+1)}, \mathbf{X}^\Omega_{(j)}, X_i)$.

5.3.4 Estimation of p-values for the permutation tests

We use the p-value estimation in the permutation test for the proposed overall association measure as an example. Note that the same procedures can be applied to the permutation tests for the proposed marginal and conditional association measures.

To estimate the p-value associated with the observed test statistic $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ computed from the observed contingency table, we first obtain the value of the permuted test statistic from each permutation denoted by $\tilde{\rho}^2_{(\mathbf{X}_{-i}\to X_i)_b}$ where $1 \leq b \leq B$ and B is the number of generated permutations, and then calculate the proportion of $\tilde{\rho}^2_{(\mathbf{X}_{-i}\to X_i)_b}$ that is larger than or equal to $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ among all the permutations. The resulting proportion is the estimated p-value, $\hat{p}$. That is,

$$\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\left(\tilde{\rho}^2_{(\mathbf{X}_{-i}\to X_i)_b} \geq \hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}\right),$$

where $I(\cdot)$ is an indicator function. In addition, we calculate the relative error given by $\sqrt{Var(\hat{p})}/E(\hat{p}) = \sqrt{(1-\hat{p})/(B\hat{p})}$ to quantify the precision of $\hat{p}$. Note that we make B equal to the total number of distinct permutations of the observed contingency table when it is no greater than $1,000,000$ and hence the estimated p-value is exact in this case. Otherwise, we let $B \geq 1,000,000$ and estimate the p-value using the Monte Carlo approximation.

5.3.5 Relationships of the permutation tests for the overall, marginal and conditional association measures

If the estimated p-value for $H_0 : \rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ is larger than a predetermined significance level α, the null hypothesis is not rejected, which indicates by Proposition 4.1(c) that the independent variables $\mathbf{X}_{-i}$ has no significant influence on the mean subcop-

ula scores for the dependent variable X_i. This also implies, by Theorem 4.2 (a), that the null hypotheses of zero values of the marginal and conditional association measures obtained from any sequential decomposition of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ are not rejected either. Therefore, one does not need to proceed with the permutations tests for any marginal and conditional association measures in this case.

On the other hand, when $H_0 : \rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ is rejected, it means that the independent variables $\mathbf{X}_{-i}$ has significant influence on the mean subcopula scores for the dependent variable X_i in a certain way. This also implies that at least one null hypothesis concerning some marginal or conditional association measure obtained from any sequential decomposition of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is rejected. Therefore, if one is further interested in discovering which null hypotheses related to the marginal and/or conditional association measures are rejected under a specific sequential decomposition, one can proceed with the permutation tests for the marginal and conditional association measures as explained in Section 5.3.2 and 5.3.3. Note that one may need to control the family-wise error rate or false discovery rate to test the null hypotheses of marginal and conditional association measures simultaneously. Such adjustment methods include the Bonferroni correction (Bonferroni, 1936), the Holm–Bonferroni method (Holm, 1979), the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) and the Benjamini–Yekutieli procedure (Benjamini and Yekutieli, 2001), where the last three are regarded to be less conservative. Due to the dependence between the tests of the marginal and conditional association measures, one may use the Benjamini–Yekutieli procedure that has been proved to be valid under arbitrary dependence.

5.4 Use of the interval estimation and hypothesis testing

For the statistical inference concerning the proposed subcopula regression based association measures, one can choose the interval estimation methods explained in Section 5.2 or the permutation test procedures discussed in Section 5.3. However, we argue that

it is still beneficial to quantify a small magnitude of association in practice by forming the confidence intervals for the proposed association measures, even if the null hypotheses concerning the (overall/marginal/conditional) association measures are not rejected. This is because that the hypothesis testing only provides statistical significance on whether a single association measure is zero or multiple association measures are simultaneously zero. On the other hand, the interval estimation provides practical significance which indicates how strong or weak the actual association is between the ordinal dependent variable and a set of the independent variables.

CHAPTER 6

SIMULATION STUDY

In this chapter, we conduct Monte Carlo experiments to examine the finite-sample properties of the proposed subcopula regression based association measures for multi-dimensional contingency tables with ordinal variables under various experimental settings. By the finite-sample performance, we mean the sampling distributions, the sampling variances and the biases for the proposed association measures. We will also investigate the sensitivity of the proposed association measure to the number of categories in the ordinal dependent variable and compare it to *Gray-Williams' index* known to be sensitive to the number of categories in the dependent variable.

6.1 Simulation settings

In the simulation study, we consider the following factors:

1. the number of the variables in a contingency table denoted by d (the dimension of the contingency table),
2. the marginal distribution of each ordinal variable,
3. the number of categories in each variable denoted by k,
4. the sample size of the contingency table denoted by n,
5. the type of association among the ordinal variables,

6. the strength of each type of association.

To comprehensively assess the finite-sample performance of the proposed association measures, we consider four scenarios with different purposes:

- S1 The goal is to investigate the sensitivity of the proposed overall association measure compared to *Gray-Williams' index* to the number of categories in the dependent variable for a given sample size, number of categories in each independent variable, type of association and its strength. Note that this investigation is limited to the case of three-dimensional contingency tables because *Gray-Williams' index* is available only for three-dimensional tables.

- S2 The objective is to evaluate the performance of the proposed overall, marginal and conditional association measures as the number of categories in the dependent variable increases for different combinations of sample size, type of association and its strength while the number of categories in each independent variable is fixed.

- S3 The focus is to assess the performance of the proposed overall, marginal and conditional association measures as the number of categories in each variable (both the dependent and independent variables) increases for different combinations of sample size, type of association and its strength.

- S4 The intention is to examine the performance of the proposed overall, marginal and conditional association measures over various sample sizes for a given number of categories in each variable, type of association and its strength.

The difference between Scenario 3 and 4 is that different sets of sample sizes in Scenario 4 are considered for each number of categories in every ordinal variable (both the dependent and independent variables), unlike in Scenario 3.

In the simulation study, the dimension and the marginal distribution are fixed both within every scenario and across four scenarios; the type of association and its strength

are fixed across four scenarios but vary within every scenario; the number of categories k and sample size n vary both within every scenario and across four scenarios. Since the dimension, marginal distribution, type of association and its strength are fixed across scenarios, we list their values in Table 6.1. Note that X_d is the dependent variable and $(X_1, \ldots, X_{d-1})$ are the independent variables. Because the number of categories k and sample size n vary across scenarios and so we will provide their values along with the simulation results for each scenario in Section 6.3.

Regarding the values in Table 6.1, first note that the marginal distribution is fixed to be discrete uniform on $[0, 1]$. This is because we want to investigate the performance of the proposed association measures free from the effect of the marginal distribution of each variable under different types of association among the variables. Secondly, the partial correlation between X_d and X_1 is used to define the types of association, but the magnitude of the correlation between X_d and X_1 is used to define the strength of association. This is because the change of strength with respect to the partial correlation becomes impossible when it is zero. In addition, for a case of non-zero partial correlation between X_d and X_1, we make the degree of correlation between X_d and X_1 similar to that of the partial correlation between X_d and X_1. Table 6.2 and 6.3 show the partial correlation and correlation matrix of (X_3, X_1, X_2) and $(X_5, X_1, X_2, X_3, X_4)$ for every combination of type of association and its strength when $d = 3$ and $d = 5$, respectively.

The simulation process for each scenario starts with simulating the "population" contingency table with the "true" joint p.m.f. for the ordinal variables. Here we say the "population" contingency table by referring to the d-dimensional table of sample size N. In this study, the "population" contingency table for $d = 3$ and $d = 5$ have the sample sizes of $N = 10^6$ and $N = 5 \times 10^8$, respectively.

Next, the simulation process continues by generating B contingency tables of sample size $n < N$ from each "population" table according to the marginal distribution, number of categories k, type of association and its strength. For the simulation study, the number

Factor Name	Values
dimension d	3, 5
marginal distribution	discrete uniform on $[0, 1]$
type of association	$pcorr(X_d, X_1 \vert \mathbf{X}_{-d,-1}) > 0$, $pcorr(X_d, X_1 \vert \mathbf{X}_{-d,-1}) < 0$, $pcorr(X_d, X_1 \vert \mathbf{X}_{-d,-1}) = 0$, $auto1corr(X_i, X_j) = \phi^{\vert(i \mod d)-(j \mod d)\vert}$
strength of association	weak: $\vert corr(X_d, X_1)\vert = 0.3$, moderate: $\vert corr(X_d, X_1)\vert = 0.5$, strong: $\vert corr(X_d, X_1)\vert = 0.7$

Table 6.1: Fixed simulation factors across four scenarios. For type of association, $\mathbf{X}_{-d,-1} = \{X_2, \dots, X_{d-1}\}$, $pcorr(X_d, X_1|\mathbf{X}_{-d,-1})$ is the partial correlation between X_d and X_1 given $\mathbf{X}_{-d,-1}$ and $auto1corr(X_i, X_j)$ is the lag-1 autocorrelation between X_i and X_j with $|\phi| < 1$. For strength of association, $corr(X_d, X_1)$ is the correlation between X_d and X_1.

of sample contingency tables to generate is $B = 1000$. The method for generating the "population" and sample contingency tables will be introduced in Section 6.2.

For each generated "population" or sample contingency table, we calculate five subcopula regression based association measures: the overall association measure $\rho^2_{(\mathbf{X}_{-d} \to X_d)}$, two marginal association measures $\rho^2_{(X_1 \to X_d)}$ and $\rho^2_{(\mathbf{X}_{-d,-1} \to X_d)}$, and two conditional association measures $\rho^2_{(X_1 \to X_d|\mathbf{X}_{-d,-1})}$ and $\rho^2_{(\mathbf{X}_{-d,-1} \to X_d|X_1)}$, where $\mathbf{X}_{-d,-1} = \mathbf{X}_d - \{X_1, X_d\} = \{X_2, \dots, X_{d-1}\}$. Note that only the overall association measure $\rho^2_{(\mathbf{X}_{-d} \to X_d)}$ and *Gray-Williams' index* $\tau^{GW}_{X_d}$ will be computed in Scenario 1.

Finally, the simulation results for each scenario are presented in side-by-side boxplots with respect to the combinations of the sample size n and number of categories k in each ordinal variable. In Scenario 1, every red boxplot represents the sampling distribution of the estimator of the overall association measure using 1000 contingency tables of size n generated from the "population" contingency table. The red dashed line in the boxplot is the mean of the 1000 estimates of the overall association measure and the corresponding red triangle is the "true" value computed from the "population" contingency table. On the other hand, every blue boxplot represents the sampling distribution of the estimator of *Gray-Williams' index*. The blue dashed line in the boxplot is the mean of the 1000 estimates of *Gray-Williams' index* and the corresponding blue star is the "true" value computed from

Type of association	Strength of association	Partial correlation matrix	Correlation matrix
$pcorr(X_3, X_1 \vert X_2) > 0$	weak: $\vert corr(X_3, X_1)\vert = 0.3$	$\begin{bmatrix} 1.000 & 0.313 & 0.137 \\ 0.313 & 1.000 & -0.137 \\ 0.137 & -0.137 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.100 \\ 0.300 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) > 0$	moderate: $\vert corr(X_3, X_1)\vert = 0.5$	$\begin{bmatrix} 1.000 & 0.515 & 0.174 \\ 0.515 & 1.000 & -0.174 \\ 0.174 & -0.174 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.100 \\ 0.500 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) > 0$	strong: $\vert corr(X_3, X_1)\vert = 0.7$	$\begin{bmatrix} 1.000 & 0.717 & 0.239 \\ 0.717 & 1.000 & -0.239 \\ 0.239 & -0.239 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.100 \\ 0.700 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) < 0$	weak: $\vert corr(X_3, X_1)\vert = 0.3$	$\begin{bmatrix} 1.000 & -0.293 & 0.074 \\ -0.293 & 1.000 & -0.074 \\ 0.074 & -0.074 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.300 & 0.100 \\ -0.300 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) < 0$	moderate: $\vert corr(X_3, X_1)\vert = 0.5$	$\begin{bmatrix} 1.000 & -0.495 & 0.058 \\ -0.495 & 1.000 & -0.058 \\ 0.058 & -0.058 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.500 & 0.100 \\ -0.500 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) < 0$	strong: $\vert corr(X_3, X_1)\vert = 0.7$	$\begin{bmatrix} 1.000 & -0.697 & 0.042 \\ -0.697 & 1.000 & -0.042 \\ 0.042 & -0.042 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.700 & 0.100 \\ -0.700 & 1.000 & -0.100 \\ 0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) = 0$	weak: $\vert corr(X_3, X_1)\vert = 0.3$	$\begin{bmatrix} 1.000 & 0.000 & 0.419 \\ 0.000 & 1.000 & 0.545 \\ 0.419 & 0.545 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.500 \\ 0.300 & 1.000 & 0.600 \\ 0.500 & 0.600 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) = 0$	moderate: $\vert corr(X_3, X_1)\vert = 0.5$	$\begin{bmatrix} 1.000 & 0.000 & 0.566 \\ 0.000 & 1.000 & 0.589 \\ 0.566 & 0.589 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.700 \\ 0.500 & 1.000 & 0.714 \\ 0.700 & 0.714 & 1.000 \end{bmatrix}$
$pcorr(X_3, X_1 \vert X_2) = 0$	strong: $\vert corr(X_3, X_1)\vert = 0.7$	$\begin{bmatrix} 1.000 & 0.000 & 0.542 \\ 0.000 & 1.000 & 0.735 \\ 0.542 & 0.735 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.800 \\ 0.700 & 1.000 & 0.875 \\ 0.800 & 0.875 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\vert (i \bmod 3) - (j \bmod 3)\vert}$, $i, j = 1, 2, 3$	weak: $\vert\phi\vert = \vert corr(X_3, X_1)\vert = 0.3$	$\begin{bmatrix} 1.000 & 0.287 & 0.000 \\ 0.287 & 1.000 & 0.287 \\ 0.000 & 0.287 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.090 \\ 0.300 & 1.000 & 0.090 \\ 0.090 & 0.300 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\vert (i \bmod 3) - (j \bmod 3)\vert}$, $i, j = 1, 2, 3$	moderate: $\vert\phi\vert = \vert corr(X_3, X_1)\vert = 0.5$	$\begin{bmatrix} 1.000 & 0.447 & 0.000 \\ 0.447 & 1.000 & 0.447 \\ 0.000 & 0.447 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.250 \\ 0.500 & 1.000 & 0.500 \\ 0.250 & 0.500 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\vert (i \bmod 3) - (j \bmod 3)\vert}$, $i, j = 1, 2, 3$	strong: $\vert\phi\vert = \vert corr(X_3, X_1)\vert = 0.7$	$\begin{bmatrix} 1.000 & 0.573 & 0.000 \\ 0.573 & 1.000 & 0.573 \\ 0.000 & 0.573 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.490 \\ 0.700 & 1.000 & 0.700 \\ 0.490 & 0.700 & 1.000 \end{bmatrix}$

Table 6.2: Partial correlation and correlation matrix of (X_3, X_1, X_2) for each combination of type of association and its strength

Type of association	Strength of association	Partial correlation matrix	Correlation matrix
$pcorr(X_5, X_1 \| X_2, X_3, X_4) > 0$	weak: $\|corr(X_5, X_1)\| = 0.3$	$\begin{bmatrix} 1.000 & 0.351 & 0.177 & 0.177 & 0.177 \\ 0.351 & 1.000 & -0.177 & -0.177 & -0.177 \\ 0.177 & -0.177 & 1.000 & -0.152 & -0.152 \\ 0.177 & -0.177 & -0.152 & 1.000 & -0.152 \\ 0.177 & -0.177 & -0.152 & -0.152 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.100 & 0.100 & 0.100 \\ 0.300 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.100 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) > 0$	moderate: $\|corr(X_5, X_1)\| = 0.5$	$\begin{bmatrix} 1.000 & 0.558 & 0.229 & 0.229 & 0.229 \\ 0.558 & 1.000 & -0.229 & -0.229 & -0.229 \\ 0.229 & -0.229 & 1.000 & -0.171 & -0.171 \\ 0.229 & -0.229 & -0.171 & 1.000 & -0.171 \\ 0.229 & -0.229 & -0.171 & -0.171 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.100 & 0.100 & 0.100 \\ 0.500 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.100 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) > 0$	strong: $\|corr(X_5, X_1)\| = 0.7$	$\begin{bmatrix} 1.000 & 0.766 & 0.325 & 0.325 & 0.325 \\ 0.766 & 1.000 & -0.325 & -0.325 & -0.325 \\ 0.325 & -0.325 & 1.000 & -0.217 & -0.217 \\ 0.325 & -0.325 & -0.217 & 1.000 & -0.217 \\ 0.325 & -0.325 & -0.217 & -0.217 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.100 & 0.100 & 0.100 \\ 0.700 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.100 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) < 0$	weak: $\|corr(X_5, X_1)\| = 0.3$	$\begin{bmatrix} 1.000 & -0.273 & 0.094 & 0.094 & 0.094 \\ -0.273 & 1.000 & -0.094 & -0.094 & -0.094 \\ 0.094 & -0.094 & 1.000 & -0.133 & -0.133 \\ 0.094 & -0.094 & -0.133 & 1.000 & -0.133 \\ 0.094 & -0.094 & -0.133 & -0.133 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.300 & 0.100 & 0.100 & 0.100 \\ -0.300 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.1 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) < 0$	moderate: $\|corr(X_5, X_1)\| = 0.5$	$\begin{bmatrix} 1.000 & -0.480 & 0.074 & 0.074 & 0.074 \\ -0.480 & 1.000 & -0.074 & -0.074 & -0.074 \\ 0.074 & -0.074 & 1.000 & -0.130 & -0.130 \\ 0.074 & -0.074 & 1.000 & -0.130 & -0.130 \\ 0.074 & -0.074 & -0.130 & 1.000 & -0.130 \\ 0.074 & -0.074 & -0.130 & -0.130 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.500 & 0.100 & 0.100 & 0.100 \\ -0.500 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.100 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) < 0$	strong: $\|corr(X_5, X_1)\| = 0.7$	$\begin{bmatrix} 1.000 & -0.688 & 0.054 & 0.054 & 0.054 \\ -0.688 & 1.000 & -0.054 & -0.054 & -0.054 \\ 0.054 & -0.054 & 1.000 & -0.128 & -0.128 \\ 0.054 & -0.054 & -0.128 & 1.000 & -0.128 \\ 0.054 & -0.054 & -0.128 & -0.128 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & -0.700 & 0.100 & 0.100 & 0.100 \\ -0.700 & 1.000 & -0.100 & -0.100 & -0.100 \\ 0.100 & -0.100 & 1.000 & -0.100 & -0.100 \\ 0.100 & -0.100 & -0.100 & 1.000 & -0.100 \\ 0.100 & -0.100 & -0.100 & -0.100 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) = 0$	weak: $\|corr(X_5, X_1)\| = 0.3$	$\begin{bmatrix} 1.000 & 0.000 & 0.062 & 0.062 & 0.324 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.480 \\ 0.062 & 0.000 & 1.000 & 0.062 & 0.324 \\ 0.062 & 0.000 & 0.062 & 1.000 & 0.324 \\ 0.324 & 0.480 & 0.324 & 0.324 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.300 & 0.300 & 0.500 \\ 0.300 & 1.000 & 0.300 & 0.300 & 0.6 \\ 0.300 & 0.300 & 1.000 & 0.300 & 0.500 \\ 0.300 & 0.300 & 0.300 & 1.000 & 0.500 \\ 0.500 & 0.600 & 0.500 & 0.500 & 1.000 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) = 0$	moderate: $\|corr(X_5, X_1)\| = 0.5$	$\begin{bmatrix} 1.000 & 0.000 & 0.019 & 0.019 & 0.430 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.465 \\ 0.019 & 0.000 & 1.000 & 0.019 & 0.430 \\ 0.019 & 0.000 & 0.019 & 1.000 & 0.430 \\ 0.430 & 0.465 & 0.430 & 0.430 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.500 & 0.500 & 0.700 \\ 0.500 & 1.000 & 0.500 & 0.500 & 0.714 \\ 0.500 & 0.500 & 1.000 & 0.500 & 0.700 \\ 0.500 & 0.500 & 0.500 & 1.000 & 0.700 \\ 0.700 & 0.714 & 0.700 & 0.700 & 1.00 \end{bmatrix}$
$pcorr(X_5, X_1 \| X_2, X_3, X_4) = 0$	strong: $\|corr(X_5, X_1)\| = 0.7$	$\begin{bmatrix} 1.000 & 0.000 & 0.143 & 0.143 & 0.339 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.629 \\ 0.143 & 0.000 & 1.000 & 0.143 & 0.339 \\ 0.143 & 0.000 & 0.143 & 1.000 & 0.339 \\ 0.339 & 0.629 & 0.339 & 0.339 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.700 & 0.700 & 0.800 \\ 0.700 & 1.000 & 0.700 & 0.700 & 0.875 \\ 0.700 & 0.700 & 1.000 & 0.700 & 0.800 \\ 0.700 & 0.700 & 0.700 & 1.000 & 0.800 \\ 0.800 & 0.875 & 0.800 & 0.800 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\|(i \mod 5)-(j \mod 5)\|}$, $i, j = 1, \ldots, 5$	weak: $\|\phi\| = \|corr(X_5, X_1)\| = 0.3$	$\begin{bmatrix} 1.000 & 0.287 & 0.000 & 0.000 & 0.000 \\ 0.287 & 1.000 & 0.275 & 0.000 & 0.000 \\ 0.000 & 0.275 & 1.000 & 0.275 & 0.000 \\ 0.000 & 0.000 & 0.275 & 1.000 & 0.287 \\ 0.000 & 0.000 & 0.000 & 0.287 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.300 & 0.090 & 0.027 & 0.008 \\ 0.300 & 1.000 & 0.090 & 0.027 & 0.008 \\ 0.090 & 0.300 & 1.000 & 0.300 & 0.090 \\ 0.027 & 0.090 & 0.300 & 1.000 & 0.300 \\ 0.008 & 0.027 & 0.090 & 0.300 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\|(i \mod 5)-(j \mod 5)\|}$, $i, j = 1, \ldots, 5$	moderate: $\|\phi\| = \|corr(X_5, X_1)\| = 0.5$	$\begin{bmatrix} 1.000 & 0.447 & 0.000 & 0.000 & 0.000 \\ 0.447 & 1.000 & 0.447 & 0.000 & 0.000 \\ 0.000 & 0.400 & 1.000 & 0.400 & 0.000 \\ \\ 0.000 & 0.000 & 0.400 & 1.000 & 0.447 \\ 0.000 & 0.000 & 0.000 & 0.447 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.500 & 0.250 & 0.125 & 0.062 \\ 0.500 & 1.000 & 0.500 & 0.250 & 0.125 \\ 0.250 & 0.500 & 1.000 & 0.500 & 0.250 \\ 0.125 & 0.250 & 0.500 & 1.000 & 0.500 \\ 0.062 & 0.125 & 0.250 & 0.500 & 1.000 \end{bmatrix}$
$auto1corr(X_i, X_j) = \phi^{\|(i \mod 5)-(j \mod 5)\|}$, $i, j = 1, \ldots, 5$	strong: $\|\phi\| = \|corr(X_5, X_1)\| = 0.7$	$\begin{bmatrix} 1.000 & 0.573 & 0.000 & 0.000 & 0.000 \\ 0.573 & 1.000 & 0.470 & 0.000 & 0.000 \\ 0.000 & 0.470 & 1.000 & 0.47070.000 & \\ 0.000 & 0.000 & 0.470 & 1.000 & 0.573 \\ 0.000 & 0.000 & 0.000 & 0.573 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.700 & 0.490 & 0.343 & 0.240 \\ 0.700 & 1.000 & 0.700 & 0.490 & 0.343 \\ 0.490 & 0.700 & 1.000 & 0.700 & 0.490 \\ 0.343 & 0.490 & 0.700 & 1.000 & 0.700 \\ 0.240 & 0.343 & 0.490 & 0.700 & 1.000 \end{bmatrix}$

Table 6.3: Partial correlation and correlation matrix of $(X_5, X_1, X_2, X_3, X_4)$ for each combination of type of association and its strength

the "population" contingency table. In Scenario 2, 3 and 4, every boxplot represents the sampling distribution of the estimator of a selected association measure using 1000 contingency tables of size n generated from the "population" contingency table. The black dashed line in the boxplot is the mean of the 1000 estimates of a selected association measure and the corresponding red star is the "true" value of the selected association measure computed from the "population" contingency table.

6.2 Simulation method for generating "population" and sample contingency tables

In the simulation study, each "population" contingency table of size N ($N = 10^6$ for $d = 3$ and $N = 10^9$ for $d = 5$) is generated by the simulation algorithm proposed by Ferrari and Barbiero (2012). The core idea is to simulate raw data from a multivariate normal distribution with zero mean vector and a correlation matrix $\mathbf{\Sigma}^*$ and then produce the ordinal variables with the desired correlation matrix $\mathbf{\Sigma}$ and their desired marginal cumulative distributions by discretizing each of the continuous variables in the simulated raw data. The correlation matrix $\mathbf{\Sigma}^*$ is obtained by adjusting the desired correlation matrix $\mathbf{\Sigma}$ of the ordinal variables to take into account the effect of discretization. Then, to discretize a normal variable Z into the desired ordinal variable X, the category of X is determined by the quantiles of Z corresponding to the probabilities specified by the desired marginal cumulative distribution function. For example, suppose that q_1 and q_2 ($q_1 < q_2$) are the two quantiles of Z corresponding to the desired probabilities 0.3 and 0.7 in the cumulative distribution function of X. Then $Z = z$ falls into the first, second or third category of X if $z < q_1$, $q_1 \le z < q_2$ or $q_2 \le z$, respectively. Note that the simulation algorithm is available in the R package "GenOrd" (Barbiero and Ferrari, 2015).

Given a "population" contingency table of size N, we employ the strategy of conditional paired simulation to generate sample contingency tables of size n for every scenario

listed in Section 6.1. First, it fixes the counts in the marginal contingency table of the independent variables and converts the marginal contingency table of the dependent variable into a case-form data set with respect to each combination of the categories of the independent variables. It then simulates new case-form data sets and integrates them back to the table form.

There are two main advantages of the conditional paired simulation over the unconditional one. First, because it fixes the joint distribution of the independent variables, Scenario 1 and 2 benefit from this case because the variability of the estimated association measures only attributes to that of the dependent variable across different numbers of categories. Furthermore, the conditional paired simulation reduces the extent of sparseness in generating multi-dimensional sample contingency tables, especially when some of the independent variables have rare categories.

6.3 Simulation results

Before we present the details for the outputs and results from the simulation, we give a brief summary of the simulation results as follows. First, we find in Scenario 1 that the proposed overall association measure is insensitive to the number of categories in the dependent variable in that its magnitude does not decrease as the number of categories increases, unlike *Gray-Williams' index*. In Scenario 2, 3 and 4, we obtain similar simulation results for non-sparse and sparse contingency tables, respectively. When the true values are far from zero, the sampling distributions of the proposed (overall/marginal/conditional) association measures are symmetrical about the estimated means; when the true values are getting closer to zero, the sampling distributions become more asymmetrical (e.g. right-skewed) about the estimated means. Specifically, we notice that the skewness in the sampling distributions of the proposed (overall/marginal/conditional) association measures is more notable for sparse contingency tables especially when the true values are relatively close to zero. The biases are also more pronounced for sparse contingency tables than

those for non-sparse ones. Note that such a bias issue for sparse contingency tables is common for many other existing methods (Agresti, 2002, p. 391–398). Finally, even when the contingency tables are not sparse, the biases seem to be larger for five-dimensional case ($d = 5$) than for three-dimensional case ($d = 3$) with respect to the overall and conditional association measures.

6.3.1 Three-dimensional case ($d = 3$)

Let X_1, X_2, X_3 be the ordinal variables in a three-dimensional contingency table, where X_3 is the dependent variable and (X_1, X_2) are the independent variables.

6.3.1.1 Scenario 1

Factor Name	Values
the number of categories for X_1 (k_{X_1})	3
the number of categories for X_2 (k_{X_2})	3
the number of categories for X_3 (k_{X_3})	(3, 5, 7, 9, 11, 13, 15, 17, 19)
sample size (n)	(855, 1710, 3420, 6840)

Table 6.4: Simulation factors for Scenario 1 in three-dimensional case

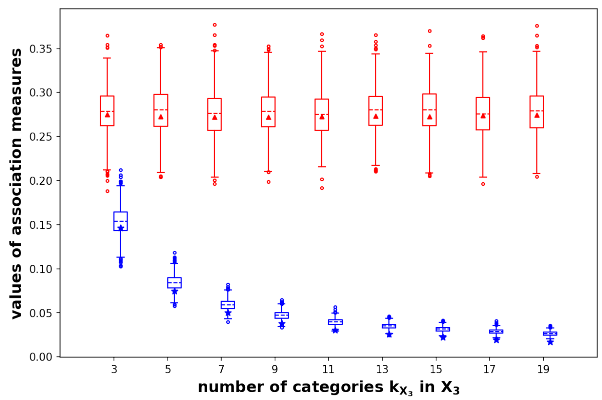
Table 6.4 provides the settings for the number of categories in each variable denoted by k_{X_1}, k_{X_2} and k_{X_3}, and the set of sample sizes n used in this scenario. Note that k_{X_1} and k_{X_2} are fixed to be 3, k_{X_3} is varied from 3 to 19, and the sample sizes are fixed to be $n = (855, 1710, 3420, 6840)$ for each level of k_{X_3} by taking into account the maximum number of cells ($3 \times 3 \times 19 = 171$) and four different conceptual average cell counts, $(5, 10, 20, 40)$, in the contingency table. We estimate the proposed overall association measure $\rho^2_{(X_1, X_2 \to X_3)}$ and Gray-Williams' index $\tau^{GW}_{X_3}$ over a range of the number of the categories k_{X_3} in the dependent variable at a fixed sample size n to investigate their sensitivity with respect to k_{X_3}. Figure 6.1 and 6.2 shows the selected simulation results regarding the two types

of association: $pcorr(X_3, X_1|X_2) > 0$ and $pcorr(X_3, X_1|X_2) = 0$, with the moderate strength of association $|corr(X_3, X_1)| = 0.5$. Note that the remaining simulation results for Scenario 1 are provided in Appendix J. Based on the boxplots of all the simulation results for Scenario 1, we made the following observations:

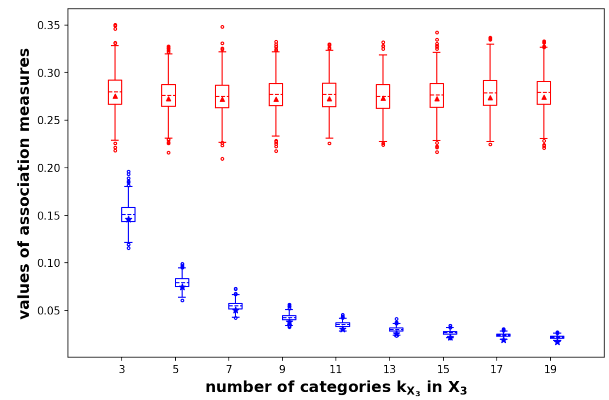
1. The "true" values (red colored asterisks inside the red boxplots) of the proposed overall association measure and the sampling distributions (variances/biases) of its estimator stay the same as the number of categories in the dependent variable increases at a fixed sample size, regardless of the type of association and its strength. In addition, the variances of the estimator of the overall association measure decrease and the biases appear to be almost zero as the sample size increases at a fixed number of categories in the dependent variable, regardless of the type of association and its strength.

2. The "true" values (blue colored asterisks inside the blue boxplots) of *Gray-Williams' index* and the sampling distributions (variances/biases) of its estimator decrease and then converge to a small value near zero as the number of categories in the dependent variable increases, regardless of the sample size, the type of association and its strength. In particular, the variances of the estimator of *Gray-Williams' index* decrease and its biases become zero as the number of categories in the dependent variable (sample size) increases at fixed sample size (number of categories in the dependent variable), regardless of the type of association and its strength.

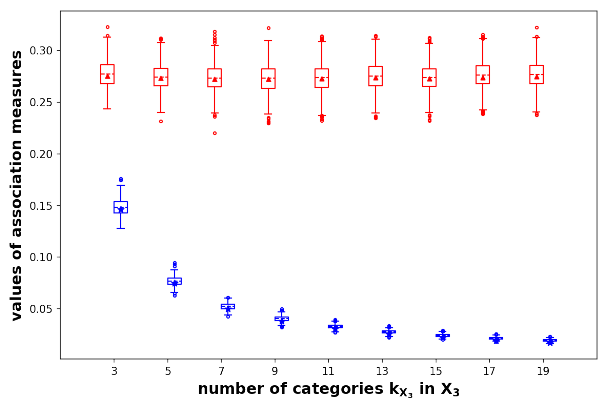
3. The proposed overall association measure appears to have a consistent performance and be insensitive to the number of categories in the dependent variable, regardless of the sample size, the type of association and its strength.



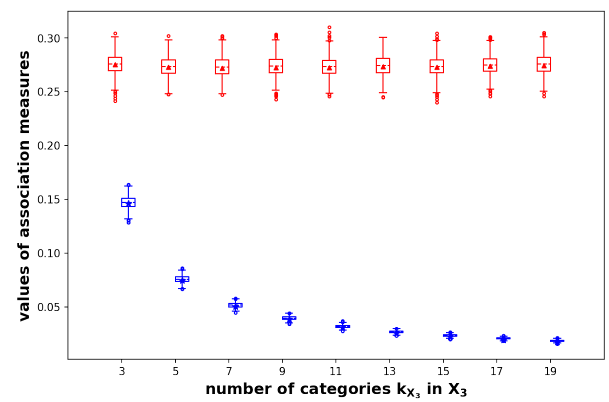
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

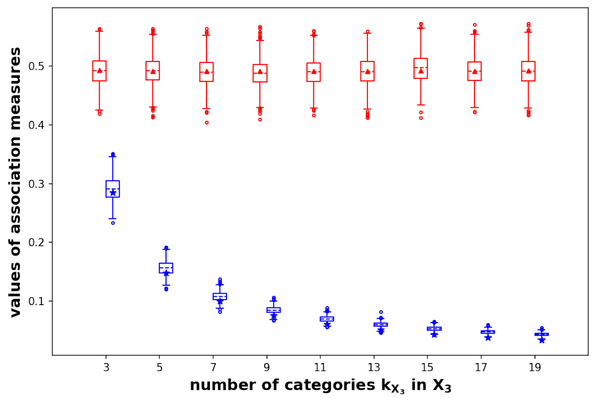


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

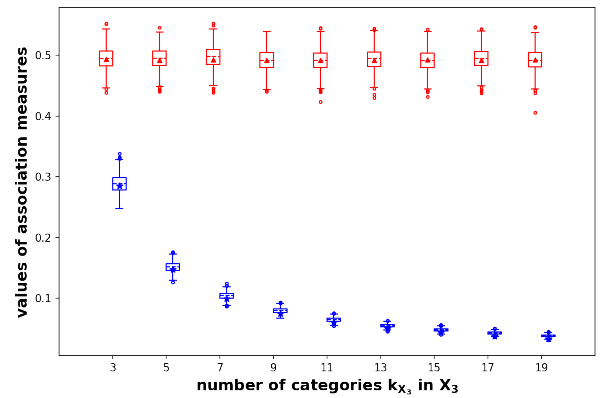


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

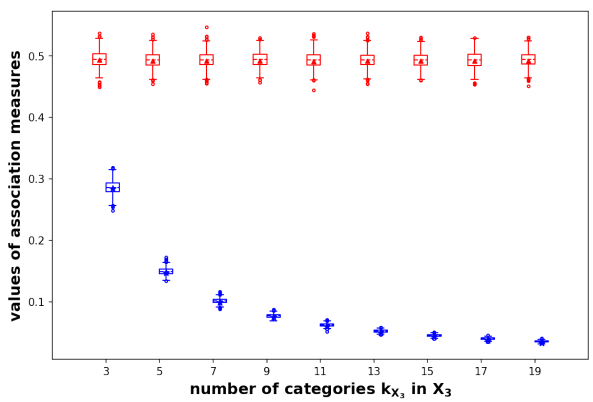
Figure 6.1: The association measure and *Gray-Williams' index* for the partial correlation of $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



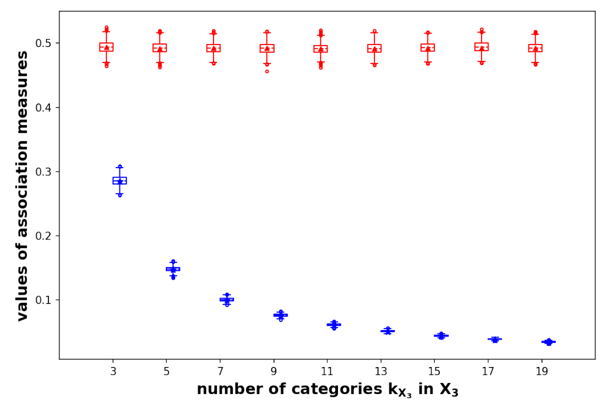
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$



(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$



(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

Figure 6.2: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

6.3.1.2 Scenario 2

Factor Name	Values
the number of categories for X_1 (k_{X_1})	3
the number of categories for X_2 (k_{X_2})	3
the number of categories for X_3 (k_{X_3})	(3, 5, 7)
sample size (n)	(630, 1260, 2520, 5040)

Table 6.5: Simulation factors for Scenario 2 in three-dimensional case

Table 6.5 provides the settings for the number of categories in each variable denoted by k_{X_1}, k_{X_2} and k_{X_3}, and the set of sample sizes n used in this scenario. Note that the k_{X_1} and k_{X_2} are fixed to be 3, k_{X_3} varies from 3 to 7, and the sample sizes are fixed to be $n = (630, 1260, 2520, 5040)$ for each level of k_{X_3} by considering the maximum number of cells $(3 \times 3 \times 7 = 63)$ and four different conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.3 to 6.6 shows the selected simulation results regarding the two types of association: $pcorr(X_3, X_1|X_2) > 0$ and $pcorr(X_3, X_1|X_2) = 0$ with the moderate strength of association $|corr(X_3, X_1)| = 0.5$. Note that the remaining simulation results for Scenario 2 are given in Appendix J. From all the simulation results for Scenario 2, the following observations were made:

1. The observations made in Scenario 1 about the overall association measures are applicable in this case. Moreover, they are also applied to the marginal and conditional association measures.

2. For each combination of type of association and its strength, the marginal association measures reflect the associations between the dependent variable and the independent variables represented by the corresponding correlation matrix in Table 6.2, while the conditional association measures capture the associations represented by the partial correlation matrix in Table 6.2. For example, we can see from Figure

6.6 that when the type of association is $pcorr(X_3, X_1|X_2) = 0$, the respective estimates of the marginal association measures $\rho^2_{(X_1 \to X_3)}$ and $\rho^2_{(X_2 \to X_3)}$ for each value of k_{X_3} stay away from 0, given that $|corr(X_3, X_1)| = 0.5$. The estimate of the conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$ for each value of k_{X_3} is far from 0 because $pcorr(X_3, X_2|X_1) = 0.566$ while the estimate of the conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$ for each value of k_{X_3} is very close to 0 since $pcorr(X_3, X_1|X_2) = 0$.

3. When the "true" values of the (overall/marginal/conditional) association measures stay away from zero (larger than 0.05), the sampling distributions of the estimators of the association measures appear to be symmetrical about the estimated means. In the cases where the "true" values of the (overall/marginal/conditional) association measures is relatively close to zero (between 0.01 and 0.05), the sampling distributions of the estimators appear to be somewhat asymmetrical. When the "true" values of the (overall/marginal/conditional) association measures are very close to zero (less than 0.01), the sampling distribution of the estimators of the association measures appear to be quite skewed to the right. Note that when the "true" value of an association measure is almost 0, its estimator may be upward-biased relative to its "true" one.

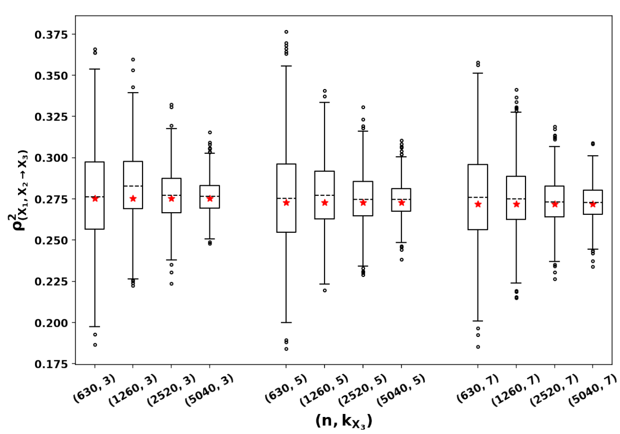
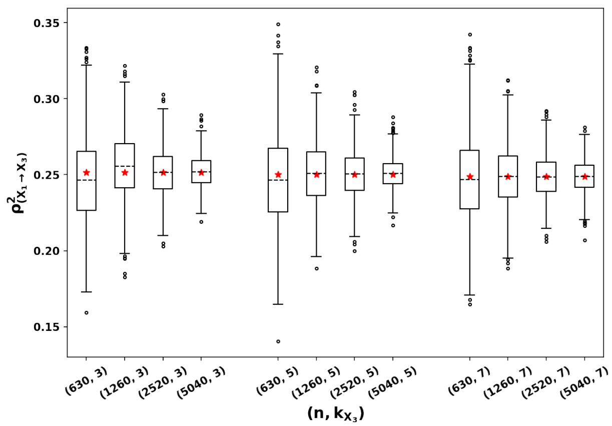
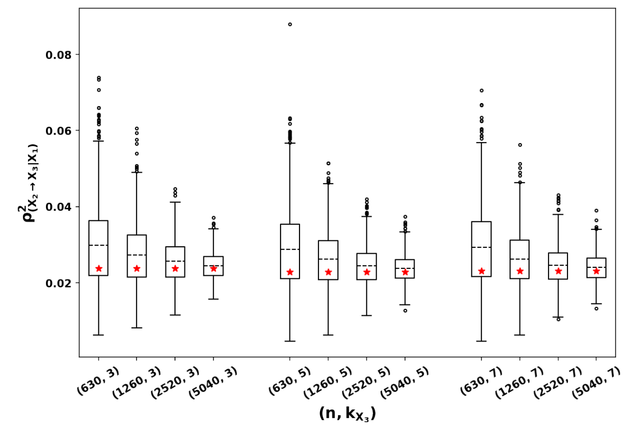


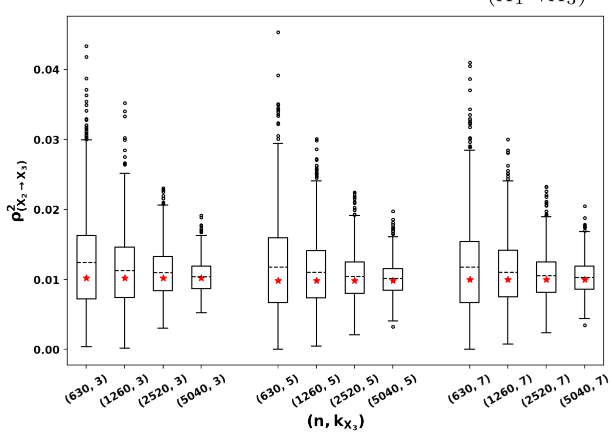
Figure 6.3: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



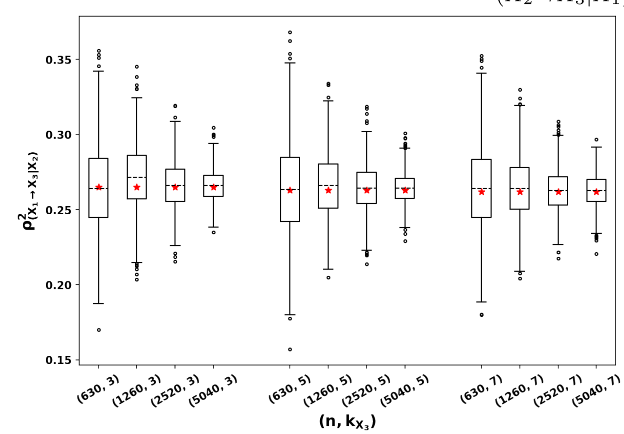
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.4: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

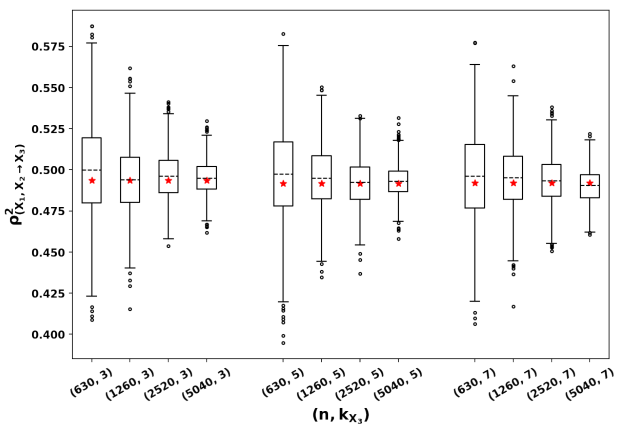
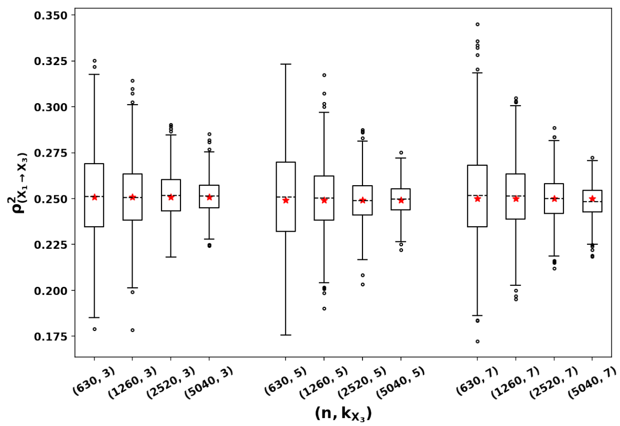
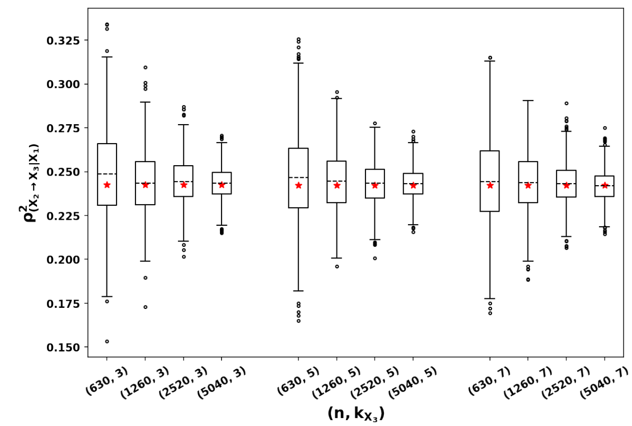


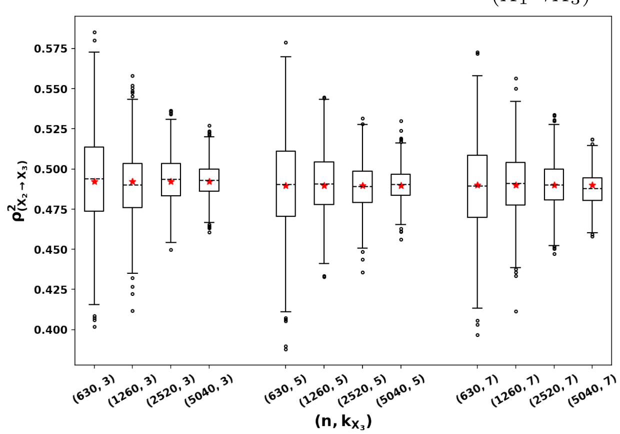
Figure 6.5: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



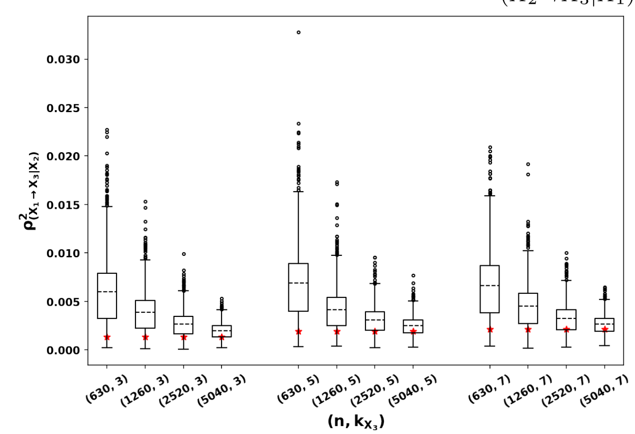
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure 6.6: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

6.3.1.3 Scenario 3

Factor Name	Values
the number of categories in each variable ($k_{X_1} = k_{X_2} = k_{X_3}$)	3, 5, 7
sample size (n)	(3430, 6860, 13720, 27440)

Table 6.6: Simulation factors for Scenario 3 in three-dimensional case

Table 6.6 provides the settings for the number of categories in each variable denoted by k_{X_1}, k_{X_2} and k_{X_3}, and the set of sample sizes n considered in this scenario. Note that, given that $k_{X_1} = k_{X_2} = k_{X_3} = k$, the sample sizes are $n = (3430, 6860, 13720, 27440)$ for all the combinations of $(k_{X_1}, k_{X_2}, k_{X_3})$, by considering the maximum number of cells $(7^3 = 343)$ and four conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.7 to 6.12 shows the selected simulation results regarding the three types of association: $pcorr(X_3, X_1|X_2) > 0$, $pcorr(X_3, X_1|X_2) = 0$ and $auto1corr(X_i, X_j) = \phi^{|(i \mod 3)-(j \mod 3)|}$ with moderate strength of association $|corr(X_3, X_1)| = 0.5$ where $i, j = 1, 2, 3$. Note that the remaining simulation results for Scenario 3 are provided in Appendix J. The following observations were made upon the simulation results for Scenario 3:

1. The observations made in Scenario 2 about the overall association measures, as well as those for the marginal and conditional association measures, are all applicable in this case.

2. When the "true" values of the (overall/marginal/conditional) measures stay away from zero (larger than 0.05), the sampling distributions of the estimators of the association measures appears to be symmetrical about the estimated means. Furthermore, their variances stay almost the same as the number of categories increases at a fixed sample size but decrease as the sample size increases at a fixed number of categories, regardless of the type of association and its strength. However, the biases slightly in-

crease as the number of categories increases at a fixed sample size, and these biases become less marked as the sample size increases, regardless of the type of association and its strength.

3. When the "true" values of the (overall/marginal/conditional) measures are relatively close to zero (between 0.01 and 0.05), the sampling distributions of the estimators appear to be symmetrical about the estimated means. Their variances and biases behave similarly to those in the case when the "true" values of the (overall/marginal/conditional) measures stay away from zero (larger than 0.05), but the biases are more notable.

4. For the cases where the "true" values of the (overall/marginal/conditional) measures are very close to zero (less than 0.01), the sampling distributions of the estimators of the association measures appear to be skewed to the right, and they become symmetrical as the sample size increases, regardless of the number of categories, the type of association and its strength. Their variances and biases behave similarly to those when the "true" values of the (overall/marginal/conditional) measures are relatively close to zero (between 0.01 and 0.05) but the biases appear more pronounced.

5. When the number of categories in each variable increases from 3 to 7, the number of cells in a contingency table increases from $3^3 = 27$ to $7^3 = 343$, which results in the sparseness of the contingency table with the sample sizes fixed. Therefore, the biases increase as the number of categories increases at a fixed sample size. Note that Agresti (2002, p. 391–398) pointed out that this bias issue is a common problem of the methods for the contingency tables such as odds ratio and the parameter estimates in log-linear models.

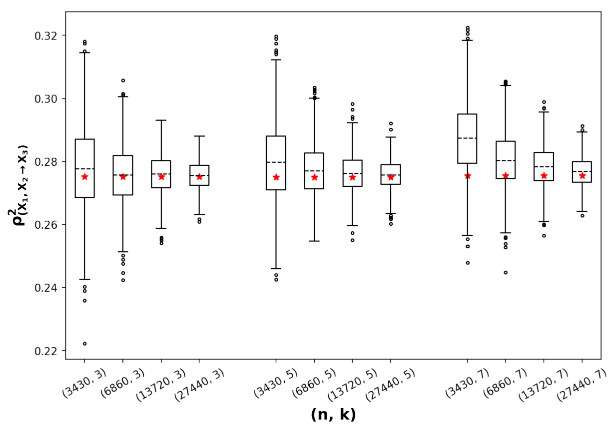
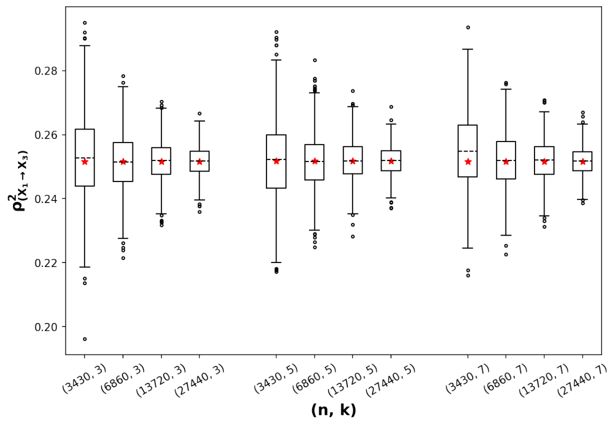
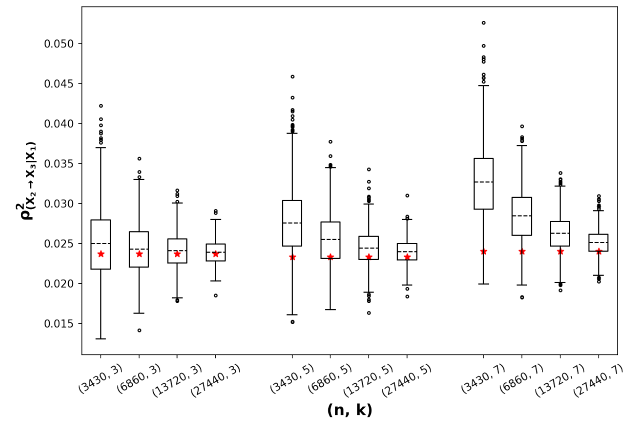


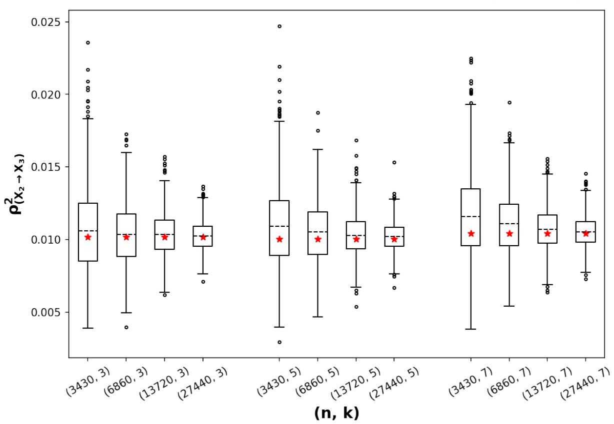
Figure 6.7: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



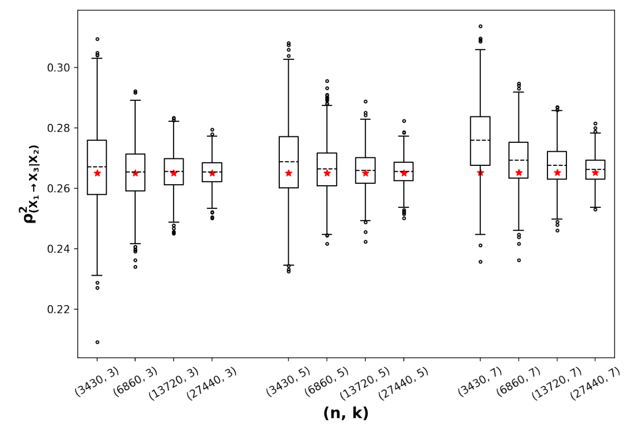
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.8: the marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

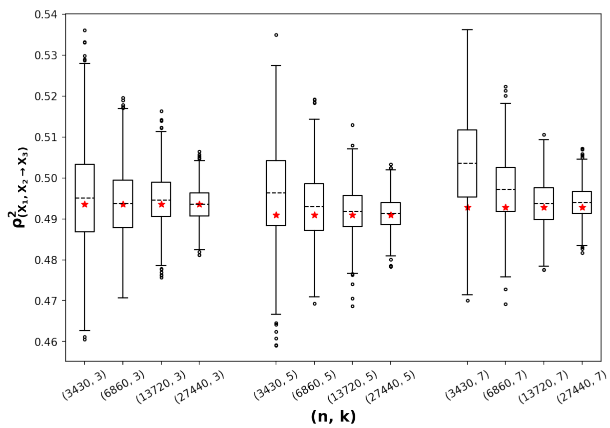
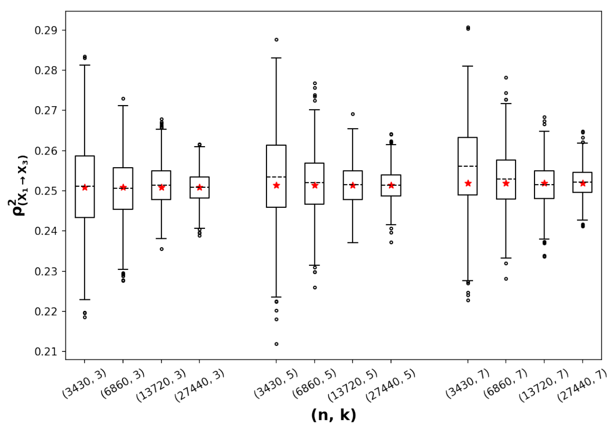
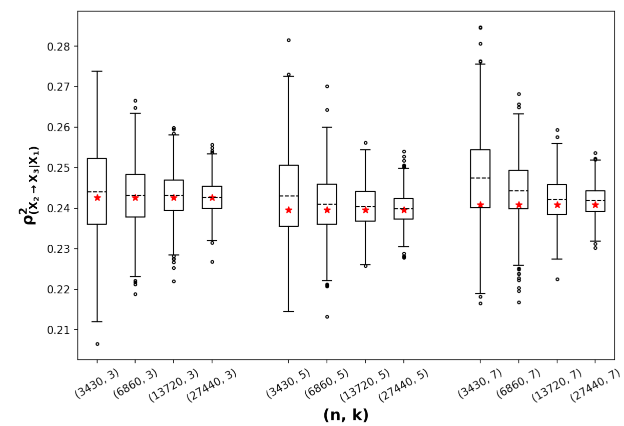


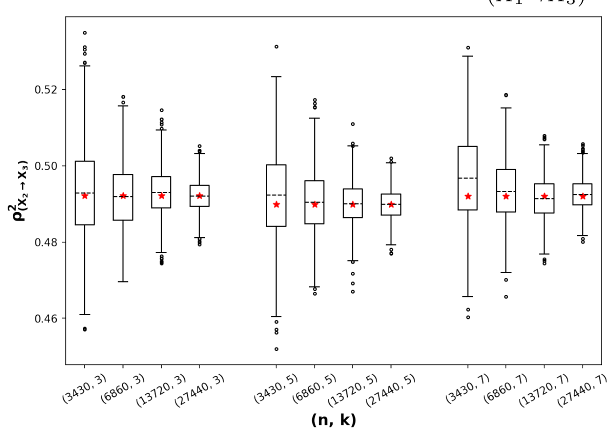
Figure 6.9: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



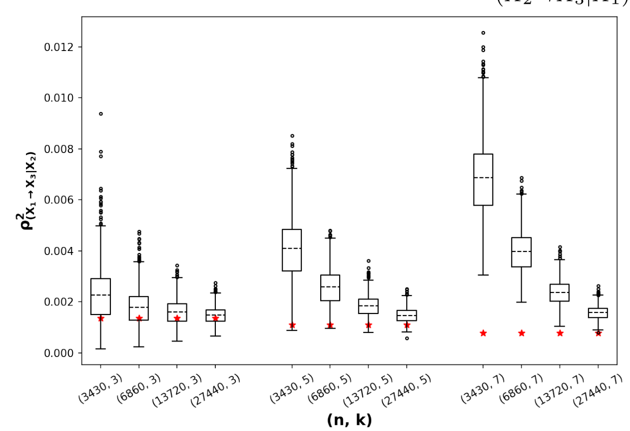
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure 6.10: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

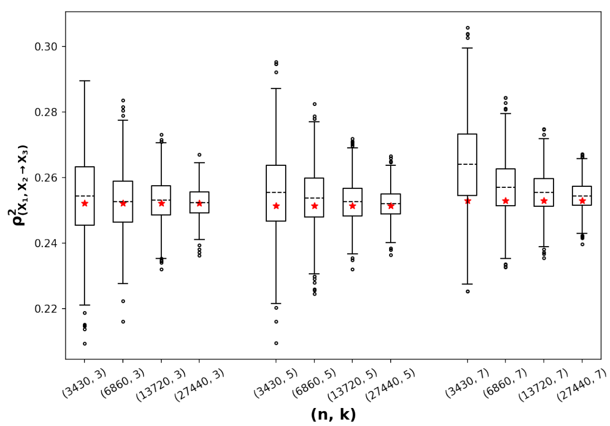
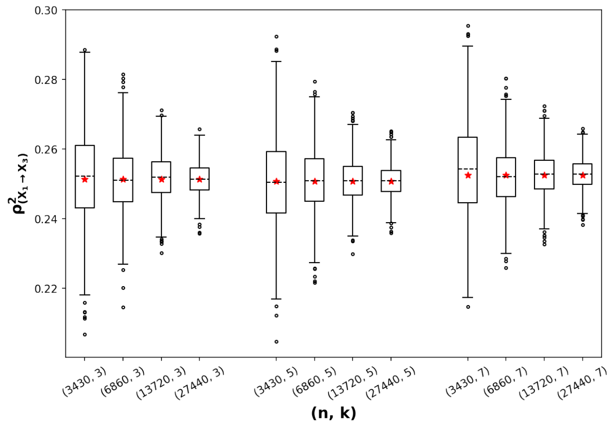
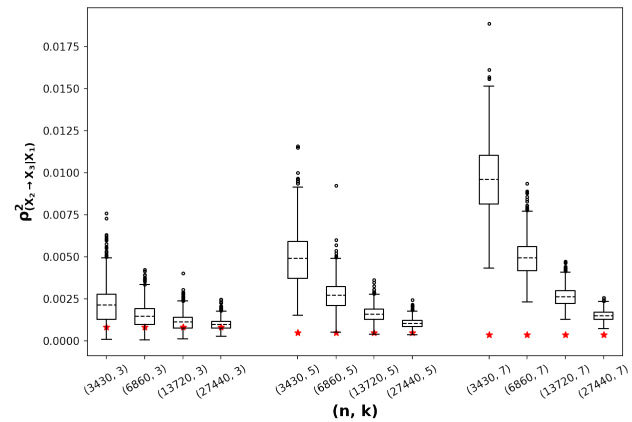


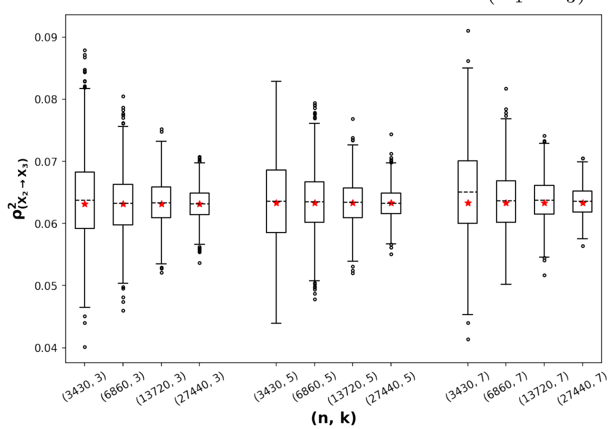
Figure 6.11: the overall association measure for $auto1corr(X_i, X_j) = \phi^{|(i \mod 3)-(j \mod 3)|}$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$



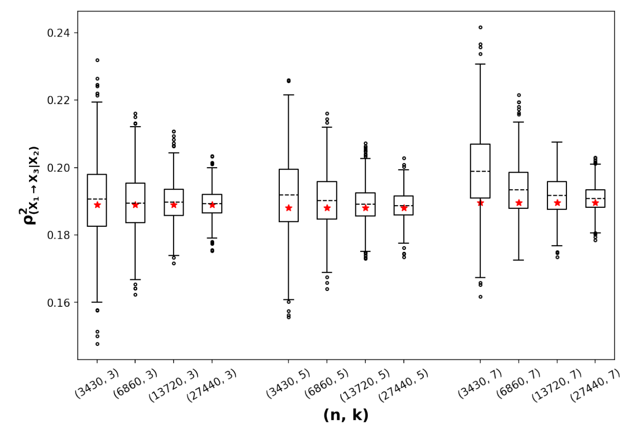
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.12: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$

6.3.1.4 Scenario 4

Factor Name	Values
the number of categories in each variable ($k_{X_i} = k$)	3, 5, 7
sample size (n)	(270, 540, 1080, 2160) for $k_{X_i} = 3$, (1250, 2500, 5000, 10000) for $k_{X_i} = 5$, (3430, 6860, 13720, 27440) for $k_{X_i} = 7$

Table 6.7: Simulation factors for Scenario 4 in three-dimensional case

Table 6.7 provides the values for the number of categories in each variable denoted by k_{X_1}, k_{X_2} and k_{X_3}, and the set of sample sizes n employed in this scenario. Note that, given that $k_{X_1} = k_{X_2} = k_{X_3} = k$, the sample sizes are $n = (10k^3, 20k^3, 40k^3, 80k^3)$ for $k = 3, 5, 7$, by taking into consideration four different conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.13 to 6.18 shows the selected simulation results regarding the three types of association: $pcorr(X_3, X_1|X_2) > 0$, $pcorr(X_3, X_1|X_2) = 0$ and $auto1corr(X_i, X_j) = \phi^{|(i \mod 3)-(j \mod 3)|}$ with moderate association $|corr(X_3, X_1)| = 0.5$ where $i, j = 1, 2, 3$. Note that the remaining simulation results for Scenario 4 are given in Appendix J. The summary of the simulation results for Scenario 4 was made as follows:

1. The observations made in Scenario 2 about the overall association measures, as well as those for the marginal and conditional association measures, are all applicable in this case.

2. When the "true" values of the (overall/marginal/conditional) measures stay away from zero (larger than 0.05), the sampling distributions of the estimators of the association measures appear to be symmetrical about the estimated means. When the "true" values of the (overall/marginal/conditional) measures are relatively close to 0 (between 0.01 and 0.05), the sampling distributions of the estimators appear to be

asymmetrical about the estimated means for (small/moderate) sample sizes and/or the small number of categories (i.e. $k = 3$). When the "true" values of the (overall/marginal/conditional) measures are very close to 0 (less than 0.01), the sampling distributions of the estimators of the association measures appear to be quite asymmetrical (especially when the number of categories is small), and they become symmetrical and are concentrated towards zero as the sample size increases.

3. The variances of the estimators of the (overall/marginal/conditional) association measures decrease and their biases become zero as the sample size increases at a fixed number of categories, regardless of the type of association and its strength. Hence, when the "true" value of an association measure is almost 0, its estimator may be upward-biased with respect to its true one.

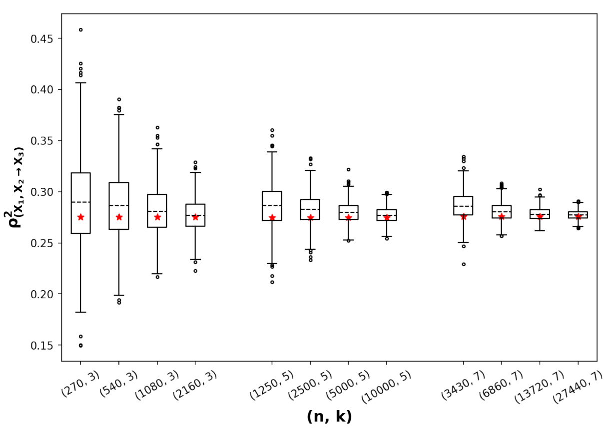
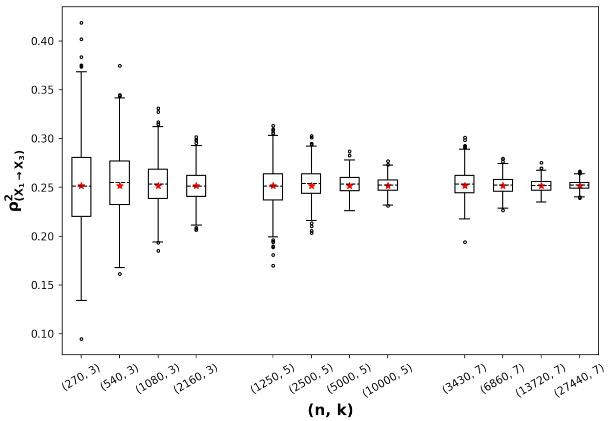
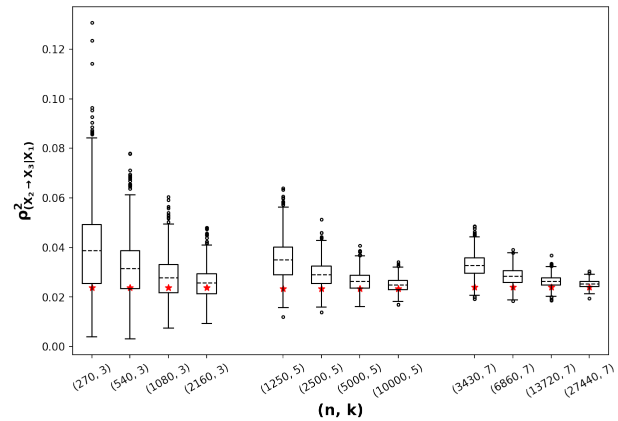


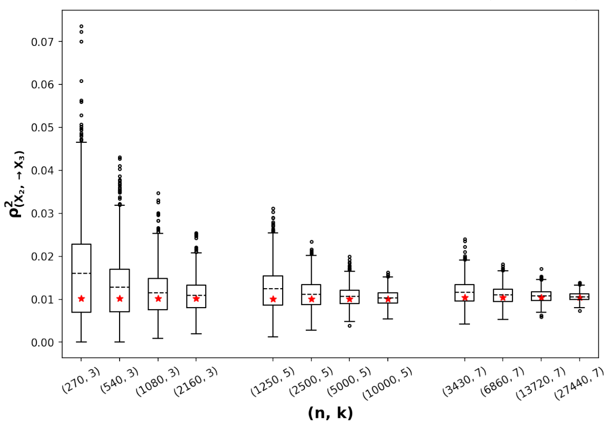
Figure 6.13: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



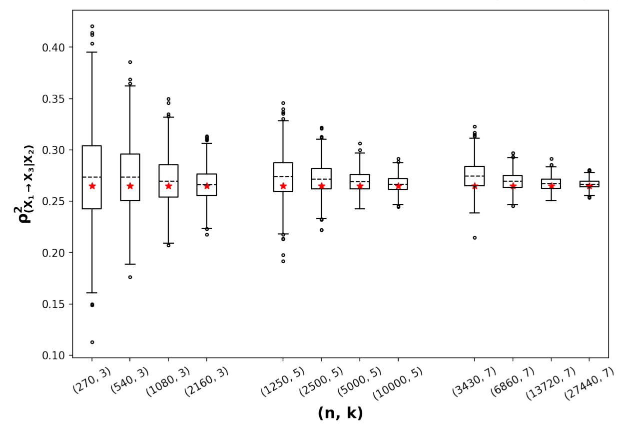
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.14: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

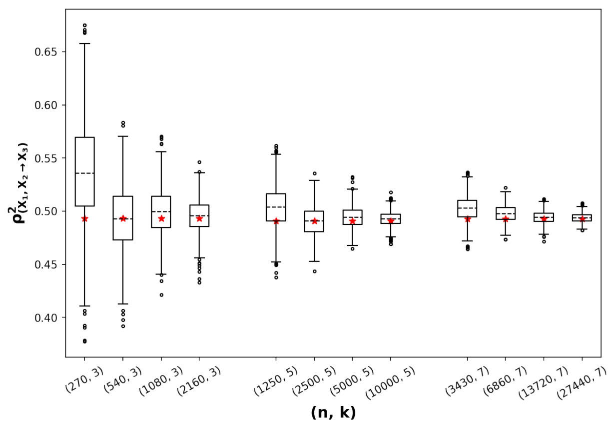
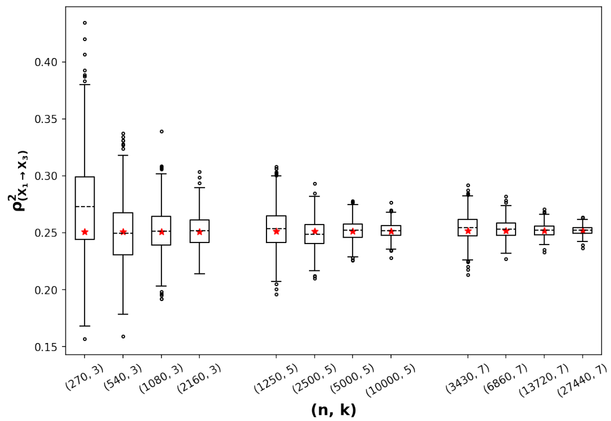
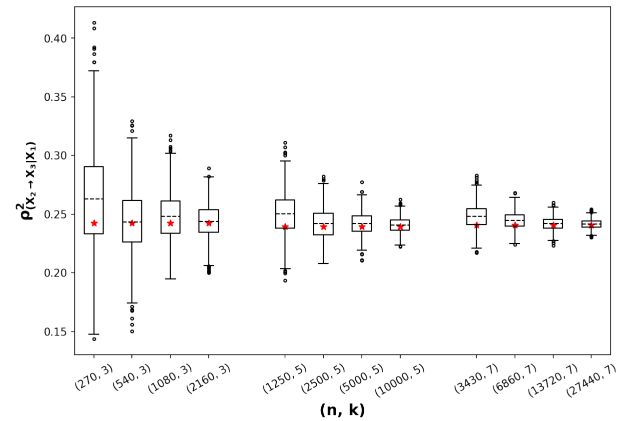


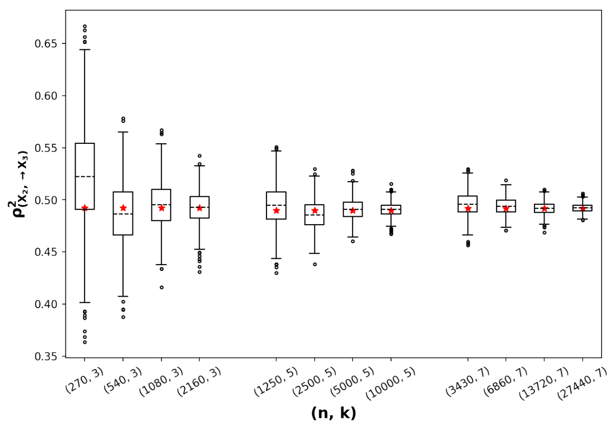
Figure 6.15: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$



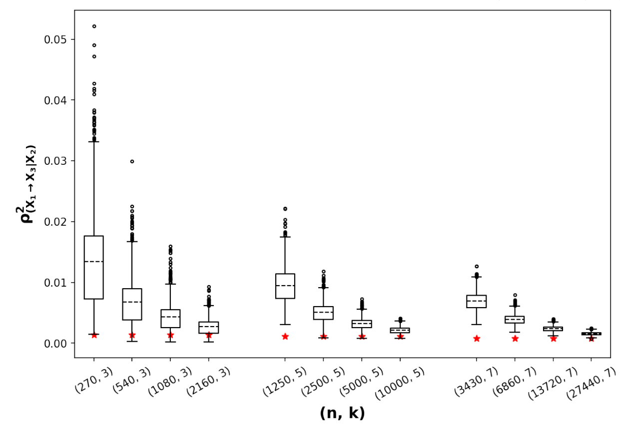
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.16: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

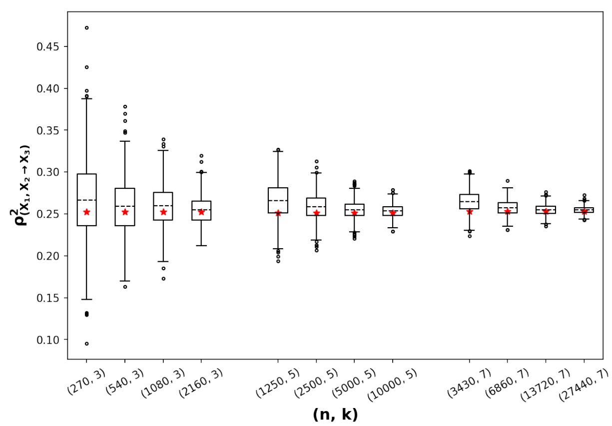
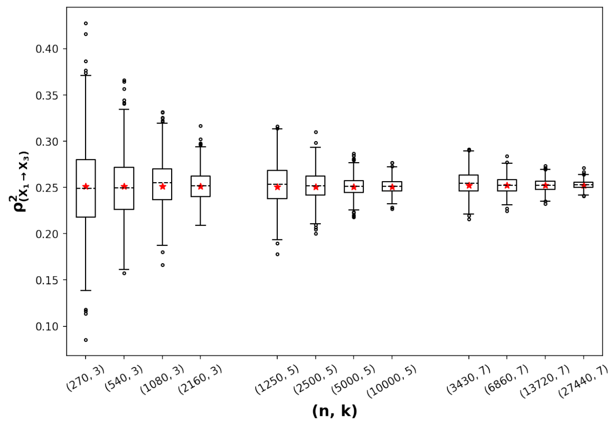
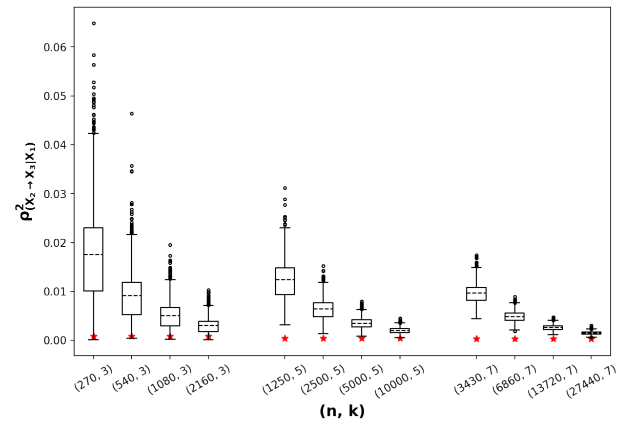


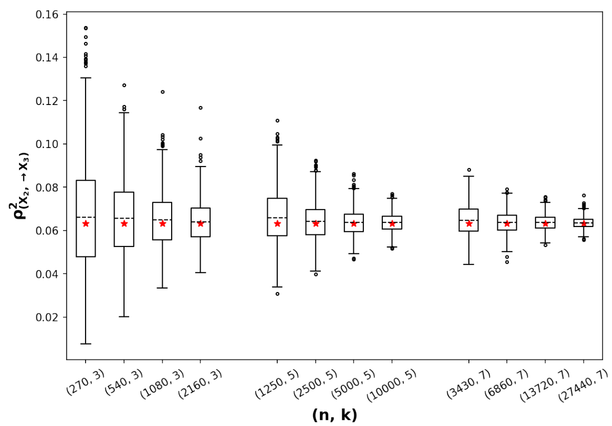
Figure 6.17: The overall association measure for $auto1corr(X_i, X_j) = \phi^{|(i \mod 3)-(j \mod 3)|}$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$



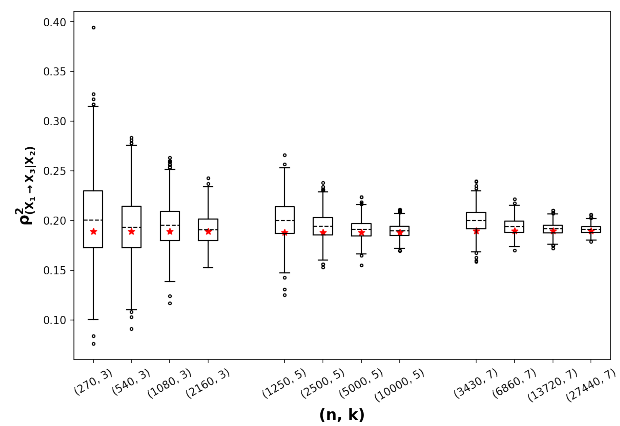
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure 6.18: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$

6.3.2 Five-dimensional case ($d = 5$)

Let $X_1, \ldots, X_5$ be the ordinal variables in a five-dimensional contingency table, where X_5 is the dependent variable and $(X_1, \ldots, X_4)$ are the independent variables.

6.3.2.1 Scenario 2

Factor Name	Values
the number of categories for $X_1, \ldots, X_4$ ($k_{X_1} = \ldots = k_{X_4} = k$)	3
the number of categories for X_5 (k_{X_5})	(3, 5, 7)
sample size (n)	(5670, 11340, 22680, 45360)

Table 6.8: Simulation factors for Scenario 2 in five-dimensional case

Table 6.8 provides the settings for the number of categories in each variable denoted by $k_{X_1}, \ldots, k_{X_5}$, and the set of sample sizes n used in this scenario. Note that the $k_{X_1}, \ldots, k_{X_4}$ are fixed to be 3, k_{X_5} varies from 3 to 7, and the sample sizes are fixed to be $n = (5670, 11340, 22680, 45360)$ for each level of k_{X_5} by considering the maximum number of cells $(3^4 \times 7 = 567)$ and four different conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.19 to 6.22 shows the selected simulation results regarding the two types of association: $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ with the moderate strength of association $|corr(X_5, X_1)| = 0.5$. Note that the remaining simulation results for Scenario 2 are provided in Appendix J. Based on the simulation results for Scenario 2 below, we make similar observations to those for the three-dimensional ($d = 3$) case provided in Section 6.3.1.2.

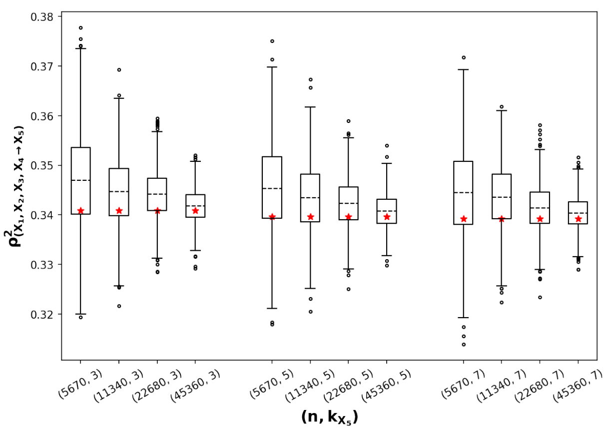
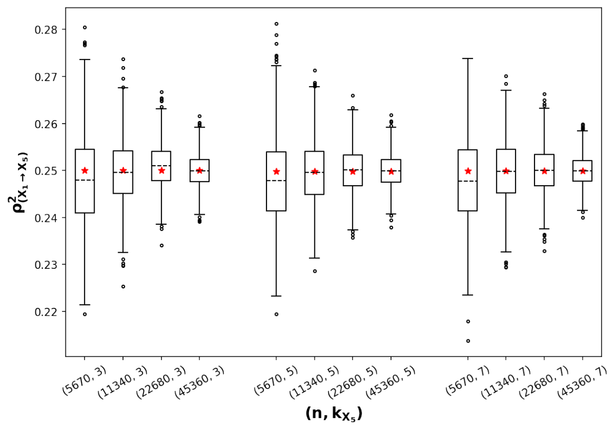
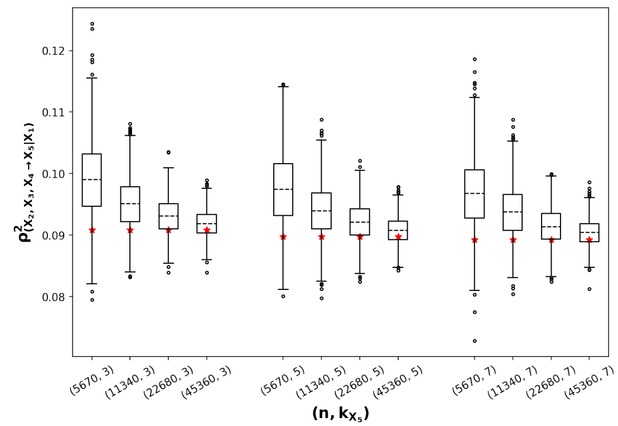


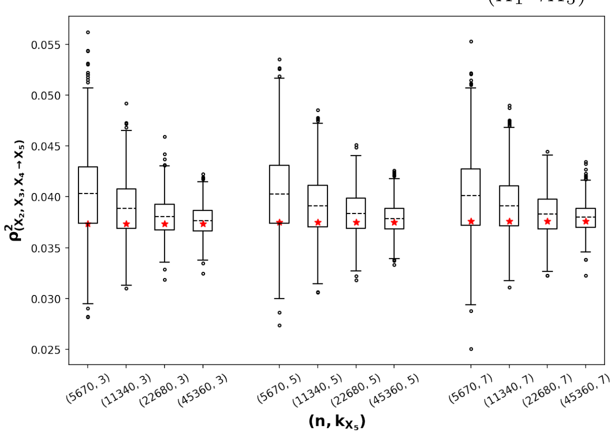
Figure 6.19: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



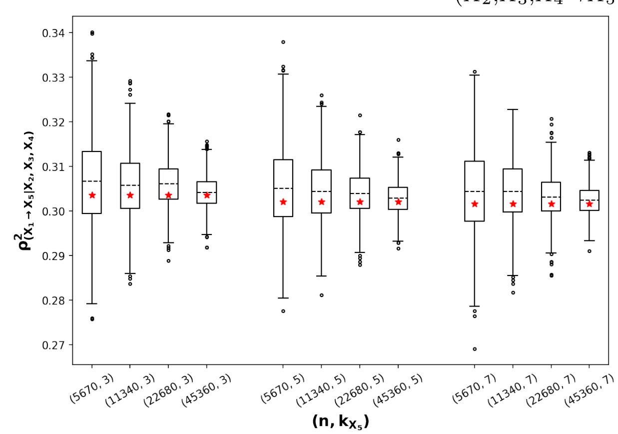
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2,X_3,X_4 \to X_5|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2,X_3,X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5|X_2,X_3,X_4)}$

Figure 6.20: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$

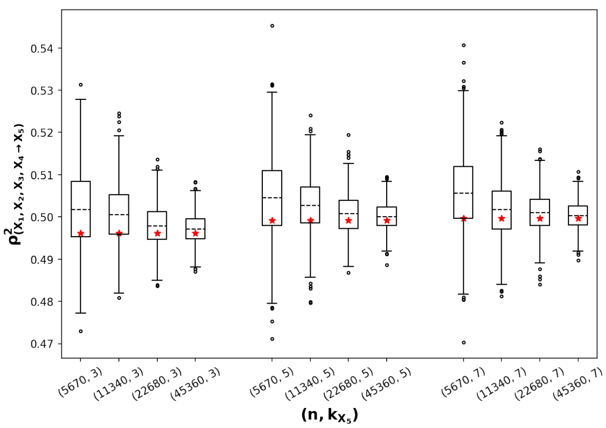
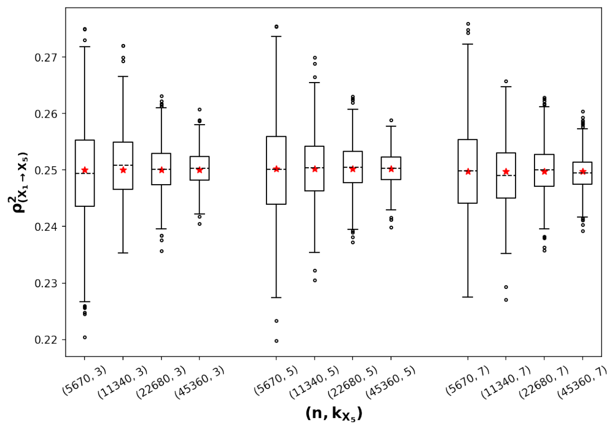
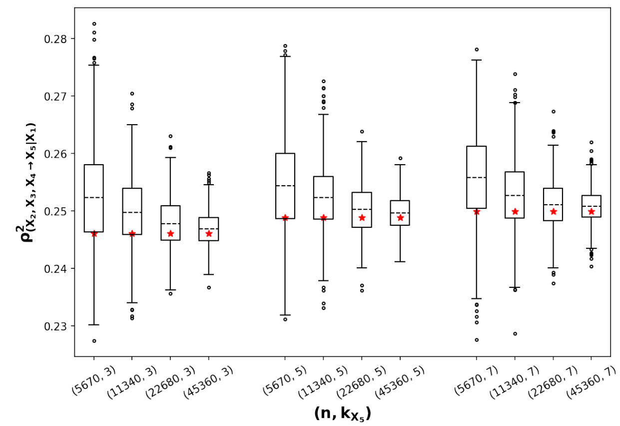


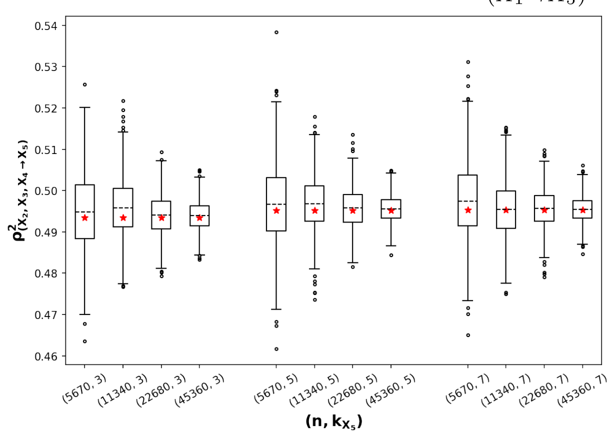
Figure 6.21: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



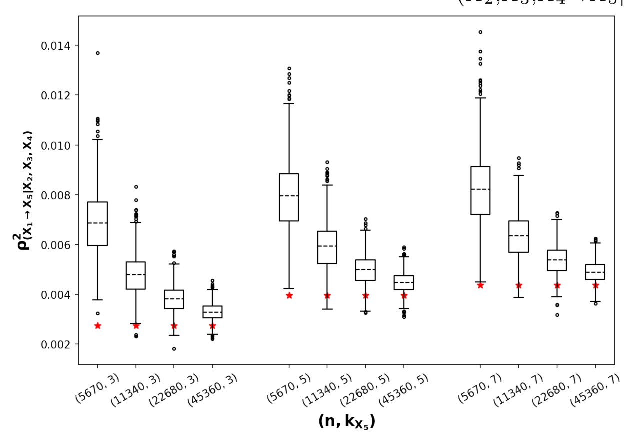
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2, X_3, X_4 \to X_5 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2, X_3, X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2, X_3, X_4)}$

Figure 6.22: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) = 0$ and moderate association $|corr(X_5, X_1)| = 0.5$

6.3.2.2 Scenario 3

Factor Name	Values
the number of categories in each variable ($k_{X_i} = k$)	3, 5, 7
sample size (n)	(168070, 336140, 672280, 1344560)

Table 6.9: Simulation factors for Scenario 3 in five-dimensional case

Table 6.9 provides the settings for the number of categories in each variable denoted by $k_{X_1} \ldots, k_{X_5}$, and the set of sample sizes n considered in this scenario. Note that, given that $k_{X_1} = \cdots = k_{X_5} = k$, the sample sizes are equal to $n = (168070, 336140, 672280, 1344560)$ for all the combinations of $(k_{X_1}, \ldots, k_{X_5})$, by considering the maximum number of cells ($7^5 = 16807$) and four different conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.23 to 6.28 shows the selected simulation results regarding the three types of association: $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$, $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and $auto1corr(X_i, X_j) = \phi^{|(i \mod 5)-(j \mod 5)|}$ with moderate strength of association $|corr(X_5, X_1)| = 0.5$ where $i, j = 1, \ldots, 5$. Note that the remaining simulation results for Scenario 3 are given in Appendix J. For the simulation results below, we observe similar patterns to those for the three-dimensional ($d = 3$) case given in Section 6.3.1.3.

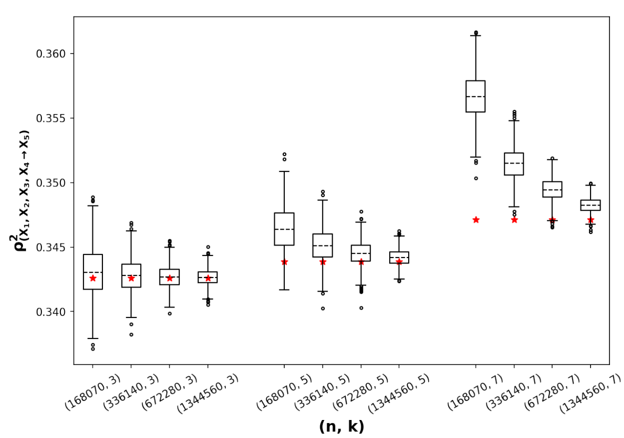
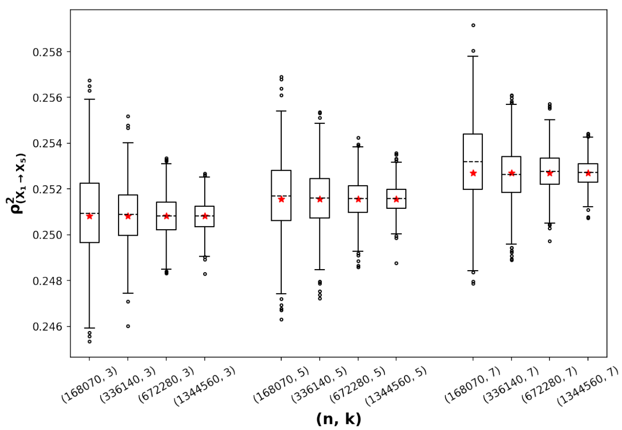
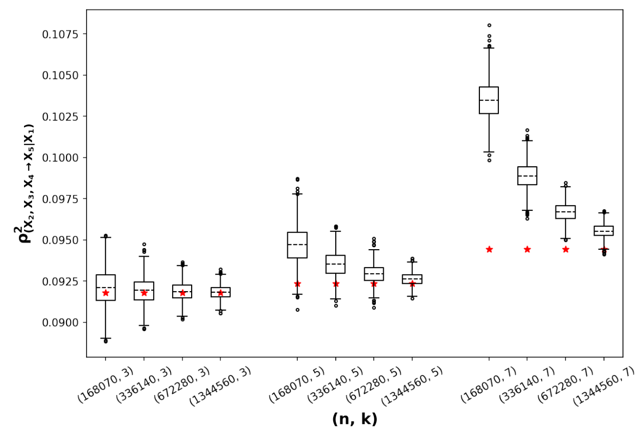


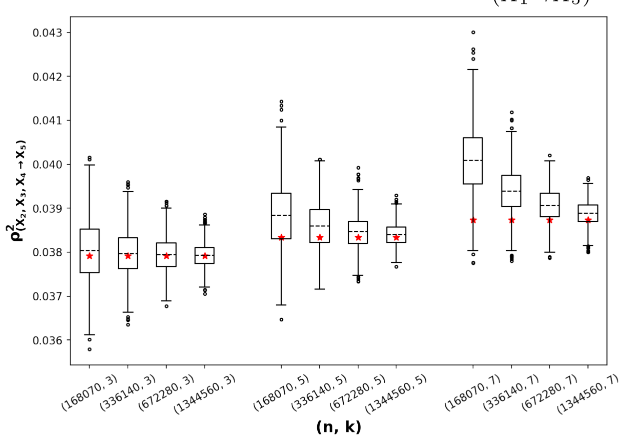
Figure 6.23: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



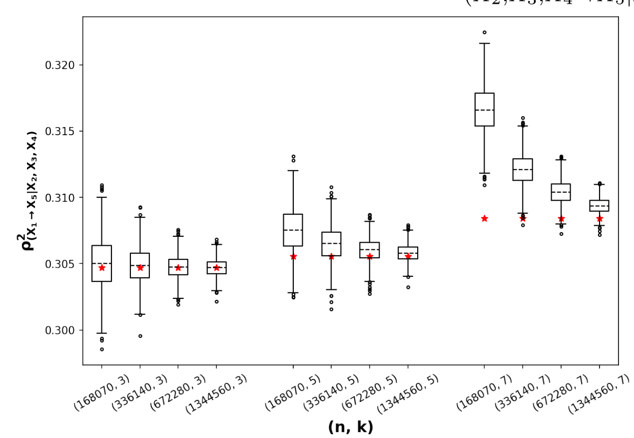
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2, X_3, X_4 \to X_5 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2, X_3, X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5 | X_2, X_3, X_4)}$

Figure 6.24: the marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$

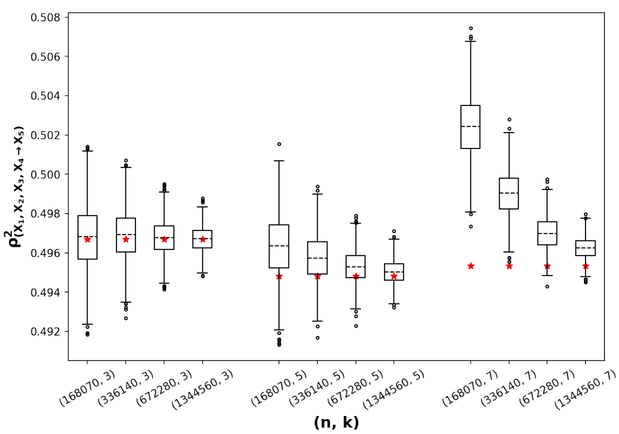
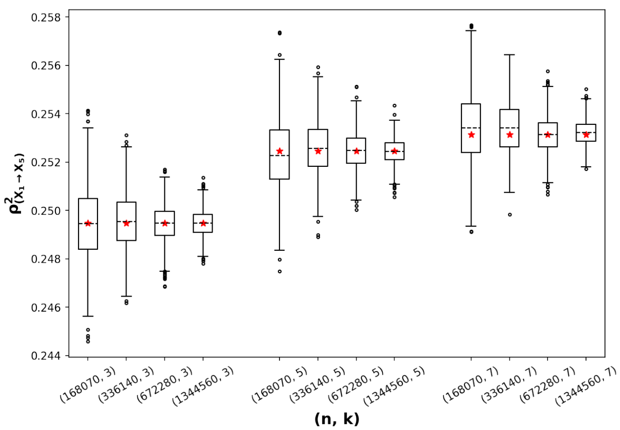
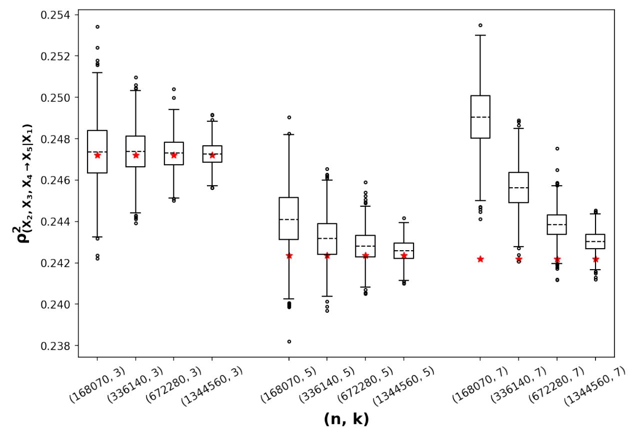


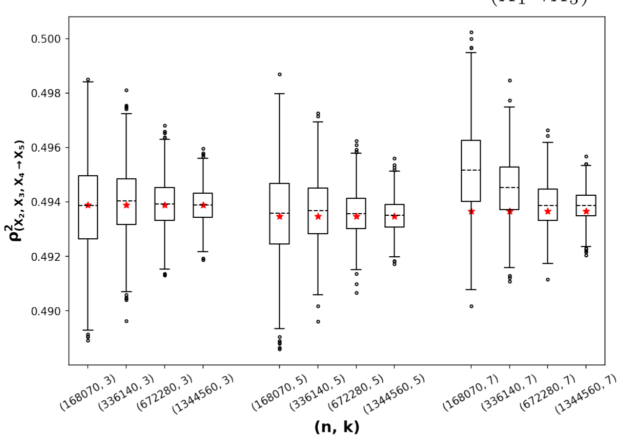
Figure 6.25: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



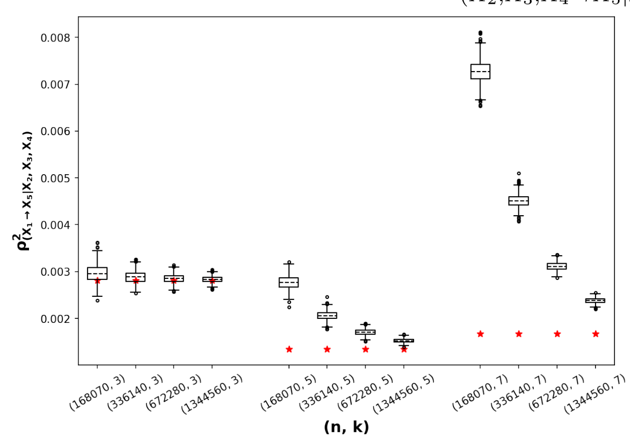
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2,X_3,X_4 \to X_5|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2,X_3,X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5|X_2,X_3,X_4)}$

Figure 6.26: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and moderate association $|corr(X_5, X_1)| = 0.5$

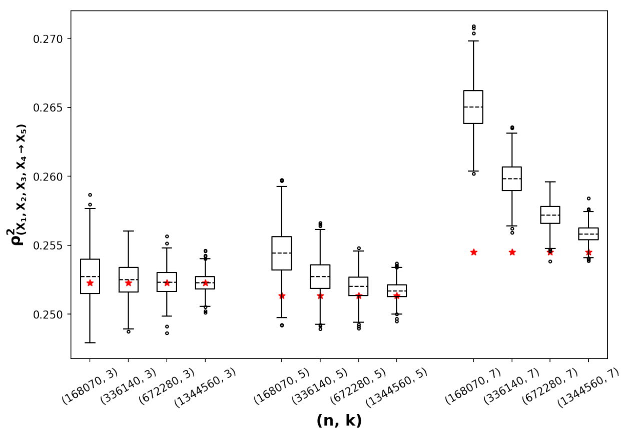
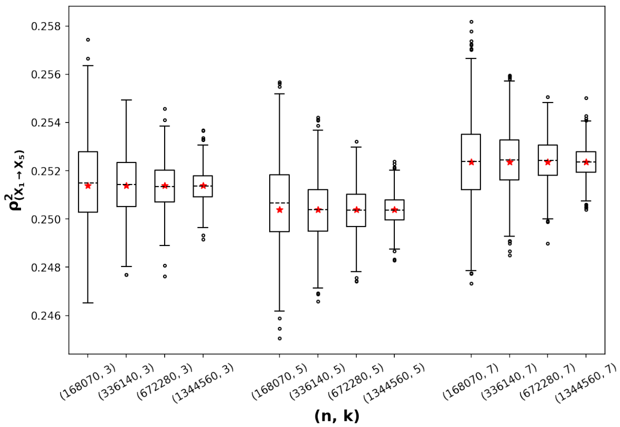
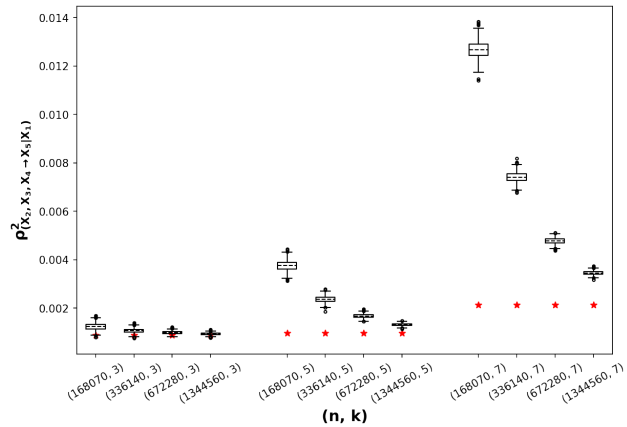


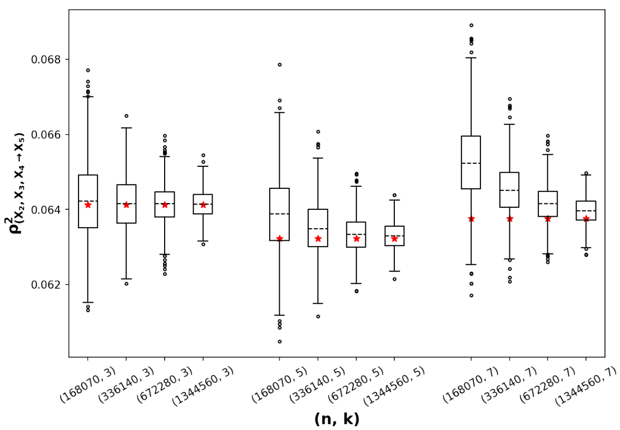
Figure 6.27: the overall association measure for $auto1corr(X_i, X_j) = \phi^{|(i \mod 5)-(j \mod 5)|}$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$



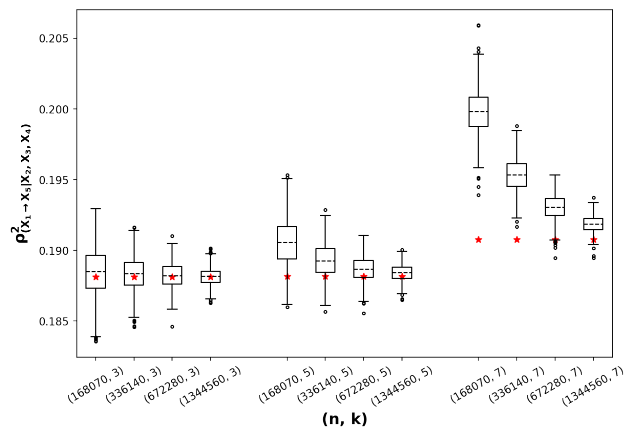
(a) The marginal association measure $\rho^2_{(X_1,X_2,X_3,X_4 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2,X_3,X_4 \to X_5|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2,X_3,X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5|X_2,X_3,X_4)}$

Figure 6.28: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$

6.3.2.3 Scenario 4

Factor Name	Values
the number of categories in each variable ($k_{X_i} = k$)	3, 5, 7
sample size (n)	(2430, 4860, 9720, 19440) for $k_{X_i} = 3$, (31250, 62500, 125000, 250000) for $k_{X_i} = 5$, (168070, 336140, 672280, 1344560) for $k_{X_i} = 7$

Table 6.10: Simulation factors for Scenario 4 in five-dimensional case

Table 6.10 provides the values for the number of categories in each variable denoted by $k_{X_1}, \ldots, k_{X_5}$, and the set of sample sizes n employed in this scenario. Note that, given that $k_{X_1} = \ldots = k_{X_5} = k$, the sample sizes are equal to $n = (10k^3, 20k^3, 40k^3, 80k^3)$ for $k = 3, 5, 7$, by taking into consideration four different conceptual average cell counts $(10, 20, 40, 80)$ in the contingency table. Figure 6.29 to 6.34 shows the selected simulation results regarding the three types of association: $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$, $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and $auto1corr(X_i, X_j) = \phi^{|(i \mod 5)-(j \mod 5)|}$ with moderate association $|corr(X_5, X_1)| = 0.5$ where $i, j = 1, \ldots, 5$. Note that the remaining simulations for Scenario 4 are provided in Appendix J. We notice that the simulation results below are similar to those for the three-dimensional ($d = 3$) case presented in Section 6.3.1.4.

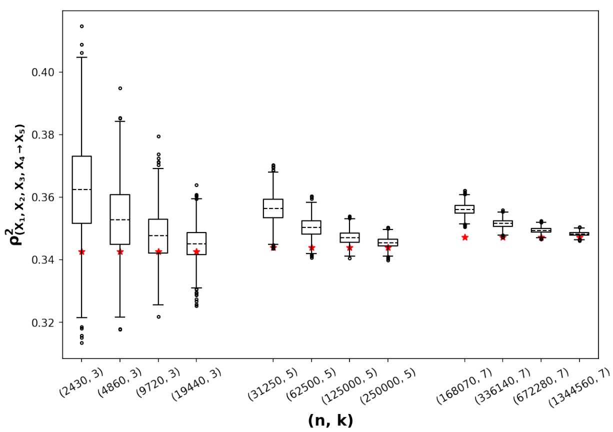
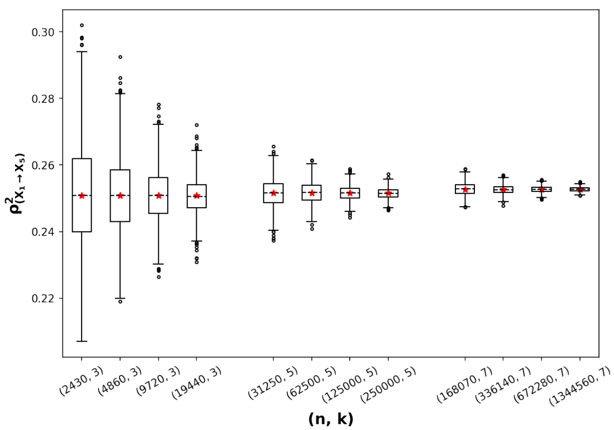
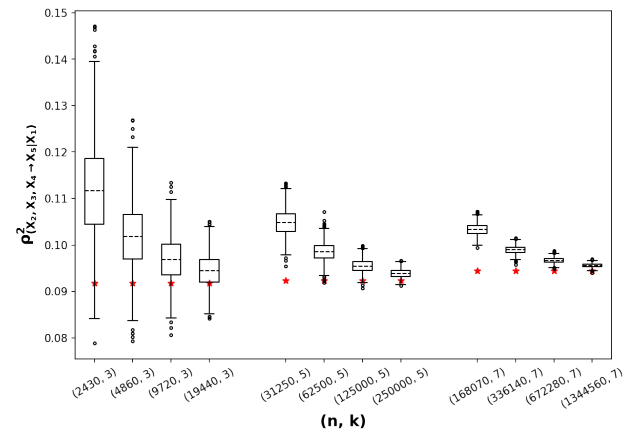


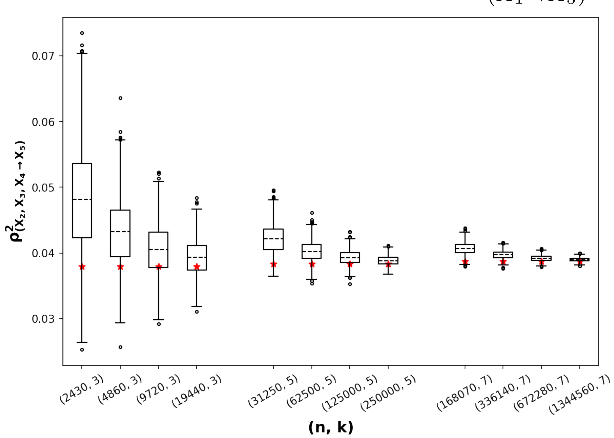
Figure 6.29: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



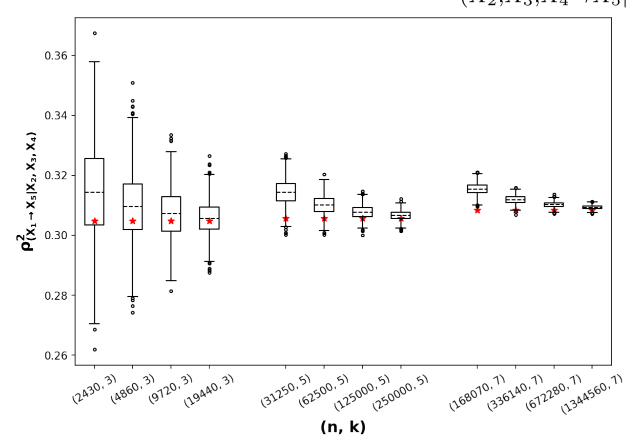
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2,X_3,X_4 \to X_5|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2,X_3,X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5|X_2,X_3,X_4)}$

Figure 6.30: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and moderate association $|corr(X_5, X_1)| = 0.5$

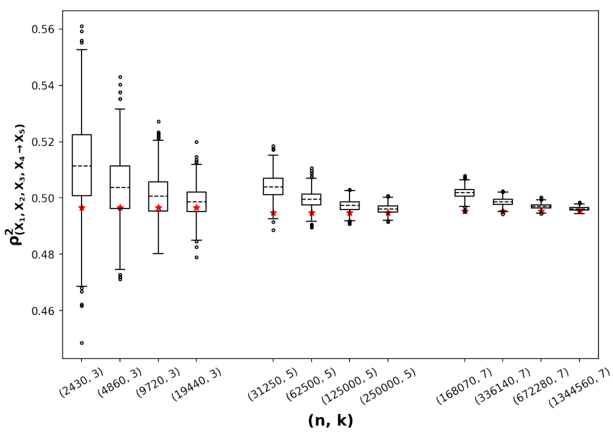
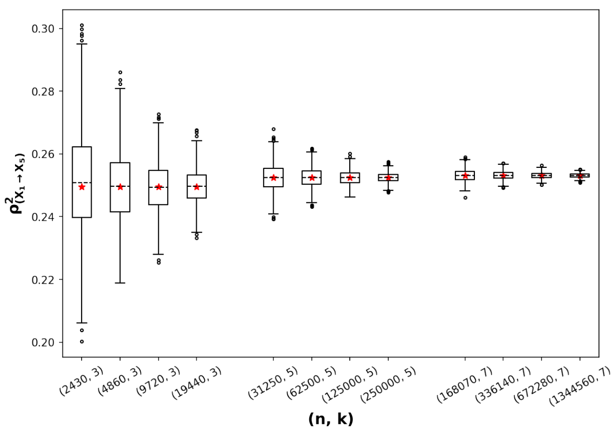
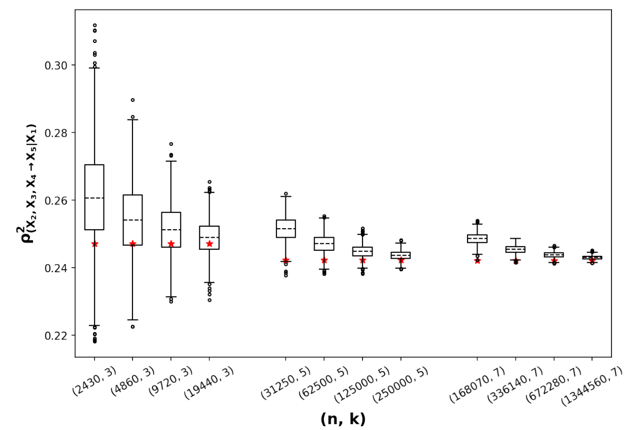


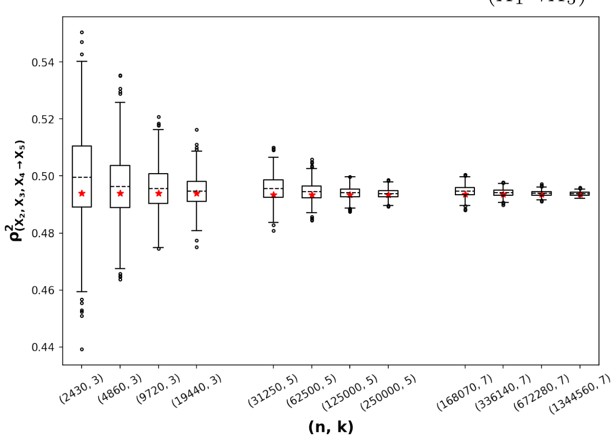
Figure 6.31: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and moderate association $|corr(X_5, X_1)| = 0.5$



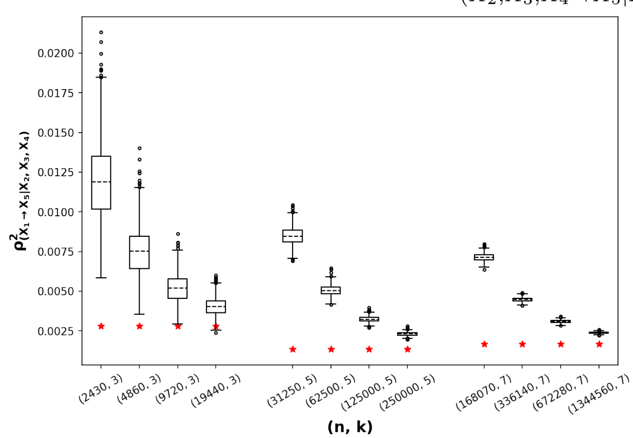
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2,X_3,X_4 \to X_5|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2,X_3,X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2,X_3,X_4)}$

Figure 6.32: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and moderate association $|corr(X_3, X_1)| = 0.5$

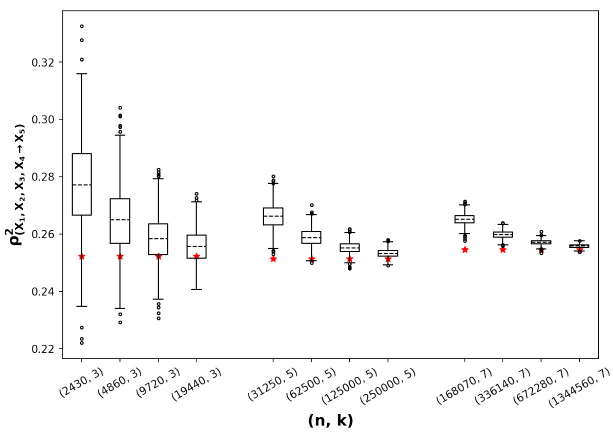
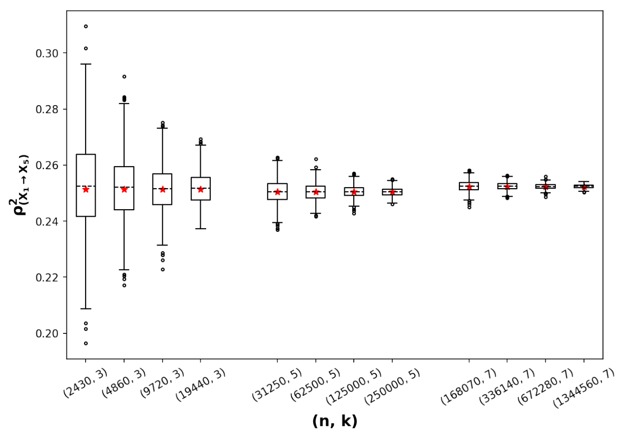
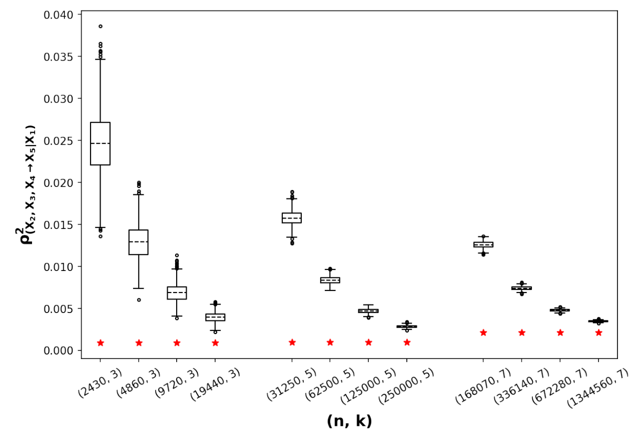


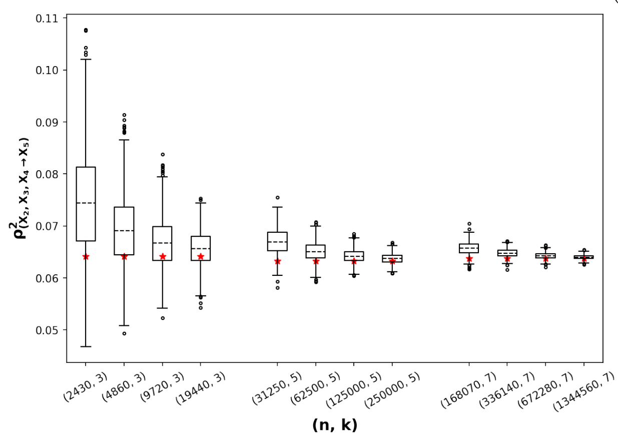
Figure 6.33: The overall association measure for $auto1corr(X_i, X_j) = \phi^{|(i \mod 5)-(j \mod 5)|}$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$



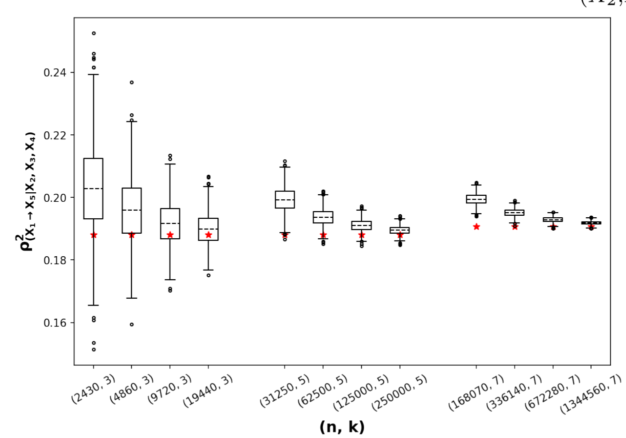
(a) The marginal association measure $\rho^2_{(X_1 \to X_5)}$



(b) The conditional association measure $\rho^2_{(X_2, X_3, X_4 \to X_5 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2, X_3, X_4 \to X_5)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_5 | X_2, X_3, X_4)}$

Figure 6.34: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$

CHAPTER 7

DATA ANALYSIS

In this chapter, we demonstrate the performance of the proposed subcopula regression, subcopula regression based association measure and its two decompositions using four real data sets: the *ice cream* data set (*The Ice Cream Study at Penn State*, 2012), *acute migraine* data set (Vandenhende and Lambert, 2000), *nuclear accident* data set (Fienberg et al., 1985), *post-operative patients* data set (Budihardjo et al., 1991). Each of these real data sets is adopted for a specific analytic goal as stated below.

First, the goal of analyzing the *ice cream* data is to assess the performance of the proposed association measure and other existing non-model based measures on a non-monotone nonlinear (quadratic) association structure between the dependent variable (rating on the ice cream) and the independent variable (fat level in ice cream).

Secondly, the intent of the analysis of the *acute migraine* data is to describe the association between the dependent variable (Pain score) and two independent variables (Treatment Groups and Occasions) by way of the proposed overall association measure and the prediction of the proposed subcopula regression. We also quantify the contribution of the two independent variables to the overall association measure through its sequential and non-sequential decompositions.

Thirdly, from the analysis of the *nuclear accident* data, we intend to understand a time-dependent association between the dependent variable (stress level on the fourth interview) and a set of independent variables (stress levels from the first three interviews and the distance from the nuclear plant). To this end, we focus on the overall association mea-

sure of the independent variables on the dependent variable and the marginal/conditional association measures resulting from the sequential decomposition. Note that we also obtain the prediction of the dependent variable using the subcopula regression to facilitate the understanding of the time-dependent association structure.

Finally, the purpose of the analysis of the *post-operative patients* data set is to illustrate the utility of the proposed overall association measure on the variable selection problem for a multi-dimensional contingency table, which is regarded as one of the critical preprocessing steps in statistical learning. Thus, we initiate the process of variable selection by performing the *all-possible-subset subcopula regressions* to choose the candidate sets of independent variables. Then we search for a potentially important subset of the independent variables with the maximum proportional increment in the overall association measure by considering the contribution of each newly added independent variable to the overall association measure given that all other independent variables selected previously are in the subcopula regression.

7.1 Ice cream data

The Pennsylvania State University (Berkey) Creamery is the largest university creamery in the U.S., manufacturing its famous ice cream, along with cheese, milk and other products. All the products are made through the Department of Food Science and researchers are constantly interested in finding the optimal amount of fat to produce buttery, creamy and rich flavor of ice cream. In this study, 493 randomly selected subjects tasted and rated the ice cream with different fat levels. The rating (*R*) was ordinal with a *Likert scale* from 1 (didn't like it all) to 9 (yum yum!). The fat level (*F*) consisted of an increasing sequence of 8 values ranged from 0.00 to 0.28 with the increment 0.04. Table 7.1 presents the data set in a 8×9 contingency table. As the purpose of this study was to find the optimal amount of fat in the ice cream to attain the highest average rating, *R* and *F* are the dependent and independent variable, respectively.

Fat level (F) \ Rating (R)	1	2	3	4	5	6	7	8	9
0.00	4	17	8	16	5	6	4	2	1
0.04	1	1	5	6	7	9	21	12	0
0.08	0	2	2	2	4	13	16	21	3
0.12	1	1	1	3	4	11	15	23	4
0.16	0	3	2	6	3	7	17	17	5
0.20	0	1	3	8	4	13	14	11	8
0.24	1	5	4	14	2	13	13	7	2
0.28	4	6	9	11	5	9	7	8	3

Table 7.1: The ice cream study at the Pennsylvania State University with 493 subjects rating the ice cream with different fat levels

Table 7.1 shows that the pattern between the rating and the fat level appears to be quadratic in the sense that when the fat level becomes too low or too high, rating decreases, compared with ratings with the moderate fat level. In order to take into account that rating is the ordinal response and its relationship with the fat level is quadratic, Dr. William Harkness, Professor Emeritus of Statistics, in the Department of Statistics at the Pennsylvania State University employed the proportional odds cumulative logit model with the linear and quadratic terms of fat level as the predictors:

$$\log\left(\frac{Pr(R \le r_i)}{1 - Pr(R \le r_i)}\right) = \alpha_i + \beta_1 F + \beta_2 F^2,$$

where r_i is the i-th ($i = 1, \ldots, 9$) category of R, and α_i, β_1 and β_2 are the parameters. Note that the proportional odds assumption was reasonable based on an insignificant p-value of the score test (0.4173). From the fitted proportional odds cumulative logit model, the optimal amount of fat in ice cream was found to be 0.14372. According to the Berkey Creamery's website (`https://creamery.psu.edu/customer-service`), over 100 ice cream flavors have a butterfat content of 14.1%.

Because we are interested in the performance of the proposed association measure and other existing non-model based measures on a non-monotone nonlinear (quadratic) association structure between the rating and fat level, we first calculate five association

Name of measure	Estimate	95% asymptotic C.I.
Goodman-Kruskal's tau	0.044	(0.029, 0.058)
Goodman-Kruskal's lambda	0.080	(0.022, 0.137)
Theil's uncertainty coefficient	0.080	(0.058, 0.102)
Somers' D	0.035	(-0.038, 0.109)
$\rho^2_{(F\to R)}$	0.208	(0.142, 0.274)

Table 7.2: The five association measures quantifying the strength of association between the rating of ice cream and fat level

measures: *Goodman-Kruskal's tau*, *lambda*, *Theil's uncertainty coefficient*, *Somers' D* and the proposed subcopula regression based association measure, $\rho^2_{(F\to R)}$. Table 7.2 shows the estimate of each measure with the corresponding 95% asymptotic confidence interval. The results indicate that there is a significant association between *R* and *F*, except for *Somers' D*. Note that *Goodman-Kruskal's tau*, *lambda* and *Theil's uncertainty coefficient* do not account for the ordinality of the rating variable. The proposed association measure appears to have larger magnitude than the other four measures do. In addition, the estimate of the proposed association measure implies that 20.8% of variance for the subcopula scores of *R* can be explained by *F* via the subcopula regression on average. The 95% bootstrap confidence interval for $\rho^2_{F\to R}$ is $(0.141, 0.263)$. The estimated p-value from the permutation test of $H_0 : \rho^2_{(F\to R)} = 0$ is 0.0010 and its relative error is 0.0316.

To highlight the usefulness of the proposed subcopula regression, we predict the category of *R* for each level of *F* through the subcopula regression. Table 7.3 contains the predicted category of *R* for a given level of *F* with the bootstrap estimation uncertainty. Note that the number for each combination of the categories of *F* and *R* is the percentage of times that a category of *R* is predicted for an observed *F* over 1000 bootstrap samples. The prediction result reveals the pattern between *R* and *F* is quadratic in nature, which is consistent with the result from the fitted proportional odds cumulative logit model mentioned above. Moreover, the amount of fat in the ice cream that yields the highest average rating appears to be between 0.08 and 0.16 which includes the optimal value of fat level (0.14372) estimated by the proportional odds cumulative logit model.

Fat level (F)	Rating (R)								
	1	2	3	4	5	6	7	8	9
0.00				**52.1**	46.0	1.6			
0.04						3.9	**96.1**		
0.08							**87.8**	12.2	
0.12							**73.2**	26.8	
0.16							**93.4**	6.6	
0.20						0.1	**99.4**	0.5	
0.24						**67.0**	33.0		
0.28					3.0	**92.7**	4.3		

Table 7.3: The predicted category of the rating for each fat level with bootstrap estimation of uncertainty. Note that the category of R in the column with bold, red colored number are the category of R predicted by the subcopula regression.

7.2 Acute migraine data

Table 7.4 summarizes the data from a longitudinal study on acute migraine. In this study, 39 subjects with moderate to severe migraine were randomly assigned to one of the three treatment groups with an investigational drug LY334370 (1–placebo, 2–5 mg of LY334370, 3–20 mg of LY334370) and the severity of migraine pain was recorded at eight occasions (0.5, 1, 1.5, 2, 3, 4, 6, 24 hours) on four levels (1–none, 2–mild, 3–moderate, 4–severe). In particular, 12, 13 and 14 subjects were assigned to the placebo, 5 mg and 20 mg group, respectively, and then 6, 2, and 2 subjects dropped the study before the 24-hour assessment. Hence the total number of responses over eight occasions from the 39 subjects is 290 (less than $39 \times 8 = 312$). The main interest of this study was to compare the pain scores at eight occasions across different treatment groups. Thus it is reasonable to consider *Pain scores* (P) to be the dependent variable, and *Occasions* (O) and *Treatment Groups* (T) to be the independent variables.

Vandenhende and Lambert (2000) proposed a parametric copula based generalized linear model to analyze the longitudinal ordinal response data in Table 7.4. For the best fit model found in the paper, the marginal distribution of the i-th subject's pain score at the j-th occasion, P_{ij}, was first modeled in terms of the corresponding logarithm of occasion

Pain scores (P)	Occasions (O)							
	0.5h	1h	1.5h	2h	3h	4h	6h	24h
Treatment Groups (T)	Placebo							
1 (none)	0	0	1	1	2	2	3	4
2 (mild)	1	4	5	4	4	4	3	2
3 (moderate)	6	5	2	3	2	3	3	0
4 (severe)	5	3	4	4	2	0	0	0
	5mg							
1 (none)	1	0	1	1	4	5	6	9
2 (mild)	1	5	7	8	6	7	5	2
3 (moderate)	10	8	5	4	3	0	0	0
4 (severe)	2	0	0	0	0	0	0	0
	20mg							
1 (none)	1	2	3	6	4	5	6	9
2 (mild)	3	7	8	6	10	6	5	3
3 (moderate)	7	5	3	2	0	3	1	0
4 (severe)	3	0	0	0	0	0	0	0

Table 7.4: A longitudinal ordered categorical data collected from a double-blind clinical trial involving 39 subjects suffering acute migraine.

O_{ij} and treatment group T_{ij} using a standard cumulative regression model:

$$g[Pr(P_{ij} \leq k)] = \alpha_k + \beta \log(O_{ij}) + \delta_1 I(T_{ij} = 5 \text{ mg}) + \delta_2 I(T_{ij} = 20 \text{ mg}),$$

where g is the complementary log-log link function, I is the indicator function, $(\alpha_k, \beta, \delta_1, \delta_2)$ are the parameters, $i = 1, \ldots, 39$, $j = 1, \ldots, n_i \leq 8$ and $k = 1, 2, 3$. Note that the best fit model assumes that the parameters are identical for different subjects and occasions, and the values of n_i are less than 8 for those subjects that early dropped the study. Then the autoregressive structure was applied to the i-th patient's pain scores at different occasions via a first-order Markov model with the bivariate single-parameter Frank copula C_θ:

$$Pr(P_{i1} = k_{i1}, \ldots, P_{in_i} = k_{in_i}) = Pr(P_{i1} = k_{i1}) \prod_{j=2}^{n_i} Pr(P_{ij} = k_{ij} | P_{i,j-1} = k_{i,j-1}),$$

where $Pr(P_{i1} = k_{i1}) = Pr(P_{i1} \leq k_{i1}) - Pr(P_{i1} \leq k_{i1} - 1)$, $Pr(P_{ij} = k_{ij} | P_{i,j-1} = k_{i,j-1}) = Pr(P_{ij} \leq k_{ij} | P_{i,j-1} = k_{i,j-1}) - Pr(P_{ij} \leq k_{ij} - 1 | P_{i,j-1} = k_{i,j-1})$ and $Pr(P_{ij} \leq k_{ij} | P_{i,j-1} = k_{i,j-1})$ is given by

$$\frac{C_\theta[Pr(P_{ij} \leq k_{ij}), Pr(P_{i,j-1} \leq k_{i,j-1})] - C_\theta[Pr(P_{ij} \leq k_{ij}), Pr(P_{i,j-1} \leq k_{i,j-1} - 1)]}{Pr(P_{i,j-1} \leq k_{i,j-1}) - Pr(P_{i,j-1} \leq k_{i,j-1} - 1)}.$$

Since the goal of analyzing this data using the proposed methods is to describe the association structure between the Pain scores and (Treatment group, Occasions) in a model-free manner, we assess the explanatory power of *T* and *O* on *P* in Table 7.5 by computing the overall association measure $\rho^2_{(O,T\to P)}$, and the marginal, conditional, interactive and correlative association measures using the sequential and non-sequential decompositions of $\rho^2_{(O,T\to P)}$, along with their 95% asymptotic and bootstrap confidence intervals. We also carry out the permutation tests for the overall, marginal and conditional association measures and provide the corresponding estimated p-values and their relative errors in Table 7.6. For the multiple hypotheses testing regarding the marginal and conditional association measures, we compute the adjusted p-values based on the Benjamini–Yekutieli procedure (Benjamini and Yekutieli, 2001).

The point estimates for $\rho^2_{(O\to P)}$, $\rho^2_{(T\to P)}$ and $\rho^2_{(O,T\to P)}$ show that the average proportions of variance for the subcopula scores of *P* explained by the subcopula regressions on *O*, *T* and (*O*, *T*) are 32.1%, 6.2% and 39.1%, respectively. We also see that the joint contribution by (O,T) is larger than the sum of their marginal contribution. Note that the asymptotic and bootstrap confidence intervals for $\rho^2_{(O\to P)}$ and $\rho^2_{(T\to P)}$ do not overlap, which indicates the marginal contribution by *O* can be significantly larger than that by *T*.

From the sequential decomposition of the overall association measure $\rho^2_{(O,T\to P)}$ that starts with *O*, $\rho^2_{(O\to P)}$ and $\rho^2_{(T\to P|O)}$, we can see that the percentages of the contribution to $\rho^2_{(O,T\to P)}$ by *O* and then *T* are 82.1% ($= 0.321/0.391 * 100\%$) and 17.9% ($= 0.070/0.391 * 100\%$), respectively. The point estimates for $\rho^2_{(T\to P)}$ and $\rho^2_{(O\to P|T)}$ in the sequential decomposition of the overall association measure that starts with *T* show that the percentages of the contribution to $\rho^2_{(O,T\to P)}$ by *T* and then *O* are 15.8% ($= 0.062/0.391*100\%$) and 84.2% ($= 0.329/0.391 * 100\%$), respectively.

The contribution of both *T* and *O* to the overall association measure appears to be statistically significant according to the permutation tests. At the significance level of 0.05, we reject the null hypothesis $H_0 : \rho^2_{(T,O\to P)} = 0$ based on the estimated p-value in Table 7.6

and proceed with the testing for the two sets of multiple hypotheses: 1) $H_0^{11}: \rho^2_{(T\to P)} = 0$, $H_0^{12}: \rho^2_{(O\to P|T)} = 0$ and 2) $H_0^{21}: \rho^2_{(O\to P)} = 0$, $H_0^{22}: \rho^2_{(T\to P|O)} = 0$. It turns out that both the null hypotheses are rejected at the significance level of 0.05 based on the corresponding adjusted estimated p-values.

Regarding the interactive and correlative association measures in the non-sequential decomposition of the overall association measure, we can see that the contribution by the interaction between *O* and *T* and the (unnormalized) correlation between *O* and *T* (i.e. $\rho^2_{(OT)\to P} = 0.024$ and $\gamma_{(O\to P,T\to P)} = 0.008$) to the overall association measure on *P* are relatively smaller than the marginal contribution by *O* or *T* alone. Note that the contribution by the interaction between *O* and *T* is more than double the contribution by the (unnormalized) correlation between *O* and *T*. In addition, according to Eq. (4.5), both O and T appear to provide beneficial information in estimating the mean subcopula score of P: $\rho^2_{(O\to P)} + \rho^2_{(T\to P)} - 2\gamma_{(O\to P,T\to P)} = 0.321 + 0.062 - 2 * 0.008 = 0.367 > 0$.

To further investigate the dependence structure between the dependent variable and two independent variables, the prediction of the category of *P* for each combination of the categories of *O* and *T* is presented in Table 7.7 and Figure 7.1. Note that *T* enters the subcopula regression function first and *O* does next in Figure 7.1. Overall it appears that the level of pain scores decreases from “severe” to “moderate” to “mild” as the time progresses, and such a decreasing pattern in the level of pain scores depends on the dosage used in a treatment group. This suggests that there exists a meaningful interaction between *P* and (*O*, *T*). For the “placebo” group, the pain scores are “severe” in the first 0.5 hour, stay at “moderate” for another 2.5 hours, and decrease to “mild” afterwards. For the “5mg” group, the pain scores stay at “moderate” in the first 2 hours and decrease to “mild” afterwards. For the “20mg” group, the pain scores stay at “moderate” in the first 1 hour and decrease to “mild” afterwards. This indicates that the investigational drug with a higher dosage may be alleviate the acute migraine pain.

We find that there was a large extent of agreement between the results presented

Explanatory Variable	Association Measure	Point Estimate	95% Asymptotic C.I.	95% Bootstrap C.I.
Occasions	$\rho^2_{(O\to P)}$	0.321	(0.229, 0.413)	(0.222, 0.389)
Treatment Groups	$\rho^2_{(T\to P)}$	0.062	(0.006, 0.117)	(0.020, 0.108)
$\mathit{Treatment\ Groups} \mid \mathit{Occasions}$	$\rho^2_{(T\to P\mid O)}$	0.070	–	(0.035, 0.091)
$\mathit{Occasions} \mid \mathit{Treatment\ Groups}$	$\rho^2_{(O\to P\mid T)}$	0.329	–	(0.244, 0.365)
$\mathit{Occasions} * \mathit{Treatment\ grooups} \mid \mathit{Occasions}, \mathit{Treatment\ Groups}$	$\rho^2_{(OT\to P\mid O,T)}$	0.024	–	(0.014, 0.026)
$(\mathit{Occasions}, \mathit{Treatment\ Groups})$	$\gamma_{(O\to P,T\to P)}$	0.008	–	(0.005, 0.011)
$\mathit{Occasions}, \mathit{Treatment\ Groups}$	$\rho^2_{(O,T\to P)}$	0.391	(0.298, 0.483)	(0.313, 0.422)

Table 7.5: Subcopula regression based association measures for all possible subsets of the independent variables, sequential and non-sequential decomposition of the overall association measure in the acute migraine data set

H_0	Estimate	P-value	Relative error
$\rho^2_{(T,O\to P)} = 0$	0.391	0.0009	0.0333
$\rho^2_{(T\to P)} = 0$	0.062	0.0003	0.0936
$\rho^2_{(O\to P\mid T)} = 0$	0.329	0.0015	0.0316
$\rho^2_{(O\to P)} = 0$	0.321	0.0030	0.0316
$\rho^2_{(T\to P\mid O)} = 0$	0.070	0.0344	0.0065

Table 7.6: Results of the permutation tests for the overall, marginal and conditional association measures in the acute migraine data set. The estimated p-values for $H_0^{11} : \rho^2_{(T\to P)} = 0$ and $H_0^{12} : \rho^2_{(O\to P\mid T)} = 0$, are adjusted using the Benjamini–Yekutieli procedure. So are The estimated p-values for $H_0^{21} : \rho^2_{(O\to P)} = 0$ and $H_0^{22} : \rho^2_{(T\to P\mid O)} = 0$, are also adjusted using the Benjamini–Yekutieli procedure.

above and those from the parametric copula based generalized linear model (Vandenhende and Lambert, 2000) reviewed above. That is, the investigational drug, LY334370, is effective in relieving the acute migraine pain at the dosage of 5 or 20 mg, compared to the placebo, and the dosage of 5 and 20 mg show a similar effect relieving the pain at any occasion.

Explanatory Variables		Predicted Category	Bootstrap Proportion			
Occasions	Treatment Groups	Pain scores	1–none	2–mild	3–moderate	4–severe
0.5h	placebo	4–severe	0.0	0.0	31.6	68.4
0.5h	5mg	3–moderate	0.0	0.0	86.6	13.4
0.5h	20mg	3–moderate	0.0	0.5	95.9	3.6
1h	placebo	3–moderate	0.0	0.0	96.0	4.0
1h	5mg	3–moderate	0.0	0.0	100.0	0.0
1h	20mg	3–moderate	0.0	24.9	75.1	0.0
1.5h	placebo	3–moderate	0.0	3.9	94.5	1.6
1.5h	5mg	3–moderate	0.0	9.3	90.7	0.0
1.5h	20mg	2–mild	0.0	66.1	33.9	0.0
2h	placebo	3–moderate	0.0	2.7	94.2	3.1
2h	5mg	3–moderate	0.0	15.0	85.0	0.0
2h	20mg	2–mild	0.0	98.0	2.0	0.0
3h	placebo	3–moderate	0.1	29.4	70.2	0.4
3h	5mg	2–mild	0.0	80.0	20.0	0.0
3h	20mg	2–mild	0.0	100.0	0.0	0.0
4h	placebo	2–mild	0.0	50.3	49.7	0.0
4h	5mg	2–mild	0.0	100.0	0.0	0.0
4h	20mg	2–mild	0.0	91.4	8.6	0.0
6h	placebo	2–mild	0.0	65.7	34.3	0.0
6h	5mg	2–mild	0.1	99.9	0.0	0.0
6h	20mg	2–mild	0.0	99.7	0.3	0.0
24h	placebo	2–mild	9.3	90.7	0.0	0.0
24h	5mg	2–mild	12.2	87.8	0.0	0.0
24h	20mg	2–mild	3.1	96.9	0.0	0.0

Table 7.7: Predicted category of *Pain scores* for each combination of the categories of *Occasions* and *Treatment Groups* with its uncertainty estimated by the bootstrap method

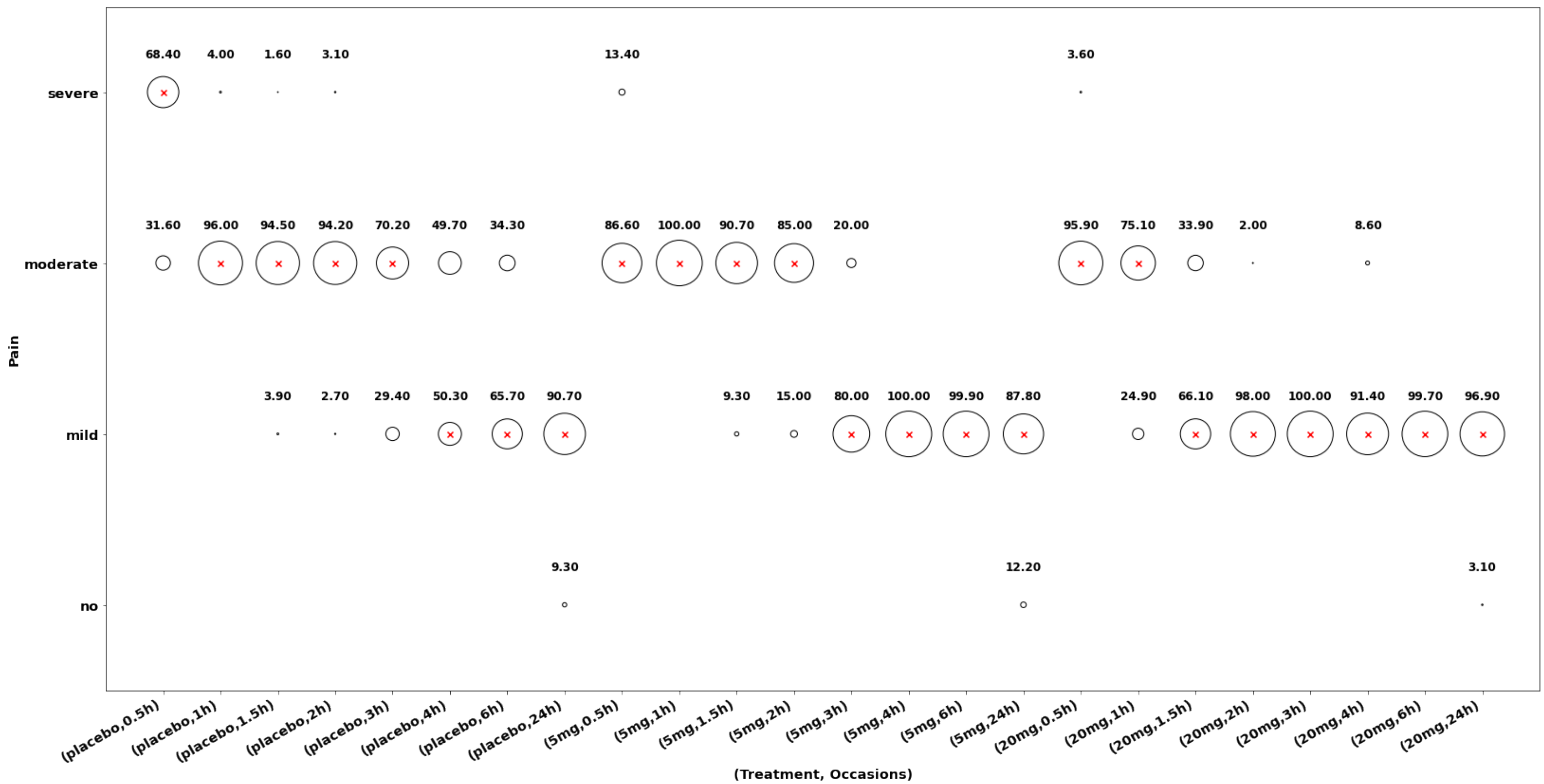


Figure 7.1: Bubble plot of the predicted category of the dependent variable *Pain* for each combination of the categories of the independent variables *Treatment group* and *Occasions* with its uncertainty estimated by the bootstrap method. The radius of each circle is proportional to the number of times that a category of *Pain scores* is predicted for a given combination of the categories of *Treatment group* and *Occasions* over 1000 bootstrap samples.

7.3 Nuclear accident data

Table 7.8 is the five-dimensional data from a longitudinal study containing the interview responses of the stress levels of mothers living around the Three Mile Island nuclear power plant, after an accident occurred in Spring 1979. The four waves of interviews carried out in Winter 1970, Spring 1980, Autumn 1980 and Autumn 1982 are represented by the four ordinal variables *WAVE1* (*W1*), *WAVE2* (*W2*), *WAVE3* (*W3*) and *WAVE4* (*W4*). For each wave of interview, the stress levels of mothers of young children living within ten miles of the Three Mile Island nuclear power plant are recorded as 1–L (Low), 2–M (Medium) and 3–H (High). In addition, the mothers are classified by distance from the nuclear plant represented by the ordinal variable *DISTANCE* (*D*) with two levels 1–"< 5 miles" and 2–"> 5 miles": 115 mothers lived within 5 miles of the plant and the other 152 mothers lived 6 or more miles away from the plant. The total number of responses from the four waves of interviews were 267. The main interest of this study was to see how the level of stress changed over time and differed from the distance to the nuclear plant. In this case, we consider *W4* to be the dependent variable and *W1*, *W2*, *W3*, *D* to be the independent variables.

Fienberg et al. (1985) proposed three sets of hierarchical additive log continuation parametric ratio models of *D*, *W1*, *W2* and *W3* on *W4*:

1. Set 1:

$$\log\left(\frac{m_{ijk1d}}{m_{ijk2d}+m_{ijk3d}}\right) = w + w_{1(i)} + w_{2(j)} + w_{3(k)} + w_{5(d)}, \tag{7.1}$$

$$\log\left(\frac{m_{ijk2d}}{m_{ijk3d}}\right) = w + w_{1(i)} + w_{2(j)} + w_{3(k)} + w_{5(d)}. \tag{7.2}$$

2. Set 2:

$$\log\left(\frac{m_{ijk1d}}{m_{ijk2d}+m_{ijk3d}}\right) = w + w_{1(i)} + w_{2(j)} + w_{3(k)}, \tag{7.3}$$

			< 5 miles			> 5 miles		
Wave			Wave 4					
1	2	3	L	M	H	L	M	H
L	L	L	2	0	0	1	2	0
L	L	M	2	3	0	2	0	0
L	L	H	0	0	0	0	0	0
L	M	L	0	1	0	1	0	0
L	M	M	2	4	0	0	3	0
L	M	H	0	0	0	0	0	0
L	H	L	0	0	0	0	0	0
L	H	M	0	0	0	0	0	0
L	H	H	0	0	0	0	0	0
M	L	L	5	1	0	4	4	0
M	L	M	1	4	0	5	15	1
M	L	H	0	0	0	0	0	0
M	M	L	3	2	0	2	2	0
M	M	M	2	38	4	6	53	6
M	M	H	0	2	3	0	5	1
M	H	L	0	0	0	0	0	0
M	H	M	0	2	0	0	1	1
M	H	H	0	1	1	0	2	1
H	L	L	0	0	0	0	0	1
H	L	M	0	0	0	0	0	0
H	L	H	0	0	0	0	0	0
H	M	L	0	0	0	0	0	0
H	M	M	0	4	4	1	13	0
H	M	H	0	1	4	0	0	0
H	H	L	0	0	0	0	0	0
H	H	M	1	2	0	1	7	2
H	H	H	0	5	12	0	2	7

Table 7.8: A longitudinal study on four waves of interviews about the stress levels of mothers of young children living within ten miles of the Three Mile Island nuclear power plant

$$\log\left(\frac{m_{ijk2d}}{m_{ijk3d}}\right) = w + w_{1(i)} + w_{2(j)} + w_{3(k)}, \tag{7.4}$$

3. Set 3:

$$\log\left(\frac{m_{ijk1d}}{m_{ijk2d} + m_{ijk3d}}\right) = w + w_{3(k)}, \tag{7.5}$$

$$\log\left(\frac{m_{ijk2d}}{m_{ijk3d}}\right) = w + w_{3(k)}, \tag{7.6}$$

where m_{ijkhd} is the expected count for the combination of the i-th, j-th, k-th, h-th and d-th category of *W1*, *W2*, *W3*, *W4* and *D*, and w, $w_{1(i)}$, $w_{2(j)}$, $w_{3(k)}$ and $w_{5(d)}$ are the overall, *W1*, *W2*, *W3* and *D* effect, respectively.

Fienberg et al. (1985) investigated the predictive power of *D* on *W4* via the difference between the likelihood ratio statistics of the models in Eq. (7.1) and Eq. (7.3) as well as that of the models in Eq. (7.2) and Eq. (7.4). It reported that the predictive power of *D* is negligible, due to the insignificant difference between the likelihood ratio statistics of the models in Eq. (7.1) and Eq. (7.3) as well as that of the models in Eq. (7.2) and Eq. (7.4). Moreover, Fienberg et al. (1985) also suggested that there is some residual effects of psychological symptomatology from *W1* and *W2* on *W4* even after it is adjusted for *W3*, according to the significant difference between the likelihood ratio statistics of the models in Eq. (7.3) and Eq. (7.5) as well as that of the models in Eq. (7.4) and Eq. (7.6).

The purpose of applying the proposed methods to the data in Table 7.8 is to illustrate how the proposed methods capture information on time-dependent association structure between the stress level of *W4* and those from previous waves and distance in a model-free manner.

Explanatory Variable	Association Measure	Point Estimate	95% Asymptotic C.I.	95% Bootstrap C.I.
DISTANCE	$\rho^2_{(\mathrm{D}\to\mathrm{W4})}$	0.001	(-0.007, 0.009)	(0.000, 0.012)
WAVE1	$\rho^2_{(\mathrm{W1}\to\mathrm{W4})}$	0.115	(0.032, 0.199)	(0.062, 0.177)
WAVE2	$\rho^2_{(\mathrm{W2}\to\mathrm{W4})}$	0.166	(0.064, 0.268)	(0.093, 0.268)
WAVE3	$\rho^2_{(\mathrm{W3}\to\mathrm{W4})}$	0.246	(0.127, 0.365)	(0.155, 0.343)
WAVE3|*DISTANCE*	$\rho^2_{(\mathrm{W3}\to\mathrm{W4}\mid\mathrm{D})}$	0.259	–	(0.159, 0.349)
WAVE2|*DISTANCE*, *WAVE3*	$\rho^2_{(\mathrm{W2}\to\mathrm{W4}\mid\mathrm{D,W3})}$	0.058	–	(0.018, 0.103)
WAVE1|*DISTANCE*, *WAVE3*, *WAVE2*	$\rho^2_{(\mathrm{W1}\to\mathrm{W4}\mid\mathrm{D,W3,W2})}$	0.080	–	(0.050, 0.096)
WAVE2|*WAVE3*	$\rho^2_{(\mathrm{W2}\to\mathrm{W4}\mid\mathrm{W3})}$	0.040	–	(0.005, 0.093)
WAVE1|*WAVE3*, *WAVE2*	$\rho^2_{(\mathrm{W1}\to\mathrm{W4}\mid\mathrm{W3,W2})}$	0.047	–	(0.024, 0.075)
DISTANCE|*WAVE3*, *WAVE2*, *WAVE1*	$\rho^2_{(\mathrm{D}\to\mathrm{W4}\mid\mathrm{W3,W2,W1})}$	0.065	–	(0.032, 0.084)
DISTANCE, *WAVE1*, *WAVE2*|*WAVE3*	$\rho^2_{(\mathrm{D,W1,W2}\to\mathrm{W4}\mid\mathrm{W3})}$	0.152	–	(0.116, 0.156)
WAVE1, *WAVE2*|*WAVE3*	$\rho^2_{(\mathrm{W1,W2}\to\mathrm{W4}\mid\mathrm{W3})}$	0.087	–	(0.042, 0.118)
DISTANCE, *WAVE1*, *WAVE2*, *WAVE3*	$\rho^2_{(\mathrm{D,W1,W2,W3}\to\mathrm{W4})}$	0.398	(0.277, 0.518)	(0.322, 0.431)

Table 7.9: Subcopula regression based marginal and overall association measures, and the selected sequential decomposition of the overall association measure for the nuclear accident data set

H_0	Estimate	P-value	Relative error
$\rho^2_{(D,W1,W2,W3\to W4)}=0$	0.398	0.0010	0.0316
$\rho^2_{(D\to W4)}=0$	0.001	1.000	0.0008
$\rho^2_{(W3\to W4\mid D)}=0$	0.259	0.0083	0.0316
$\rho^2_{(W2\to W4\mid D,W3)}=0$	0.058	0.1263	0.0056
$\rho^2_{(W1\to W4\mid D,W3,W2)}=0$	0.079	0.3308	0.0027
$\rho^2_{(W3\to W4)}=0$	0.246	0.0083	0.0316
$\rho^2_{(W2\to W4\mid W3)}=0$	0.0392	0.1171	0.0059
$\rho^2_{(W1\to W4\mid W3,W2)}=0$	0.047	0.4483	0.0023
$\rho^2_{(D\to W4\mid W3,W2,W1)}=0$	0.065	0.6475	0.0015
$\rho^2_{(W2,W1,D\to W4\mid W3)}=0$	0.152	0.0153	0.0080
$\rho^2_{(W2,W1\to W4\mid W3)}=0$	0.087	0.0215	0.0068

Table 7.10: Results of the permutation tests for the overall, marginal and conditional association measures in the nuclear accident data set. The estimated p-values for $H_0^{11}:\rho^2_{(D\to W4)}=0$, $H_0^{12}:\rho^2_{(W3\to W4\mid D)}=0$, $H_0^{13}:\rho^2_{(W2\to W4\mid D,W3)}=0$ and $H_0^{14}:\rho^2_{(W1\to W4\mid D,W3,W2)}=0$ are adjusted using the Benjamini–Yekutieli procedure. So are the estimated p-values for $H_0^{21}:\rho^2_{(W3\to W4)}=0$, $H_0^{22}:\rho^2_{(W2\to W4\mid W3)}=0$, $H_0^{23}:\rho^2_{(W1\to W4\mid W3,W2)}=0$ and $H_0^{24}:\rho^2_{(D\to W4\mid W3,W2,W1)}=0$.

First, we evaluate the explanatory power of *W1*, *W2*, *W3* and *D* on *W4* by computing the overall association measure $\rho^2_{(W3,W2,W1,D\to W4)}$, the marginal association measures for each independent variable and the conditional association measures in the sequential decomposition of the overall association measure using two orders of the independent variables: $\rho^2_{(W3\to W4|D)}$, $\rho^2_{(W2\to W4|D,W3)}$ and $\rho^2_{(W1\to W4|D,W3,W2)}$ for $(D, W3, W2, W1)$ and $\rho^2_{(W2\to W4|W3)}$, $\rho^2_{(W1\to W4|W3,W2)}$ and $\rho^2_{(D\to W4|W3,W2,W1)}$ for $(W3, W2, W1, D)$. Table 7.9 contains the point estimates for the aforementioned association measures, followed by their 95% asymptotic and bootstrap confidence intervals. Additionally, Table 7.10 contains the results for the permutation tests for the corresponding overall, marginal and conditional association measures in the nuclear accident data set. For the multiple hypotheses testing concerning the marginal and conditional association measures, we present the adjusted p-values based on the Benjamini–Yekutieli procedure (Benjamini and Yekutieli, 2001).

Regarding the point estimates for $\rho^2_{(D\to W4)}$, $\rho^2_{(W1\to W4)}$, $\rho^2_{(W2\to W4)}$, $\rho^2_{(W3\to W4)}$ and $\rho^2_{(D,W1,W2,W3\to W4)}$, we can see that the average proportions of variance for the subcopula scores of *W4* explained by the subcopula regressions of *D*, *W1*, *W2*, *W3* and (*D*, W1, *W2*, W3) are 0.1%, 11.5%, 16.6%, 24.6% and 39.8%. Note that the corresponding asymptotic and bootstrap confidence intervals appear to be wide, which may be resulted from the small sample size (total cell counts) relative to the size of the contingency table (depending on the number of variables and the number of categories in each variable).

The estimates of $\rho^2_{(D\to W4)}$, $\rho^2_{(W3\to W4|D)}$, $\rho^2_{(W2\to W4|D,W3)}$ and $\rho^2_{(W1\to W4|D,W3,W2)}$ obtained from the sequential decomposition based on the order of $(D, W3, W2, W1)$ show that the percentages of the contribution by *D*, *W3*, *W2* and *W1* to the overall association measure are 0.2% ($= 0.001/0.398 * 100\%$), 65.1% ($= 0.259/0.398 * 100\%$), 14.6% ($= 0.058/0.398 * 100\%$) and 20.1% ($= 0.080/0.398 * 100\%$), respectively. On the other hand, the point estimates for $\rho^2_{(W3\to W4)}$, $\rho^2_{(W2\to W4|W3)}$, $\rho^2_{(W1\to W4|W3,W2)}$ and $\rho^2_{(D\to W4|W3,W2,W1)}$ obtained from the sequential decomposition of the overall association measure based on the order of $(W3, W2, W1, D)$ indicate that the percentages of the contribution by *W3*, *W2*, *W1*

and *D* to the overall association measure are 61.8% ($= 0.246/0.398 * 100\%$), 10.1% ($= 0.040/0.398 * 100\%$), 11.8% ($= 0.047/0.398 * 100\%$) and 16.3% ($= 0.065/0.398 * 100\%$), respectively.

According to the sequential decomposition of the overall association measure, we remark the following observations. First, it appears that the time-dependent relationship among *W1*, *W2* and *W3* may affect the contribution of *W1* and *W2* to the association with *W4* because the contribution of *W2* and *W1* decrease when *W3* and $(W3, W2)$ are taken into account, respectively: $\rho^2_{(W2 \to W4)} = 0.166 > \rho^2_{(W2 \to W4|W3)} = 0.040$ and $\rho^2_{(W1 \to W4)} = 0.115 > \rho^2_{(W1 \to W4|W3,W2)} = 0.047$. Secondly, the conditional association measures of *D* on *W4* given *W3*, *W2* and *W1* (6.5%) is larger than the marginal association measures of *D* on *W4* (i.e. 0.1%). This indicates that the stress levels at the fourth wave of interview may differ with the distance from the nuclear plant, when the stress levels for the previous three waves of interviews are taken into account. Thirdly, the conditional association measures of *W3*, *W2* and *W1* on *W4* tend to slightly increase when *D* is taken into consideration: $\rho^2_{(W3 \to W4|D)} = 0.259 > \rho^2_{(W3 \to W4)} = 0.246$, $\rho^2_{(W2 \to W4|D,W3)} = 0.058 > \rho^2_{(W2 \to W4|W3)} = 0.040$ and $\rho^2_{(W1 \to W4|D,W3,W2)} = 0.080 > \rho^2_{(W1 \to W4|W3,W2)} = 0.047$. We argue that such a pattern is attributed to the potential association of *W3*, *W2* and *W1* with *D*.

The notable contribution of *W3* to the overall association measure can also be confirmed by the permutations tests. At the significance level of 0.05, we reject the null hypothesis $H_0 : \rho^2_{(D,W1,W2,W3 \to W4)} = 0$ according to the estimated p-value in Table 7.10 and proceed with the testing for two sets of hypotheses:

1. $H_0^{11} : \rho^2_{(D \to W4)} = 0$, $H_0^{12} : \rho^2_{(W3 \to W4|D)} = 0$, $H_0^{13} : \rho^2_{(W2 \to W4|D,W3)} = 0$, $H_0^{14} : \rho^2_{(W1 \to W4|D,W3,W2)} = 0$.
2. $H_0^{21} : \rho^2_{(W3 \to W4)} = 0$, $H_0^{22} : \rho^2_{(W2 \to W4|W3)} = 0$, $H_0^{23} : \rho^2_{(W1 \to W4|W3,W2)} = 0$, $H_0^{24} : \rho^2_{(D \to W4|W3,W2,W1)} = 0$.

It appears that the two null hypotheses related to *W3*, $H_0^{12} : \rho^2_{(W3 \to W4|D)} = 0$ and H_0^{21} :

$\rho^2_{(W3 \to W4)} = 0$, are rejected at the significance level of 0.05 based on the adjusted estimated p-values. In addition, the other two null hypotheses $H_0 : \rho^2_{(W2,W1,D \to W4|W3)} = 0$ and $H_0 : \rho^2_{(W2,W1 \to W4|W3)} = 0$ are also rejected at the significance level of 0.05.

To further deepen the understanding of the dependence structure between the dependent variable *W4* and the set of independent variables (*D*, *W1*, *W2*, *W3*), we present in Table 7.11 the prediction of the category of *W4* for each combination of the categories of *W3*, *W2*, *W1* and *D* with its uncertainty estimated by the bootstrap method, and we visualize them in Figure 7.2 using the double decker plot. We first observe that the predicted stress level of *W4* appears to be associated with that of *W3*. That is, when the stress level of *W3* is medium and high, the predicted level of *W4* also tends to be medium and high, respectively. We also find that when the stress level of *W3* is high, the predicted category of *W4* is more likely to be high if the levels of *W2* and *W1* happen to be medium or high, regardless of the level of *D*. Note that, when the levels of *W3* is medium, the stress level of *W4* is dominantly predicted to be high for $(W2, W1, D) =$ (medium, high, $<$ 5 miles) and (high, medium, $>$ 5 miles), unlike the other combinations of categories of $(W2, W1, D)$.

According to the results by the proposed model-free methods, we obtain the following information which may be useful for the subsequent explanatory or predictive modeling. First, the stress level of *W3* has the dominant effect on the level of *W4*, which seems to be a Markov chain like pattern. Secondly, *D* has marginally negligible explanatory power relative to *W3*, *W2* and *W1*. Lastly, there appears to be some psychological residual effects from *D*, *W1* and *W2* (*W1* and *W2*) on the stress level of *W4* after taking into account the level of *W3*, which can be confirmed by the conditional measures of *D*, *W1* and *W2* (*W1* and *W2*) on *W4* given *W3*, $\rho^2_{(D,W1,W2 \to W4|W3)} = 0.152$ ($\rho^2_{(W1,W2 \to W4|W3)} = 0.087$) and the statistical significance found in the permutation test for $\rho^2_{(D,W1,W2 \to W4|W3)}$ and $\rho^2_{(W1,W2 \to W4|W3)}$. Note that these findings are consistent with those obtained when the set of the additive log continuation ratio models (Fienberg et al., 1985) briefly described above was used.

Explanatory Variables				Predicted Category	Bootstrap Proportion		
WAVE3	WAVE2	WAVE1	DISTANCE (miles)	WAVE4	1–L	2–M	3–H
L	L	L	< 5	1–L	100	0.0	0.0
L	L	L	> 5	2–M	3.3	96.7	0.0
L	L	M	< 5	2–M	33.5	66.5	0.0
L	L	M	> 5	2–M	0.1	99.9	0.
L	L	H	< 5	1–L	100.0	0.0	0.0
L	L	H	> 5	3–H	0.0	0.0	100.0
L	M	L	< 5	2–M	0.0	100.0	0.0
L	M	L	> 5	1–L	100.0	0.0	0.0
L	M	M	< 5	2–M	7.8	92.2	0.0
L	M	M	> 5	2–M	6.5	93.5	0.0
L	M	H	< 5	1–L	100.0	0.0	0.0
L	M	H	> 5	1–L	100.0	0.0	0.0
L	H	L	< 5	1–L	100.0	0.0	0.0
L	H	L	> 5	1–L	100.0	0.0	0.0
L	H	M	< 5	1–L	100.0	0.0	0.0
L	H	M	> 5	1–L	100.0	0.0	0.0
L	H	H	< 5	1–L	100.0	0.0	0.0
L	H	H	> 5	1–L	100.0	0.0	0.0
M	L	L	< 5	2–M	0.8	99.2	0.0
M	L	L	> 5	1–L	100.0	0.0	0.0
M	L	M	< 5	2–M	0.1	99.9	0.0
M	L	M	> 5	2–M	0.0	99.8	0.2
M	L	H	< 5	1–L	100.0	0.0	0.0
M	L	H	> 5	1–L	100.0	0.0	0.0
M	M	L	< 5	2–M	0.2	99.8	0.0
M	M	L	> 5	2–M	0.0	100.0	0.0
M	M	M	< 5	2–M	0.0	70.5	29.5
M	M	M	> 5	2–M	0.0	95.8	4.2
M	M	H	< 5	3–H	0.0	0.4	99.6
M	M	H	> 5	2–M	0.0	100.0	0.0
M	H	L	< 5	1–L	100.0	0.0	0.0
M	H	L	> 5	1–L	100.0	0.0	0.0
M	H	M	< 5	2–M	0.0	100.0	0.0
M	H	M	> 5	3–H	0.0	22.3	77.7
M	H	H	< 5	2–M	4.4	95.6	0.0
M	H	H	> 5	2–M	0.0	63.6	36.4
H	L	L	< 5	1–L	100.0	0.0	0.0
H	L	L	> 5	1–L	100.0	0.0	0.0
H	L	M	< 5	1–L	100.0	0.0	0.0
H	L	M	> 5	1–L	100.0	0.0	0.0
H	L	H	< 5	1–L	100.0	0.0	0.0
H	L	H	> 5	1–L	100.0	0.0	0.0
H	M	L	< 5	1–L	100.0	0.0	0.0
H	M	L	> 5	1–L	100.0	0.0	0.0
H	M	M	< 5	3–H	0.0	0.4	99.6
H	M	M	> 5	3–H	0.0	33.5	66.5
H	M	H	< 5	3–H	0.0	0.1	99.9
H	M	H	> 5	1–L	100.0	0.0	0.0
H	H	L	< 5	1–L	100.0	0.0	0.0
H	H	L	> 5	1–L	100.0	0.0	0.0
H	H	M	< 5	3–H	0.0	26.1	73.9
H	H	M	> 5	3–H	0.0	30.0	70.0
H	H	H	< 5	3–H	0.0	0.0	100.0
H	H	H	> 5	3–H	0.0	0.0	100.0

Table 7.11: The tabular representation of the predicted category of *WAVE4* for each combination of the categories of *WAVE3*, *WAVE2*, *WAVE1* and *DISTANCE* with its uncertainty estimated by the bootstrap method

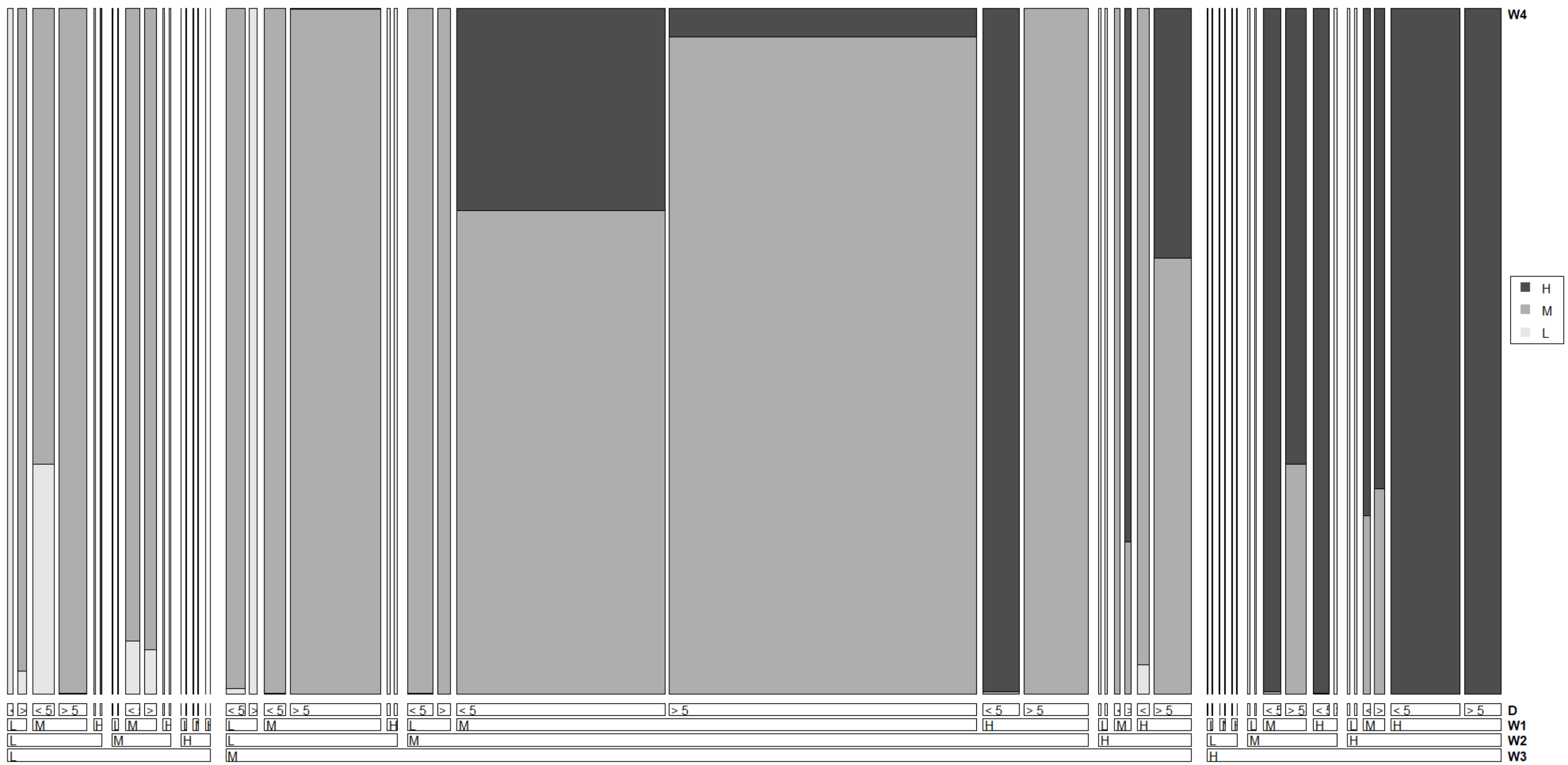


Figure 7.2: The double decker plot of the predicted category of the dependent variable *WAVE4* for each combination of the categories of the independent variables *WAVE3*, *WAVE2*, *WAVE1* and *DISTANCE*. The layers of labels along x-axis indicate the order in which the independent variables are entered into the subcopula regression, i.e *WAVE3* and DISTANCE are the first and last one, respectively. The width of each bar is proportional to the frequency of the corresponding combination of the categories of the independent variables. Finally, the blocks of three levels of gray color in each bar are proportional to the numbers of times that each category of $WAVE4$ is predicted over 1000 bootstrap samples. That is, the longer the height of each grey-colored block in a bar, the less the relative uncertainty in predicting the corresponding category of *WAVE4* given a combination of the categories of *WAVE3*, *WAVE2*, *WAVE1* and *DISTANCE*.

Var	Type	Name	Description	Categories
1	Independent	L-CORE	patient's internal temperature in C°	low (< 36), mid (≥ 36 and ≤ 37), high (> 37)
2	Independent	L-SURF	patient's surface temperature in C°	low (< 35), mid (≥ 35 and ≤ 36.5), high (> 36.5)
3	Independent	L-O2	oxygen saturation in $\%$	poor (< 80), fair (≥ 80 and < 90), good (≥ 90 and < 98), excellent (≥ 98)
4	Independent	L-BP	last measurement of blood pressure	low ($< 90/70$), mid ($\geq 90/70$ and $\leq 130/90$), high ($> 130/90$)
5	Independent	SURF-STBL	stability of patient's surface temperature	unstable, mod-stable, stable
6	Independent	CORE-STBL	stability of patient's core temperature	unstable, mod-stable, stable
7	Independent	BP-STBL	stability of patient's blood pressure	unstable, mod-stable, stable
8	Independent	COMFORT	patient's perceived comfort at discharge	0 (strongly discomfortable) $\vdots$ 20 (strongly comfortable)
9	Dependent	ADM-DECS	discharge decision	I (patient sent to Intensive Care Unit), A (patient sent to general hospital floor), S (patient prepared to go home)

Table 7.12: The nine ordinal variables in the *post-operative patients* data set

7.4 Post-operative patients data

The real data set that we will analyze in this section is the post-operative patients data presented in the study of knowledge acquisition for expert system development in nursing (Budihardjo et al., 1991; Woolery et al., 1991). In this study, 90 patients were examined after surgeries and 8 categorical health status measurement variables were recorded for each patient including the body temperature and blood pressure. The main purpose of Budihardjo et al. (1991) and Woolery et al. (1991) was to determine the next place where a patient in the post-operative recovery area should be sent to based on (his/her) health status, which is regarded as a classification problem in the area of statistical learning. Thus, the variable discharge decision (*ADM-DECS*) is considered to be the response variable that indicates whether a patient was sent to the intensive care unit, general hospital floor or home, while the 8 categorical health status measurement variables are considered to be the

predictors. Table 7.12 provides all the description of the 9 variables in the data.

Budihardjo et al. (1991) and Woolery et al. (1991) applied a rule-based machine learning program called LERS_LB (Learning from Examples using Rough Sets Lower Boundaries) to the classification problem of the response variable, *ADM-DECS*, using all the 8 health status measurement variables in the post-operative patients data. Specifically, for each category of *ADM-DECS*, the learning program LERS_LB will induce all the possible and certain association rules that describe the category based on the rough set theory. Note that a rule is said to be possible (certain) if there exist (do not exist) conflicting observations in the data, where two observations are said to be conflicting if they contain the same values for the predictors but different ones for the response variable. Dash and Dehuri (2013) investigated the training accuracies of five different classification methods on the post-operative patients data: Naïve Bayes, ID3, J48, Fuzzy-Rough nearest neighbors (FRNN) and Fuzzy K-nearest neighbors (FKNN). For each method, 7 health status measurement variables were used and the variable *COMFORT* was excluded due to the sparseness, i.e. the counts for 16 (out of 21) categories are zero. Note that all the variables are treated as nominal for LERS_LB in Budihardjo et al. (1991) and Woolery et al. (1991), as well as for the five methods in Dash and Dehuri (2013).

Our intention of analyzing this data is to demonstrate the utility of the proposed model-free methods on the variable selection problem for a multi-dimensional contingency table, which is regarded as one of the important preprocessing steps in statistical learning. According to Table 7.12, we argue that it is more appropriate to treat all the 9 variables to be ordinal because each of them has semantically ordered categories. In particular, the health status of a patient who was sent to the intensive care unit after surgeries (*I*) should be worse than that of a patient who was sent to the general hospital floor (*A*), while the health status of a patient who was sent to the general hospital floor after surgeries (*A*) should be worse than that of a patient who was sent directly to home (*S*). Therefore, we consider the variable *ADM-DECS* to be ordinal with the ordering $I < A < S$. In addition, for the same

reason in Dash and Dehuri (2013), we choose the 7 health status measurement variables other than *COMFORT* as the initial set of predictors for variable selection.

Our process of variable selection starts with the *all-possible-subset subcopula regressions* by estimating the overall association measures for all-possible subsets of independent variables. Figure 7.13 visualizes the resulting association measures by sorting them from the highest to the lowest within the same number of independent variables involving in the subcopula regressions.

Next, we choose one or more subsets of independent variables with the highest or higher values of the overall association measure within the same number of independent variables, which can be used as a pool of candidate variables for the subsequent step. For example, Table 7.14 shows seven subsets of independent variables where each one has the highest value of the overall association measure for each number of independent variables. Note that the asymptotic confidence interval for each overall association measure in Table 7.14 is much wider than the corresponding bootstrap confidence interval. This may be due to the small sample size (the total cell counts) relative to the size of the contingency table (depending on the number of variables and the number of categories in each variable). Furthermore, the bootstrap confidence intervals for two consecutive overall association measures do not overlap, except for those with 5, 6 and 7 variables involved in the subcopula regression.

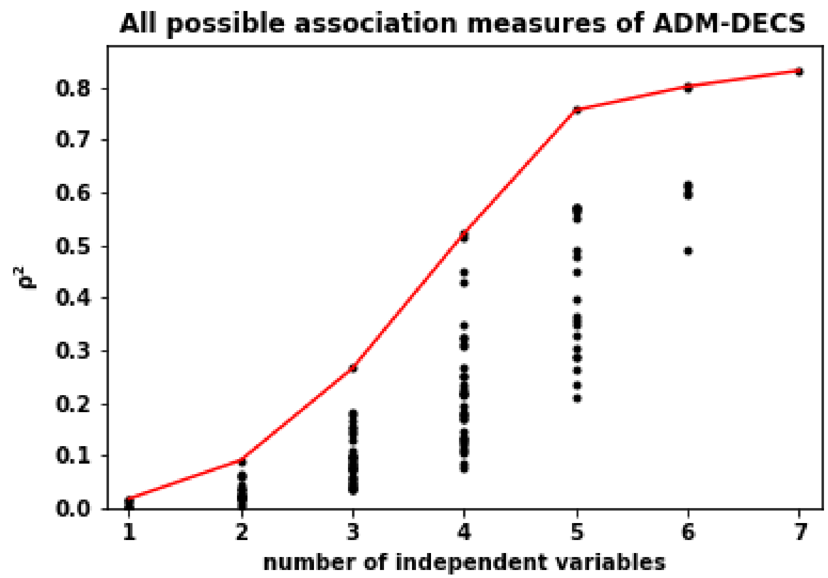


Table 7.13: Overall association measures of all-possible subsets of independent variables on the dependent variable *ADM-DECS*

Number of Vars	L-CORE	L-SURF	L-O2	L-BP	SURF-STBL	CORE-STBL	BP-STBL	ρ^2	95% Asy. CI	95% Boot. CI
7	✗	✗	✗	✗	✗	✗	✗	0.831	(0.680, 0.982)	(0.788, 0.828)
6	✗	✗	✗	✗	✗		✗	0.801	(0.629, 0.973)	(0.763, 0.799)
5	✗	✗	✗	✗			✗	0.756	(0.550, 0.962)	(0.708, 0.784)
4	✗	✗		✗			✗	0.522	(0.246, 0.798)	(0.475, 0.553)
3	✗			✗			✗	0.265	(-0.013, 0.544)	(0.236, 0.292)
2	✗						✗	0.091	(0.001, 0.181)	(0.060, 0.123)
1				✗				0.018	(-0.033, 0.069)	(0.006, 0.050)

Table 7.14: Best subset of independent variables with highest values of the overall association measure at each number of the independent variables involved in the subcopula regression

Then, we may identify a potentially important subset of independent variables from the candidate pool across different number of independent variables involving in the subcopula regressions. As the overall association measure itself cannot be used to determine the best one because of its non-decreasing property, different metrics based on the overall association measure should be adopted. For example, we consider the following metric, $\nabla\rho^2(i-1,i)$:

$$\nabla\rho^2(i-1,i) = \frac{\left[(1-\rho_{i-1}^2)-(1-\rho_i^2)\right]}{1-\rho_{i-1}^2} = \frac{\rho_i^2-\rho_{i-1}^2}{1-\rho_{i-1}^2}, \tag{7.7}$$

where ρ_i^2 is the association measure for the subcopula regression with i independent variables and $i = 2, \ldots, d-1$. This metric measures the proportion of variation explained by a new independent variable added to the existing set of independent variables in the subcopula regression. Note that when the subcopula regression with $(i-1)$ independent variables is nested within the subcopula regression with i independent variables, then $\rho_i^2 - \rho_{i-1}^2 \geq 0$ due to the non-decreasing property and hence $\nabla\rho^2(i-1,i) \geq 0$. Therefore, the potential subset of independent variables from the candidate pool can be the one with the highest value of $\nabla\rho^2(i-1,i)$ and can be considered for subsequent explanatory or predictive modeling.

As shown in Table 7.15, the potential important subset of independent variables identified by the proposed metric in Eq. (7.7) for the subsequent classification task is *L-CORE*, *L-SURF*, *L-O2*, *L-BP* and *BP-STBL*, with the corresponding association measure $\rho^2_{(\textit{L-CORE,L-SURF,L-O2,L-BP,BP-STBL}\to\textit{ADM-DECS})} = 0.75619$ and $\nabla\rho^2(i-1,i) = 0.49003$.

Number of Vars	L-CORE	L-SURF	L-O2	L-BP	SURF-STBL	CORE-STBL	BP-STBL	$\nabla\rho^2(i-1,i)$
7	✗	✗	✗	✗	✗	✗	✗	0.151
6	✗	✗	✗	✗	✗		✗	0.184
5	✗	✗	✗	✗			✗	**0.490**
4	✗	✗		✗			✗	0.349
3	✗			✗			✗	0.192
2	✗						✗	0.074
1				✗				–

Table 7.15: The value of $\nabla\rho^2(i-1,i)$ computed from the seven best subsets of the independent variables given in Table 7.14

For the rest of this section, we will illustrate the performance of the set of independent variables selected by our proposed metric in Eq. (7.7) for the classification methods used in Budihardjo et al. (1991), Woolery et al. (1991) and Dash and Dehuri (2013). We will also compare it with the performance of the sets of independent variables used in these three papers. Specifically, we will be more focused on the comparisons related to Dash and Dehuri (2013) for two reasons. First, the classification performance was not the main interest in Budihardjo et al. (1991) and Woolery et al. (1991), and hence they did not report any performance of LERS_LB. Moreover, the code of LERS_LB is not available either in those papers or in commonly used data analysis softwares and hence we adopt a similar implementation in R called LEM2 (Grzymala-Busse, 1997). In contrast, all the five classification methods considered in Dash and Dehuri (2013) are well-implemented in a data mining software called Weka (Hall et al., 2009) and their classification performances on the post-operative patients data were provided. As Dash and Dehuri (2013) employed the training accuracy to evaluate the classification performance, we will also use it for the comparison below.

Table 7.16 shows the training accuracy obtained from LEM2 using the subset of predictors selected by our proposed metric in Eq. (7.7), and the entire set of predictors as in Budihardjo et al. (1991) and Woolery et al. (1991). We can see that the training accuracy using the set of predictors selected by our proposed metric is about 0.092 lower than the training accuracy using all the predictors. That is, LEM2 predicts about 8 more observations incorrectly using the set of predictors selected by our proposed metric than using all the predictors.

Predictors	LEM2
L-CORE, L-SURF, L-O2, L-BP, BP-STBL	0.839
L-CORE, L-SURF, L-O2, L-BP, SURF-STBL, CORE-STBL, BP-STBL, COMFORT	0.931

Table 7.16: The respective training accuracies of LEM2 using the selected subset and the entire set of predictors

Table 7.17 shows the training accuracy obtained from each of the five methods studied in Dash and Dehuri (2013) using the set of predictors selected by our proposed metric in Eq. (7.7), and the entire set of predictors except for *COMFORT*. For each classification method, we observe that the training accuracies using two different sets of predictors are similar. Note that for FKNN, the training accuracy using the set of predictors selected by our proposed metric slightly increases by about 0.022. These results suggest that the variable selection approach based on our proposed methods may identify a potentially important set of predictors that contains most of the information for classification.

From the analysis of the post-operative patients data, we conclude that the proposed model-free methods have the potential to be useful tools in the variable selection process for a multi-dimensional contingency table with the ordinal dependent variable.

Methods	Predictors	
	L-CORE, L-SURF, L-O2, L-BP, BP-STBL	L-CORE, L-SURF, L-O2, L-BP, SURF-STBL, CORE-STBL, BP-STBL
Naïve Bayes	0.711	0.733
ID3	0.844	0.889
J48	0.711	0.711
FRNN	0.833	0.889
FKNN	0.711	0.689

Table 7.17: The respective training accuracies of Naïve Bayes, ID3, J48, FRNN and FKNN using the selected subset of predictors and the entire set of predictors except for *COMFORT*

CHAPTER 8

DISCUSSION AND FUTURE WORKS

In this dissertation, we proposed the descriptive modeling methodology, consisting of subcopula score, subcopula regression and subcopula regression based association measure, to delineate and quantify the association structure of multi-dimensional contingency tables with ordinal variables in a model-free manner. We first studied the subcopula score and its theoretical properties as a tool to assign data-dependent numerical scores to an ordinal variable. We then formally examined the subcopula regression and its prediction to explore and identify the association pattern between the ordinal dependent variable and a set of categorical independent variables. In order to quantify the contribution of a set of independent categorical variables of interest, we then investigated the theoretical properties of the subcopula regression-based association measure and derived its sequential and non-sequential decompositions, and proposed the marginal, conditional, interactive and correlative association measures. Then we developed in Chapter 5 a variety of statistical inference methods including (point/interval) estimation and hypothesis testing for the proposed methods. Finally, we demonstrated the performance of the proposed association measures in Chapter 6 and 7 via both simulation study and real data analysis.

In this section, we will discuss some future work in the following that can improve the performance of the proposed methods and broaden the horizon of their applications in descriptive modeling and other statistical analyses.

8.1 Large-sample distribution for the overall association measure

We knew from Chapter 5 and 6 that the asymptotic distribution of the estimator $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ in Theorem 5.1 may not hold when the true value of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is zero, a boundary of the parameter space, even if the sample size is sufficiently large. This prompted us to develop the permutation test for the null hypothesis of "no overall association". In order to further enrich the theoretical properties of the proposed estimator, we need to derive the asymptotic distribution of $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ under $H_0 : \rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and develop the asymptotic test which would be a future work direction. Instead, we would like to investigate below the empirical sampling distribution of $\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)}$ under $H_0 : \rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$.

We notice that the proposed association measure $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ and the population version of the coefficient of determination, denoted by R^2, for a multiple linear regression model share similar properties. They both range from 0 to 1 and they describe the proportion of the variance for a dependent variable explained by the independent variables. Moreover, their estimators are upward-biased with respect to the true population values.

Under the null hypothesis that all the regression coefficients are zero except for the intercept in the multiple linear regression model with $(k-1)$ independent variables and normal error distribution, i.e. $R^2 = 0$, $\frac{(n-k)\hat{R}^2}{(k-1)(1-\hat{R}^2)}$ follows a Snedecor's F distribution with the degrees of freedom $(k-1, n-k)$ where $\hat{R}^2$ is the estimator of R^2 and n is the sample size, and $\hat{R}^2$ follows a Beta distribution with shape parameters $(\frac{k-1}{2}, \frac{n-k}{2})$. Therefore,

$$\frac{(n-k)\hat{R}^2}{1-\hat{R}^2} \xrightarrow{d} \chi^2_{k-1}.$$

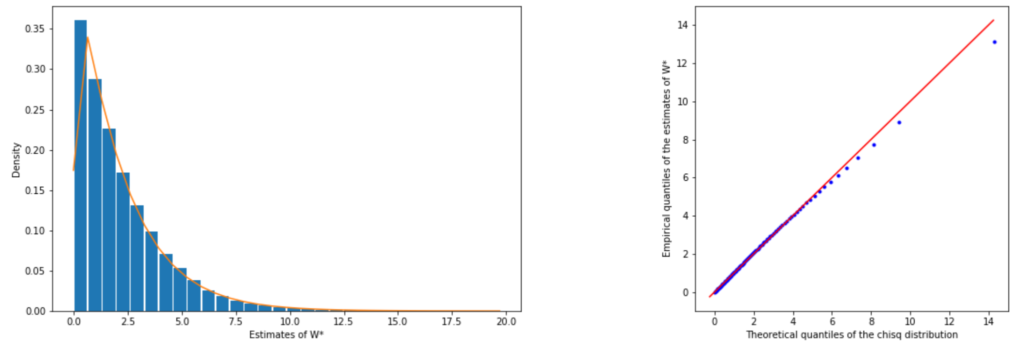
Inspired by these theoretical properties of R^2 and its estimator, we conjecture that the estimator of the proposed association measure may asymptotically follow a chi-square dis-

tribution. That is, under $H_0 : \rho^2_{(\mathbf{X}_{-i} \to X_i)} = 0$,

$$\frac{f(n, \mathbf{M}_d)\hat{\rho}^2_{(\mathbf{X}_{-i} \to X_i)}}{1 - \hat{\rho}^2_{(\mathbf{X}_{-i} \to X_i)}} \xrightarrow{d} \chi^2_{\nu(\mathbf{M}_d)},$$

as n goes to infinity, where $f(n, \mathbf{M}_d)$ and $\nu(\mathbf{M}_d)$ are two functions of the sample size n and/or the number of categories $\mathbf{M}_d$ in the d ordinal variables, and $\nu(\mathbf{M}_d)$ is the degree of freedom. Below we show empirical evidence that may support our conjecture obtained from simulations implemented in the two-dimensional contingency table with the respective ordinal dependent and independent variables X_2 and X_1.

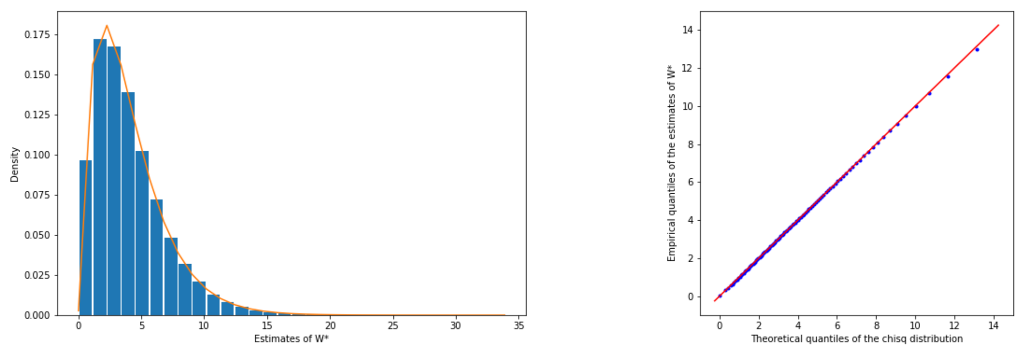
We first fix the sample size $n = 50,000,000$, vary the number of categories in each variable $k = 3, 5, 7$ and estimate the proposed association measure on $100,000$ simulated two-dimensional contingency tables with $corr(X_1, X_2) = 0$ obtained from the simulation method given in Chapter 6. Then we fit a Beta distribution to those estimates for each level of k and obtain the values of two shape parameters, $\hat{\alpha}$ and $\hat{\beta}$. Finally, we let $2\hat{\alpha} = \hat{\nu}(k)$ and $2\hat{\beta} = \hat{f}(n, k)$, fit the chi-square distribution with the degree of freedom $2\hat{\alpha}$ to the estimator $W^* = \frac{2\hat{\beta}\hat{\rho}^2_{(X_1 \to X_2)}}{1-\hat{\rho}^2_{(X_1 \to X_2)}}$ for each level of k and assess the goodness-of-fit by constructing histograms and Q-Q plots. The results below for each level of k show a good fit of the conjectured chi-square distribution to the $100,000$ estimates of the proposed association measure obtained under $H_0 : \rho^2_{(X_1 \to X_2)} = 0$. Our next future work would be to verify our conjecture by finding the suitable expression of $f(n, k)$ and $v(k)$ in the chi-square distribution for two-dimensional contingency table and extend it to a case of the multi-dimensional contingency table.



(a) The histogram of W^*

(b) The Q-Q plot of W^*

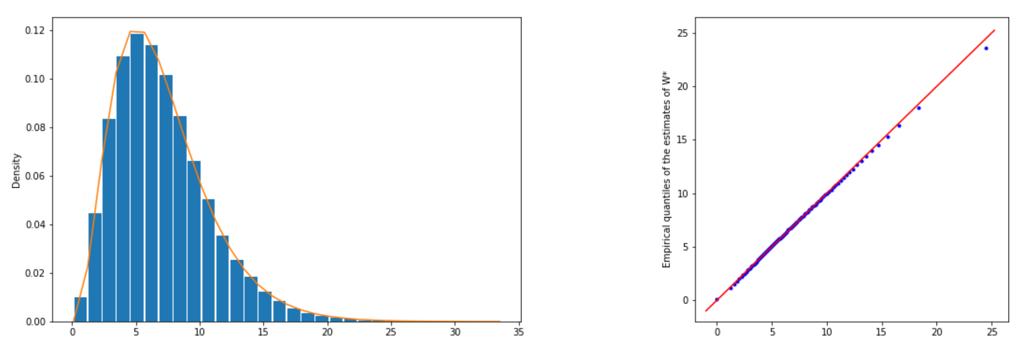
Figure 8.1: Goodness-of-fit assessment for $k = 3$



(a) The histogram of W^*

(b) The Q-Q plot of W^*

Figure 8.2: Goodness-of-fit assessment for $k = 5$



(a) The histogram of W^*

(b) The Q-Q plot of W^*

Figure 8.3: Goodness-of-fit assessment for $k = 7$

8.2 Coverage analysis of interval estimator

We provided the asymptotic theory of the proposed overall association measure in Chapter 5.2 to construct the asymptotic confidence interval when the sample size for the contingency table is sufficiently large. We also described the conditional paired non-parametric bootstrap strategy and the BCa bootstrap method in Chapter 5.2 to employ the bootstrap confidence intervals for the proposed association measures especially when the sample size is not large and/or their true values are near the boundary of their parameter spaces. To further assess the performance of the asymptotic and bootstrap confidence interval estimators, we can analyze their coverage probability for small or moderate sample sizes through the simulation study.

Similar to the design of the simulation study in Chapter 6, we can first create a “population” contingency table of size N and compute the “true” value of the proposed association measure. Next, with sufficiently large value of M, we can simulate M sample contingency tables of small and moderate size n from the “population” contingency table. Then, we compute the asymptotic confidence interval with a nominal coverage probability (e.g. 1-α) for each sample contingency table and calculate the empirical coverage probability that an interval contains the “true” value of the proposed association measure over M sample contingency tables. Finally, we can compare the empirical coverage probability to the targeted nominal coverage probability.

The investigation of the coverage probability of the bootstrap confidence interval is similar to that of the asymptotic confidence interval. The difference is that, with sufficiently large value of B, we need to generate B bootstrap replicates of each sample contingency table and compute the bootstrap confidence interval for the proposed association measure.

8.3 Hypothesis testing for interactive and correlative association measures

We developed the permutation strategies in Chapter 5 for the marginal and conditional association measures appearing in the sequential decomposition of the overall association measure. Specifically, we explained two scenarios for testing $H_0 : \rho^2_{(\mathbf{X}^P_{(j)} \to X_i)} = 0$ and $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ depending on whether $H_0 : \rho^2_{(\mathbf{X}_{-i} \to X_i)} = 0$ is rejected or not. From Definition 4.2, 4.3 and 4.4, we noticed that

$$\rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} - 2\gamma_{(\mathbf{X}^\Omega_{(j)} \to X_i, \mathbf{X}^P_{(j+1)} \to X_i)} = \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})}.$$

Hence, we may adopt a similar approach to test $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$ and $H_0 : \gamma_{(\mathbf{X}^\Omega_{(j)} \to X_i, \mathbf{X}^P_{(j+1)} \to X_i)} = 0$ depending on whether $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ and $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i)} = 0$ are rejected or not. That is, we may consider four scenarios as follows:

S1 When neither $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ nor $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i)} = 0$ is rejected, we can first test $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$. If $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$ is not rejected, then we don't need to further test $H_0 : \gamma_{(\mathbf{X}^\Omega_{(j)} \to X_i, \mathbf{X}^P_{(j+1)} \to X_i)} = 0$ further. On the other hand, if $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$ is rejected, this means that $H_0 : \gamma_{(\mathbf{X}^\Omega_{(j)} \to X_i, \mathbf{X}^P_{(j+1)} \to X_i)} = 0$ must be rejected. In all, we only need to test $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$.

S2 When $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ is not rejected but $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i)} = 0$ is rejected, this indicates that $H_0 : \gamma_{(\mathbf{X}^\Omega_{(j)} \to X_i, \mathbf{X}^P_{(j+1)} \to X_i)} = 0$ must be rejected because $\rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})}$ is non-negative. Therefore, we only need to test $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$.

S3 When $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)})} = 0$ is rejected but $H_0 : \rho^2_{(\mathbf{X}^P_{(j+1)} \to X_i)} = 0$ is not rejected, we can again start with testing $H_0 : \rho^2_{(\mathbf{X}^\Omega_{(j)} \mathbf{X}^P_{(j+1)} \to X_i | \mathbf{X}^\Omega_{(j)}, \mathbf{X}^P_{(j+1)})} = 0$ first.

If $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ is not rejected, we don't need to test further because it implies that $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$ must be rejected. Instead, if $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ is rejected, we need to further test $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$.

S4 When both $H_0 : \rho^2_{(\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)})} = 0$ and $H_0 : \rho^2_{(\mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$ are rejected, we should test both $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ and $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$ simultaneously.

For Scenario S3 and S4, we can find out a permutation strategy for testing $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$. By Definition 4.4, we knew that $r_{\mathbf{m}^{P}_{(j+1)}} = r$ for every $\mathbf{m}^{P}_{(j+1)}$ and/or $r_{\mathbf{m}^{\Omega}_{(j)}} = r$ for every $\mathbf{m}^{\Omega}_{(j)}$ if and only if $\gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$. Then we conjecture that testing $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i, \mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$ is equivalent to testing $H_0 : \gamma_{(\mathbf{X}^{P}_{(j+1)} \to X_i)} = 0$ and $H_0 : \gamma_{(\mathbf{X}^{\Omega}_{(j)} \to X_i)} = 0$ simultaneously. Thus, we can adopt the same permutation strategy for testing the marginal association measure in Chapter 5.

In contrast, it appears not straightforward to design a permutation strategy for testing $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ for Scenario S1 and S2. However, we can still find some clues by Proposition 4.3 (f), (g) and Proposition 4.4 (d) as follows:

1. If a permutation strategy is available or can be developed to simulate samples of permutations satisfying conditional independence (e.g. X_1 and X_3 are conditionally independent given X_2), then $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ could be tested.

2. If a permutation strategy is available or can be developed to simulate samples of permutations satisfying the property that the X_1 (X_2) has no influence on the mean subcopula of X_3 given X_2 (X_1), then $H_0 : \rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)} \to X_i | \mathbf{X}^{\Omega}_{(j)}, \mathbf{X}^{P}_{(j+1)})} = 0$ could be tested.

8.4 Power analysis of the permutation test

We explained in Chapter 5.3 the permutation strategies for the applications of permutation test to the respective null hypotheses of the overall, marginal and conditional association measures. To further evaluate the performance of the permutation test for each association measure, we can analyze its size and power via the simulation study.

Because the null hypothesis indicates that the true value of the proposed association measure is zero, we can first create a "population" contingency table of size N, where the "true" value of the proposed association measure is zero. Next, with sufficiently large value of M, we can simulate M sample contingency tables of size n from the "population" contingency table. Then, we can generate Q permutations of each sample contingency table, estimate the p-value and calculate the empirical size of the permutation test for the rejections of the null hypothesis over M sample contingency tables.

The investigation of the power of the permutation test is similar to that for the size of the permutation test. The difference is that we need to create a "population" contingency table of size N, where the "true" value of the proposed association measure is not zero, and compute the empirical power of the permutation test.

To better understand the choice of the number of permutations Q with respect to the performance of the permutation test, we can estimate the size and power by varying it for a sample contingency table.

8.5 Impact of sparsity of the multi-dimensional contingency table

We observed from the simulation study in Chapter 6 that the distributions of the estimators of the proposed association measures become highly right-skewed and the biases of the estimators become larger as the sizes (depending on the number of variables and the number of categories in each variable) of the contingency tables increase, even when the sample sizes are large, regardless of the closeness of the "true" values of the proposed

association measures to zero. It is often observed that the sparsity issue results in the highly-skewed sampling distribution of well-known association measures such as *Pearson's chi-square statistic* and odds ratios, even when the sample size is large (Agresti, 2002, p. 391-398).

To further investigate the impact of the sparsity issue on the proposed association measures, we can consider the simulation study with smaller sample sizes and the other settings fixed. Similar to the selection of sample sizes in the simulation study from Chapter 6, we can control the degree of sparsity by considering the contingency tables with medium size (e.g. $5^3 = 125$) and varying their conceptual average cell counts within a small range. For example, in the three-dimensional case of Scenario 3, if the conceptual average cell counts are specified to be $(1, 2, 4, 8)$, then we obtain the corresponding sample sizes $n = (125, 250, 500, 1000)$ for each combination of $(k_{X_1}, k_{X_2}, k_{X_3})$, where $k_{X_1} = k_{X_2} = k_{X_3} = k$ are the respective numbers of categories in the variables X_1, X_2 and X_3, and $k = (3, 5, 7)$. This can be beneficial in that the simulated contingency tables for $k = 3$ will be mildly sparse while those for $k = (5, 7)$ will be extremely sparse. According to the observations from Chapter 6, we conjecture that the skewness in the distributions of the estimators of the proposed association measures will become much more notable for extremely sparse contingency tables when the "true" values of the proposed association measures are close to zero. We also speculate that the biases of the estimators will become much more pronounced as the degree of sparsity increases, regardless of the closeness of the "true" values to zero.

To alleviate the sparsity issue, some remedies were proposed including dropping or collapsing a contingency table and smoothing its counts. However, dropping or collapsing the contingency table is not recommended because the effects of dropping or collapsing the zero-count cells on the subsequent statistical analysis is hard to foresee (Guo and Thompson, 1989; Baglivo et al., 1988). Regarding the smoothing methods, one ad-hoc way is to add a small constant (e.g. 0.5) to the cell counts. However, it is not clear about the choice

of a small constant. Moreover, such a remedy may have too conservative effects on the subsequent statistical inference, especially for large size of the contingency table and/or small sample size (Agresti, 2002, p.397).

More sophisticated smoothing methods can be found in Coull and Agresti (2003) and Borgoni (2004). Coull and Agresti (2003) proposed a parametric method of smoothing the cell counts in a large multi-dimensional contingency table with (nominal/ordinal) variables by fitting a generalized log-linear model (GLLM) with random effects. Borgoni (2004) proposed a two-step smoothing method for the multi-dimensional contingency table with ordinal variables. The first step adopts a multivariate kernel estimator to smooth the cell probabilities in the contingency table and the second step generates the final smoothed cell probabilities that minimize the Kullback-Leibler divergence against the result from the first step subject to a set of linear constraints imposed by the contingency table. In the context of descriptive modeling, the smoothing method proposed by Coull and Agresti (2003) may not be aligned with our motivation of creating model-free methods for delineating and quantifying the association structure. On the other hand, although the smoothing method proposed by Borgoni (2004) requires the careful choice of the kernel estimator for the multi-dimensional contingency table, it can be used to alleviate the sparsity issue before applying the proposed association measures.

8.6 *Ridit scores* based regression and association measures

We started the entire model-free methodology with the introduction of subcopula score. As defined in Chapter 3, it is a data-dependent scoring method for an ordinal variable based on the theory of subcopula. Proposition 3.1 showed that another widely used data-dependent score for ordinal variables, the *Ridit score*, has a relationship with the subcopula score in that for the i-th category of an ordinal variable, the i-th *Ridit score* is the midpoint of the $(i-1)$-th and i-th subcopula scores. The *Ridit score* is closely related to the multilinear extension copula for discrete data (Genest and Nešlehová, 2007), which is smooth

version of subcopula by the multilinear interpolation (Schweizer et al., 1974). The use of the multilinear extension copula for ordinal variables leads to the assignment of the *Ridit scores* for the ordinal variables as a data-dependent scoring method (Nešlehová, 2007; Wei and Kim, 2021). Hence, an interesting future work would be to investigate the performance of the *Ridit scores* based association measure and its decompositions for multi-dimensional contingency tables with an ordinal dependent variable.

APPENDIX A

THE PROOF OF PROPOSITION 3.1

From Chapter 2.1, we can define the *Ridit score*, $s^R_{i_{m_i}}$ and the *midrank*, $s^M_{i_{m_i}}$, for the m_i-th category of X_i as below:

$$s^R_{i_{m_i}} = \frac{F_{i_{m_i-1}} + F_{i_{m_i}}}{2}, \quad s^M_{i_{m_i}} = ns^R_{i_{m_i}} + 0.5$$

where $F_{i_0} = 0$, $F_{i_{M_i}} = 1$ and $F_{i_{m_i}} = P(X_i \leq x_{i_{m_i}})$. Then, according to Eq. (3.1), we have $u_{i_0} = 0 = F_{i_0}$, $u_{i_{M_i}} = 1 = F_{i_{M_i}}$ and

$$u_{i_{m_i}} = \sum_{r_i=1}^{m_i} p_{r_i} = P(X_i \leq x_{i_{m_i}}) = F_{i_{m_i}}.$$

Therefore, when $F_{i_{m_i-1}}$ and $F_{i_{m_i}}$ are replaced with $u_{i_{m_i-1}}$ and $u_{i_{m_i}}$, respectively, we obtain

$$s^R_{i_{m_i}} = \frac{u_{i_{m_i-1}} + u_{i_{m_i}}}{2}, \quad s^M_{i_{m_i}} = ns^R_{i_{m_i}} + 0.5 = n\frac{u_{i_{m_i-1}} + u_{i_{m_i}}}{2} + 0.5.$$

APPENDIX B

THE PROOF OF PROPOSITION 3.2

(a) Assume that the probability for the m_i-th category of X_i is p_{m_i}. If the categories $\{x_{m_i+1}, \ldots, x_{m_i+s_i}\}$ of X_i are combined into a new category $x_{m_i^*}$, the probability $p_{m_i^*}$ for $x_{m_i^*}$ is $\sum_{r_i=m_i+1}^{m_i+s_i} p_{r_i}$. Thus, according to Eq. (3.1), the subcopula score for $x_{m_i^*}$ is given by

$$u_{i_{m_i^*}} = \sum_{r_i=1}^{m_i^*} p_{r_i} = \left(\sum_{r_i=1}^{m_i} p_{r_i}\right) + p_{m_i^*} = \sum_{r_i=1}^{m_i} p_{r_i} + \sum_{r_i=m_i+1}^{m_i+s_i} p_{r_i} = \sum_{r_i=1}^{m_i+s_i} p_{r_i} = u_{i_{m_i+s_i}}.$$

For the other $M_i - s_i$ categories of X_i, the respective subcopula scores do not change because the corresponding values of cumulative distribution function do not change.

(b) When the ordering of the categories in X_i is reversed, then $\{x_{m_i+1}, \ldots, x_{M_i}\}$ will be ordered lower than x_{m_i}. Therefore, the subcopula score for x_{m_i} becomes

$$u_{i_{m_i}} = \sum_{r_i=m_i}^{M_i} p_{r_i} = \sum_{r_i=1}^{M_i} p_{r_i} - \sum_{r_i=1}^{m_i-1} p_{r_i} = 1 - u_{i_{m_i-1}}.$$

(c) According to Eq. (3.1), the mean of U_i is calculated by

$$\begin{aligned} E(U_i) &= \sum_{m_i=1}^{M_i} \left(\sum_{r_i=1}^{m_i} p_{r_i}\right) p_{m_i} = p_1 p_1 + \cdots + (p_1 + \cdots + p_{M_i}) p_{M_i} \\ &= \frac{1}{2}[2p_1 p_1 + \cdots + 2(p_1 + \cdots + p_{M_i}) p_{M_i}] = \frac{1}{2} + \frac{1}{2} \sum_{m_i=1}^{M_i} p_{m_i}^2, \end{aligned}$$

and the mean of U_i^2 is calculated by

$$
\begin{aligned}
E(U_i^2) &= \sum_{m_i=1}^{M_i}\left(\sum_{r_i=1}^{m_i} p_{r_i}\right)^2 p_{m_i} = (p_1)^2 p_1 + \cdots + (p_1 + \cdots + p_{M_i})^2 p_{M_i} \\
&= \left(\sum_{m_i=1}^{M_i} p_{m_i}^2\right) + \left(\sum\sum_{m_i<m_i^*} p_{m_i} p_{m_i^*}^2\right) + \left[2I(M_i>2)\sum\sum\sum_{m_i<m_i^*<m_i^{**}} p_{m_i} p_{m_i^*} p_{m_i^{**}}\right].
\end{aligned}
$$

Hence the variance of U_i is given by

$$
\begin{aligned}
Var(U_i) =& E(U_i^2) - [E(U_i)]^2 \\
=& \left(\frac{1}{2}\sum_{m_i=1}^{M_i} p_{m_i}^2 - \frac{1}{4}\left[\sum_{m_i=1}^{M_i} p_{m_i}^2\right]^2 - \frac{1}{4}\right) + \left(\sum\sum_{m_i<m_i^*} p_{m_i} p_{m_i^*}^2\right) \\
&+ \left[2I(M_i>2)\sum\sum\sum_{m_i<m_i^*<m_i^{**}} p_{m_i} p_{m_i^*} p_{m_i^{**}}\right].
\end{aligned}
$$

In the special case that X_i follows a discrete uniform distribution with the support $\mathbf{S}_{x_i} = \{x_{i_1}, \ldots, x_{i_{M_i}}\}$ such that each $x_{m_i} \in \mathbf{S}_{x_i}$ occurs with equal probability $1/M_i$, we find the corresponding mean and variance of U_i by using the formulas above with $p_{m_i} = 1/M_i$ for every m_i.

APPENDIX C

THE PROOF OF PROPOSITION 3.3

(a) Given that $0 \le u_{i_{m_i}} = \sum_{r_i=m_i}^{M_i} p_{r_i} \le 1$ according to Eq. (3.1), we can derive the range of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ by Eq. (3.7) as follows:

$$0 \le r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} \le \sum_{m_i=1}^{M_i} p_{m_i|\mathbf{m}_{-i}} = 1.$$

(b) According to Eq. (3.7), the mean of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ is given by,

$$\begin{aligned} E\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] &= \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i,\mathbf{m}_{-i}} \\ &= \sum_{m_i=1}^{M_i} u_{i_{m_i}} \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} p_{m_i,\mathbf{m}_{-i}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i} = E(U_i), \end{aligned}$$

and hence the variance of $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ is given by,

$$\begin{aligned} Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] &= E\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) - E\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]\right]^2 \\ &= \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left[\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right]^2 p_{\mathbf{m}_{-i}}. \end{aligned}$$

APPENDIX D

THE PROOF OF PROPOSITION 3.4

Let $\tilde{\mathbf{U}}_{-i} = \{\tilde{U}_1, \ldots, \tilde{U}_{i-1}, \tilde{U}_{i+1}, \ldots, \tilde{U}_d\}$, where $\tilde{U}_i = F_{\tilde{X}_i}(g_i(\tilde{X}_i))$ for $i = 1, \ldots, d$. It is sufficient to show that $r^C_{U_i|\tilde{\mathbf{U}}_{-i}}(\tilde{\mathbf{u}}_{-i_{\tilde{\mathbf{m}}_{-i}}}) = r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ for every $(\tilde{\mathbf{m}}_{-i}, \mathbf{m}_{-i})$. That is,

$$\sum_{m_i=1}^{M_i} u_{i_{m_i}} \tilde{p}_{m_i|\tilde{\mathbf{m}}_{-i}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}}.$$

Since

$$\tilde{p}_{m_i|\tilde{\mathbf{m}}_{-i}} = \frac{\tilde{p}_{m_i,\tilde{\mathbf{m}}_{-i}}}{\sum_{m_i=1}^{M_i} \tilde{p}_{m_i,\tilde{\mathbf{m}}_{-i}}} \quad \text{and} \quad p_{m_i|\mathbf{m}_{-i}} = \frac{p_{m_i,\mathbf{m}_{-i}}}{\sum_{m_i=1}^{M_i} p_{m_i,\mathbf{m}_{-i}}},$$

so it is equivalent to show $\tilde{p}_{m_i,\tilde{\mathbf{m}}_{-i}} = p_{m_i,\mathbf{m}_{-i}}$ for every $(\tilde{\mathbf{m}}_{-i}, \mathbf{m}_{-i})$.

Given that $g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_d$ are injective functions of $\mathbf{X}_{-i}$, then for every $j \neq i$, there exists $\tilde{m}_j \in \tilde{\mathbf{m}}_{-i}$ and $m_j \in \mathbf{m}_{-i}$ such that the joint probability $\tilde{p}_{m_i,\tilde{m}_j,\tilde{\mathbf{m}}_{-ij}} = p_{m_i,m_j,\mathbf{m}_{-ij}}$, where $\tilde{\mathbf{m}}_{-ij} = \tilde{\mathbf{m}}_{-i} - \{\tilde{m}_j\}$ and $\mathbf{m}_{-ij} = \mathbf{m}_{-i} - \{m_j\}$. Hence we have $\tilde{p}_{m_i,\tilde{\mathbf{m}}_{-i}} = p_{m_i,\mathbf{m}_{-i}}$, which leads to $\tilde{p}_{i_{m_i|\tilde{\mathbf{m}}_{-i}}} = p_{m_i|\mathbf{m}_{-i}}$. Therefore, $r^C_{U_i|\tilde{\mathbf{U}}_{-i}}(\tilde{\mathbf{u}}_{-i_{\tilde{\mathbf{m}}_{-i}}}) = r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ and the predicted category of X_i is invariant.

APPENDIX E

THE PROOF OF PROPOSITION 4.1

(a) Since

$$Var(U_i) = E[Var(U_i|\mathbf{U}_{-i})] + Var[E(U_i|\mathbf{U}_{-i})] = E[Var(U_i|\mathbf{U}_{-i})] + Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right],$$

so we have $0 \leq Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] = Var(U_i) - E[Var(U_i|\mathbf{U}_{-i})] \leq Var(U_i)$. Therefore,

$$0 = \frac{0}{Var(U_i)} \leq \rho^2_{(\mathbf{X}_{-i} \to X_i)} \leq \frac{Var(U_i)}{Var(U_i)} = 1.$$

(b) If X_i and $\mathbf{X}_{-i}$ are independent, so are U_i and $\mathbf{U}_{-i}$. Then

$$Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] = Var[E(U_i|\mathbf{U}_{-i})] = Var[E(U_i)] = 0.$$

Therefore, $\rho^2_{(\mathbf{X}_{-i} \to X_i)} = 0/Var(U_i) = 0$.

(c) Given $\rho^2_{(\mathbf{X}_{-i} \to X_i)} = 0$, then $Var[E(U_i|\mathbf{U}_{-i})] = E[E(U_i|\mathbf{U}_{-i}) - E(U_i)]^2 = 0$, which indicates that $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = E(U_i|\mathbf{U}_{-i}) = E(U_i)$. Therefore, $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ is constant over every combination of categories in $\mathbf{X}_{-i}$. Furthermore, let $\mathbf{U}_{d^*} = \{U_{j_1}, \ldots, U_{j_{d^*}}\}$ be a subset of $\mathbf{U}_{-i}$ with size d^*, where $1 \leq d^* \leq d-1$. Also let $\mathbf{m}_{d^*} = \{m_{j_1}, \ldots, m_{j_{d^*}}\}$ be a subset of $\mathbf{m}_{-i}$, $\mathbf{M}_{d^*} = \{M_{j_1}, \ldots, M_{j_{d^*}}\}$ be a subset of $\mathbf{M}_{-i}$, $\mathbf{M}_{-i-d^*} = \mathbf{M}_{-i} - \mathbf{M}_{d^*}$ and $\mathbf{m}_{-i-d^*} = \mathbf{m}_{-i} - \mathbf{m}_{d^*}$. Since $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = E(U_i)$ implies $\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}$ for every $\mathbf{m}_{-i}$, then for every

$\mathbf{m}_{-i}$ and $\mathbf{m}_{d^*}$,

$$\frac{p_{\mathbf{m}_{-i}}}{p_{\mathbf{m}_{d^*}}}\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} = \frac{p_{\mathbf{m}_{-i}}}{p_{\mathbf{m}_{d^*}}}\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}, \tag{E.1}$$

where $p_{\mathbf{m}_{d^*}}$ is the joint p.m.f of $\mathbf{U}_{d^*}$. By summing over $\mathbf{m}_{-i-d^*}$ on both sides of Eq. (E.1), we can simplify the left-hand side to be

$$\sum_{m_i=1}^{M_i} u_{i_{m_i}} \sum_{\mathbf{m}_{-i-d^*}=\mathbf{1}_{-i-d^*}}^{\mathbf{M}_{-i-d^*}} \frac{p_{m_i,\mathbf{m}_{-i}}}{p_{\mathbf{m}_{d^*}}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} \frac{p_{m_i,\mathbf{m}_{d^*}}}{p_{\mathbf{m}_{d^*}}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{d^*}},$$

and right-hand side to be

$$\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i} \sum_{\mathbf{m}_{-i-d^*}=\mathbf{1}_{-i-d^*}}^{\mathbf{M}_{-i-d^*}} \frac{p_{\mathbf{m}_{-i}}}{p_{\mathbf{m}_{d^*}}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i} \frac{p_{\mathbf{m}_{d^*}}}{p_{\mathbf{m}_{d^*}}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i},$$

where $\mathbf{1}_{-i-d^*}$ is a vector of all ones with size $d-1-d^*$. Thus $\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{d^*}} = \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}$. Finally, let $W = \prod_{r=j_1}^{j_{d^*}} U_r$ denote the product (interactions) of $U_r \in \mathbf{U}_{d^*}$. Then the covariance $Cov(W, U_i)$ between W and U_i is given by

$$\begin{aligned}
&Cov(W,U_i) = E(WU_i) - E(W)E(U_i) \\
&= \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} \sum_{m_i=1}^{M_i} u_{i_{m_i}} w_{\mathbf{m}_{d^*}} p_{m_i,\mathbf{m}_{d^*}}\right) - \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} w_{\mathbf{m}_{d^*}} p_{\mathbf{m}_{d^*}}\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right) \\
&= \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} w_{\mathbf{m}_{d^*}} p_{\mathbf{m}_{d^*}} \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{d^*}}\right) - \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} w_{\mathbf{m}_{d^*}} p_{\mathbf{m}_{d^*}}\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right) \\
&= \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} w_{\mathbf{m}_{d^*}} p_{\mathbf{m}_{d^*}}\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right) - \left(\sum_{\mathbf{m}_{d^*}=\mathbf{1}_{d^*}}^{\mathbf{M}_{d^*}} w_{\mathbf{m}_{d^*}} p_{\mathbf{m}_{d^*}}\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right) = 0,
\end{aligned}$$

where $w_{\mathbf{m}_{d^*}} = \prod_{r=j_1}^{j_{d^*}} u_{r_{m_r}}$ is the product of the m_r-th subcopula scores of $U_r \in \mathbf{U}_{d^*}$. Therefore, $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ implies that U_i is uncorrelated with every $U_j \in \mathbf{U}_{-i}$ and their interactions.

(d) Since $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 1$ indicates $Var(U_i) = Var(E(U_i|\mathbf{U}_{-i}))$. Hence,

$$
\begin{aligned}
&Var(U_i) - Var(E(U_i|\mathbf{U}_{-i})) = E(U_i^2) - [E(U_i)]^2 - E[E^2(U_i|\mathbf{U}_{-i})] + [E(U_i)]^2 \\
&= E(U_i^2) - E[E^2(U_i|\mathbf{U}_{-i})] = 0,
\end{aligned}
$$

which leads to $E(U_i^2) = E[E^2(U_i|\mathbf{U}_{-i})]$. Hence we conclude $U_i = E(U_i|\mathbf{U}_{-i} = \mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ almost surely. Therefore, $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 1$ if and only if $U_i = h(U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_d) = h(\mathbf{U}_{-i})$ almost surely for some measurable function h. Finally, since $U_j = F_{X_j}(X_j)$ where F_{X_j} is a bijective function of $X_j \in \mathbf{X}_{-i}$, so $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 1$ if and only if $X_i = h(F^{-1}_{X_1}(X_1), \ldots, F^{-1}_{X_{i-1}}(X_{i-1}), F^{-1}_{X_{i+1}}(X_{i+1}), \ldots, F^{-1}_{X_d}(X_d)) = g(\mathbf{X}_{-i})$ almost surely for some measurable function g.

(e) If $U_i = g(\mathbf{U}_{-i}) + \epsilon$, then $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = E(U_i|\mathbf{U}_{-i}) = E[g(\mathbf{U}_{-i}) + \epsilon|\mathbf{U}_{-i}] = g(\mathbf{U}_{-i}) + E(\epsilon)$ and $Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] = Var[g(\mathbf{U}_{-i}) + E(\epsilon)] = Var[g(\mathbf{U}_{-i})]$. Furthermore, given that ϵ is independent of $\mathbf{U}_{-i}$ and has finite second moment, $Var(U_i) = Var[g(\mathbf{U}_{-i}) + \epsilon] = Var[g(\mathbf{U}_{-i})] + Var(\epsilon)$. Therefore,

$$
\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \frac{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}{Var(U_i)} = \frac{Var[g(\mathbf{U}_{-i})]}{Var[g(\mathbf{U}_{-i})] + Var(\epsilon)}
$$

(f) The expectation of the product of U_i and $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ is given by

$$
\begin{aligned}
&E\left[U_i r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] = E[U_i E(U_i|\mathbf{U}_{-i})] = \sum_{\mathbf{m}_d=\mathbf{1}_d}^{\mathbf{M}_d} u_{i_{m_i}} \left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}}\right) p_{\mathbf{m}_i} \\
&= \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \sum_{m_i=1}^{M_i} u_{i_{m_i}} \left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}}\right) p_{m_i|\mathbf{m}_{-i}} p_{\mathbf{m}_{-i}} \\
&= \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}}\right) \left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}}\right) p_{\mathbf{m}_{-i}} = E[E^2(U_i|\mathbf{U}_{-i})],
\end{aligned}
$$

and hence the covariance between U_i and $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})$ is derived by

$$Cov\left[U_i, r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right] = E[U_i E(U_i|\mathbf{U}_{-i})] - E(U_i)E[E(U_i|\mathbf{U}_{-i})]$$
$$= E[E^2(U_i|\mathbf{U}_{-i})] - E^2[E(U_i|\mathbf{U}_{-i})] = Var(E(U_i|\mathbf{U}_{-i})) = Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right].$$

Thus,

$$\rho_{(\mathbf{X}_{-i}\to X_i)} = \frac{\sqrt{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}}{\sqrt{Var(U_i)}} = \frac{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}{\sqrt{Var(U_i)}\sqrt{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}}$$
$$= \frac{Cov\left[U_i, r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}{\sqrt{Var(U_i)}\sqrt{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}} = corr\left[U_i, r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right].$$

(g) By Proposition 3.4, given $\tilde{\mathbf{X}}_{-i} = g_{-i}(\mathbf{X}_{-i})$, $r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}}) = r^C_{U_i|\tilde{\mathbf{U}}_{-i}}(\tilde{\mathbf{u}}_{-i_{\tilde{\mathbf{m}}_{-i}}})$. Therefore,

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \frac{Var\left[r^C_{U_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}{Var(U_i)} = \frac{Var\left[r^C_{U_i|\tilde{\mathbf{U}}_{-i}}(\tilde{\mathbf{u}}_{-i_{\tilde{\mathbf{m}}_{-i}}})\right]}{Var(U_i)} = \rho^2_{(\tilde{\mathbf{X}}_{-i}\to X_i)},$$

which implies that $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is invariant over the permutation of the categories of every $X_j \in \mathbf{X}_{-i}$.

(h) Suppose $\tilde{X}_i$ is the permuted X_i and $\tilde{U}_i = F_{\tilde{X}_i}(\tilde{X}_i)$, where $F_{\tilde{X}_i}$ is the c.d.f. of $\tilde{X}_i$. Then,

$$\rho^2_{(\mathbf{X}_{-i}\to \tilde{X}_i)} = \frac{Var\left[r^C_{\tilde{U}_i|\mathbf{U}_{-i}}(\mathbf{u}_{-i_{\mathbf{m}_{-i}}})\right]}{Var(\tilde{U}_i)}$$
$$= \frac{\sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}}\left(\sum_{\tilde{m}_i=1}^{M_i}\tilde{u}_{i_{\tilde{m}_i}}\tilde{p}_{\tilde{m}_i|\mathbf{m}_{-i}} - \sum_{\tilde{m}_i=1}^{M_i}\tilde{u}_{i_{\tilde{m}_i}}\tilde{p}_{\tilde{m}_i}\right)^2 p_{\mathbf{m}_{-i}}}{\sum_{\tilde{m}_i=1}^{M_i}\left(\tilde{u}_{i_{\tilde{m}_i}} - \sum_{\tilde{m}_i=1}^{M_i}\tilde{u}_{i_{\tilde{m}_i}}\tilde{p}_{\tilde{m}_i}\right)^2 \tilde{p}_{\tilde{m}_i}}. \qquad \text{(E.2)}$$

By Eq. (4.1) and Eq. (E.2), if

$$\frac{\left(\sum_{\tilde{m}_i=1}^{M_i} \tilde{u}_{i_{\tilde{m}_i}} \tilde{p}_{\tilde{m}_i|\mathbf{m}_{-i}} - \sum_{\tilde{m}_i=1}^{M_i} \tilde{u}_{i_{\tilde{m}_i}} \tilde{p}_{\tilde{m}_i}\right)^2}{\sum_{\tilde{m}_i=1}^{M_i} \left(\tilde{u}_{i_{\tilde{m}_i}} - \sum_{\tilde{m}_i=1}^{M_i} \tilde{u}_{i_{\tilde{m}_i}} \tilde{p}_{\tilde{m}_i}\right)^2 \tilde{p}_{\tilde{m}_i}} = \frac{\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right)^2}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right)^2 p_{m_i}}, \tag{E.3}$$

holds for every $\mathbf{m}_{-i}$, then $\rho^2_{(\mathbf{X}_{-i}\to\tilde{X}_i)} = \rho^2_{(\mathbf{X}_{-i}\to X_i)}$. Given that X_i is a binary variable, the left-hand side of Eq. (E.3) becomes

$$\frac{[p_2 p_{2|\mathbf{m}_{-i}} + p_{1|\mathbf{m}_{-i}} - p_2^2 - p_1]^2}{(p_2 - p_2^2 - p_1)^2 p_2 + (1 - p_2^2 - p_1)^2 p_1} = \frac{(p_{1|\mathbf{m}_{-i}} p_1 - p_1^2)^2}{p_1^3 p_2} = \frac{(p_{1|\mathbf{m}_{-i}} - p_1)^2}{p_1 p_2}.$$

and the right-hand becomes

$$\frac{[p_1 p_{1|\mathbf{m}_{-i}} + p_{2|\mathbf{m}_{-i}} - p_1^2 - p_2]^2}{(p_1 - p_1^2 - p_2)^2 p_1 + (1 - p_1^2 - p_2)^2 p_2} = \frac{(p_{2|\mathbf{m}_{-i}} p_2 - p_2^2)^2}{p_1 p_2^3} = \frac{(p_{1|\mathbf{m}_{-i}} - p_1)^2}{p_1 p_2}.$$

Therefore, $\rho^2_{(\mathbf{X}_{-i}\to\tilde{X}_i)} = \rho^2_{(\mathbf{X}_{-i}\to X_i)}$.

APPENDIX F

THE PROOF OF THEOREM 4.2

(a) It is sufficient to show that $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)})}$ for $\mathbf{X}^{\Omega}_{(J-1)}$ and $\mathbf{X}^{P}_{(J)}$ in a partition P of $\mathbf{X}_{(d-1)}$, where $\mathbf{X}_{(d-1)} = \mathbf{X}^{\Omega}_{(J-1)}\cup\mathbf{X}^{P}_{(J)}$ is a permutation of $\mathbf{X}_{-i}$ defined in Section 4.2.2. This is because we can further recursively decompose $\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)}, \ldots, \rho^2_{(\mathbf{X}^{\Omega}_{(2)}\to X_i)}$ in the same way as we decompose $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$. By Eq. (4.1),

$$
\begin{aligned}
&\rho^2_{(\mathbf{X}_{-i}\to X_i)}\\
=&\frac{\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}} - r_{\mathbf{m}^{\Omega}_{(J-1)}} + r_{\mathbf{m}^{\Omega}_{(J-1)}} - r\right)^2 p_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}}{Var(U_i)}\\
=&\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)})}\\
&+\frac{2\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}} - r_{\mathbf{m}^{\Omega}_{(J-1)}}\right)\left(r_{\mathbf{m}^{\Omega}_{(J-1)}} - r\right) p_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}}{Var(U_i)}\\
=&\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)})}\\
&+\frac{2\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)}} - r\right)\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}} - r_{\mathbf{m}^{\Omega}_{(J-1)}}\right) p_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}}{Var(U_i)}\\
=&\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)})}\\
&+\frac{2\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)}} - r\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i,\mathbf{m}^{\Omega}_{(J-1)}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i,\mathbf{m}^{\Omega}_{(J-1)}}\right)}{Var(U_i)}\\
=&\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)})}.
\end{aligned}
$$

Therefore, the sequential decomposition of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is proved. In particular, if P^* is a new partition of $\mathbf{X}_{(d-1)}$ where $\mathbf{X}^{P^*} = \mathbf{X}^P_{(j+1)} \cup \mathbf{X}^P_{(j+2)}$, then

$$
\begin{aligned}
\rho^2_{(\mathbf{X}^\Omega_{(j+2)}\to X_i)} &= \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i)} + \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})} + \rho^2_{(\mathbf{X}^P_{(j+2)}\to X_i|\mathbf{X}^\Omega_{(j+1)})} \\
&= \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i)} + \rho^2_{(\mathbf{X}^{P^*}\to X_i|\mathbf{X}^\Omega_{(j)})}.
\end{aligned}
$$

Thus, $\rho^2_{(\mathbf{X}^{P^*}\to X_i|\mathbf{X}^\Omega_{(j)})} = \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})} + \rho^2_{(\mathbf{X}^P_{(j+2)}\to X_i|\mathbf{X}^\Omega_{(j+1)})}$. Furthermore, by Proposition 4.1, it is easy to see that $\rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i)}$ is invariant with respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$ for every j, i.e. $\rho^2_{(\tilde{\mathbf{X}}^\Omega_{(j)}\to X_i)} = \rho^2_{(\mathbf{X}^\Omega_{(j)}\to \tilde{X}_i)} = \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i)}$. Therefore, by Eq. (4.7), the conditional association measure $\rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})}$ is also invariant with respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$ for every j.

(b) It is sufficient to show that

$$
\begin{aligned}
\rho^2_{(\mathbf{X}_{-i}\to X_i)} =& \rho^2_{(\mathbf{X}^\Omega_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^P_{(J)}\to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(J-1)}\mathbf{X}^P_{(J)}\to X_i|\mathbf{X}^\Omega_{(J-1)},\mathbf{X}^P_{(J)})} \\
&- 2\gamma_{(\mathbf{X}^\Omega_{(J-1)}\to X_i,\mathbf{X}^P_{(J)}\to X_i)}.
\end{aligned}
$$

This is because we can further recursively decompose $\rho^2_{(\mathbf{X}^\Omega_{(J-1)}\to X_i)}, \ldots, \rho^2_{(\mathbf{X}^\Omega_{(2)}\to X_i)}$ in the same way as we decompose $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$. By Eq. (4.7),

$$
\begin{aligned}
&\rho^2_{(\mathbf{X}_{-i}\to X_i)} \\
=&\rho^2_{(\mathbf{X}^\Omega_{(J-1)}\to X_i)} + \frac{\sum_{\mathbf{m}^\Omega_{(J-1)}=\mathbf{1}^\Omega_{(J-1)}}^{\mathbf{M}^\Omega_{(J-1)}} \sum_{\mathbf{m}^P_{(J)}=\mathbf{1}^P_{(J)}}^{\mathbf{M}^P_{(J)}} \left(r_{\mathbf{m}^\Omega_{(J-1)},\mathbf{m}^P_{(J)}} - r_{\mathbf{m}^\Omega_{(J-1)}}\right)^2 p_{\mathbf{m}^\Omega_{(J-1)},\mathbf{m}^P_{(J)}}}{Var(U_i)} \\
=&\rho^2_{(\mathbf{X}^\Omega_{(J-1)}\to X_i)} \\
&+ \frac{\sum_{\mathbf{m}^\Omega_{(J-1)}=\mathbf{1}^\Omega_{(J-1)}}^{\mathbf{M}^\Omega_{(J-1)}} \sum_{\mathbf{m}^P_{(J)}=\mathbf{1}^P_{(J)}}^{\mathbf{M}^P_{(J)}} \left(r_{\mathbf{m}^\Omega_{(J-1)},\mathbf{m}^P_{(J)}} - r_{\mathbf{m}^\Omega_{(J-1)}} - r_{\mathbf{m}^P_{(J)}} + r + r_{\mathbf{m}^P_{(J)}} - r\right)^2 p_{\mathbf{m}^\Omega_{(J-1)},\mathbf{m}^P_{(J)}}}{Var(U_i)} \\
=&\rho^2_{(\mathbf{X}^\Omega_{(J-1)}\to X_i)} + \rho^2_{(\mathbf{X}^P_{(J)}\to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(J-1)}\mathbf{X}^P_{(J)}\to X_i|\mathbf{X}^\Omega_{(J-1)},\mathbf{X}^P_{(J)})}
\end{aligned}
$$

$$+\frac{2\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}-r_{\mathbf{m}^{\Omega}_{(J-1)}}-r_{\mathbf{m}^{P}_{(J)}}+r\right)\left(r_{\mathbf{m}^{P}_{(J)}}-r\right)p_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}}{Var(U_i)}$$

$$=\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)}+\rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i)}+\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)},\mathbf{X}^{P}_{(J)})}-2\gamma_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i,\mathbf{X}^{P}_{(J)}\to X_i)}$$

$$+\frac{2\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{P}_{(J)}}-r\right)\sum_{\mathbf{m}^{\Omega}_{(J-1)}=\mathbf{1}^{\Omega}_{(J-1)}}^{\mathbf{M}^{\Omega}_{(J-1)}}\left(r_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}-r_{\mathbf{m}^{P}_{(J)}}\right)p_{\mathbf{m}^{\Omega}_{(J-1)},\mathbf{m}^{P}_{(J)}}}{Var(U_i)}$$

$$=\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)}+\rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i)}+\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)},\mathbf{X}^{P}_{(J)})}-2\gamma_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i,\mathbf{X}^{P}_{(J)}\to X_i)}$$

$$+\frac{2\sum_{\mathbf{m}^{P}_{(J)}=\mathbf{1}^{P}_{(J)}}^{\mathbf{M}^{P}_{(J)}}\left(r_{\mathbf{m}^{P}_{(J)}}-r\right)\left(\sum_{m_i=1}^{M_i}u_{i_{m_i}}p_{m_i,\mathbf{m}^{P}_{(J)}}-\sum_{m_i=1}^{M_i}u_{i_{m_i}}p_{m_i,\mathbf{m}^{P}_{(J)}}\right)}{Var(U_i)}$$

$$=\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i)}+\rho^2_{(\mathbf{X}^{P}_{(J)}\to X_i)}+\rho^2_{(\mathbf{X}^{\Omega}_{(J-1)}\mathbf{X}^{P}_{(J)}\to X_i|\mathbf{X}^{\Omega}_{(J-1)},\mathbf{X}^{P}_{(J)})}-2\gamma_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i,\mathbf{X}^{P}_{(J)}\to X_i)}.$$

Thus, the non-sequential decomposition of $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$ is proved. Proposition 4.1 (g), it is easy to see that $r_{\mathbf{m}^{\Omega}_{(j)},\mathbf{m}^{P}_{(j+1)}}$, $r_{\mathbf{m}^{\Omega}_{(j)}}$, $r_{\mathbf{m}^{P}_{(j+1)}}$ and r are invariant with respect to $\tilde{\mathbf{X}}_{-i}$ for every j. Therefore, the interactive and correlative association measures $\rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)}\to X_i|\mathbf{X}^{\Omega}_{(j)},\mathbf{X}^{P}_{(j+1)})}$ and $\gamma_{(\mathbf{X}^{\Omega}_{(j)}\to X_i,\mathbf{X}^{P}_{(j+1)}\to X_i)}$ are invariant with respect to $\tilde{\mathbf{X}}_{-i}$ for every j. In fact, they are also invariant with respect to $\tilde{X}_i$ for every j. Given that $\rho^2_{(\mathbf{X}_{-i}\to X_i)}$, $\rho^2_{(\mathbf{X}^{\Omega}_{(j)}\to X_i)}$ and $\rho^2_{(\mathbf{X}^{P}_{(j+1)}\to X_i)}$ are invariant with respect to $\tilde{X}_i$ for every j by Proposition 4.1 (h), it is sufficient to show that the correlative association measure $\gamma_{(\mathbf{X}^{\Omega}_{(j)}\to X_i,\mathbf{X}^{P}_{(j+1)}\to X_i)}$ is invariant with respect to $\tilde{X}_i$ for every j. Based on the similar argument to that in Proposition 4.1 (h), for every $\mathbf{m}^{\Omega}_{(j)}$ and $\mathbf{m}^{P}_{(j+1)}$,

$$\gamma_{(\mathbf{X}^{\Omega}_{(j)}\to\tilde{X}_i,\mathbf{X}^{P}_{(j+1)}\to\tilde{X}_i)}=\frac{(p_{1|\mathbf{m}^{\Omega}_{(j)}}-p_1)(p_{1|\mathbf{m}^{P}_{(j+1)}}-p_1)}{p_1p_2}=\gamma_{(\mathbf{X}^{\Omega}_{(J-1)}\to X_i,\mathbf{X}^{P}_{(J)}\to X_i)}.$$

Therefore, the correlative association measure $\gamma_{(\mathbf{X}^{\Omega}_{(j)}\to X_i,\mathbf{X}^{P}_{(j+1)}\to X_i)}$ is invariant with respect to $\tilde{X}_i$ for every j. By Eq. (4.8), it is easy to see that the interactive association measure $\rho^2_{(\mathbf{X}^{\Omega}_{(j)}\mathbf{X}^{P}_{(j+1)}\to X_i|\mathbf{X}^{\Omega}_{(j)},\mathbf{X}^{P}_{(j+1)})}$ is also invariant with respect to $\tilde{X}_i$ for every j.

(c) Since the (overall/marginal/conditional/interactive/correlative) association measures are all invariant with respect to $\tilde{\mathbf{X}}_{-i}$ and $\tilde{X}_i$ for every j as shown above, so the sequential and non-sequential decomposition in Eq. (4.7) and (4.8) are invariant with

respect to the permutation of the categories of every independent variable $X_j \in \mathbf{X}_{-i}$ and the binary dependent variable X_i.

APPENDIX G

THE PROOF OF PROPOSITION 4.7

(a) If X_i is jointly independent of $\mathbf{X}_{-i}$, then $p_{m_i,\mathbf{m}_{-i}} = p_{m_i}p_{\mathbf{m}_{-i}}$, which implies $p_{m_i,\mathbf{m}^P_{(k)}} = p_{m_i}p_{\mathbf{m}^P_{(k)}}$ for every $\mathbf{m}^P_{(k)}$. Thus, by Definition 4.1, $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and $\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} = 0$ for $1 \le k \le d-1$.

(b) Since this is a special case of Proposition 4.7 (a) as every $X_i \in \mathbf{X}_d$ is independent of $X_k \in \mathbf{X}_d$ for $i \neq k$, so $\rho^2_{(\mathbf{X}_{-i}\to X_i)} = 0$ and $\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} = 0$ for $1 \le k \le d-1$.

(c) If X_i is marginally independent of $\mathbf{X}^\Omega_{(j)}$, then $p_{m_i,\mathbf{m}^\Omega_{(j)}} = p_{m_i}p_{\mathbf{m}^\Omega_{(j)}}$ for every $\mathbf{m}^\Omega_{(j)}$ and $p_{m_i,\mathbf{m}^\Omega_{(k)}} = p_{m_i}p_{\mathbf{m}^\Omega_{(k)}}$ for every $\mathbf{m}^\Omega_{(k)}$ with $1 \le k \le j-1$. Then, by Definition 4.1 and 4.2, $\rho^2_{(\mathbf{X}^\Omega_{(1)}\to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)}\to X_i)} = 0$ and $\rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})} = 0$ for $1 \le k \le j-1$. Thus, Eq. (4.7) becomes

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i}\to X_i)} &= \rho^2_{(\mathbf{X}^P_{(1)}\to X_i)} + \sum_{j=1}^{J-1}\rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})} = \sum_{k=j}^{J-1}\rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})}\\ &= \rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i)} + \sum_{k=j+1}^{J-1}\rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})}.\end{aligned}$$

On the other hand, $p_{m_i,\mathbf{m}^\Omega_{(k)}} = p_{m_i}p_{\mathbf{m}^\Omega_{(k)}}$ also indicates $r_{\mathbf{m}^\Omega_{(k)}} = \sum_{m_i=1}^{M_i} u_{i m_i} p_{m_i|\mathbf{m}^\Omega_{(k)}} = \sum_{m_i=1}^{M_i} u_{i m_i} p_{m_i} = r$ for every $\mathbf{m}^\Omega_{(k)}$ with $1 \le k \le j$. By Definition 4.2 and 4.3, this

implies

$$\rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} = \frac{\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(r_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} - r_{\mathbf{m}^P_{(j+1)}}\right)^2 p_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}}$$

$$= \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i|\mathbf{X}^P_{(j+1)})}.$$

Therefore, Eq. (4.8) becomes

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i}\to X_i)} = &\sum_{k=j+1}^{J} \rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(j)}\to X_i|\mathbf{X}^P_{(j+1)})} \\ &+ \sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)},\mathbf{X}^P_{(k+1)})} - 2\sum_{k=j+1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)}.\end{aligned}$$

(d) If $\mathbf{X}^P_{(j+1)}$ is marginally independent of $\mathbf{X}^\Omega_{(j)}$, so $p_{\mathbf{m}^P_{(j+1)},\mathbf{m}^\Omega_{(j)}} = p_{\mathbf{m}^P_{(j+1)}} p_{\mathbf{m}^\Omega_{(j)}}$ for every $\mathbf{m}^P_{(j+1)}$ and $\mathbf{m}^\Omega_{(j)}$. By Definition 4.4, this indicates

$$\gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)} = \frac{\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \left(r_{\mathbf{m}^\Omega_{(j)}} - r\right) p_{\mathbf{m}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(r_{\mathbf{m}^P_{(j+1)}} - r\right) p_{\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}} = 0.$$

Thus, Eq. (4.8) becomes

$$\begin{aligned}\rho^2_{(\mathbf{X}_{-i}\to X_i)} = &\sum_{k=1}^{J} \rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \sum_{k=1}^{J-1} \rho^2_{(\mathbf{X}^\Omega_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)},\mathbf{X}^P_{(k+1)})} \\ &- 2\sum_{k=1}^{j-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)} - 2\sum_{k=j+1}^{J-1} \gamma_{(\mathbf{X}^\Omega_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)}.\end{aligned}$$

(e) If X_i is conditionally independent of $\mathbf{X}^P_{(j+1)}$ given $\mathbf{X}^\Omega_{(j)}$, so $p_{m_i|\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} = p_{m_i|\mathbf{m}^\Omega_{(j)}}$ and hence $r_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} = r_{\mathbf{m}^\Omega_{(j)}}$ by Eq. (3.7) for every $\mathbf{m}^\Omega_{(j)}$ and $\mathbf{m}^P_{(j+1)}$. By Definition

4.2, this implies

$$\rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)})} = \frac{\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(r_{\mathbf{m}^\Omega_{(j)}} - r_{\mathbf{m}^\Omega_{(j)}}\right)^2 p_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}} = 0.$$

Therefore, Eq. (4.7) becomes

$$\rho^2_{(\mathbf{X}_{-i}\to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)}\to X_i)} + \sum_{k=1}^{j-1} \rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})} + \sum_{k=j+1}^{J-1} \rho^2_{(\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^\Omega_{(k)})}.$$

On the other hand, by Definition 4.1, 4.3 and 4.4, $r_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}} = r_{\mathbf{m}^\Omega_{(j)}}$ also implies

$$\begin{aligned}
&\rho^2_{(\mathbf{X}^P_{(j+1)}\to X_i)} + \rho^2_{(\mathbf{X}^\Omega_{(j)}\mathbf{X}^P_{(j+1)}\to X_i|\mathbf{X}^\Omega_{(j)},\mathbf{X}^P_{(j+1)})} - 2\gamma_{(\mathbf{X}^\Omega_{(j)}\to X_i,\mathbf{X}^P_{(j+1)}\to X_i)} \\
&= \frac{2\sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left[\left(r_{\mathbf{m}^P_{(j+1)}} - r\right)^2 - \left(r_{\mathbf{m}^\Omega_{(j)}} - r\right)\left(r_{\mathbf{m}^P_{(j+1)}} - r\right)\right] p_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}} \\
&= \frac{2\sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(r_{\mathbf{m}^P_{(j+1)}} - r\right) \sum_{\mathbf{m}^\Omega_{(j)}=\mathbf{1}^\Omega_{(j)}}^{\mathbf{M}^\Omega_{(j)}} \left(r_{\mathbf{m}^P_{(j+1)}} - r_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}\right) p_{\mathbf{m}^\Omega_{(j)},\mathbf{m}^P_{(j+1)}}}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}} \\
&= \frac{2\sum_{\mathbf{m}^P_{(j+1)}=\mathbf{1}^P_{(j+1)}}^{\mathbf{M}^P_{(j+1)}} \left(r_{\mathbf{m}^P_{(j+1)}} - r\right)\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i,\mathbf{m}^P_{(j+1)}} - \sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i,\mathbf{m}^P_{(j+1)}}\right)}{\sum_{m_i=1}^{M_i} \left(u_{i_{m_i}} - r\right)^2 p_{m_i}} = 0.
\end{aligned}$$

Thus, Eq. (4.8) becomes

$$\begin{aligned}
\rho^2_{(\mathbf{X}_{-i}\to X_i)} = & \sum_{k=1}^{j}\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \sum_{k=j+2}^{J-1}\rho^2_{(\mathbf{X}^P_{(k)}\to X_i)} + \sum_{k=1}^{j-1}\rho^2_{(\mathbf{X}^{\Omega}_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^{\Omega}_{(k)},\mathbf{X}^P_{(k+1)})} \\
& + \sum_{k=j+1}^{J-1}\rho^2_{(\mathbf{X}^{\Omega}_{(k)}\mathbf{X}^P_{(k+1)}\to X_i|\mathbf{X}^{\Omega}_{(k)},\mathbf{X}^P_{(k+1)})} - 2\sum_{k=1}^{j-1}\gamma_{(\mathbf{X}^{\Omega}_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)} \\
& - 2\sum_{k=j+1}^{J-1}\gamma_{(\mathbf{X}^{\Omega}_{(k)}\to X_i,\mathbf{X}^P_{(k+1)}\to X_i)}.
\end{aligned}$$

APPENDIX H

THE PROOF OF PROPOSITION 4.8

(a) By Definition 4.1, $\rho^2_{(\mathbf{X}^P_{(j)} \to X_i)} = 0$ indicates

$$\sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left(r_{\mathbf{X}^P_{(j)}} - r \right)^2 p_{\mathbf{m}_{-i}} = 0.$$

Since $p_{\mathbf{m}_{-i}}$ is assumed to be non-zero for every $\mathbf{m}_{-i}$, so $\left(r_{\mathbf{X}^P_{(j)}} - r \right)^2$ and hence $r_{\mathbf{X}^P_{(j)}} = r$ for every $\mathbf{m}^P_{(j)}$. Therefore, $r_{\mathbf{X}^P_{(j)}}$ is a constant for every $\mathbf{m}^P_{(j)}$ and equal to r.

(b) By Eq. (4.7), $\rho^2_{(\mathbf{X}_{-i} \to X_i)} = \rho^2_{(\mathbf{X}^P_{(1)} \to X_i)} + \sum_{k=1}^{j-1} \rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})}$ implies

$$\sum_{k=j}^{J-1} \rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})} = 0.$$

Since $\rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})}$ is non-negative for $j \le k \le J-1$ according to Definition 4.2, so $\rho^2_{(\mathbf{X}^P_{(k+1)} \to X_i | \mathbf{X}^\Omega_{(k)})} = 0$ for $j \le k \le J-1$ and hence $r_{\mathbf{m}^\Omega_{(k)}, \mathbf{m}^P_{(k+1)}} = r_{\mathbf{m}^\Omega_{(k)}}$ for every $\mathbf{m}^\Omega_{(k)}$ and $\mathbf{m}^P_{(k+1)}$.

APPENDIX I

THE PROOF OF THEOREM 5.1

Let $\hat{\mathbf{p}} = (\hat{p}_{1,1,\ldots,1}, \hat{p}_{2,1,\ldots,1}, \ldots, \hat{p}_{M_1,1,\ldots,1}, \ldots, \hat{p}_{M_1,M_2,\ldots,M_d})^T$ be the estimator for $\mathbf{p}$, we notice that $n\hat{\mathbf{p}}$ follows a multinomial distribution with parameters n and $\mathbf{p}^T$. Then the expectation of $\hat{\mathbf{p}}$ is given by

$$E(\hat{\mathbf{p}}) \equiv \mathbf{p} = (p_{1,1,\ldots,1}, p_{2,1,\ldots,1}, \ldots, p_{M_1,1,\ldots,1}, \ldots, p_{M_1,M_2,\ldots,M_d})^T.$$

Moreover, given $Var(n\hat{p}_{\mathbf{m}_d}) = np_{\mathbf{m}_d}(1 - p_{\mathbf{m}_d})$ and $Cov(n\hat{p}_{\mathbf{m}_d}, n\hat{p}_{\mathbf{m}_d^*}) = -np_{\mathbf{m}_d}p_{\mathbf{m}_d^*}$ for $\mathbf{m}_d \neq \mathbf{m}_d^*$, the covariance matrix of $\hat{\mathbf{p}}$ is given by

$$\begin{aligned}
Cov(\hat{\mathbf{p}}) &= \frac{1}{n}\left(diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^T\right) \\
&= \frac{1}{n}\begin{bmatrix}
p_{1,1,\ldots,1}(1-p_{1,1,\ldots,1}) & -p_{1,1,\ldots,1}p_{2,1,\ldots,1} & \cdots & -p_{1,1,\ldots,1}p_{M_1,M_2,\ldots,M_d} \\
-p_{2,1,\ldots,1}p_{1,1,\ldots,1} & p_{2,1,\ldots,1}(1-p_{2,1,\ldots,1}) & \cdots & -p_{2,1,\ldots,1}p_{M_1,M_2,\ldots,M_d} \\
\vdots & \vdots & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
-p_{M_1,M_2,\ldots,M_d}p_{1,1,\ldots,1} & -p_{M_1,M_2,\ldots,M_d}p_{2,1,\ldots,1} & \cdots & p_{M_1,M_2,\ldots,M_d}(1-p_{M_1,M_2,\ldots,M_d})
\end{bmatrix}
\end{aligned}$$

Since $n^{-1/2}(n\hat{\mathbf{p}} - n\mathbf{p}) \xrightarrow{D} \mathcal{N}\left(0, \left(diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^T\right)\right)$, we have

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{D} \mathcal{N}\left(0, \left(diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^T\right)\right),$$

To prove that $\nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p})$ exists and is non-zero, it is sufficient to show that

$$\frac{\partial \rho^2_{(\mathbf{X}_{-i}\to X_i)}}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = \frac{\frac{\partial A}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} B - A \frac{\partial B}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}}}{B^2} \tag{I.1}$$

exists and is non-zero, where

$$A = \sum_{\mathbf{m}_{-i}=\mathbf{1}_{-i}}^{\mathbf{M}_{-i}} \left(\sum_{m_i=1}^{M_i} u_{im_i} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right)^2 p_{\mathbf{m}_{-i}}$$

and

$$B = \sum_{m_i=1}^{M_i} \left(u_{im_i} - \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right)^2 p_{m_i} = \sum_{m_i=1}^{M_i} u_{im_i}^2 p_{m_i} - \left(\sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right)^2 .$$

Given that (1) $\frac{\partial\left(u_{im_i} p_{m_i|\mathbf{m}_{-i}}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = 0$ for $\mathbf{m}_{-i}^* \neq \mathbf{m}_{-i}$ and $m_i^* > m_i$; (2) $\frac{\partial\left(u_{im_i} p_{m_i|\mathbf{m}_{-i}}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = p_{m_i|\mathbf{m}_{-i}}$ for $\mathbf{m}_{-i}^* \neq \mathbf{m}_{-i}$ and $m_i^* \leq m_i$; (3) $\frac{\partial\left(u_{im_i} p_{m_i|\mathbf{m}_{-i}}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = -\frac{p_{m_i,\mathbf{m}_{-i}^*}}{p^2_{\mathbf{m}_{-i}^*}} u_{im_i}$ for $\mathbf{m}_{-i}^* = \mathbf{m}_{-i}$ and $m_i^* > m_i$; (4) $\frac{\partial\left(u_{im_i} p_{m_i|\mathbf{m}_{-i}}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = \frac{p_{\mathbf{m}_{-i}^*} - p_{m_i^*,\mathbf{m}_{-i}^*}}{p^2_{\mathbf{m}_{-i}^*}} u_{im_i^*} + p_{m_i^*|\mathbf{m}_{-i}^*}$ for $\mathbf{m}_{-i}^* = \mathbf{m}_{-i}$ and $m_i^* = m_i$; (5) $\frac{\partial\left(u_{im_i} p_{m_i|\mathbf{m}_{-i}}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = p_{m_i|\mathbf{m}_{-i}^*} - \frac{p_{m_i,\mathbf{m}_{-i}^*}}{p^2_{\mathbf{m}_{-i}^*}} u_{im_i}$ for $\mathbf{m}_{-i}^* = \mathbf{m}_{-i}$ and $m_i^* < m_i$, we have

$$\begin{aligned}
&\frac{\partial A}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} \\
&= \sum_{\mathbf{m}_{-i}^* \neq \mathbf{m}_{-i}} 2 \left(\sum_{m_i=1}^{M_i} u_{im_i} p_{m_i|\mathbf{m}_{-i}} - \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right) \left(\sum_{m_i=m_i^*}^{M_i} p_{m_i|\mathbf{m}_{-i}} - u_{im_i^*} - \sum_{m_i=m_i^*}^{M_i} p_{m_i} \right) p_{\mathbf{m}_{-i}} \\
&\quad + 2 \left(\sum_{m_i=1}^{M_i} u_{im_i} p_{m_i|\mathbf{m}_{-i}^*} - \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right) \left[\left(- \sum_{m_i=1}^{M_i} \frac{p_{m_i,\mathbf{m}_{-i}^*}}{p^2_{\mathbf{m}_{-i}^*}} u_{im_i} \right) + \left(\frac{u_{im_i^*}}{p_{\mathbf{m}_{-i}^*}} - u_{im_i^*} \right) \right. \\
&\quad \left. + \left(\sum_{m_i=m_i^*}^{M_i} p_{m_i|\mathbf{m}_{-i}^*} \right) - \left(\sum_{m_i=m_i^*}^{M_i} p_{m_i} \right) \right] p_{\mathbf{m}_{-i}^*} + \left(\sum_{m_i=1}^{M_i} u_{im_i} p_{m_i|\mathbf{m}_{-i}^*} - \sum_{m_i=1}^{M_i} u_{im_i} p_{m_i} \right)^2 .
\end{aligned}$$

Given that (1) $\frac{\partial\left(u_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = \frac{\partial\left(u^2_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = 0$ for $m_i^* > m_i$; (2) $\frac{\partial\left(u_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = p_{m_i^*} + u_{im_i^*}$ and $\frac{\partial\left(u^2_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = 2u_{im_i^*} p_{m_i^*} + u^2_{im_i^*}$ for $m_i^* = m_i$; (3) $\frac{\partial\left(u_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} = p_{m_i}$ and $\frac{\partial\left(u^2_{im_i} p_{m_i}\right)}{\partial p_{m_i^*,\mathbf{m}_{-i}^*}} =$

$2u_{i_{m_i}} p_{m_i}$ for $m_i^* < m_i$, we have

$$\frac{\partial B}{\partial p_{m_i,\mathbf{m}_{-i}}} = u^2_{i_{m_i^*}} + 2\sum_{m_i=m_i^*}^{M_i} u_{i_{m_i}} p_{m_i} - 2\left(\sum_{m_i=1}^{M_i} u_{i_{m_i}} p_{m_i}\right)\left(u_{i_{m_i^*}} + \sum_{m_i=m_i^*}^{M_i} p_{m_i}\right).$$

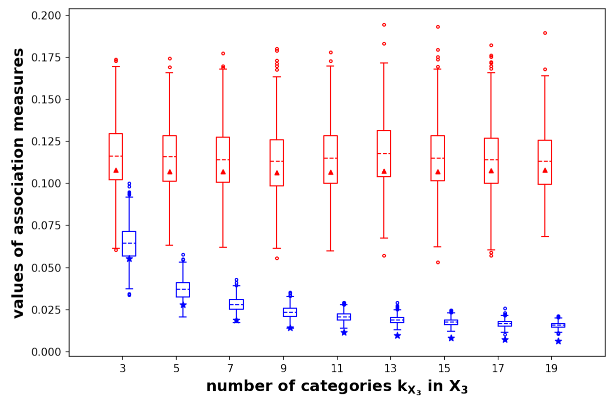
Since we find that $\nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p})$ exists and is non-zero, by delta method, we conclude

$$\sqrt{n}\left(\hat{\rho}^2_{(\mathbf{X}_{-i}\to X_i)} - \rho^2_{(\mathbf{X}_{-i}\to X_i)}\right) \xrightarrow{D} \mathcal{N}\left(0, \nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p})^T (diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^T) \nabla h_{(\mathbf{X}_{-i}\to X_i)}(\mathbf{p})\right).$$

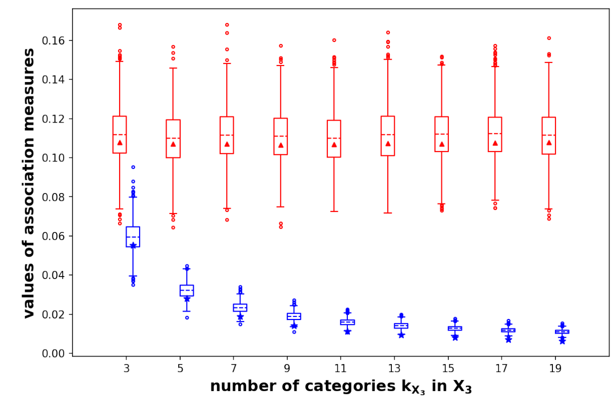
APPENDIX J

SIMULATION RESULTS FOR THREE-DIMENSION CASE

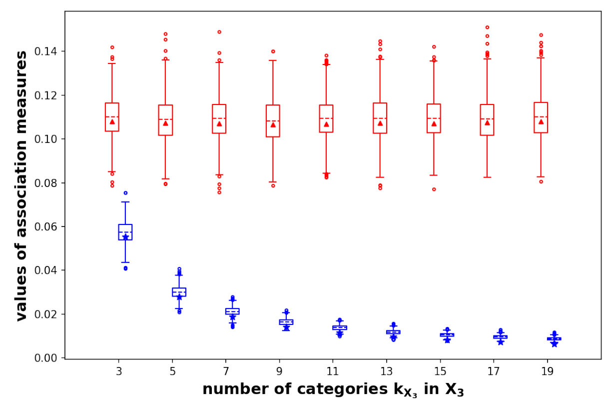
J.1 Scenario 1



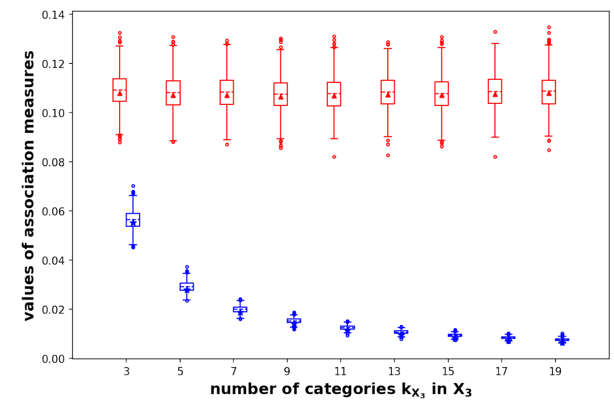
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

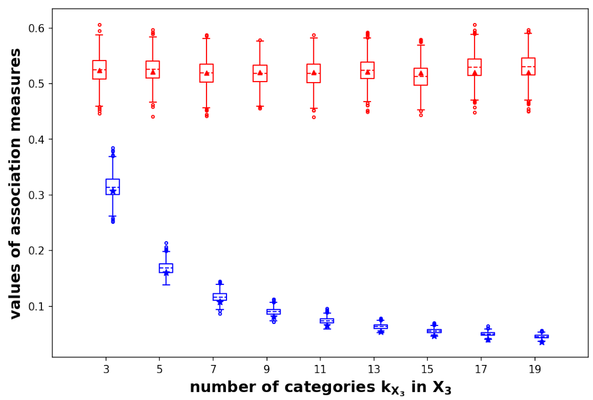


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

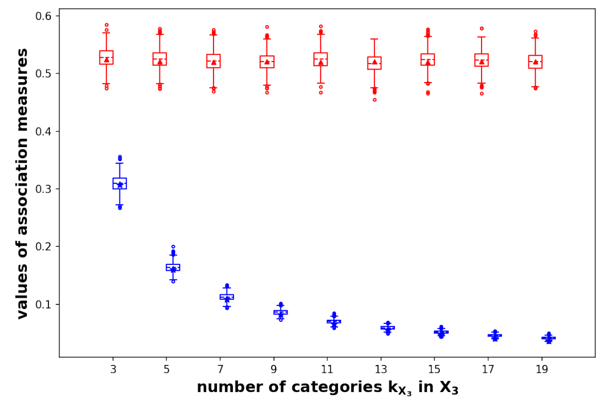


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

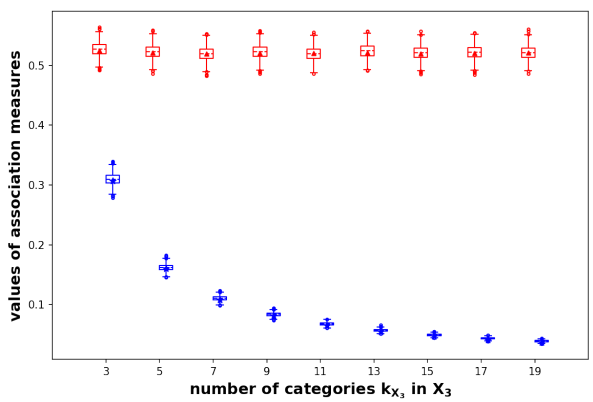
Figure J.1: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



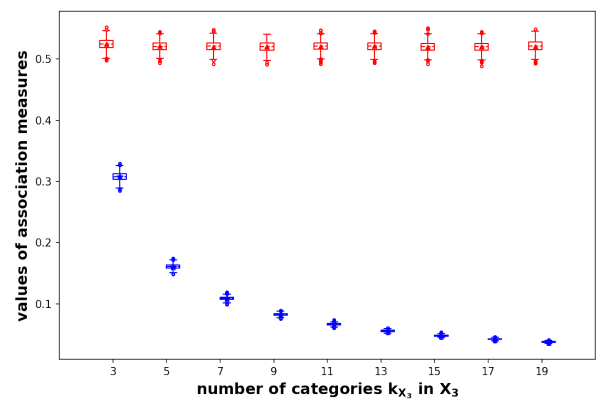
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

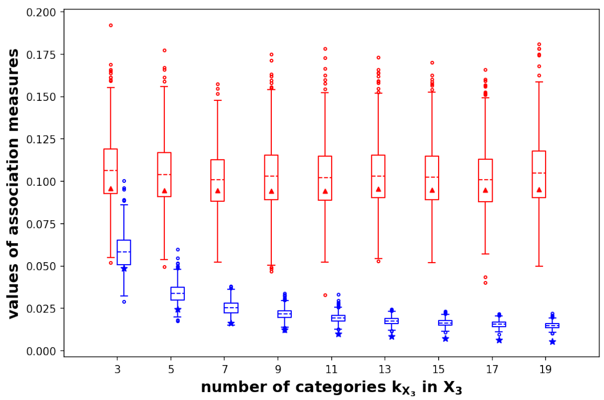


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

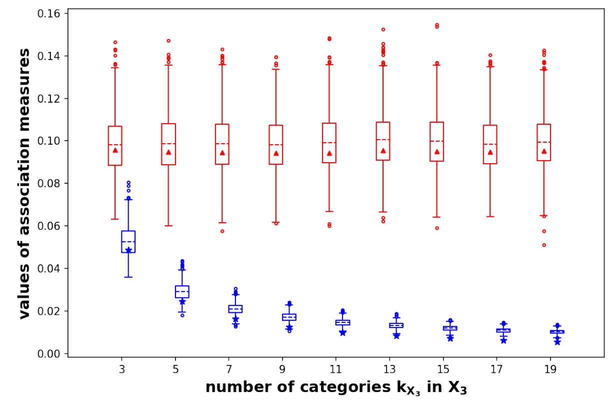


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

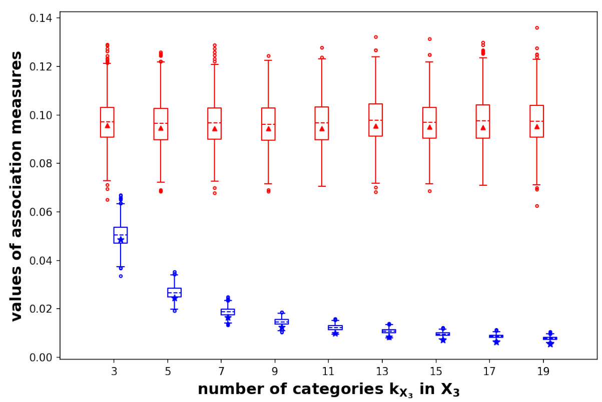
Figure J.2: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



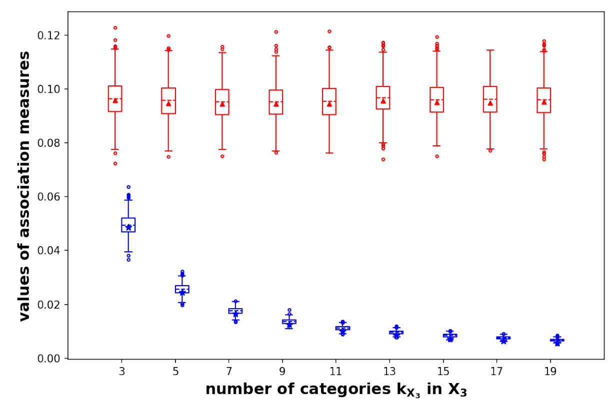
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

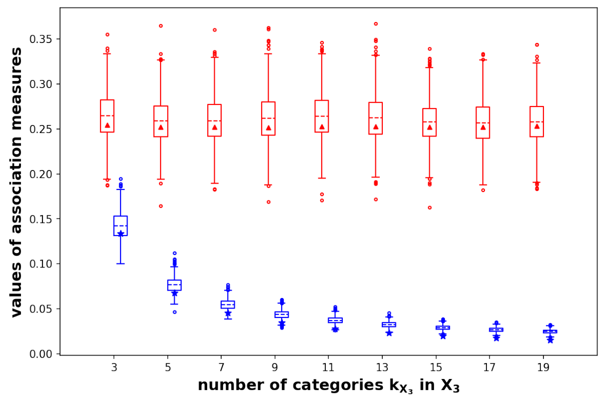


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

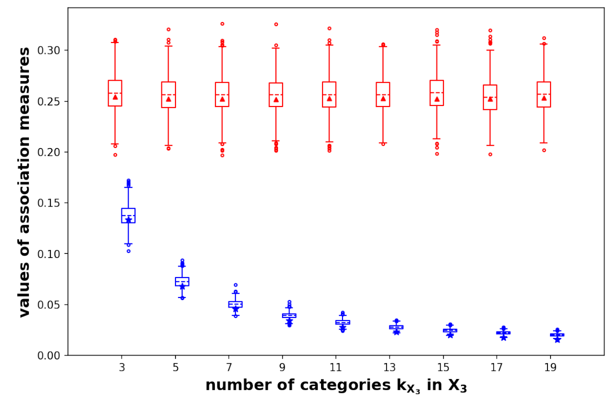


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

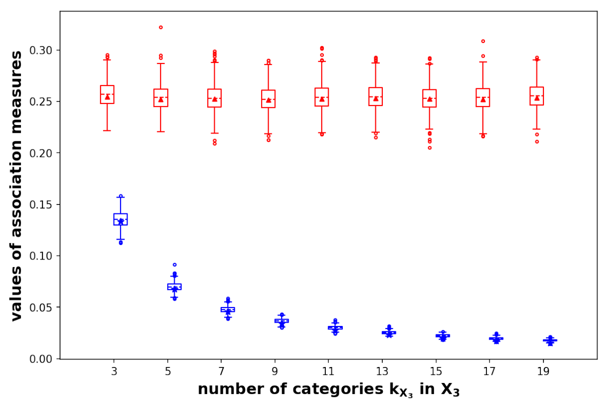
Figure J.3: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



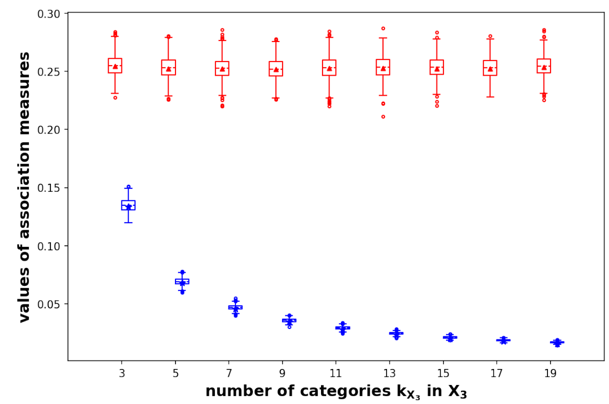
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

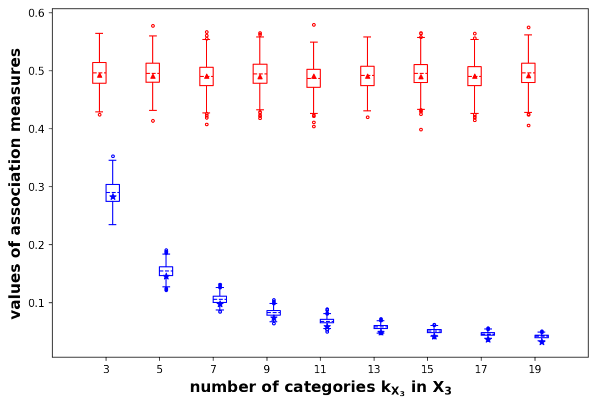


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

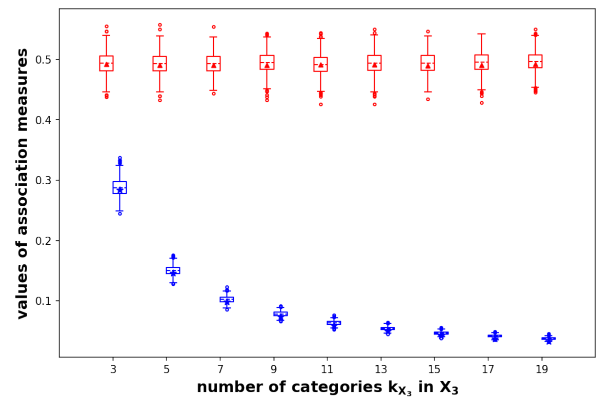


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

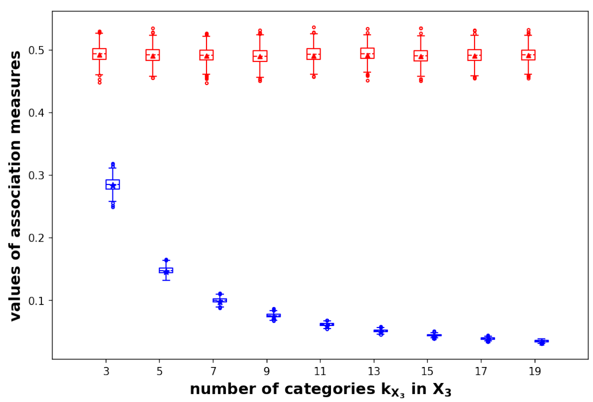
Figure J.4: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case



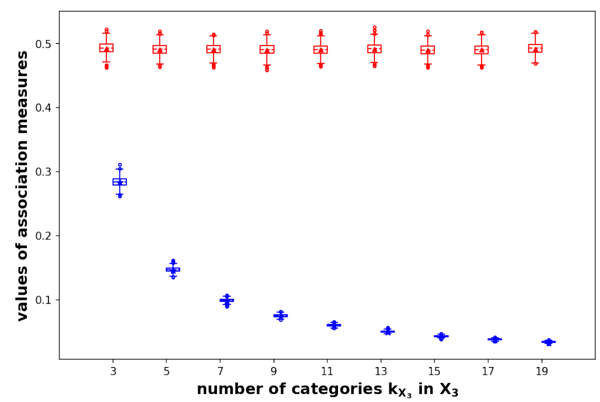
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

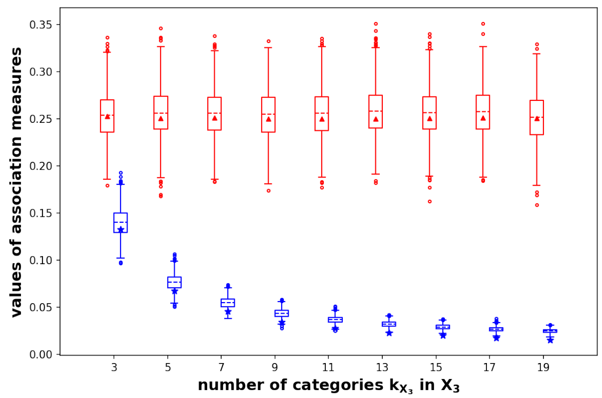


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

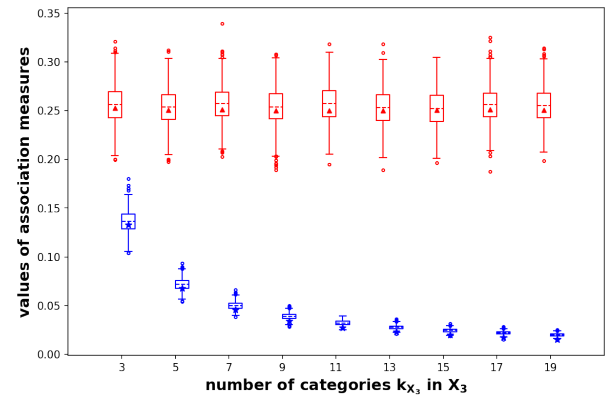


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

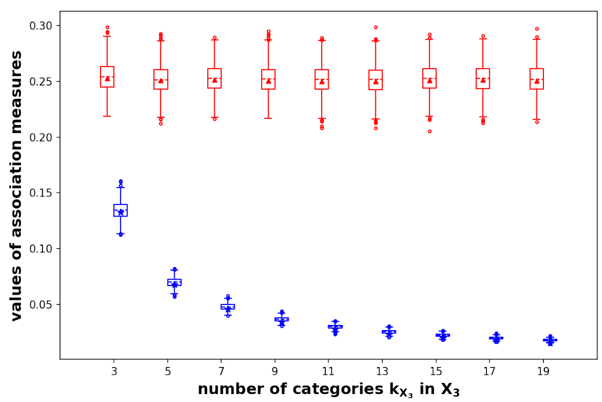
Figure J.5: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



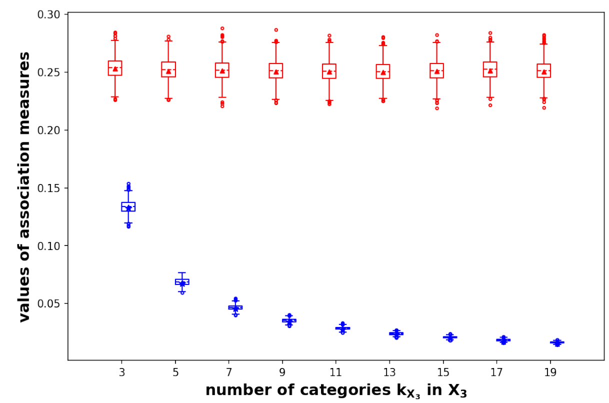
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

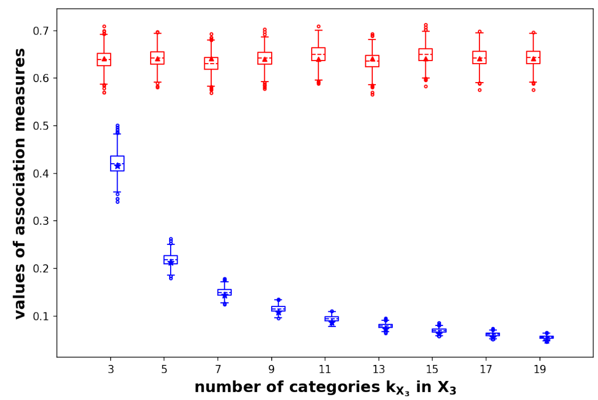


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

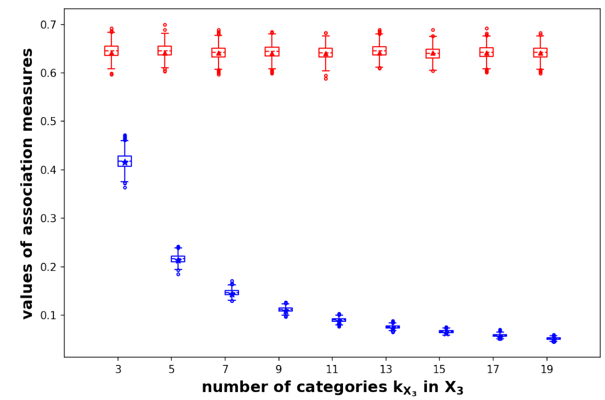


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

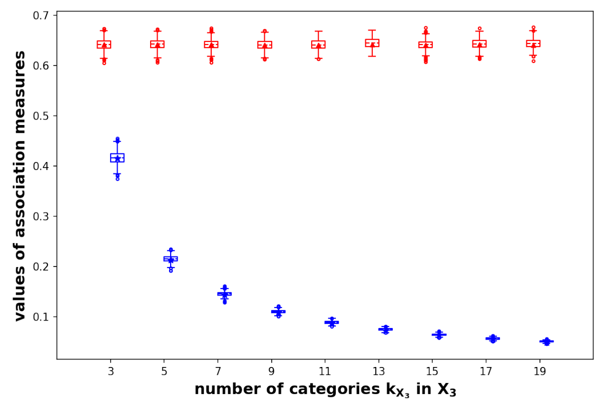
Figure J.6: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



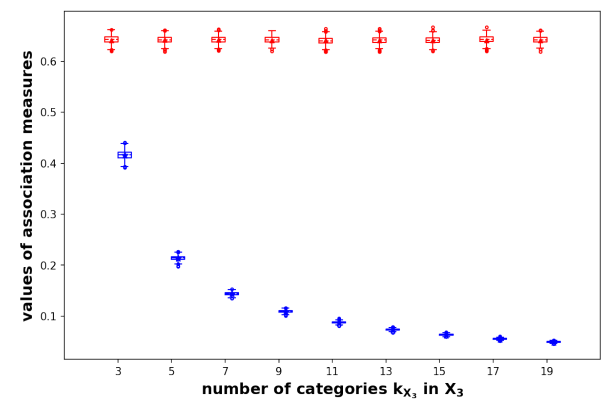
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

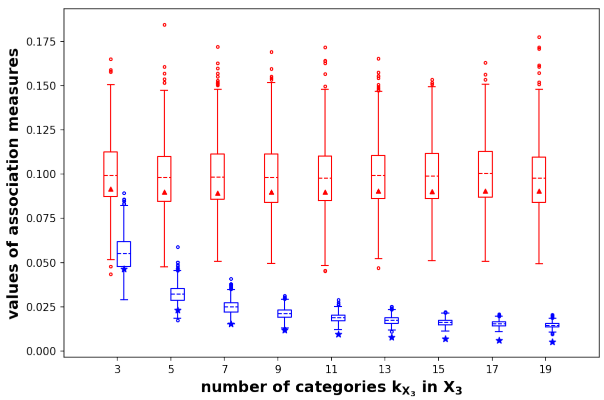


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

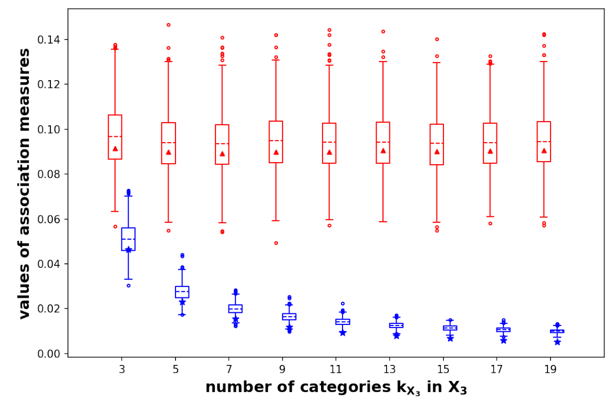


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

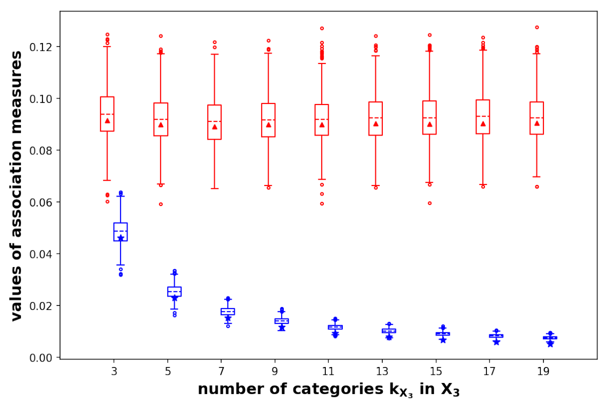
Figure J.7: The association measure and *Gray-Williams' index* for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



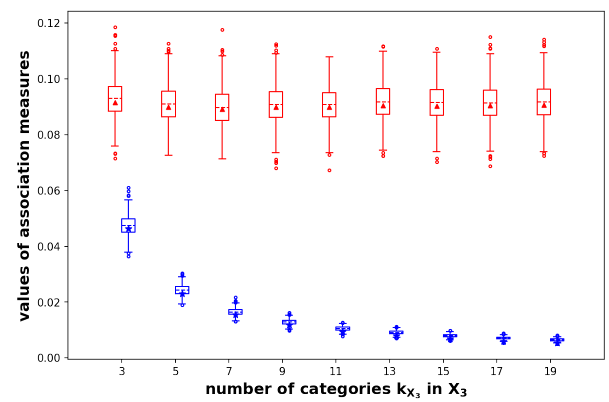
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

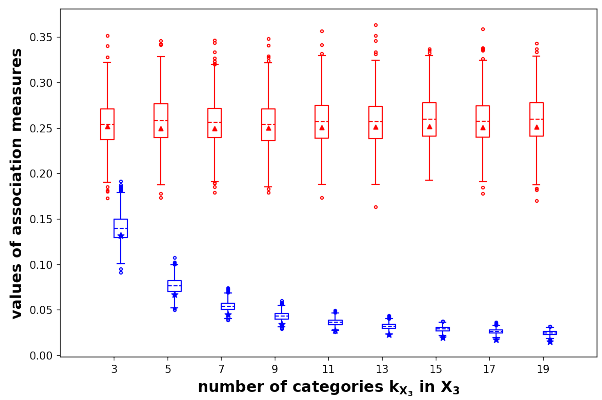


(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

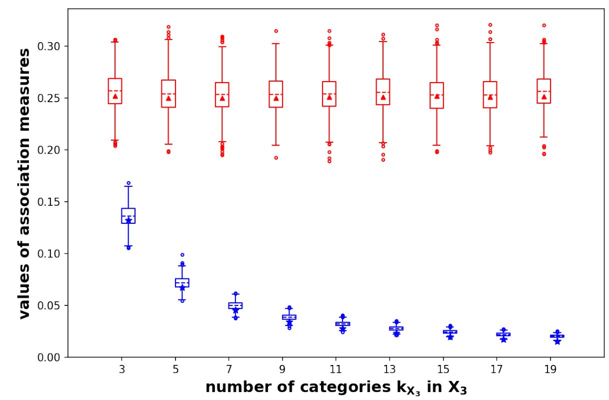


(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

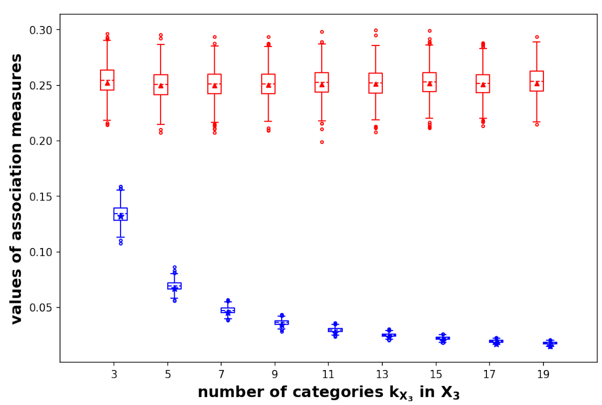
Figure J.8: The association measure and *Gray-Williams' index* for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case



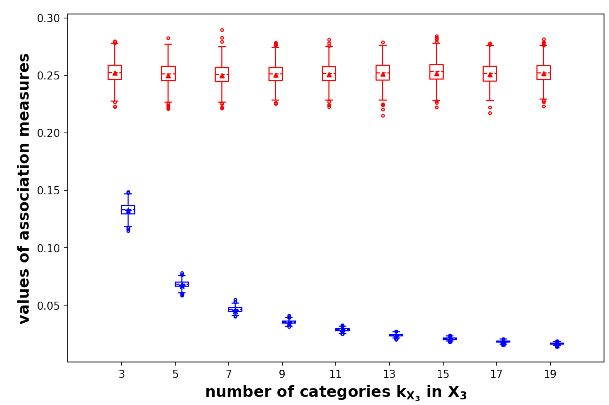
(a) True and estimated values of $\rho^2_{(X_1,X_2 \to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2 \to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$

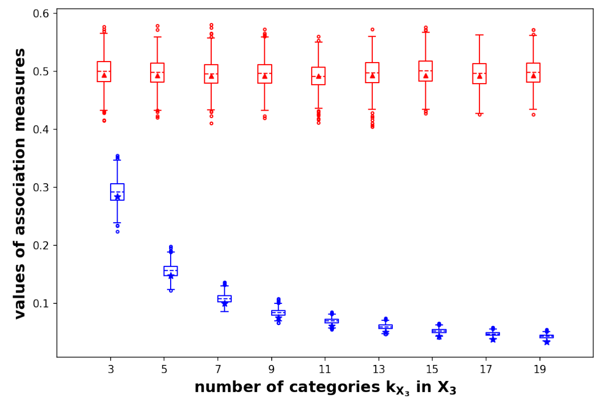


(c) True and estimated values of $\rho^2_{(X_1,X_2 \to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$

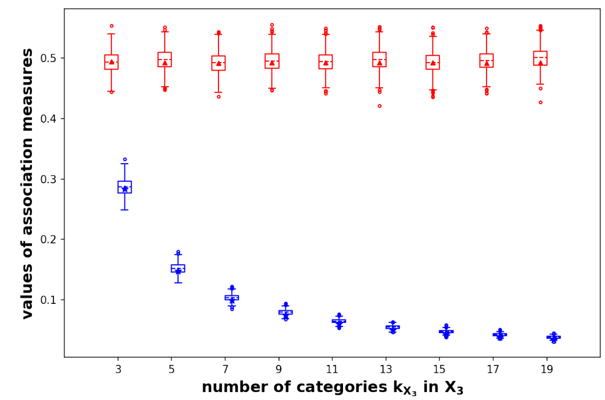


(d) True and estimated values of $\rho^2_{(X_1,X_2 \to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

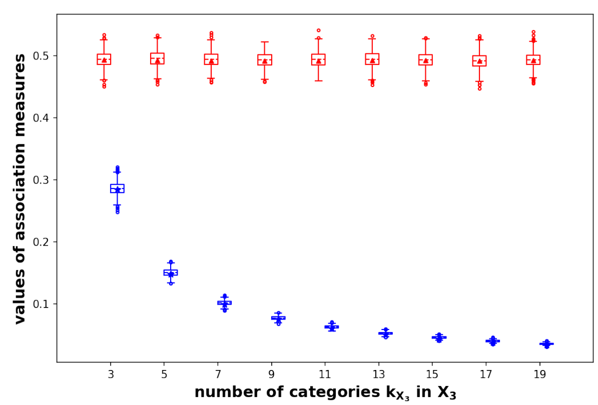
Figure J.9: The association measure and *Gray-Williams' index* for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$ in three-dimensional case



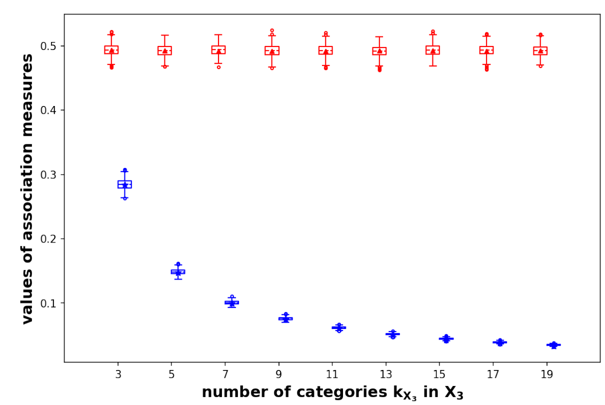
(a) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 855$



(b) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 1710$



(c) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 3420$



(d) True and estimated values of $\rho^2_{(X_1,X_2\to X_3)}$ (red) and $\tau^{GW}_{X_3}$ (blue) for $n = 6840$

Figure J.10: The overall association measure and *Gray-Williams' index* for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case

J.2 Scenario 2

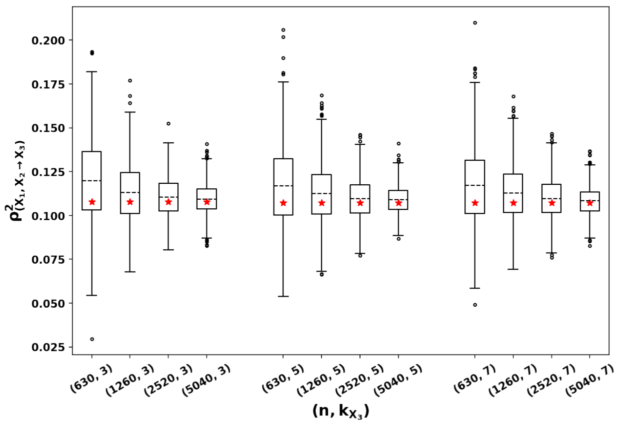
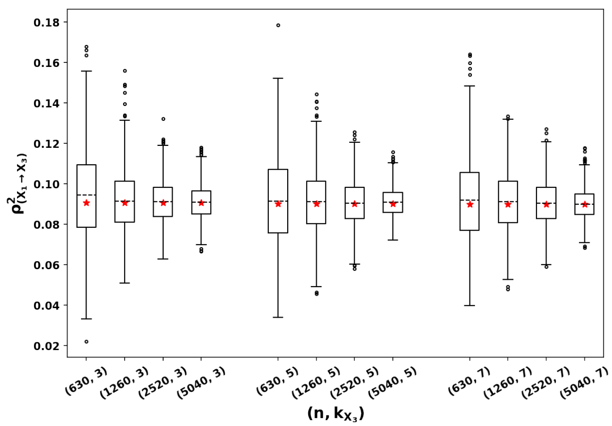
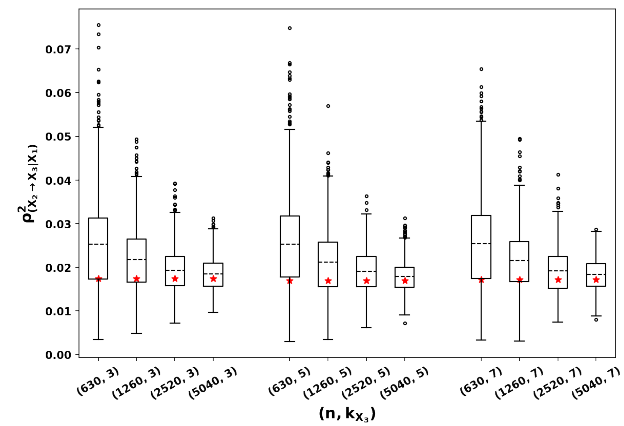


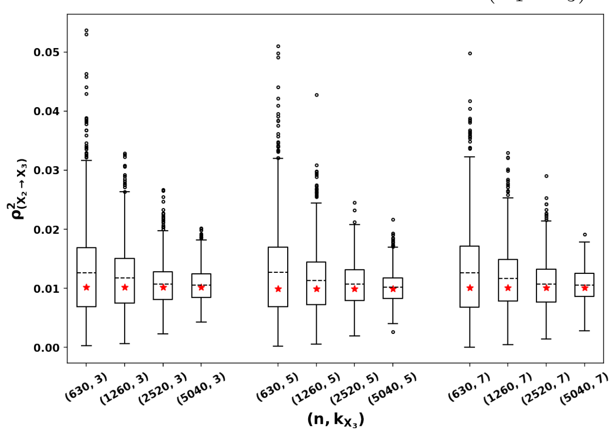
Figure J.11: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and weak association $corr(X_3, X_1) = 0.3$ in three-dimensional case



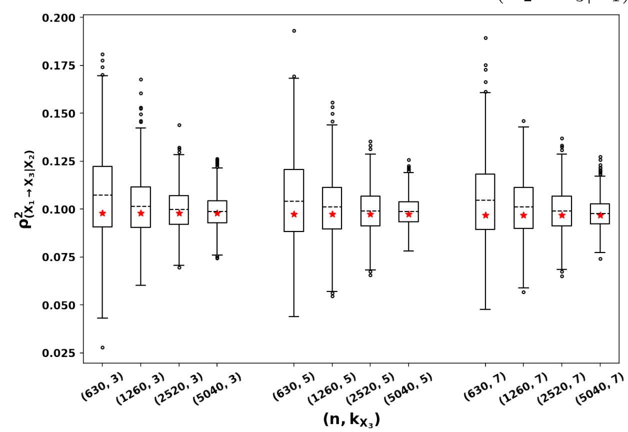
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.12: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

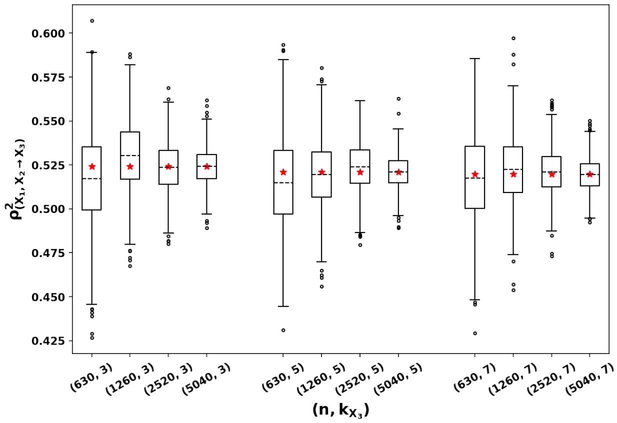
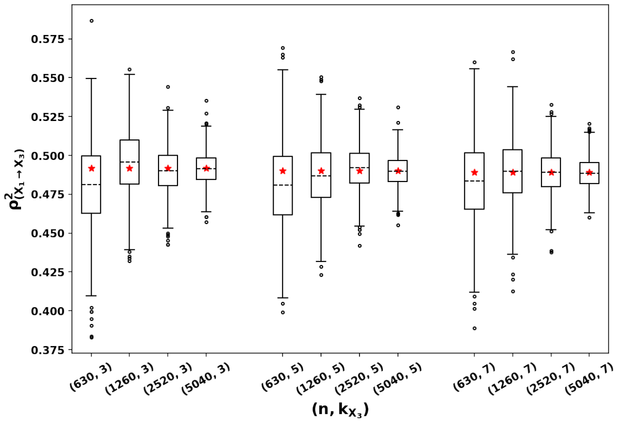
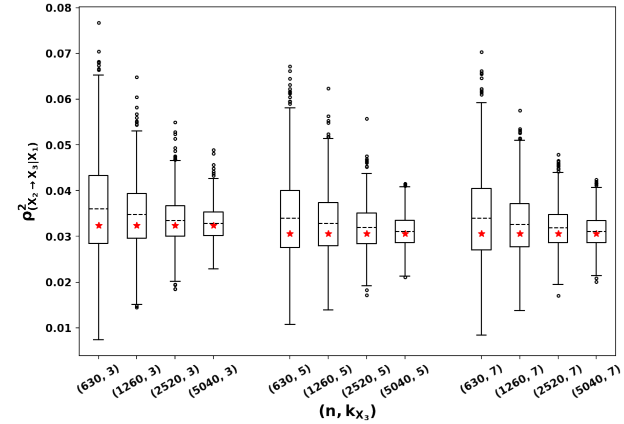


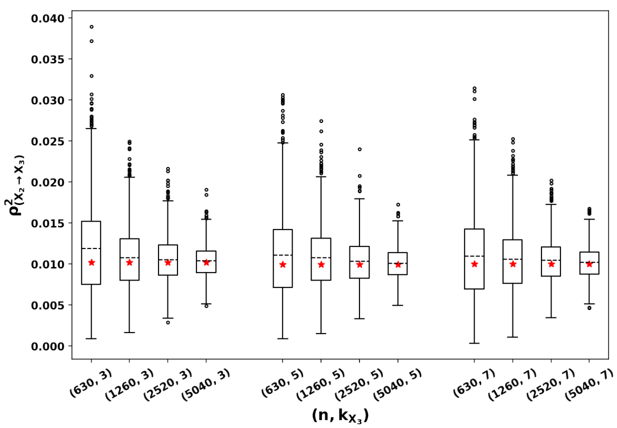
Figure J.13: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



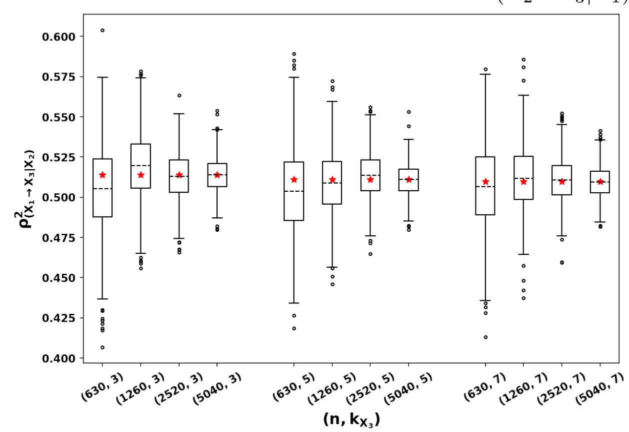
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.14: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

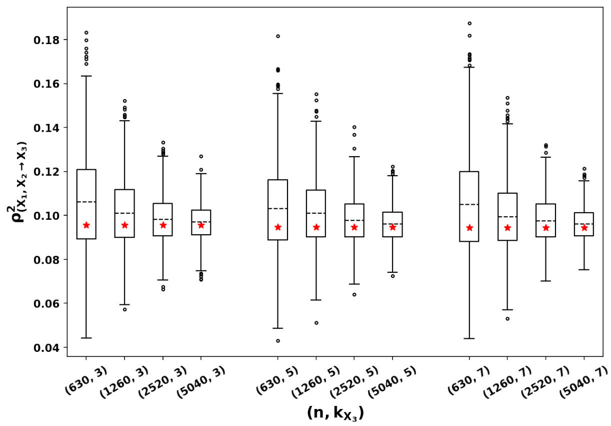
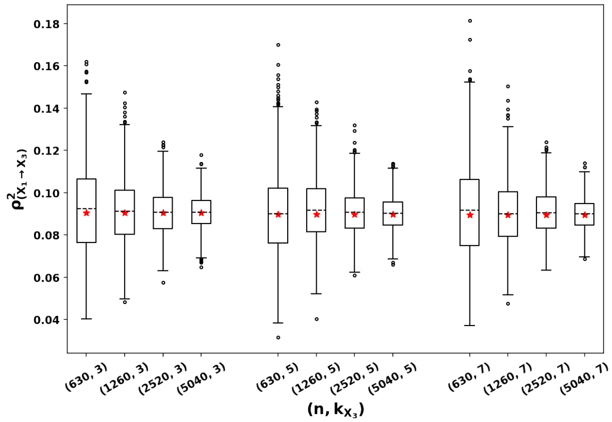
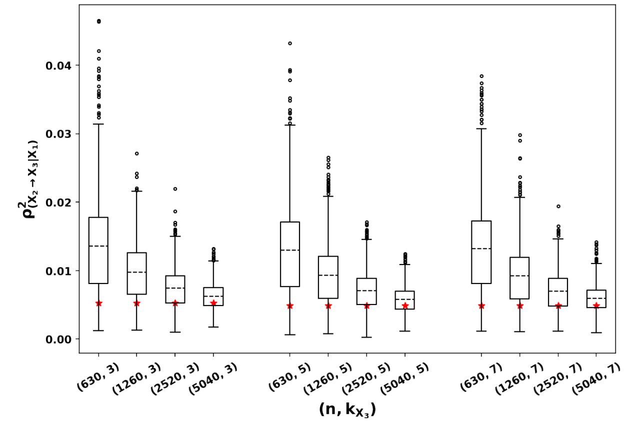


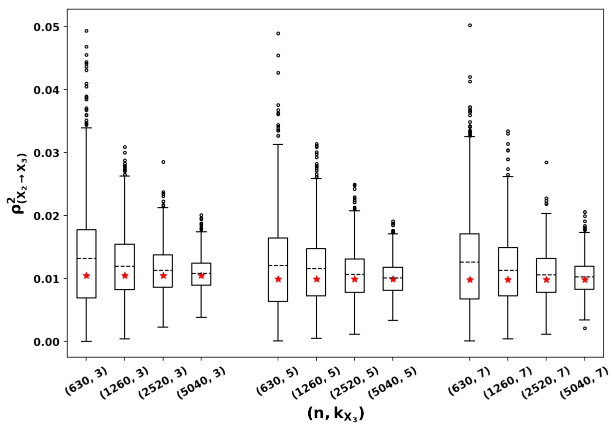
Figure J.15: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



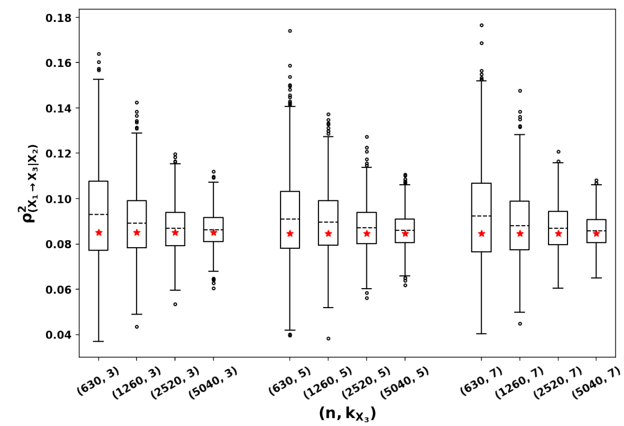
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.16: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

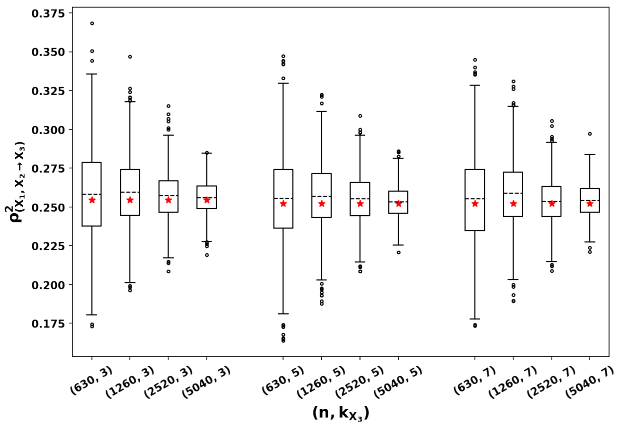
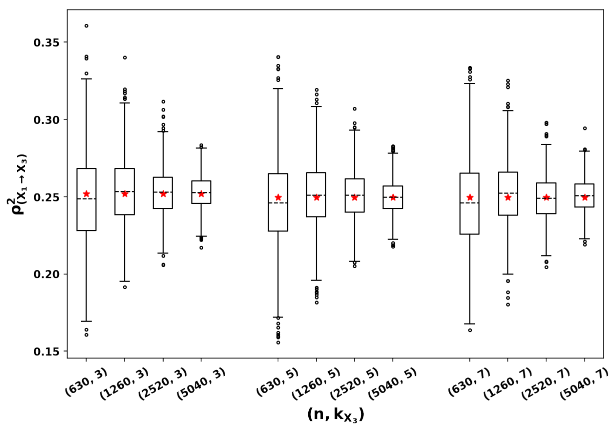
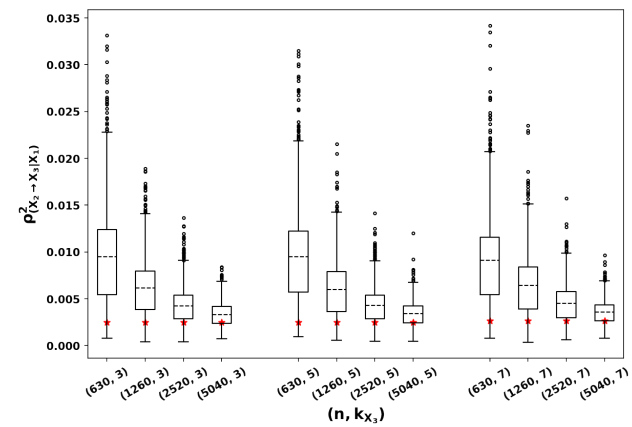


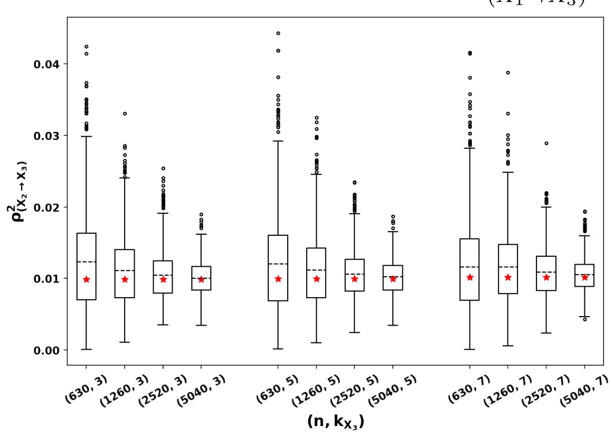
Figure J.17: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case



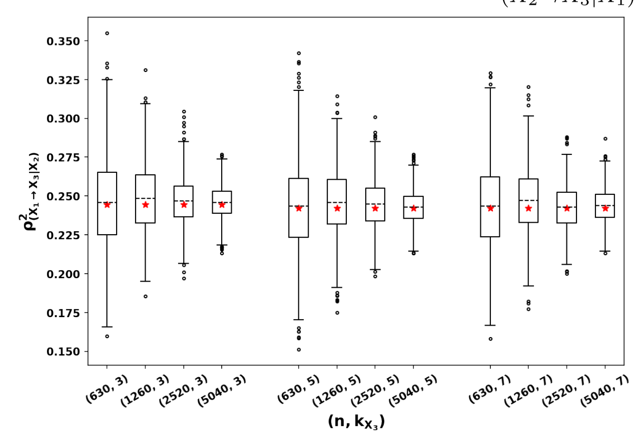
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.18: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case

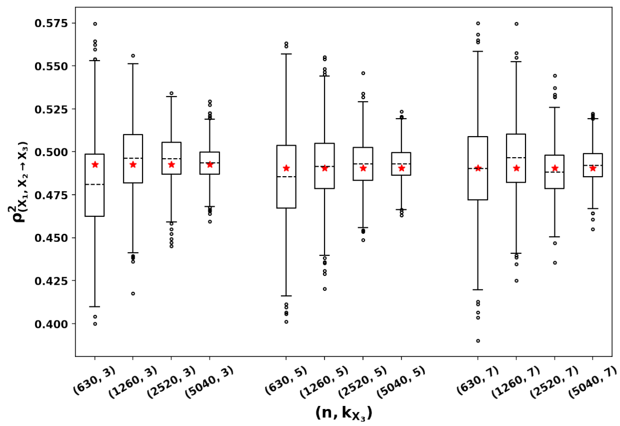
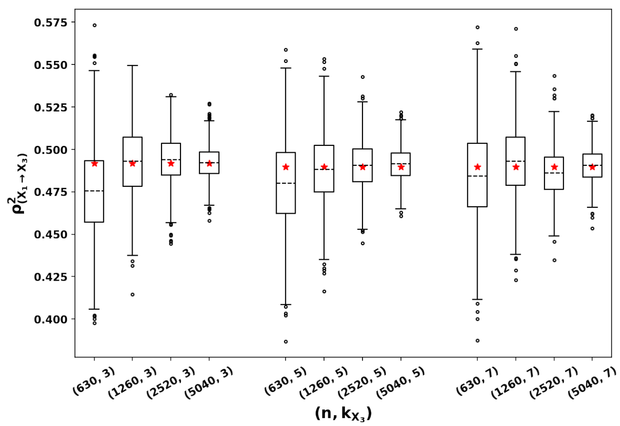
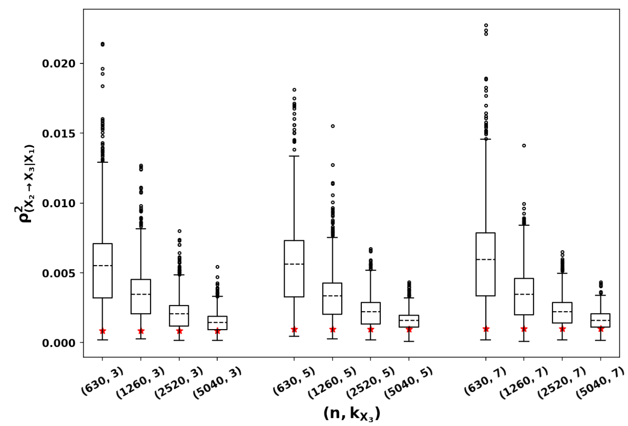


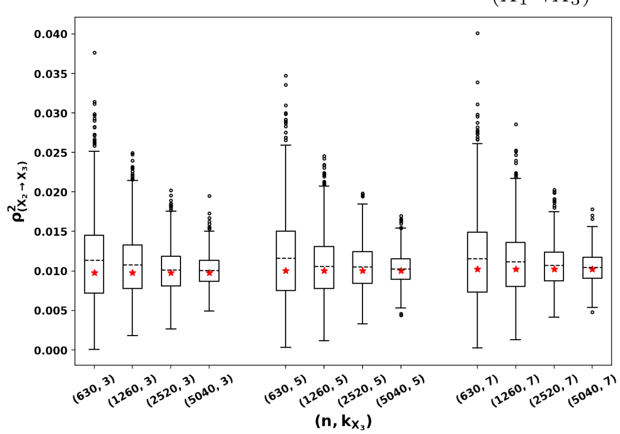
Figure J.19: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



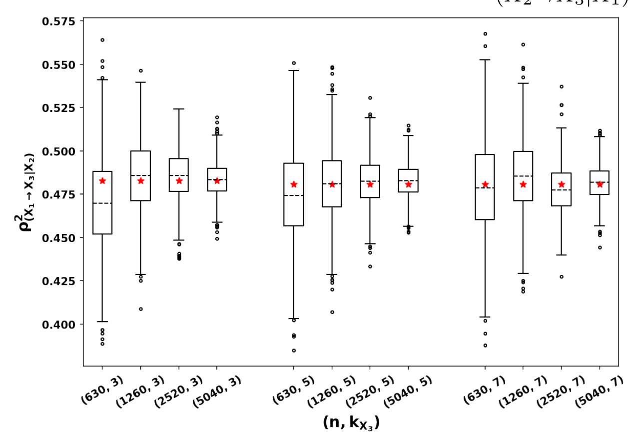
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.20: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

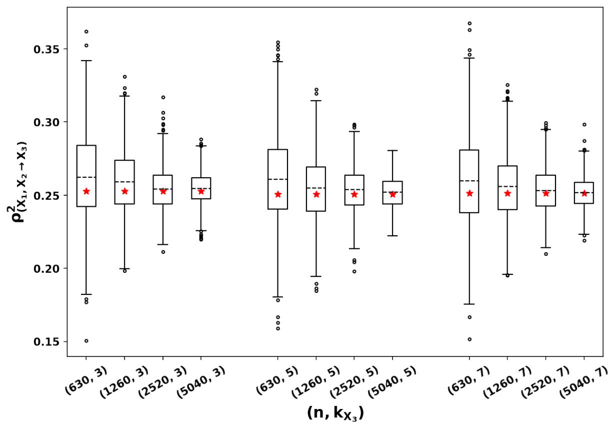
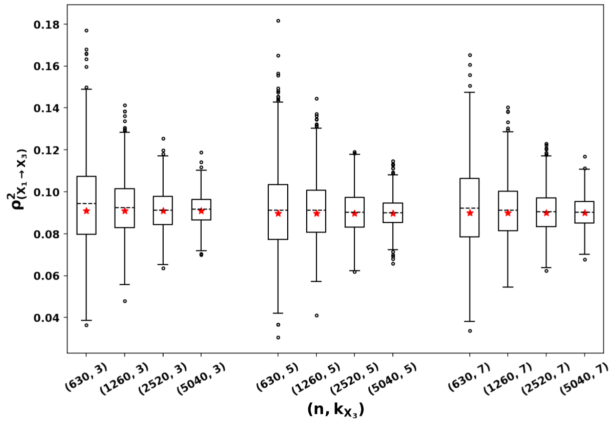
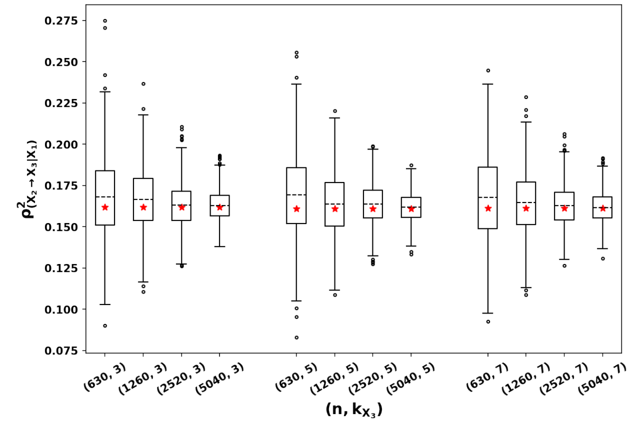


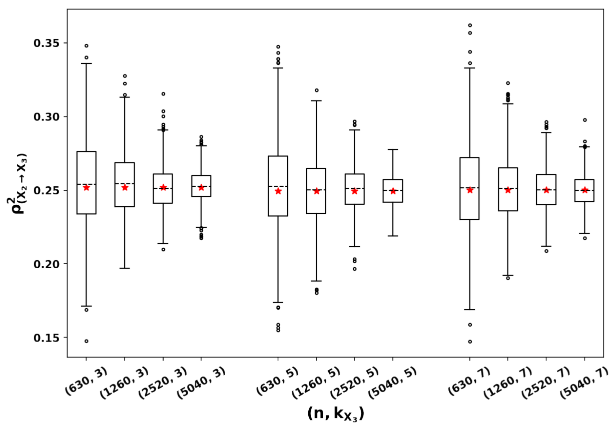
Figure J.21: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



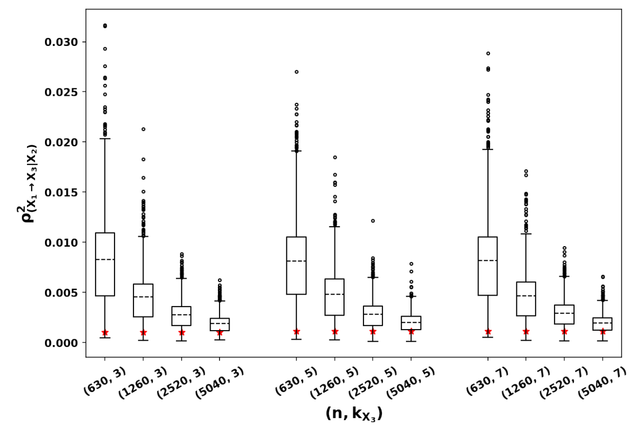
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.22: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

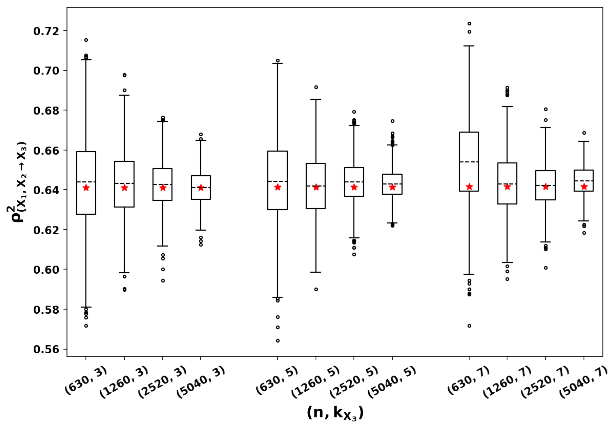
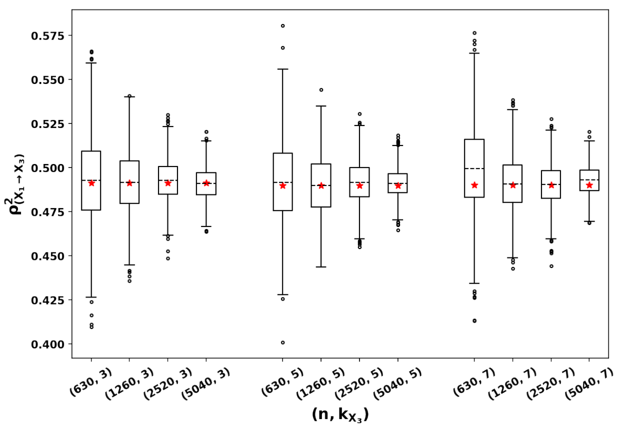
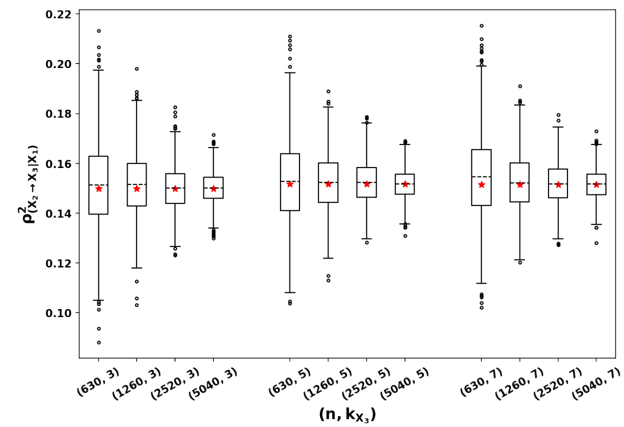


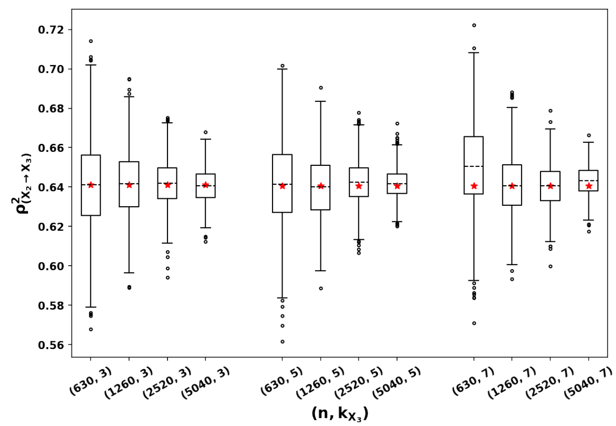
Figure J.23: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



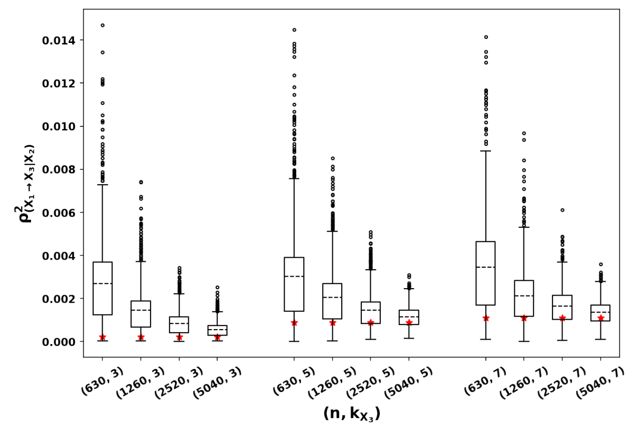
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.24: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

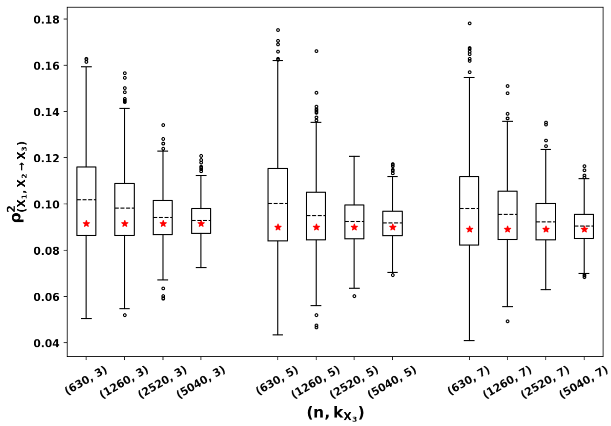
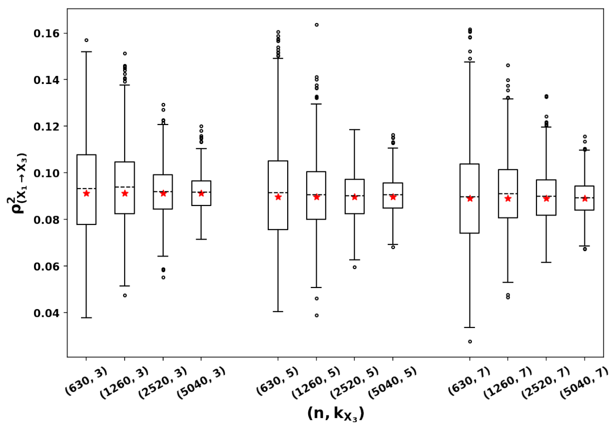
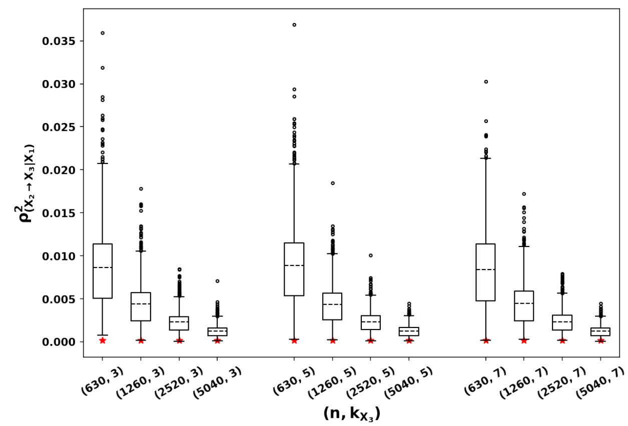


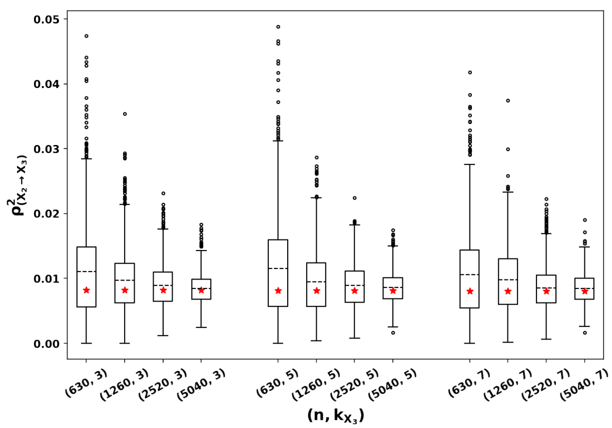
Figure J.25: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case



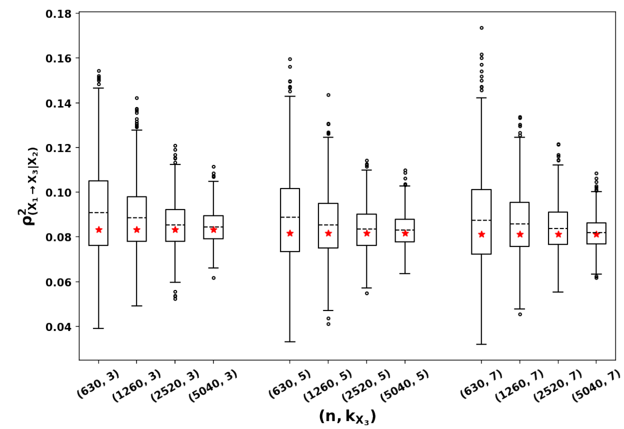
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.26: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case

J.3 Scenario 3

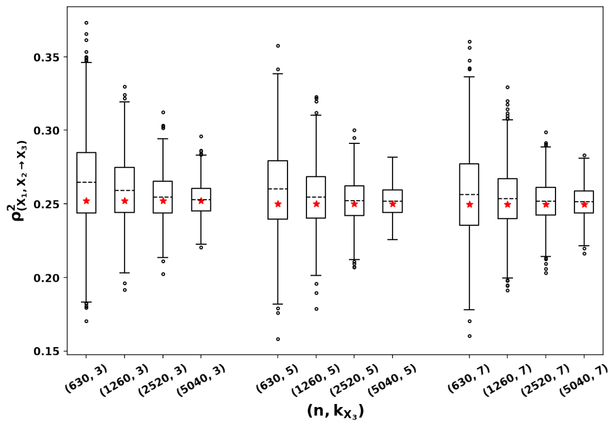
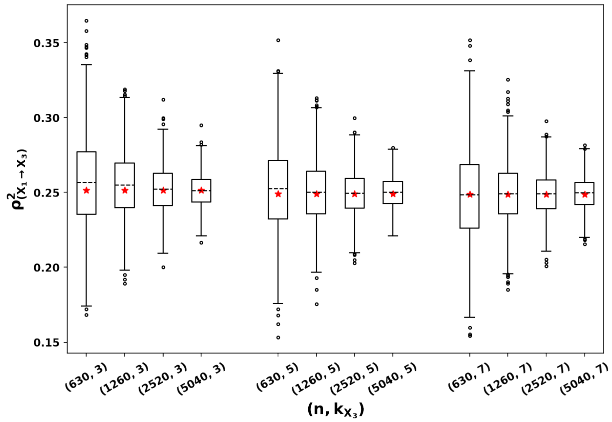
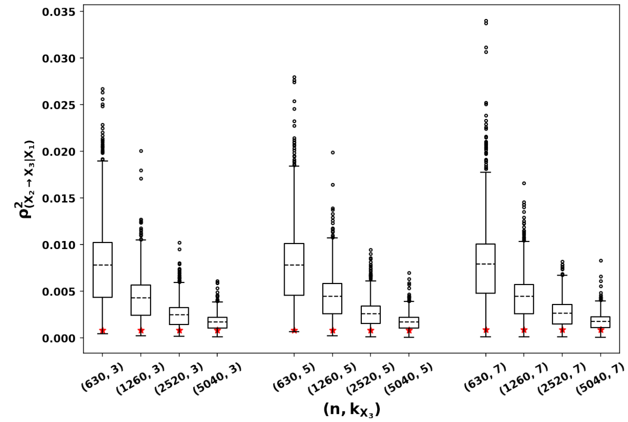


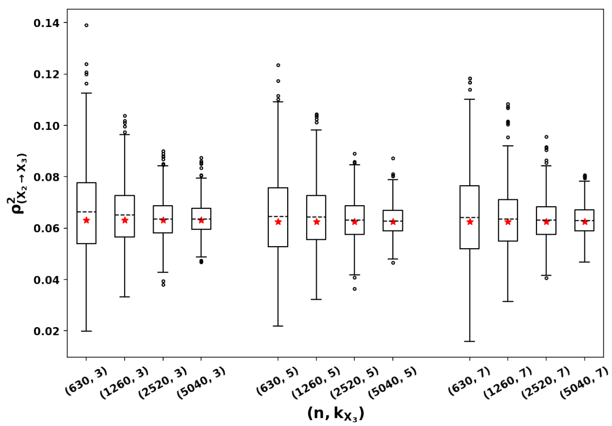
Figure J.27: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$ in three-dimensional case



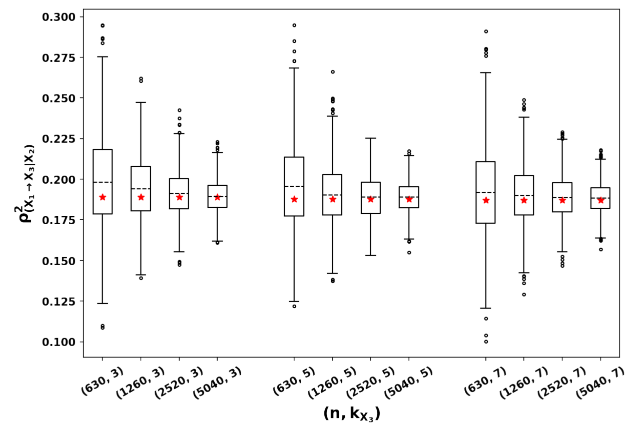
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.28: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and moderate association $|\phi| = |corr(X_3, X_1)| = 0.5$ in three-dimensional case

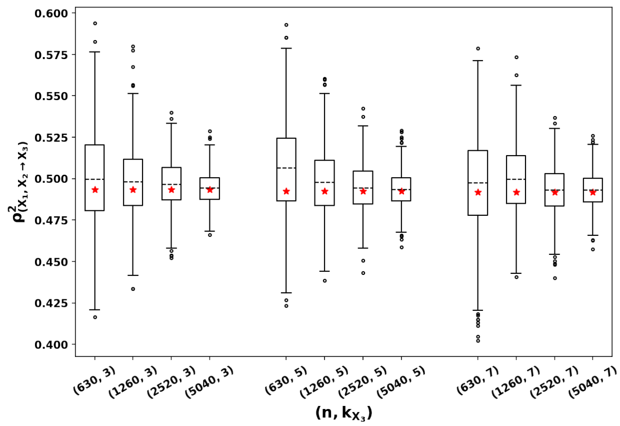
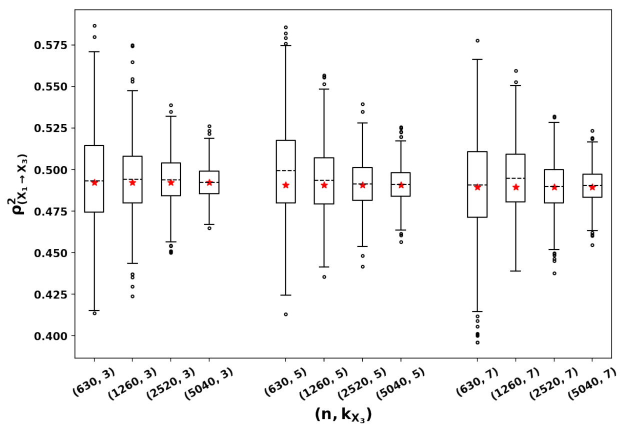
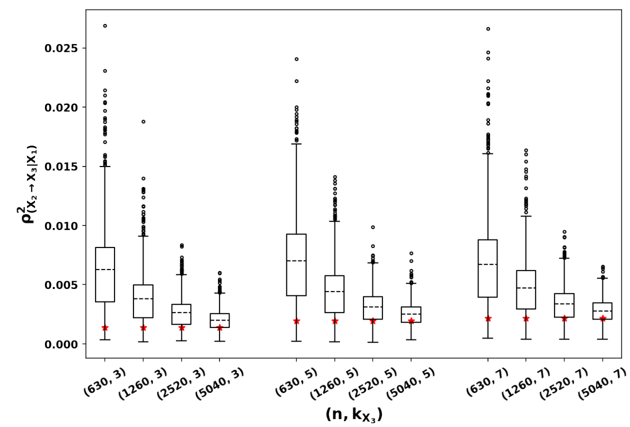


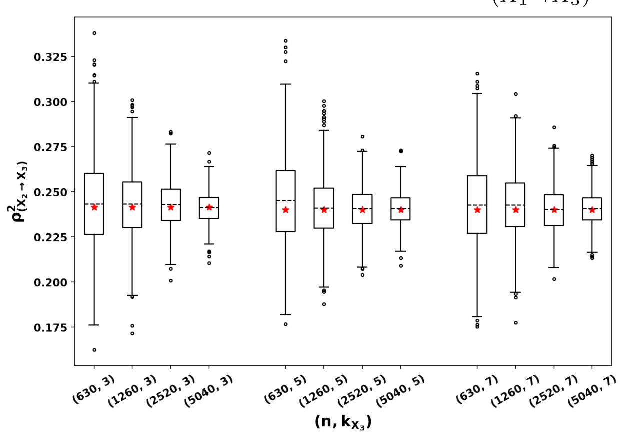
Figure J.29: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case



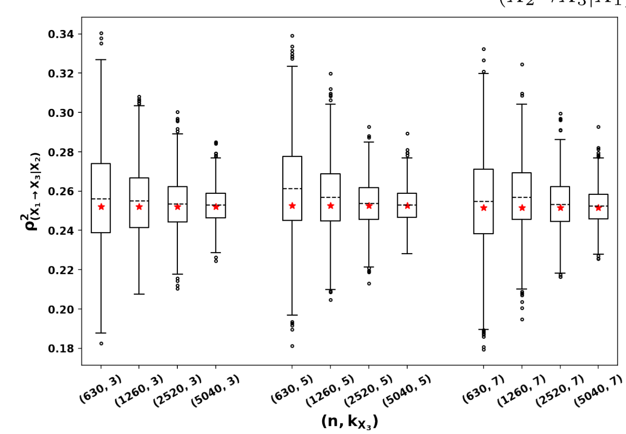
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.30: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case

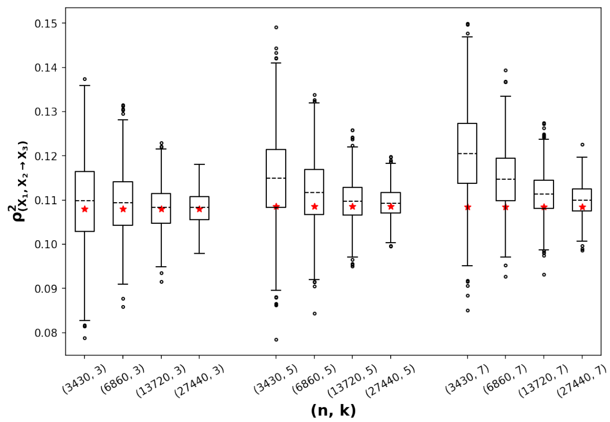
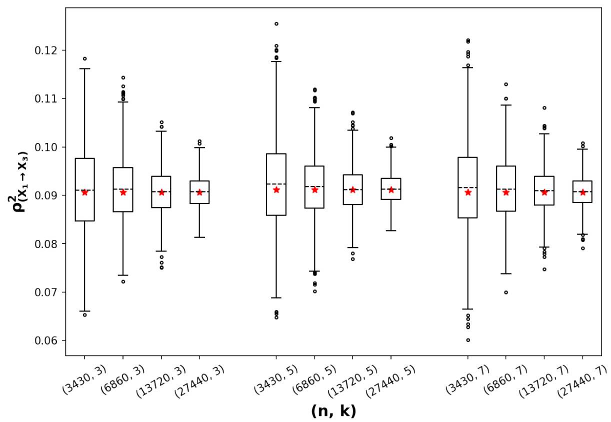
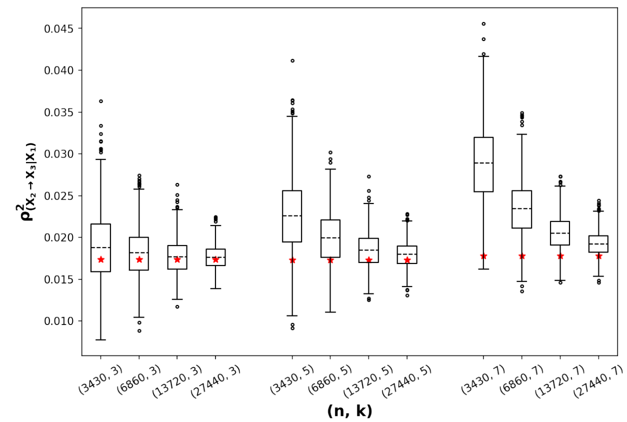


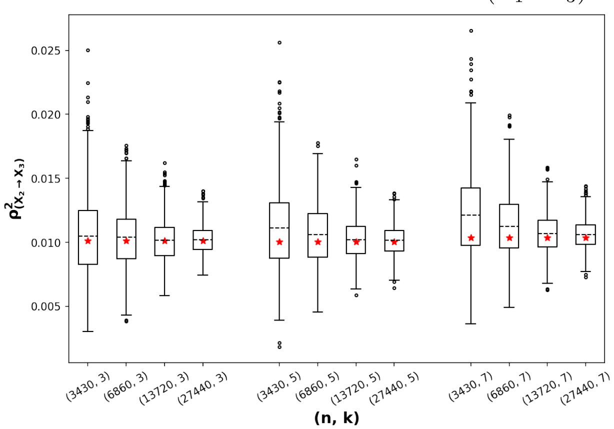
Figure J.31: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



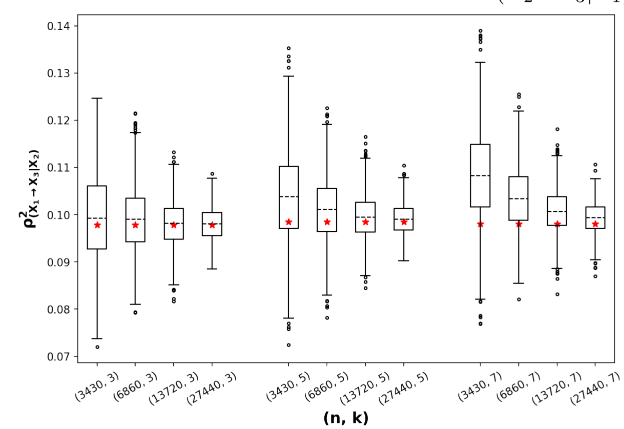
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.32: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

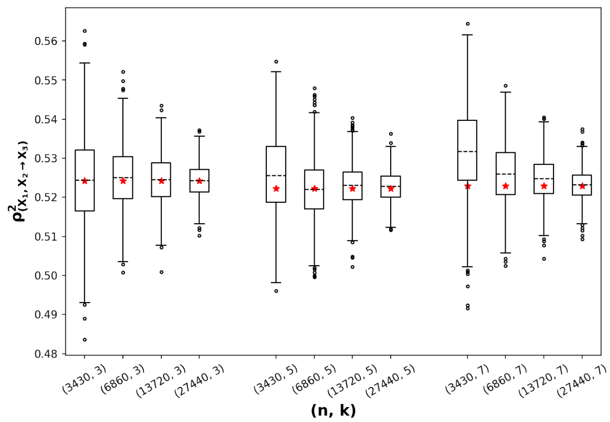
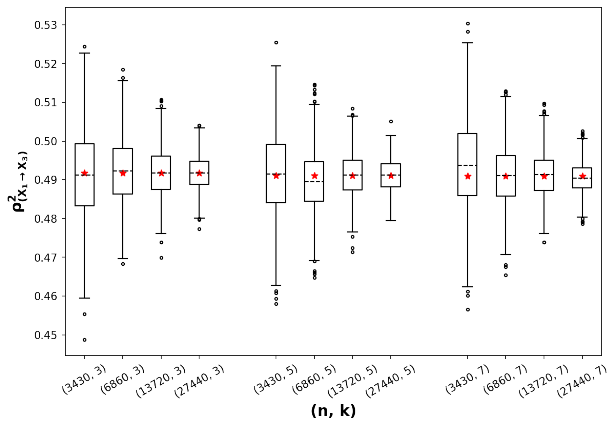
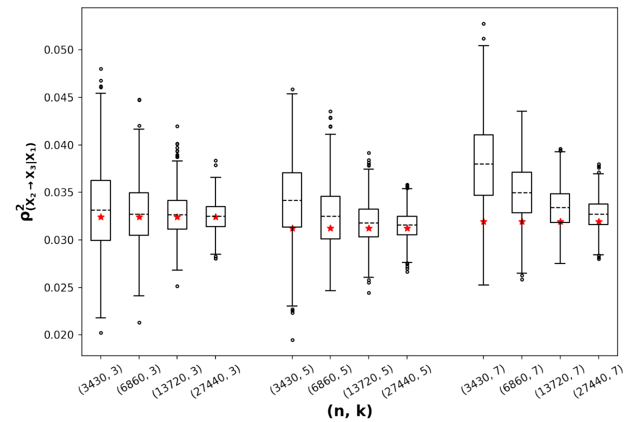


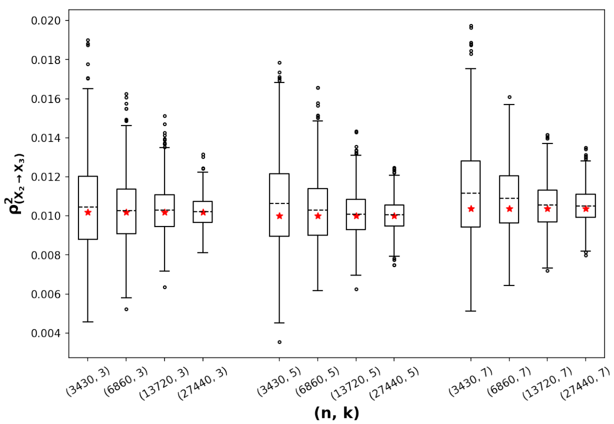
Figure J.33: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



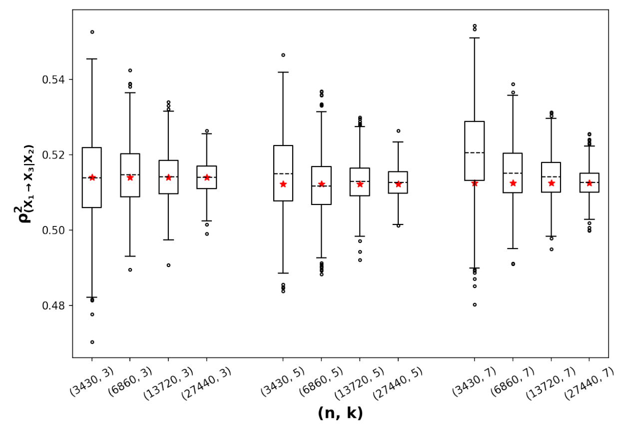
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.34: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

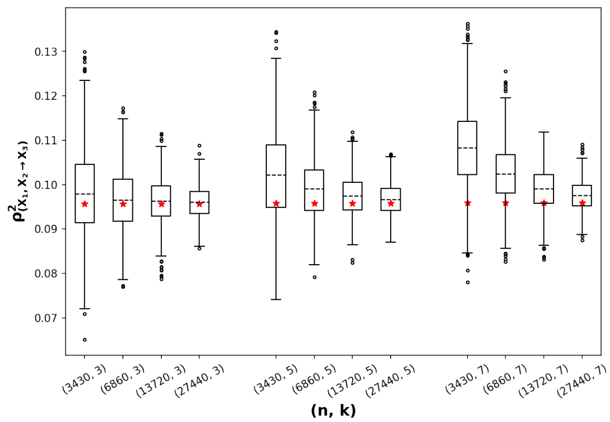
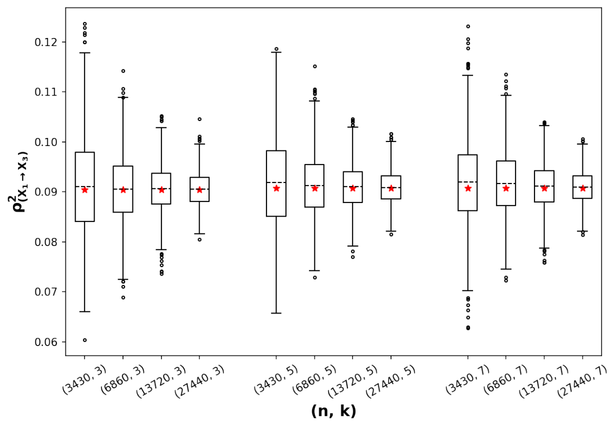
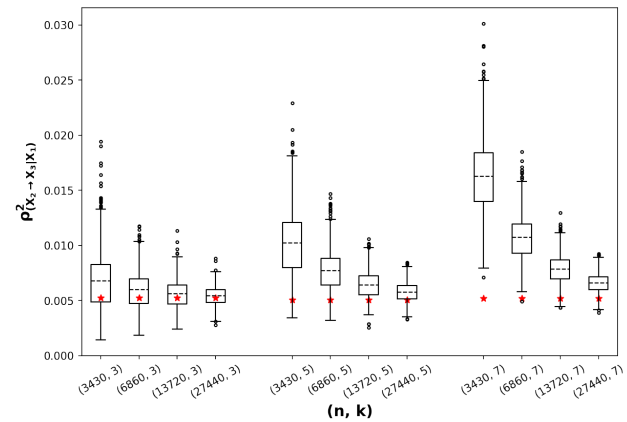


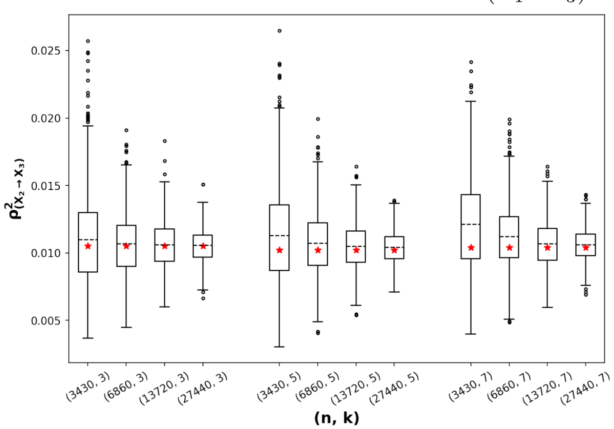
Figure J.35: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



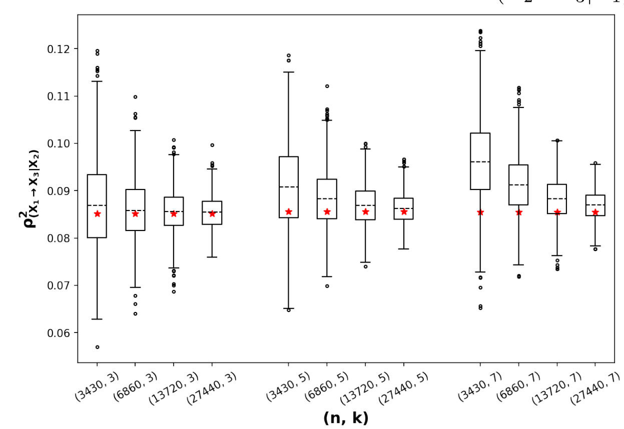
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.36: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

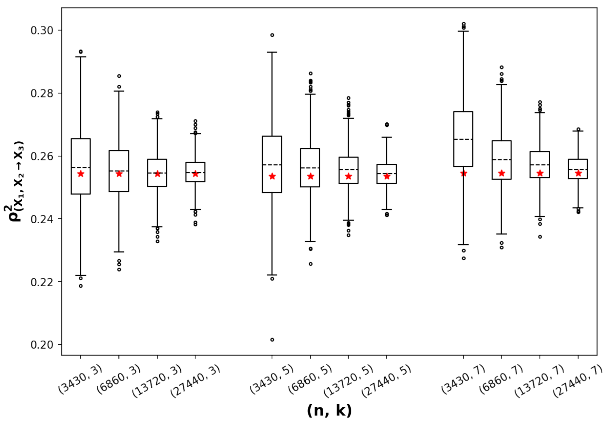
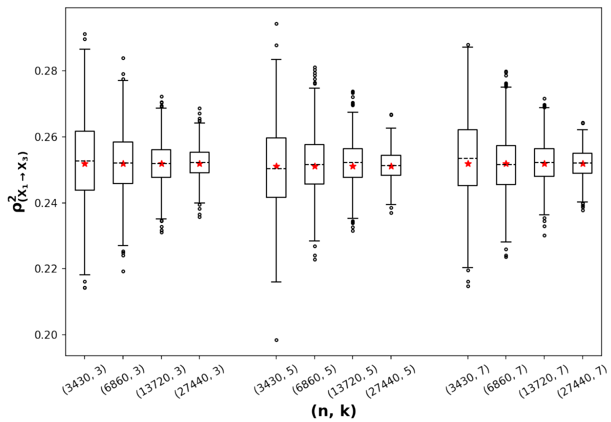
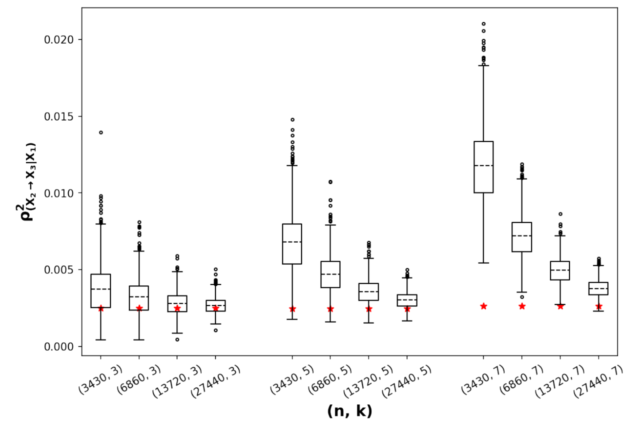


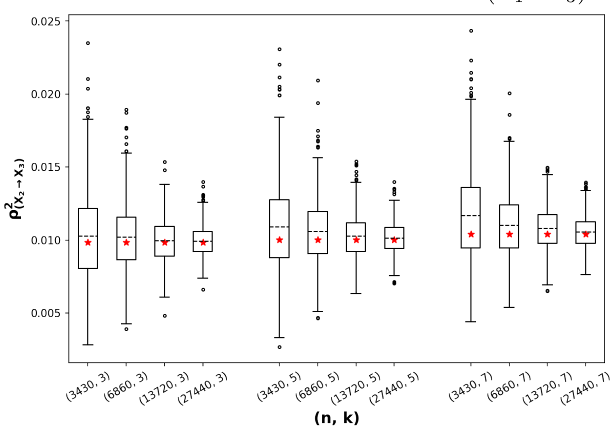
Figure J.37: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case



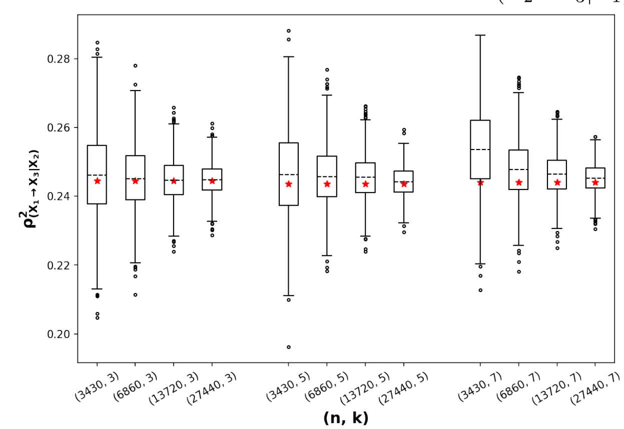
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.38: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case

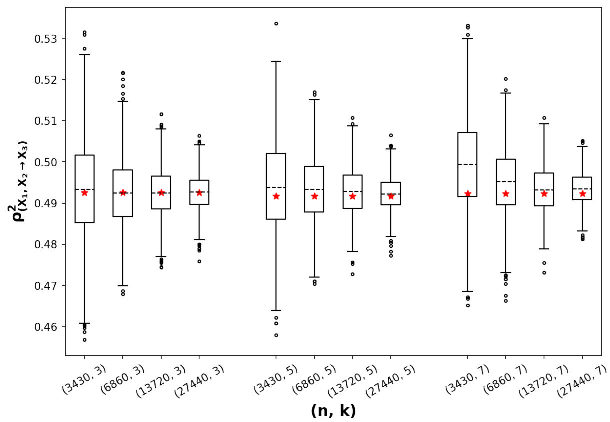
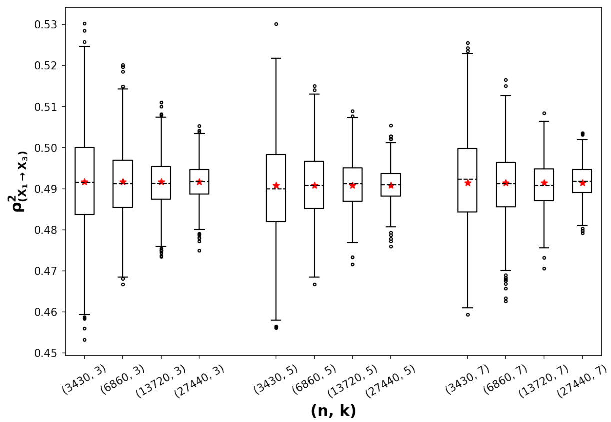
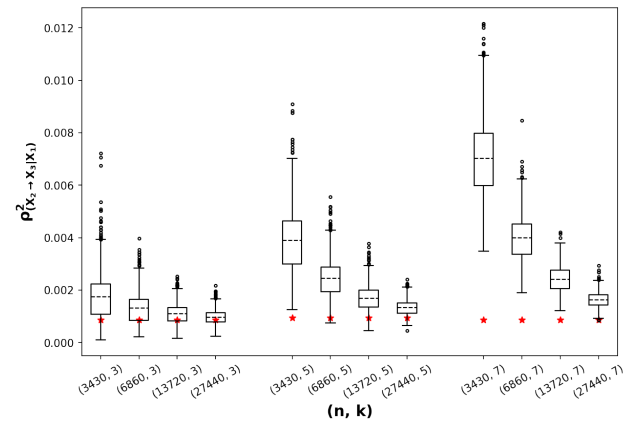


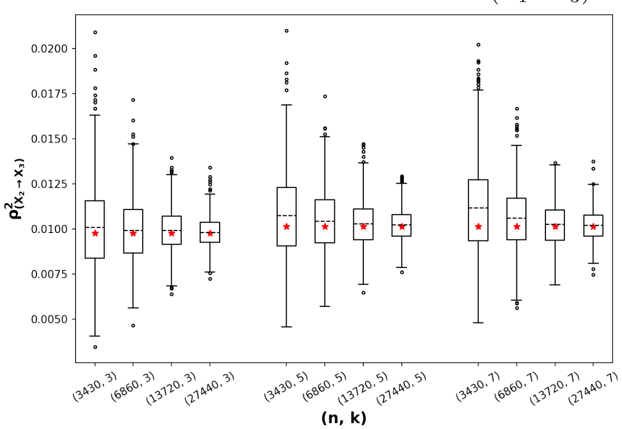
Figure J.39: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



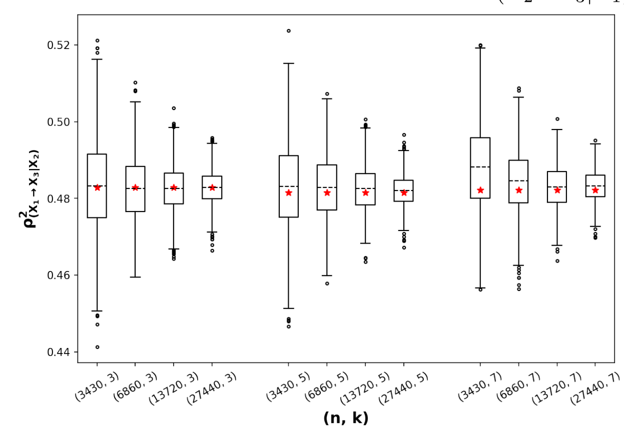
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.40: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

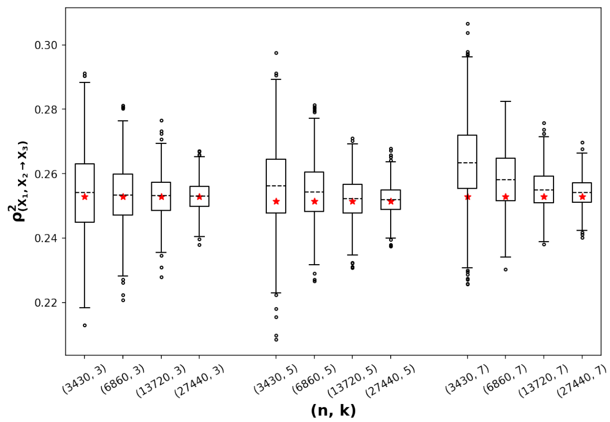
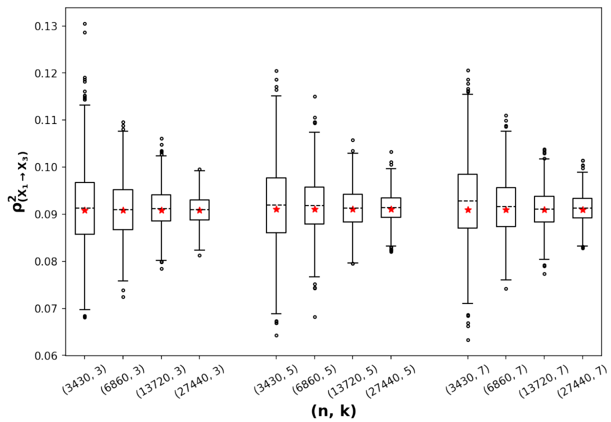
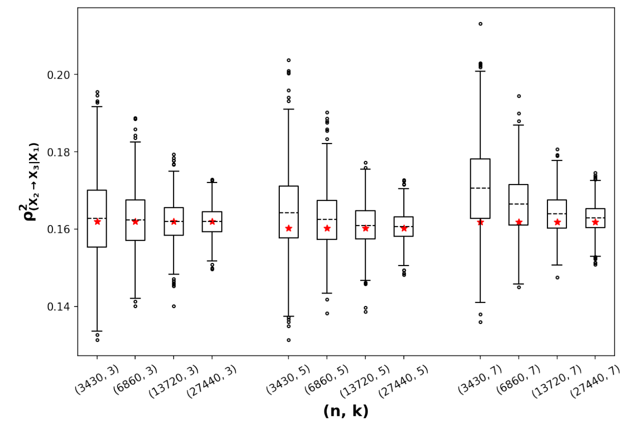


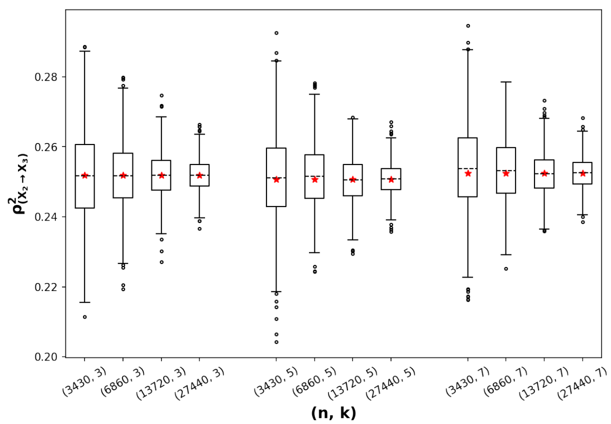
Figure J.41: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



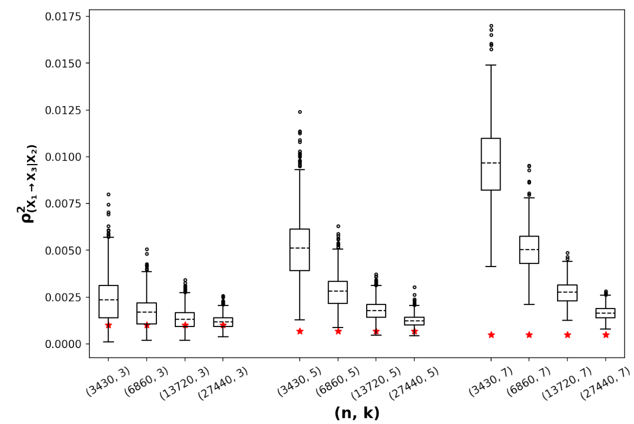
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.42: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

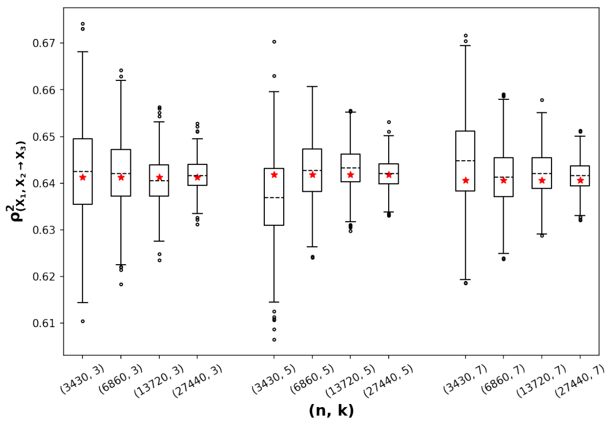
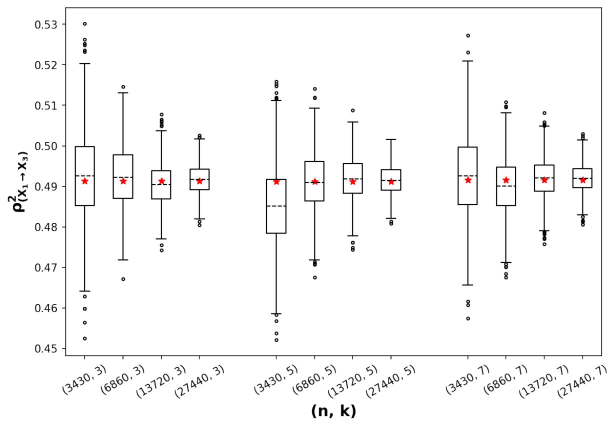
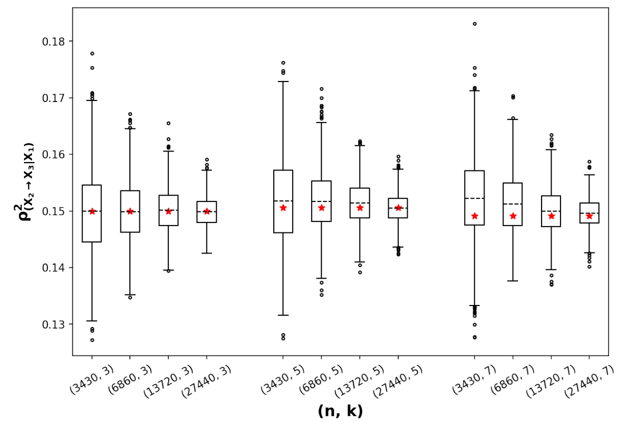


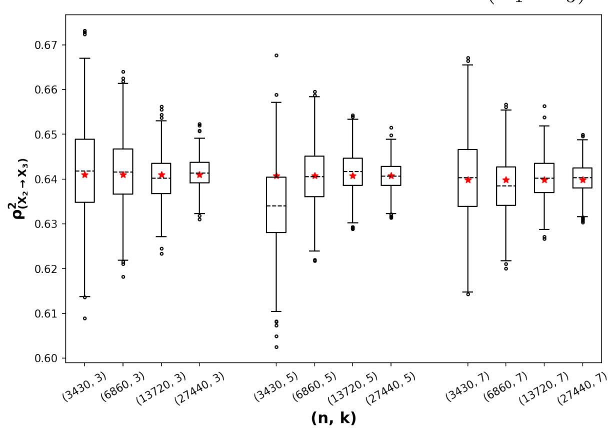
Figure J.43: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



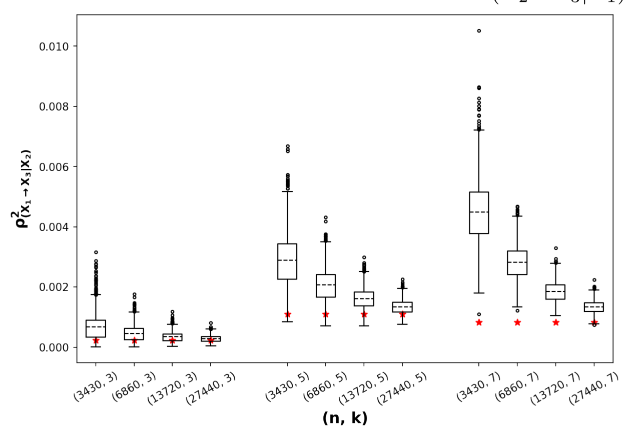
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.44: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

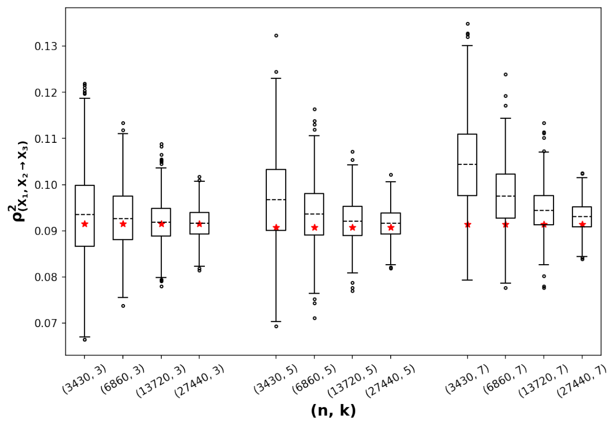
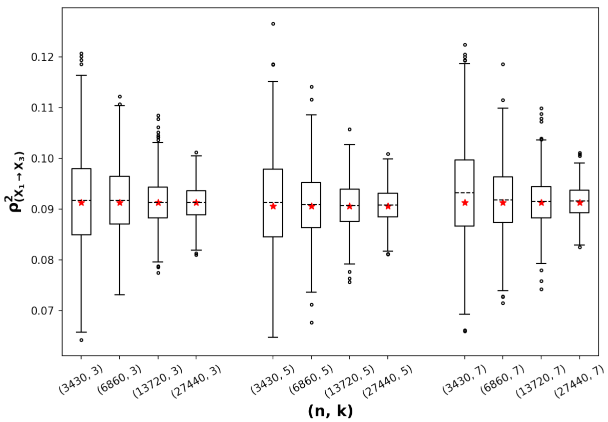
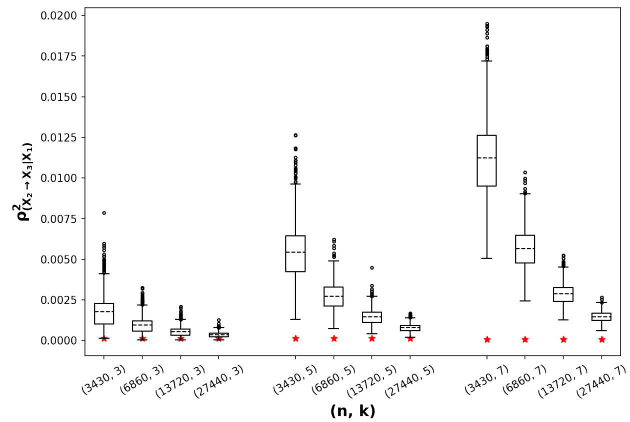


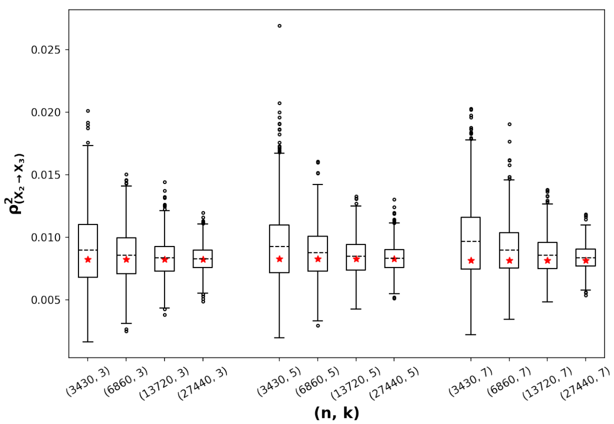
Figure J.45: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case



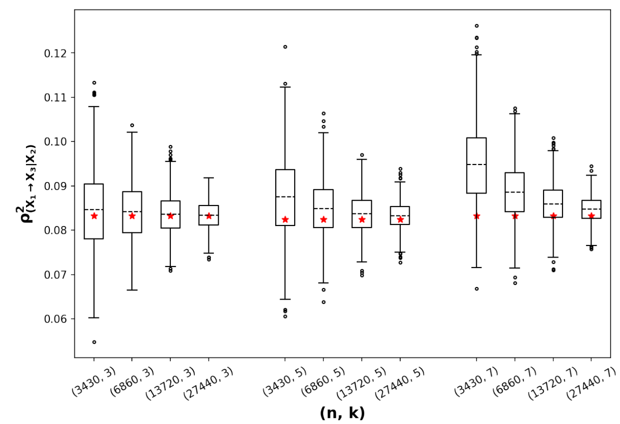
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.46: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case

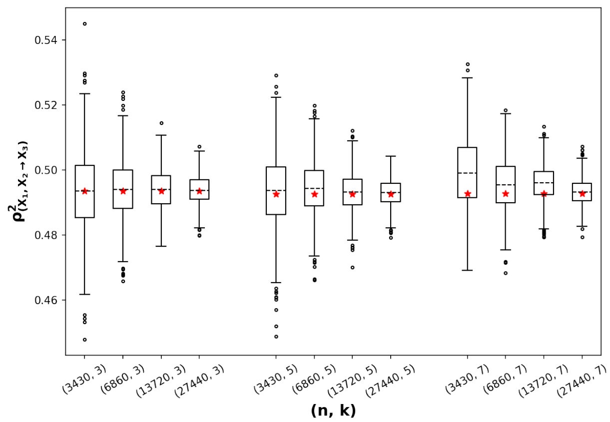
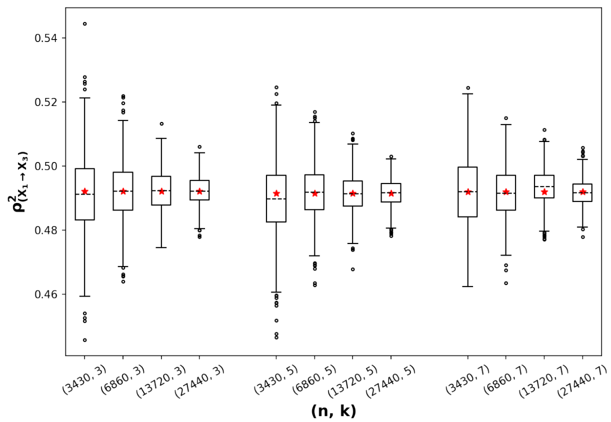
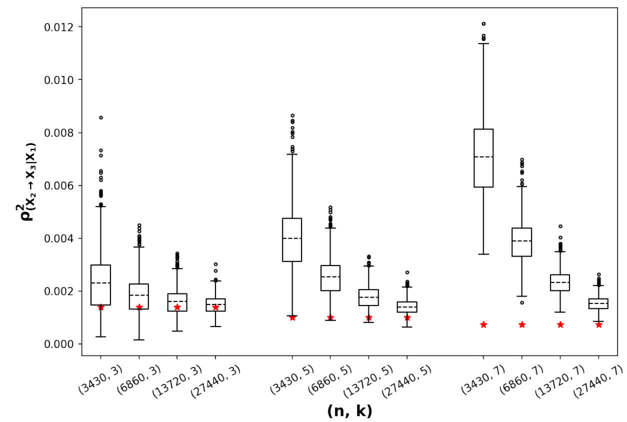


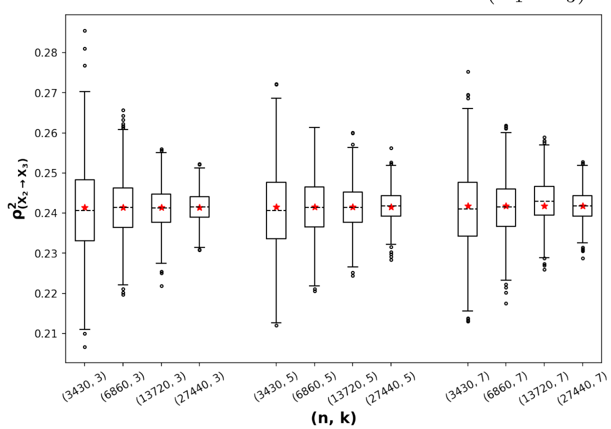
Figure J.47: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case



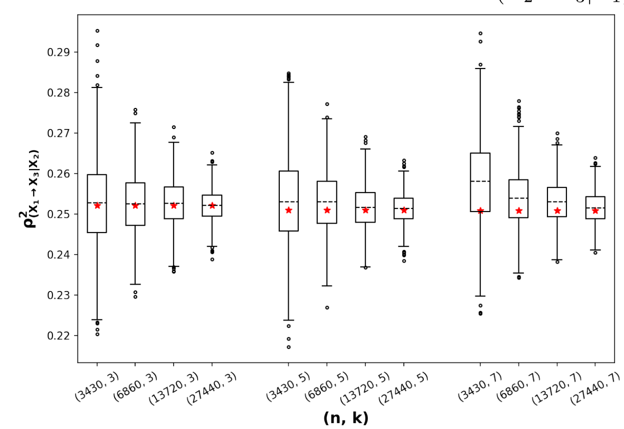
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.48: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case

J.4 Scenario 4

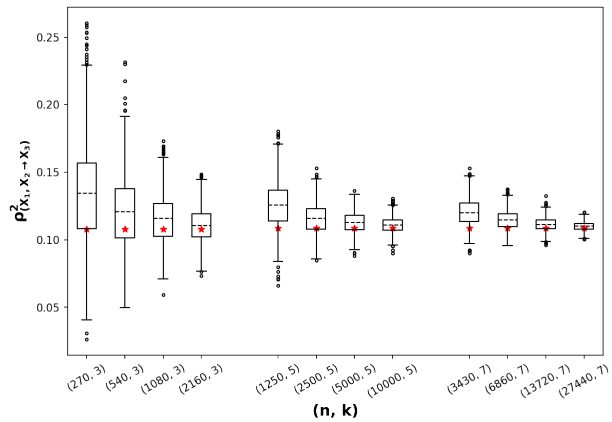
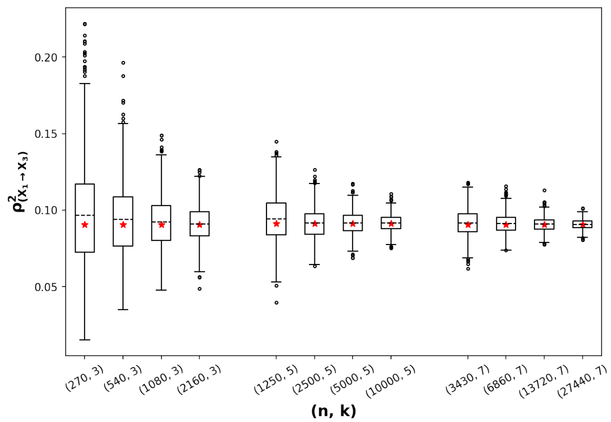
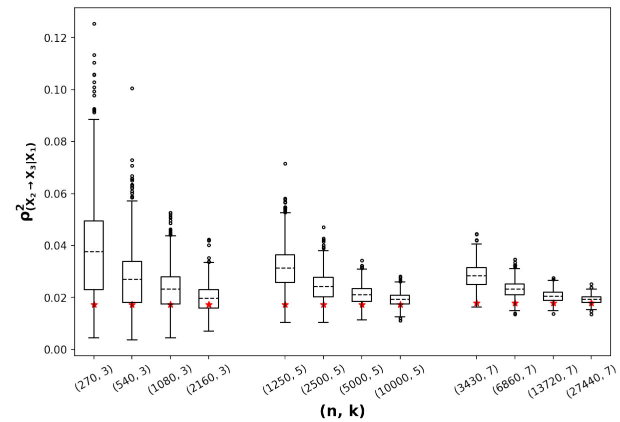


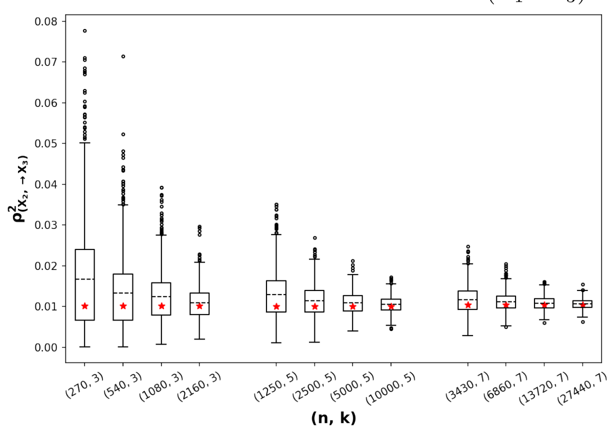
Figure J.49: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



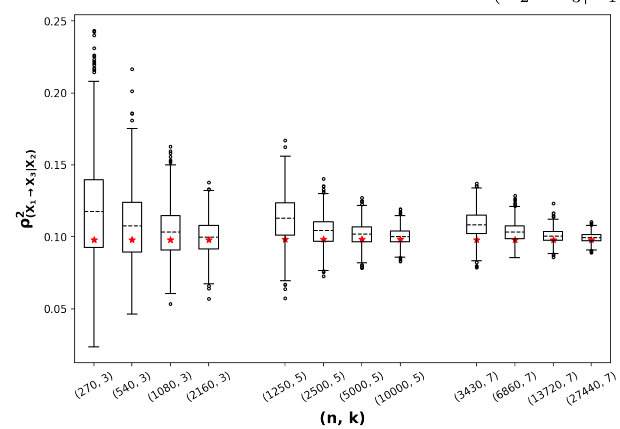
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.50: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

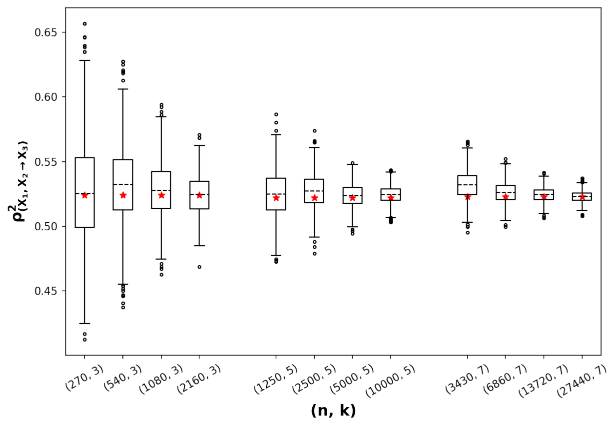
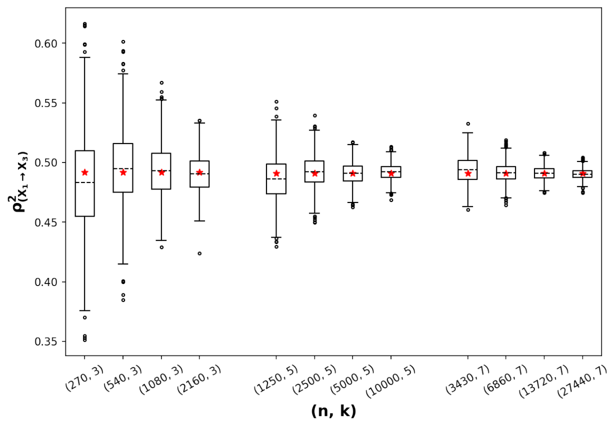
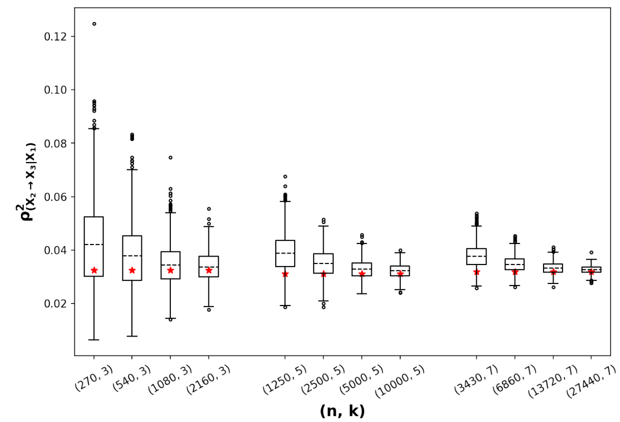


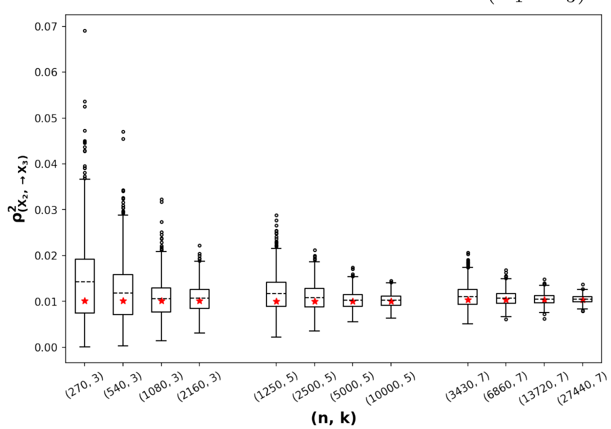
Figure J.51: The overall association measure for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_d, X_1)| = 0.7$ in three-dimensional case



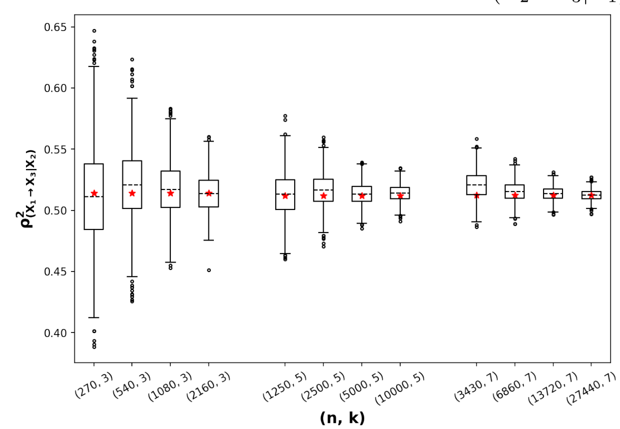
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.52: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) > 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

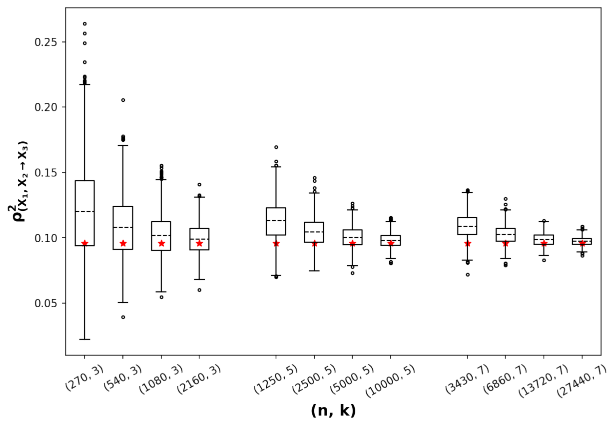
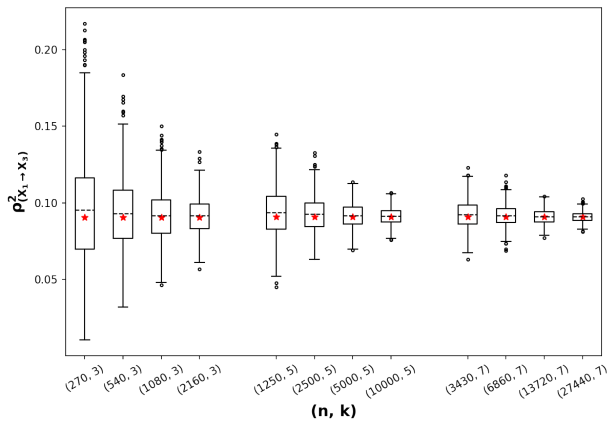
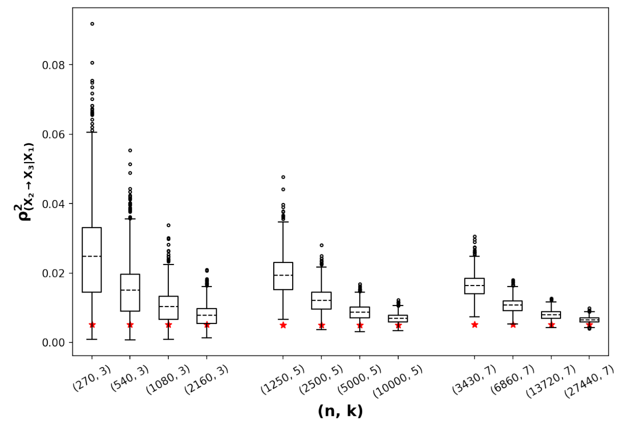


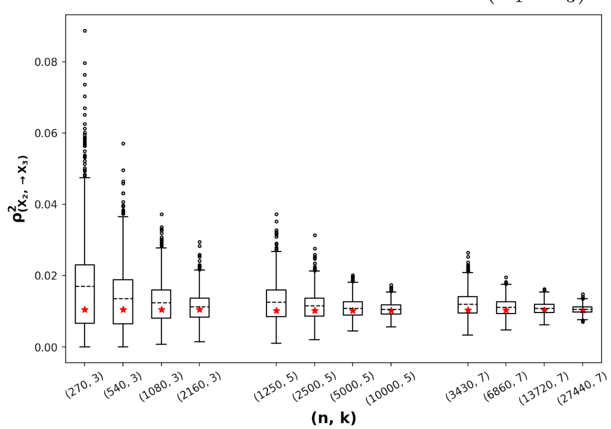
Figure J.53: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_d, X_1)| = 0.3$ in three-dimensional case



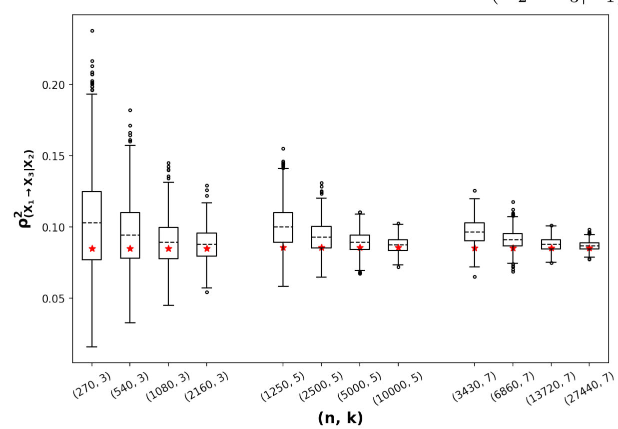
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.54: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

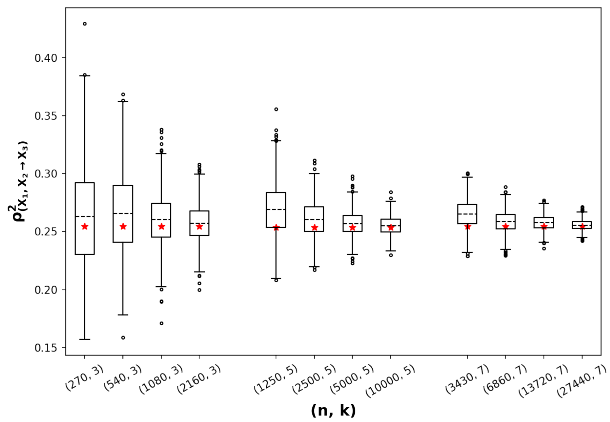
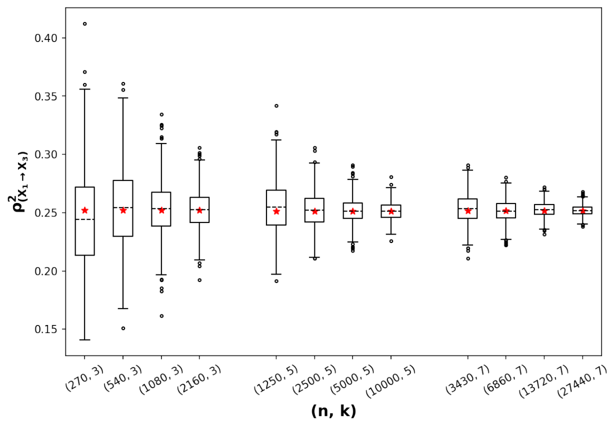
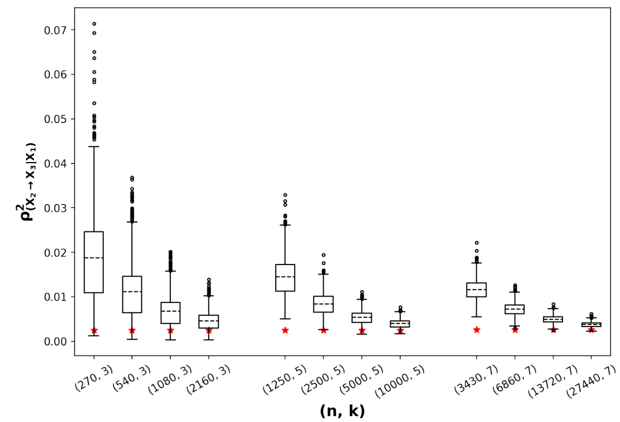


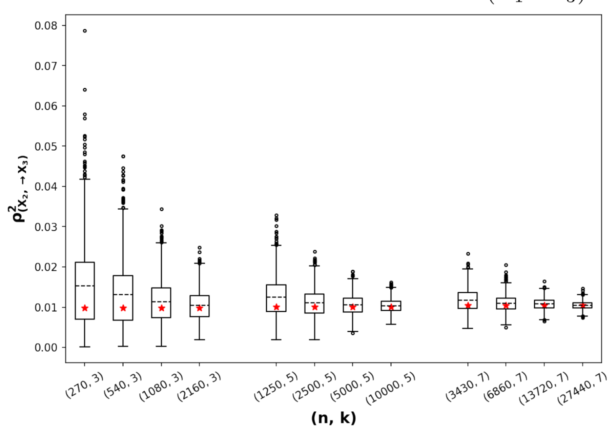
Figure J.55: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_d, X_1)| = 0.5$ in three-dimensional case



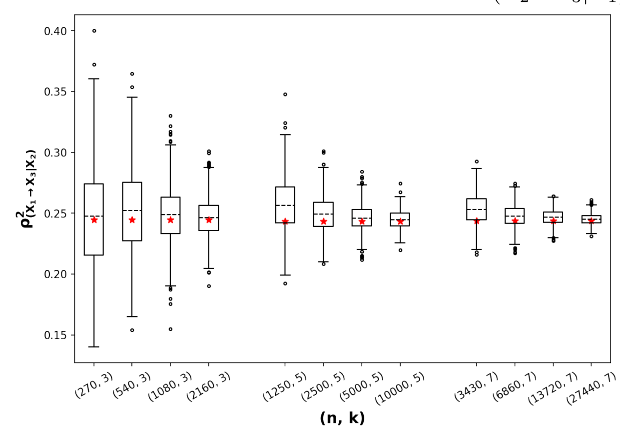
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.56: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and moderate association $|corr(X_3, X_1)| = 0.5$ in three-dimensional case

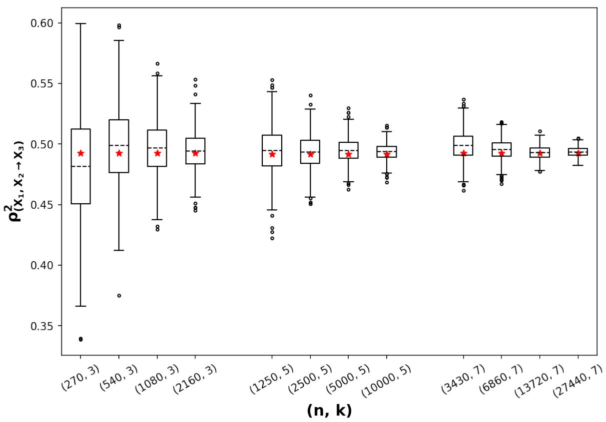
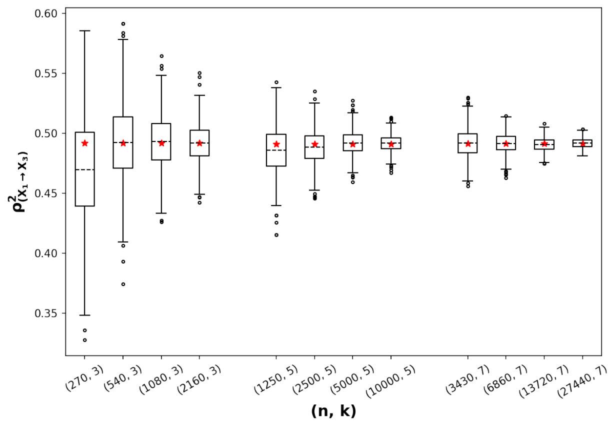
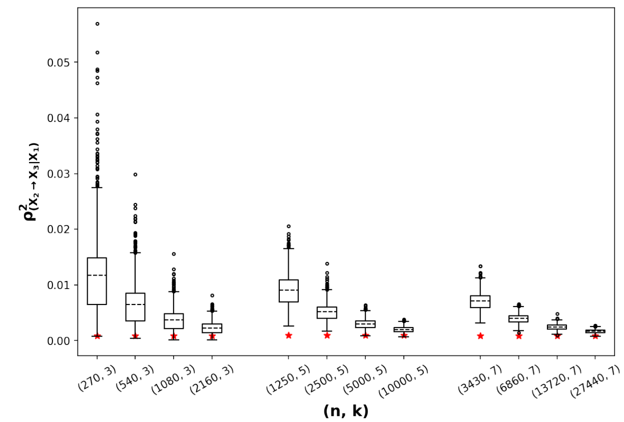


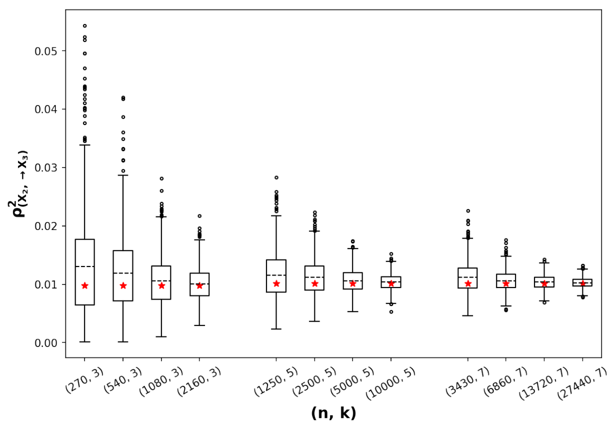
Figure J.57: The overall association measure for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_d, X_1)| = 0.7$ in three-dimensional case



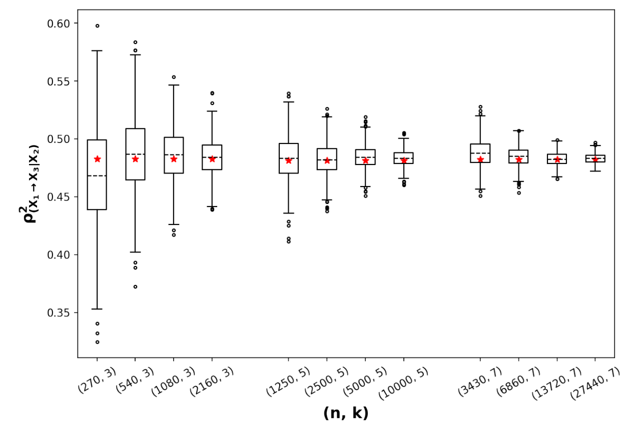
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.58: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) < 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

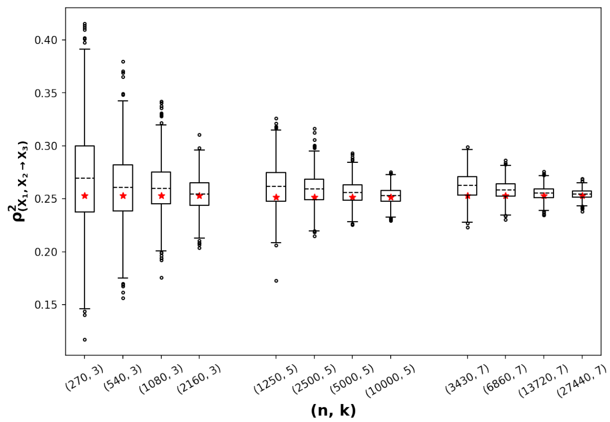
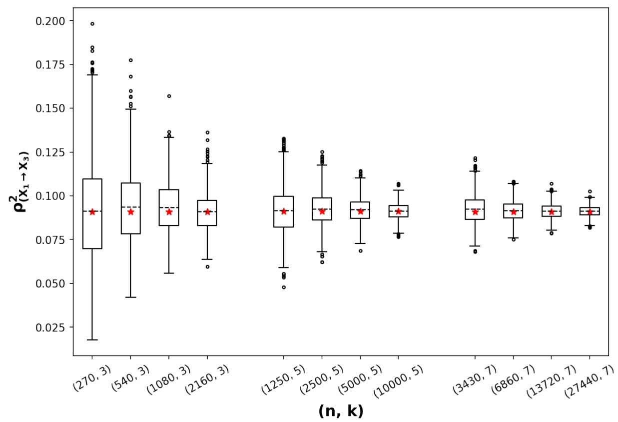
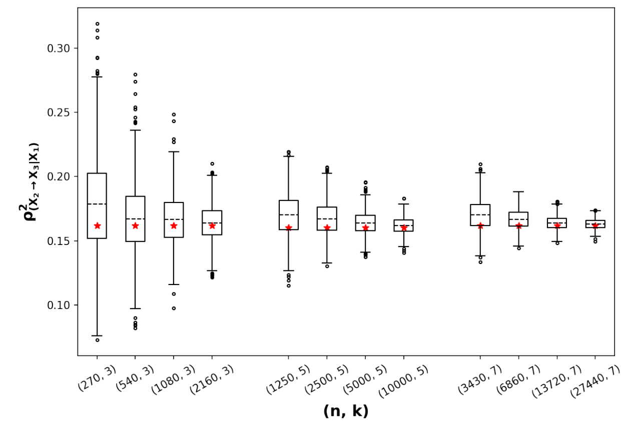


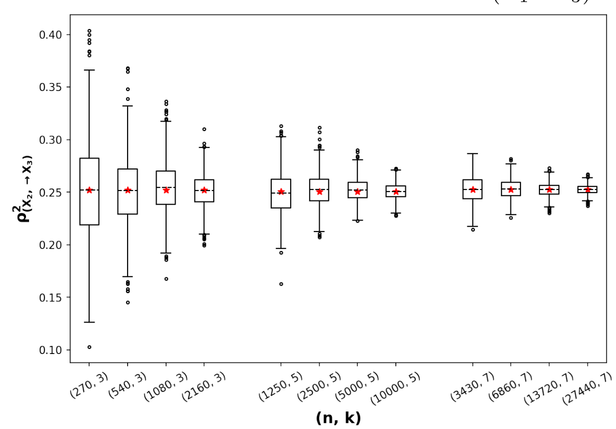
Figure J.59: the overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case



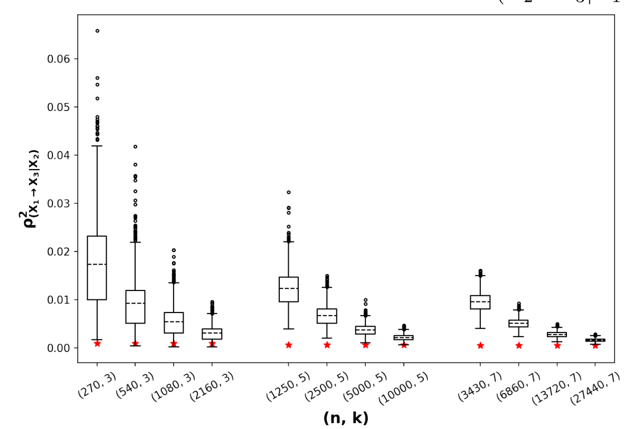
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.60: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and weak association $|corr(X_3, X_1)| = 0.3$ in three-dimensional case

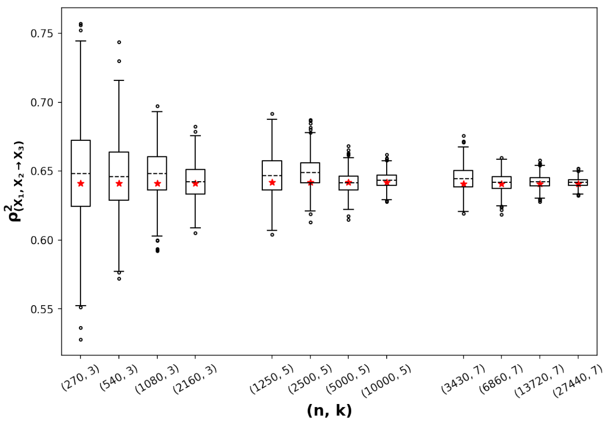
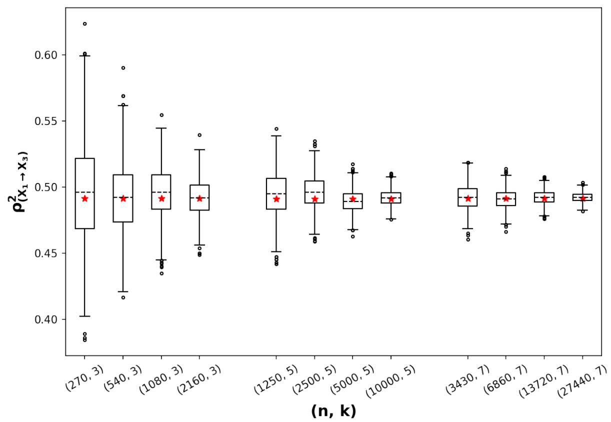
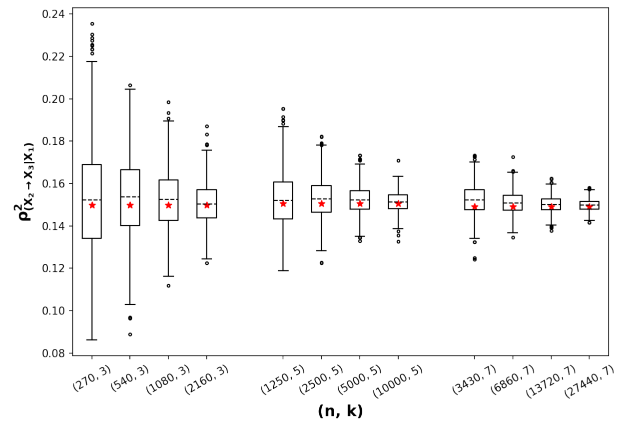


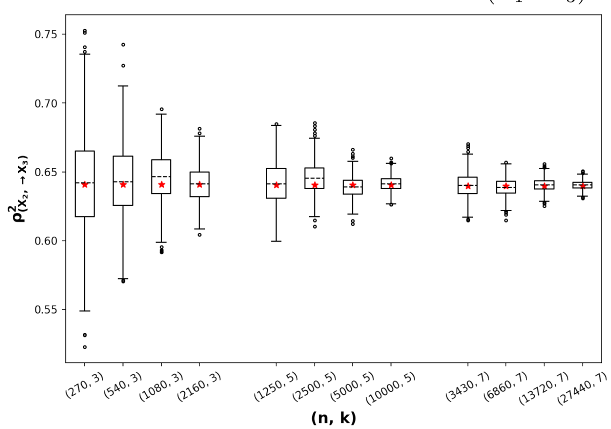
Figure J.61: The overall association measure for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case



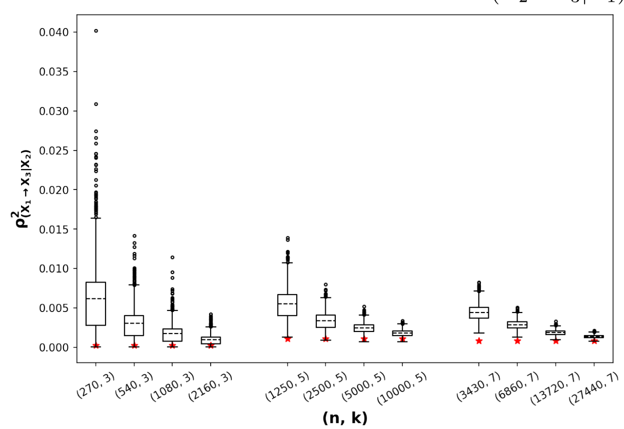
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.62: The marginal and conditional association measures for $pcorr(X_3, X_1|X_2) = 0$ and strong association $|corr(X_3, X_1)| = 0.7$ in three-dimensional case

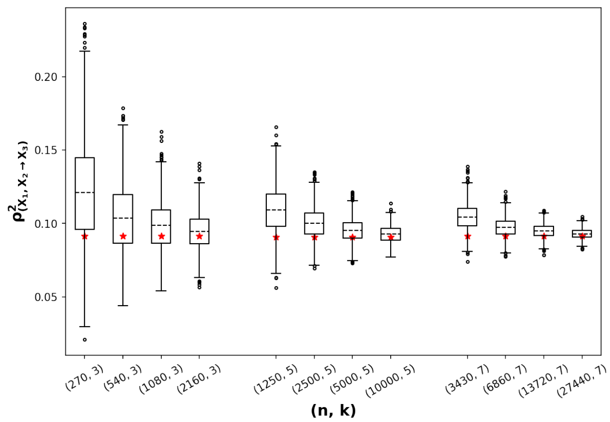
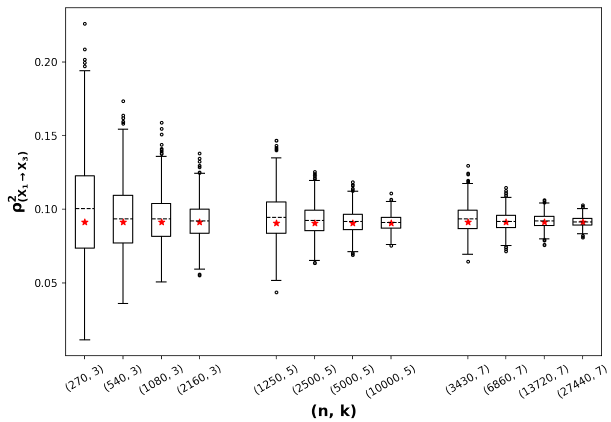
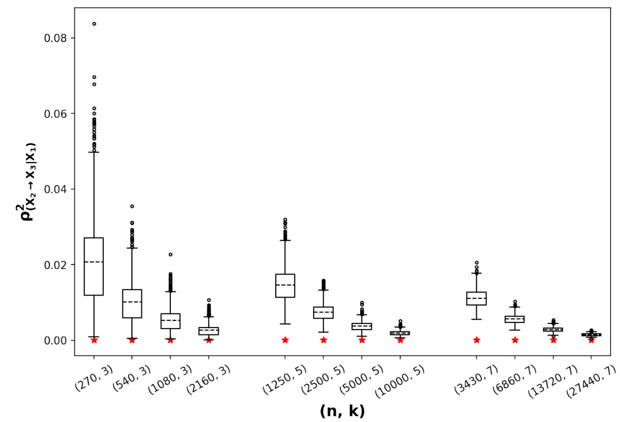


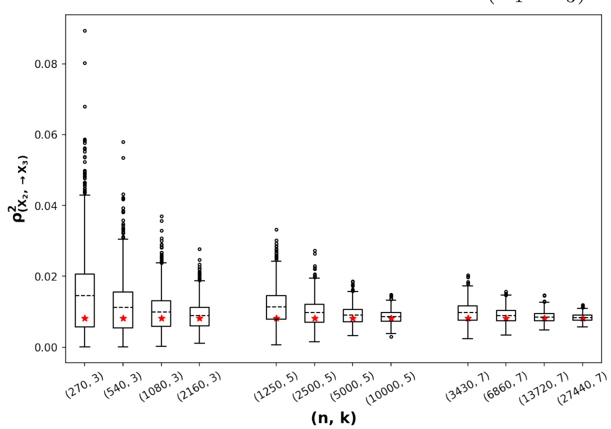
Figure J.63: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case



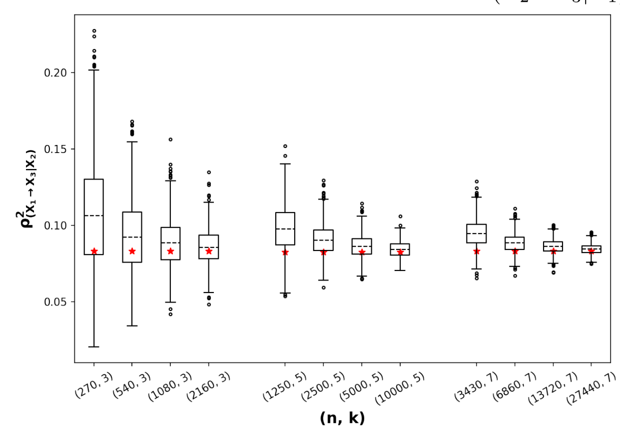
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure J.64: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and weak association $|\phi| = |corr(X_3, X_1)| = 0.3$ in three-dimensional case

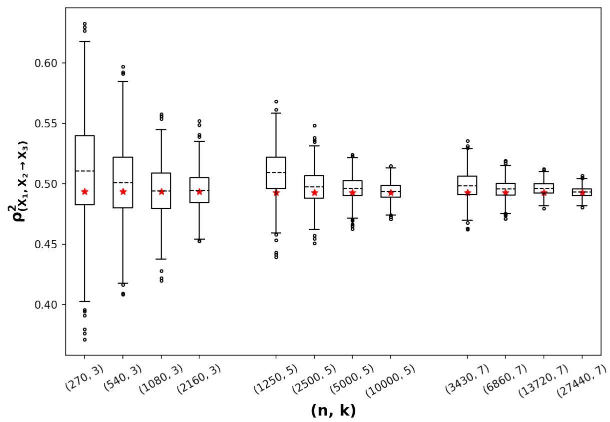
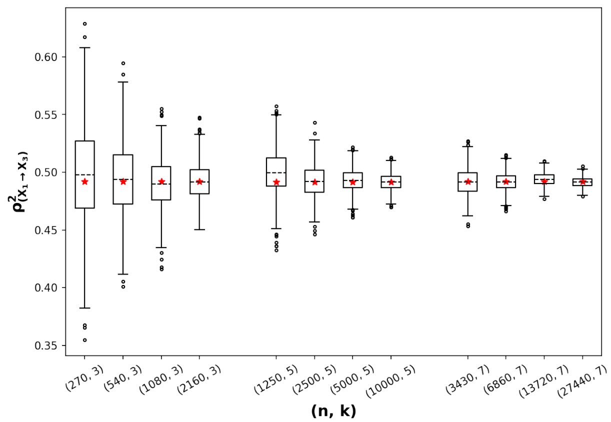
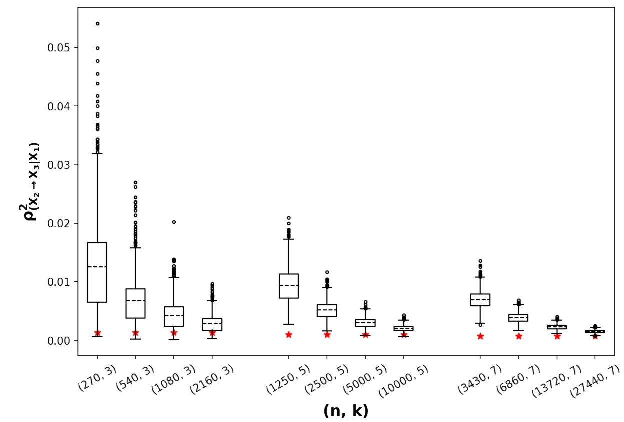


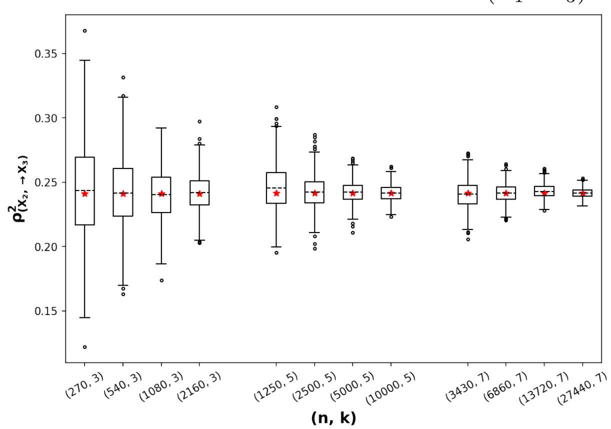
Figure J.65: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case



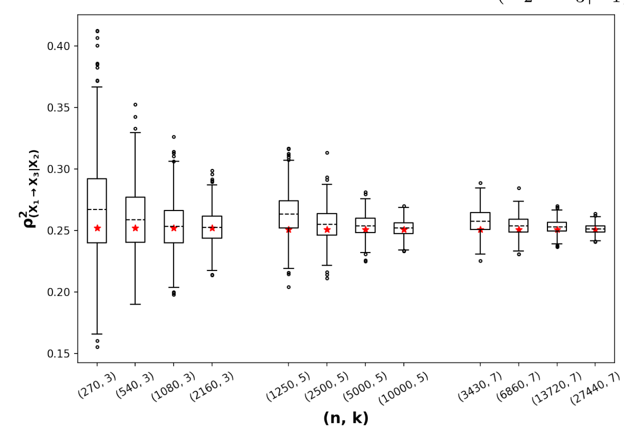
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure J.66: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, 2, 3$ and strong association $|\phi| = |corr(X_3, X_1)| = 0.7$ in three-dimensional case

APPENDIX K

SIMULATION RESULTS FOR FIVE-DIMENSIONAL CASE

K.1 Scenario 2

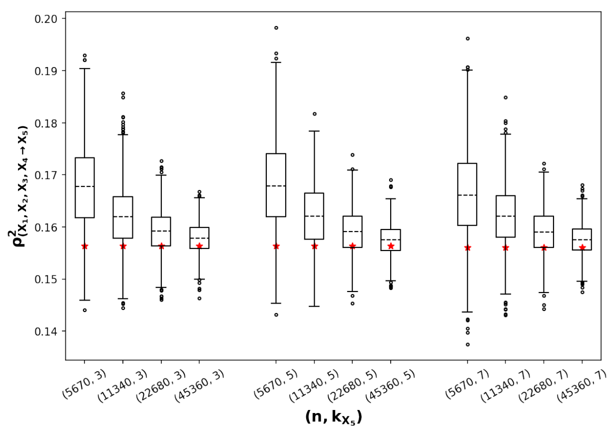
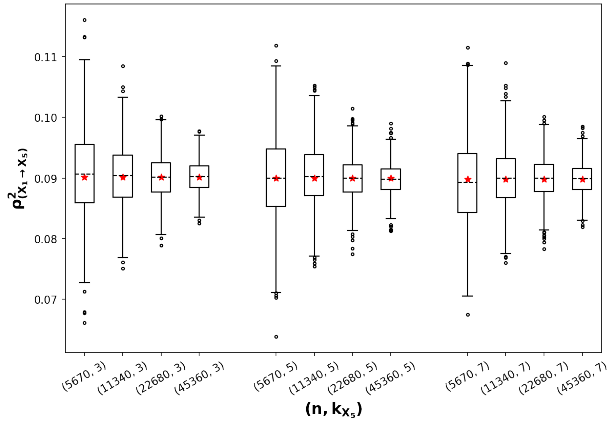
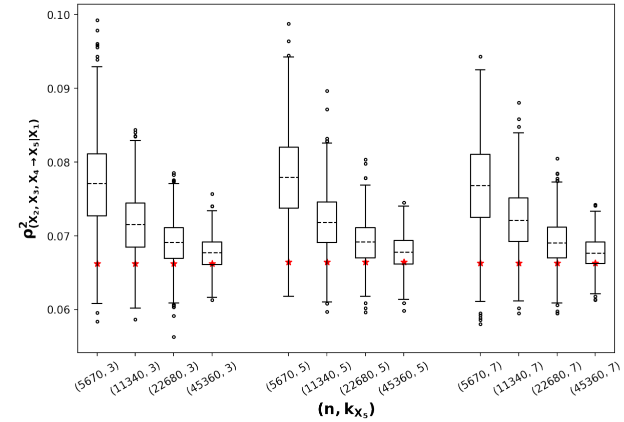


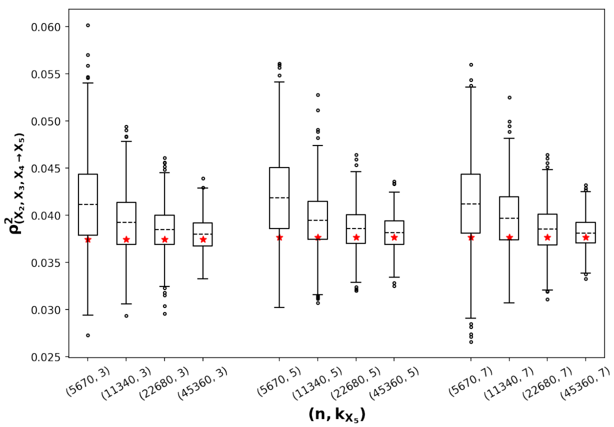
Figure K.1: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



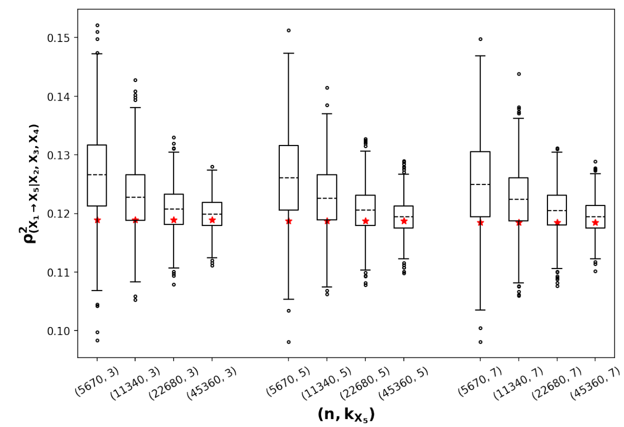
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.2: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

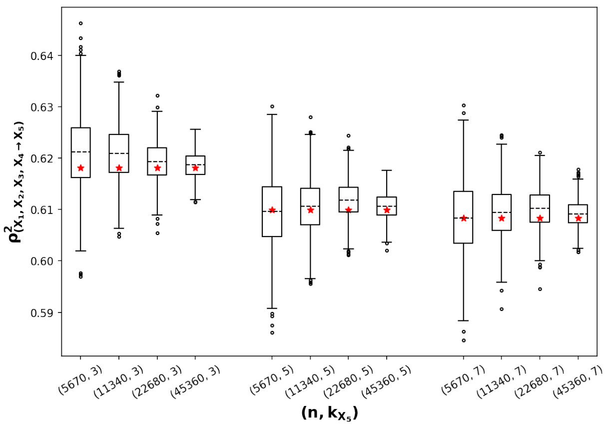
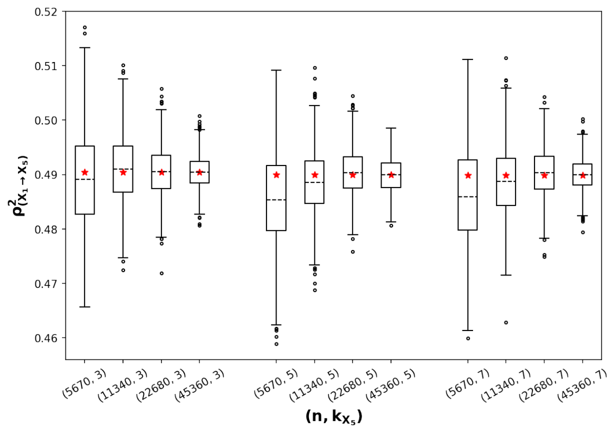
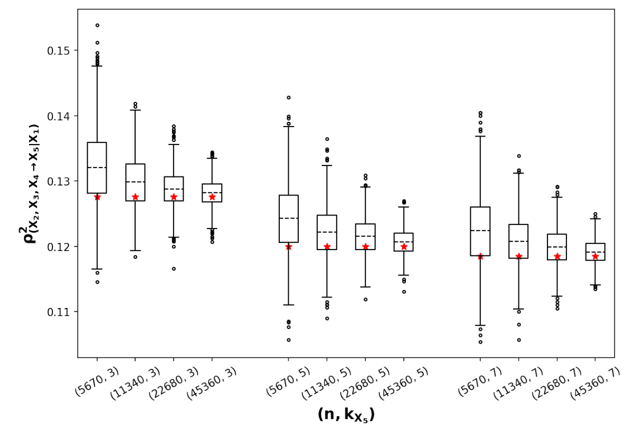


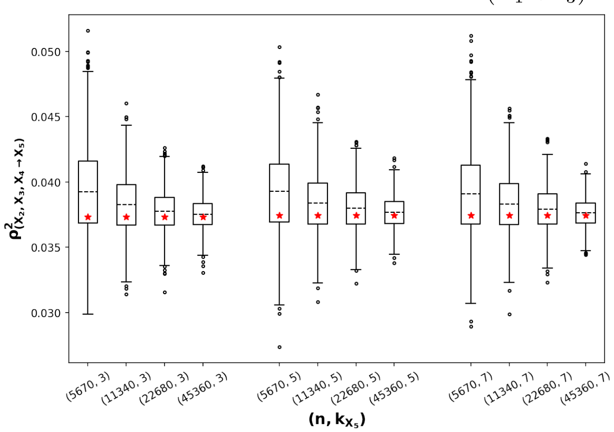
Figure K.3: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



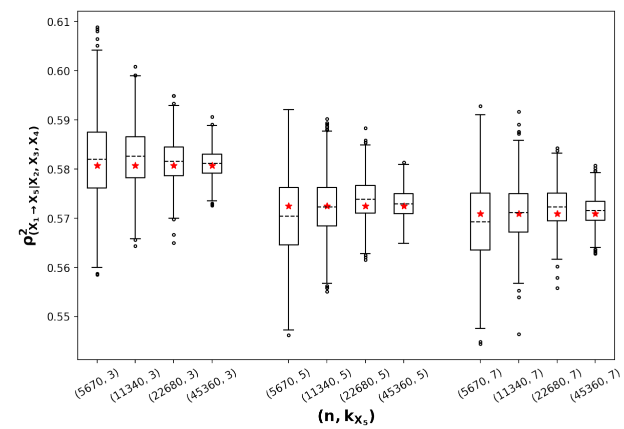
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.4: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

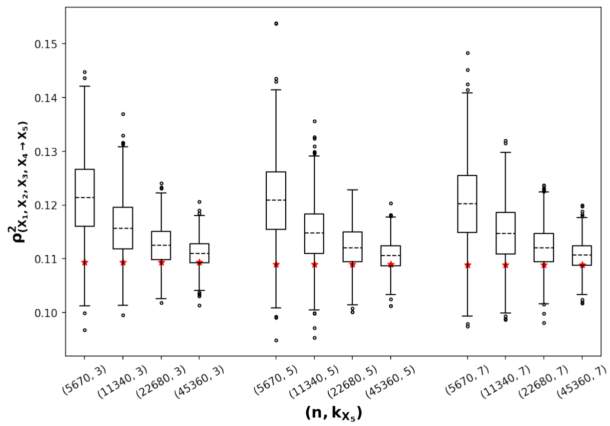
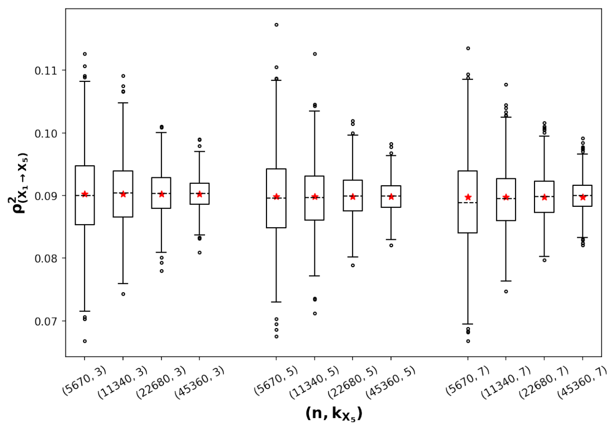
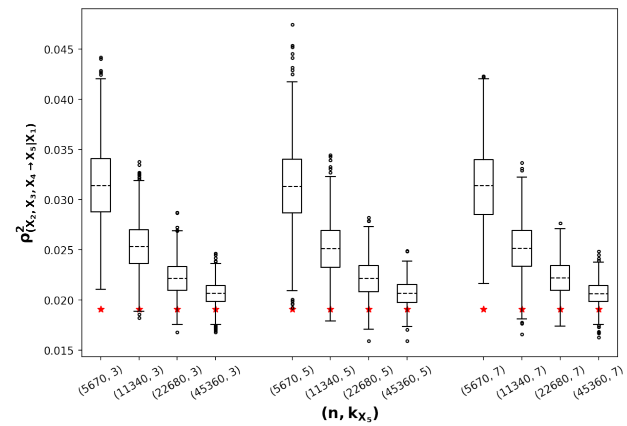


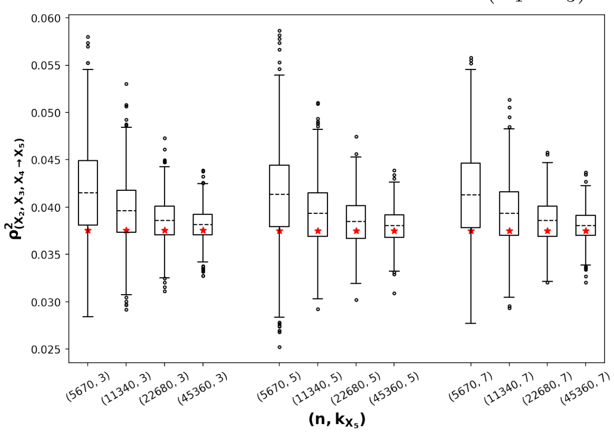
Figure K.5: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



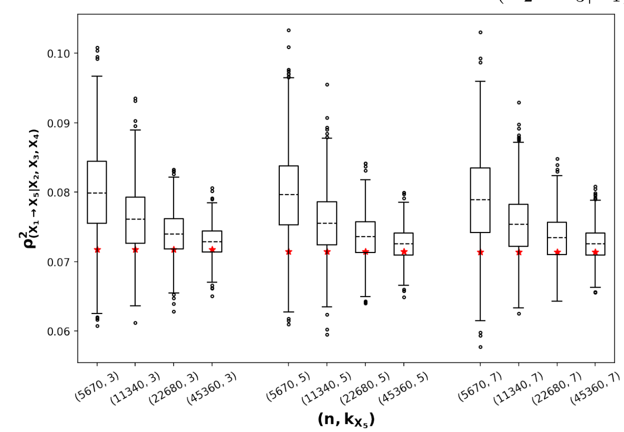
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.6: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

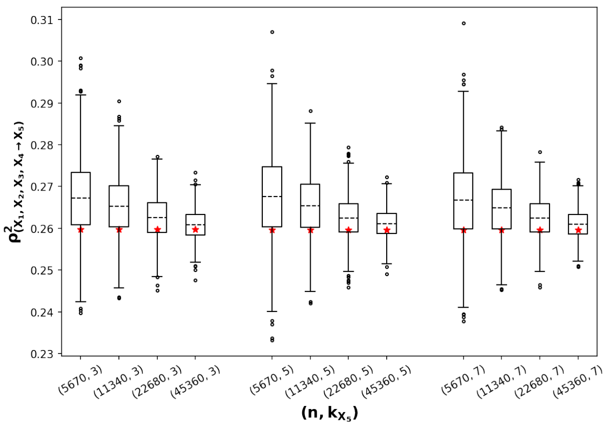
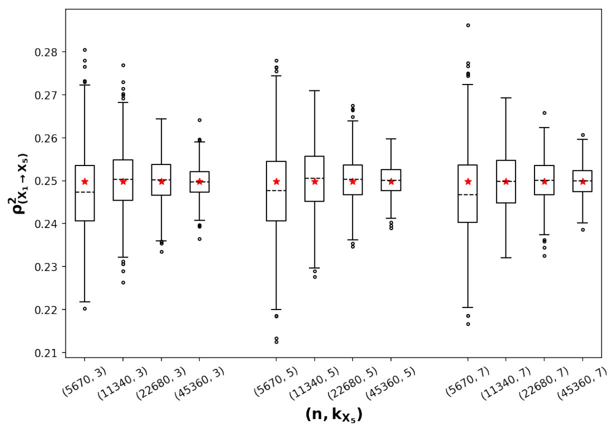
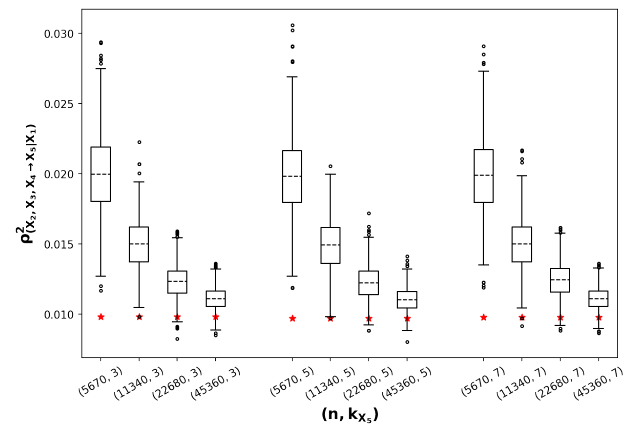


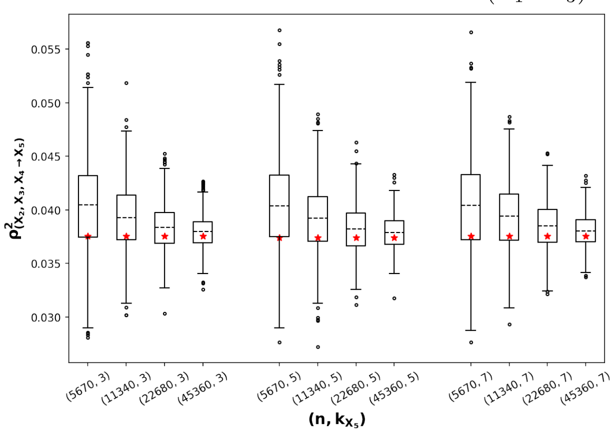
Figure K.7: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case



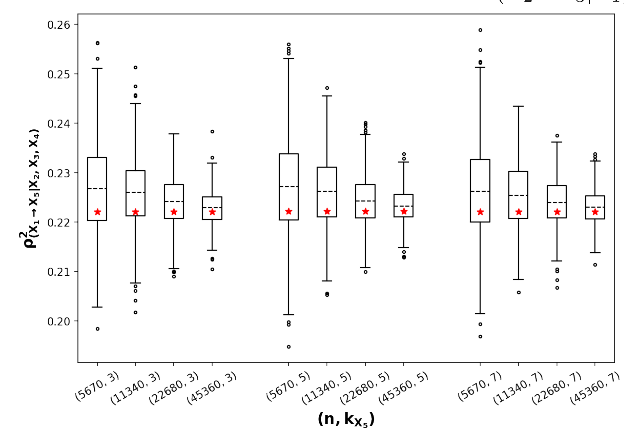
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.8: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case

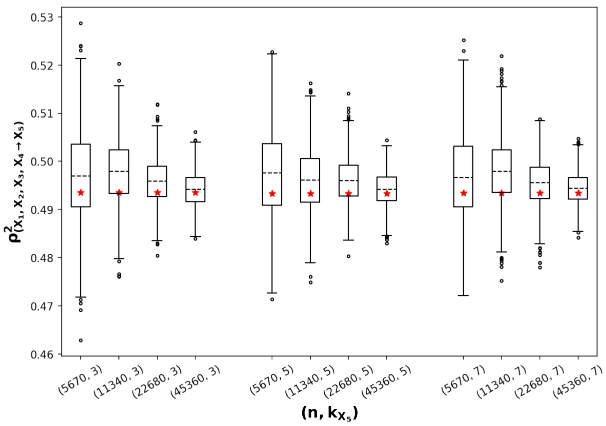
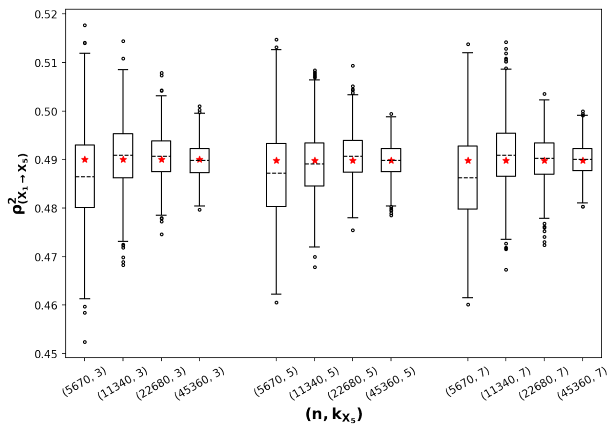
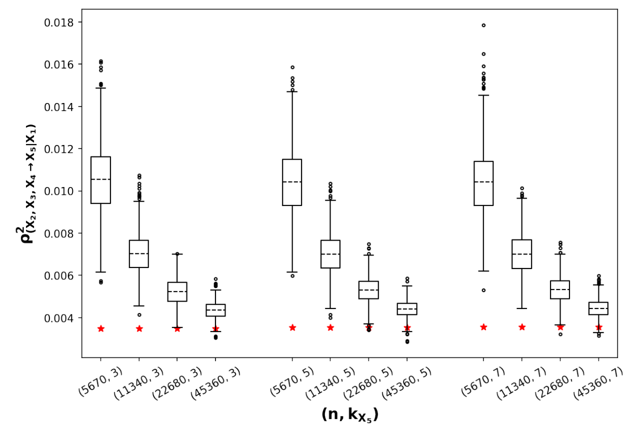


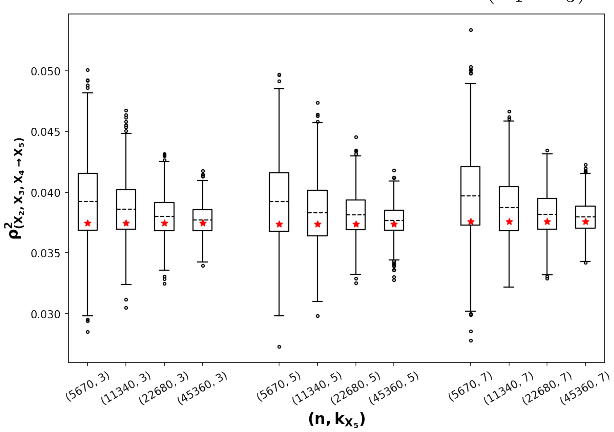
Figure K.9: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



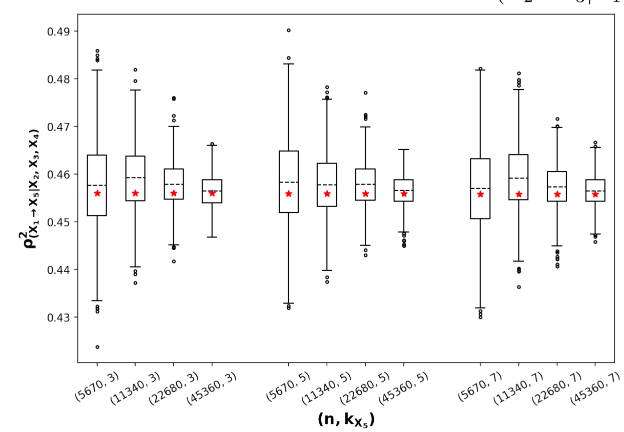
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.10: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

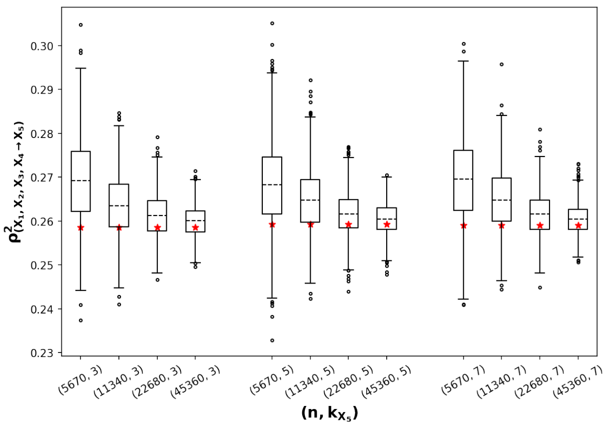
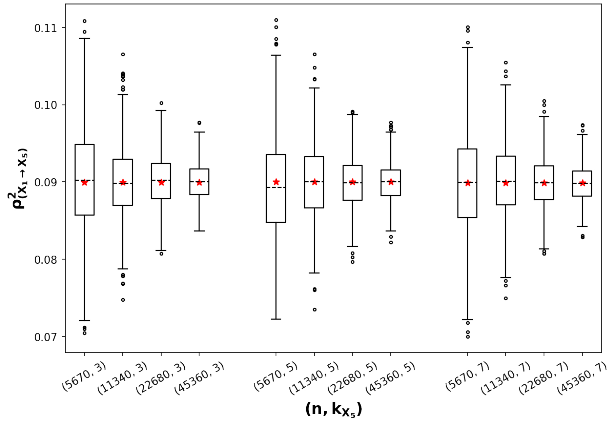
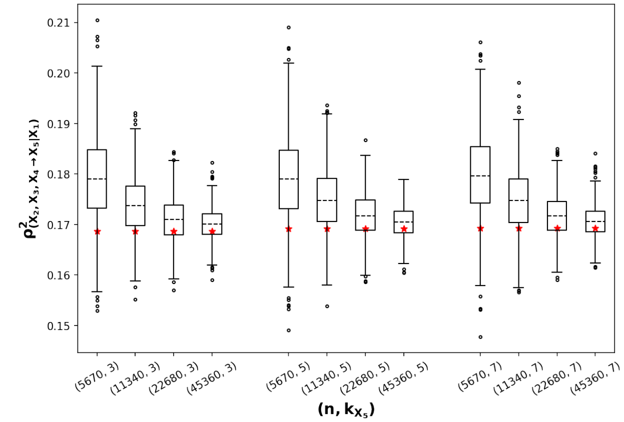


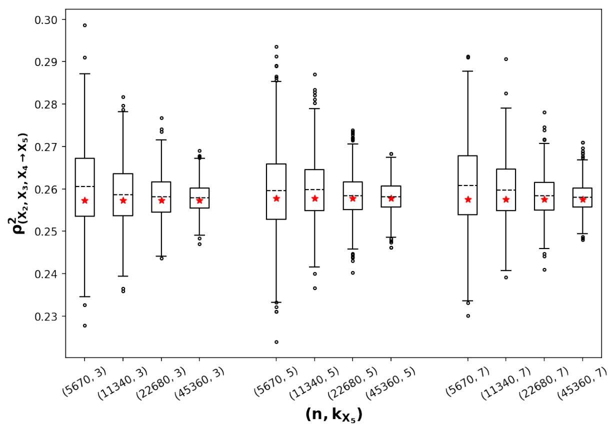
Figure K.11: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



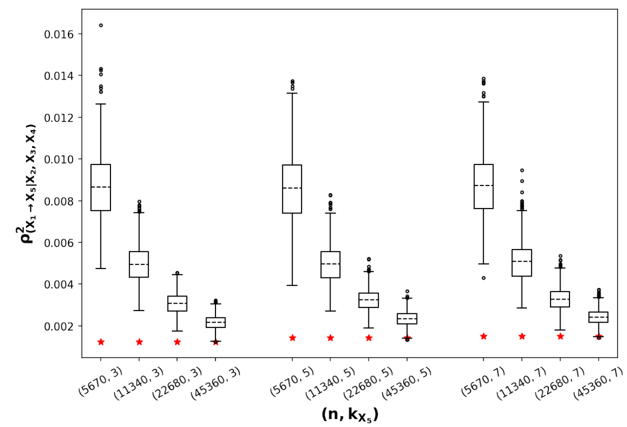
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.12: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

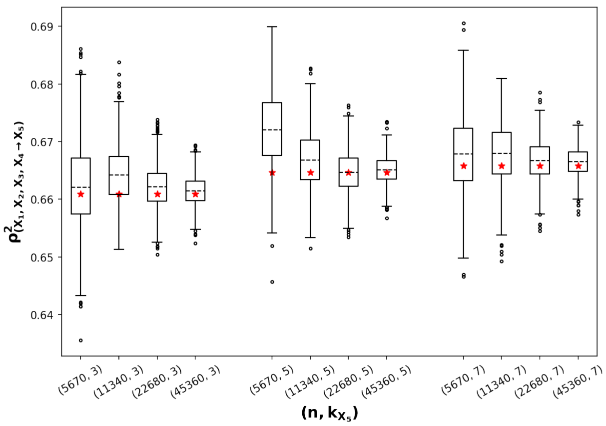
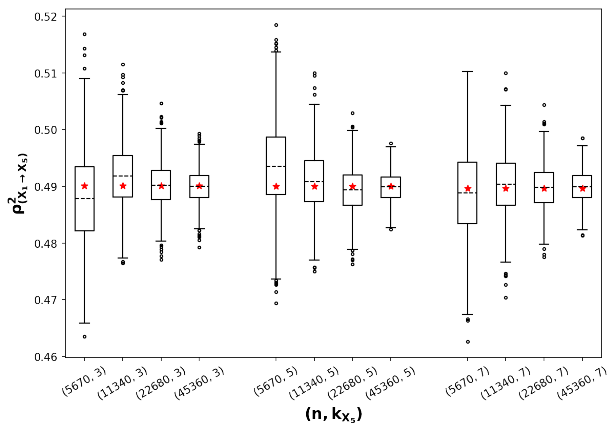
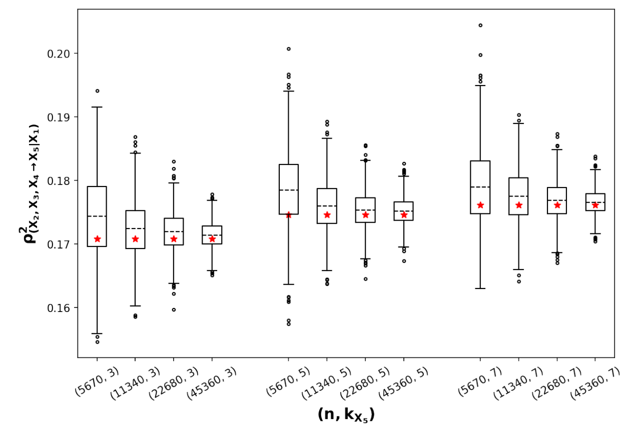


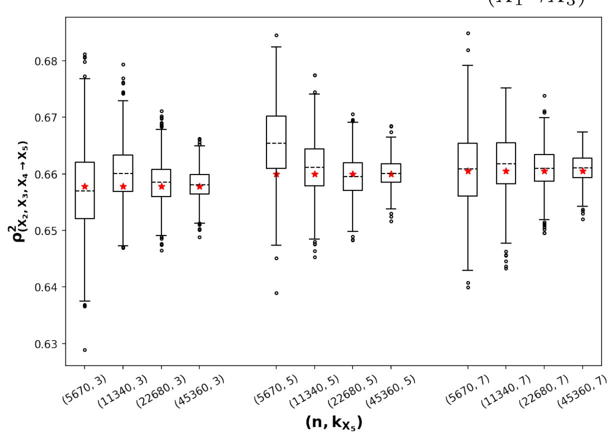
Figure K.13: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



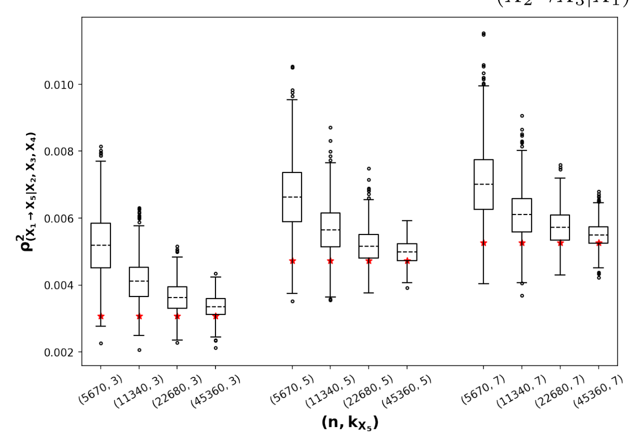
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.14: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

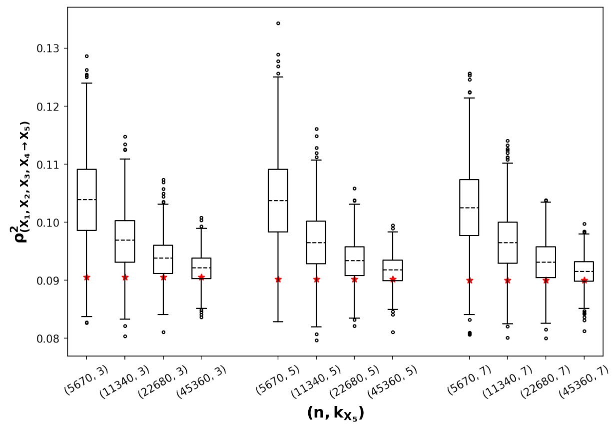


Figure K.15: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case

K.2 Scenario 3

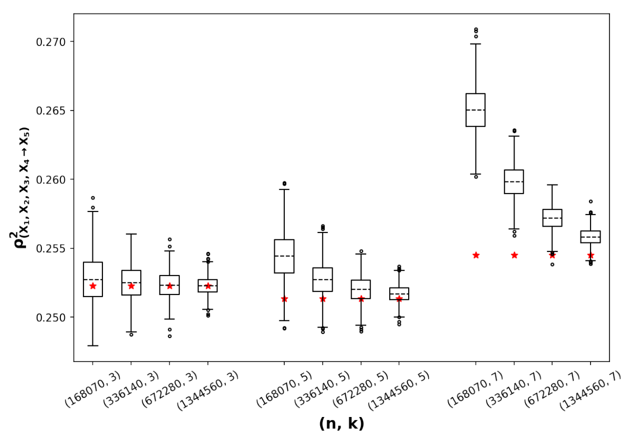
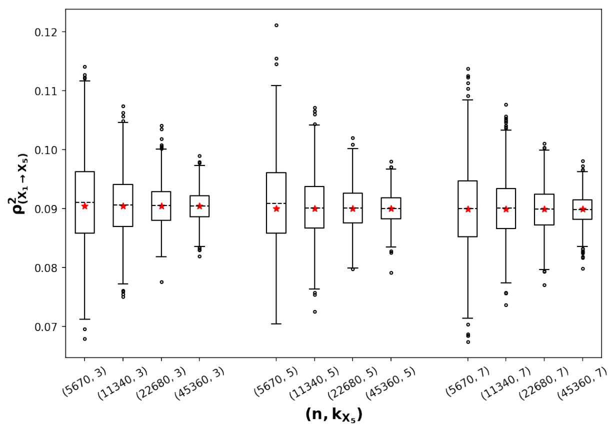
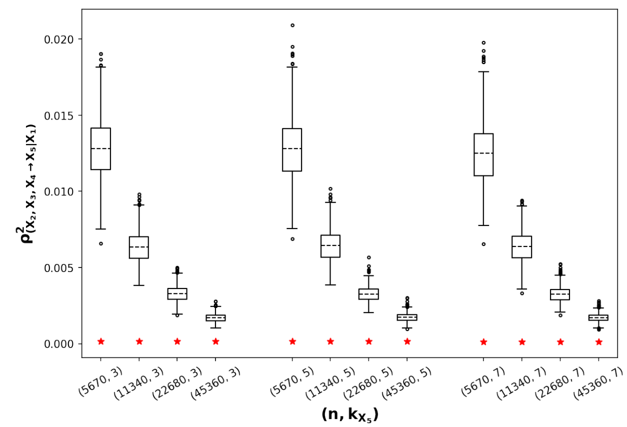


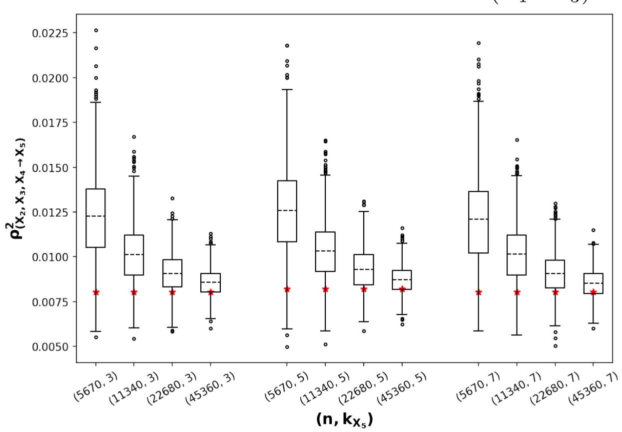
Figure K.17: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$ in five-dimensional case



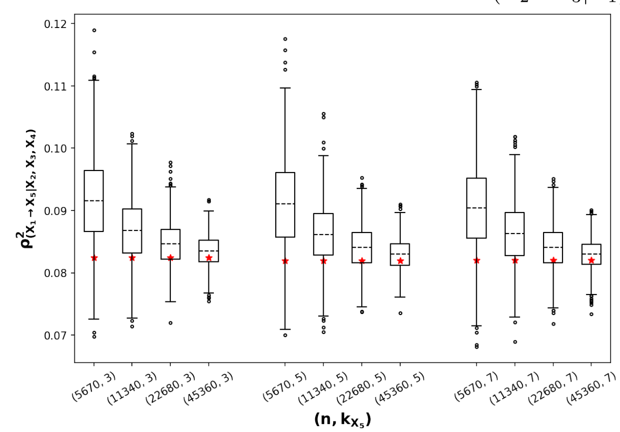
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$

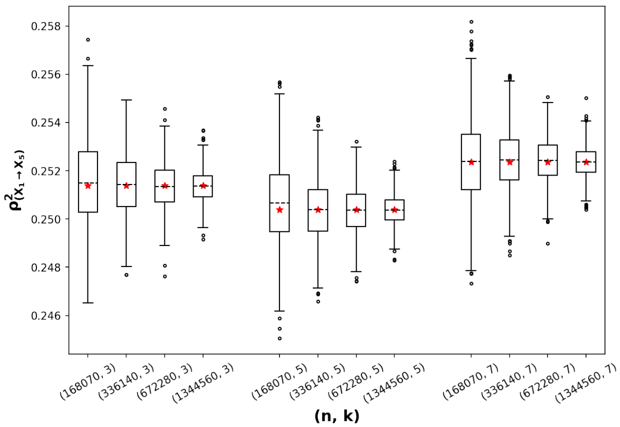


(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$

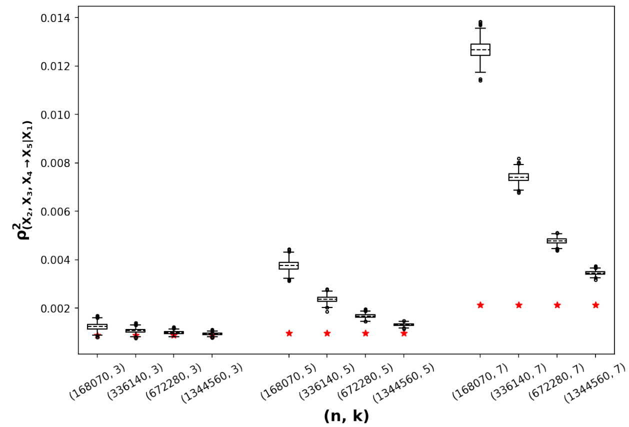


(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

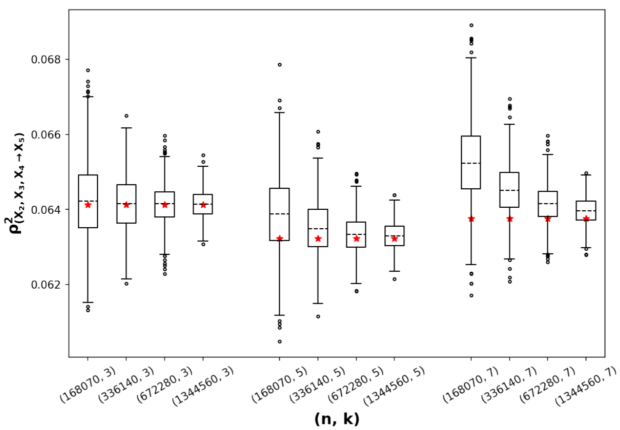
Figure K.16: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case



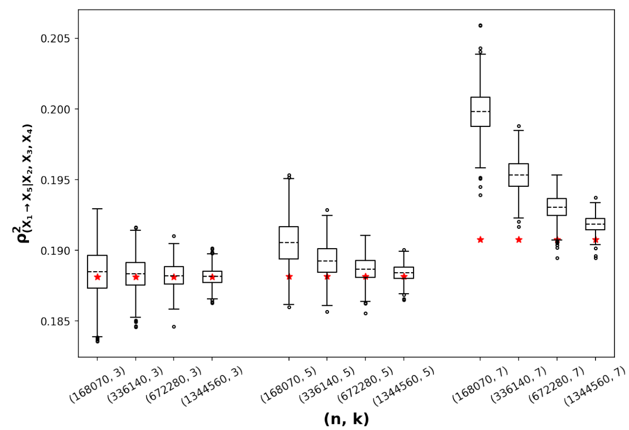
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.18: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and moderate association $|\phi| = |corr(X_5, X_1)| = 0.5$ in five-dimensional case

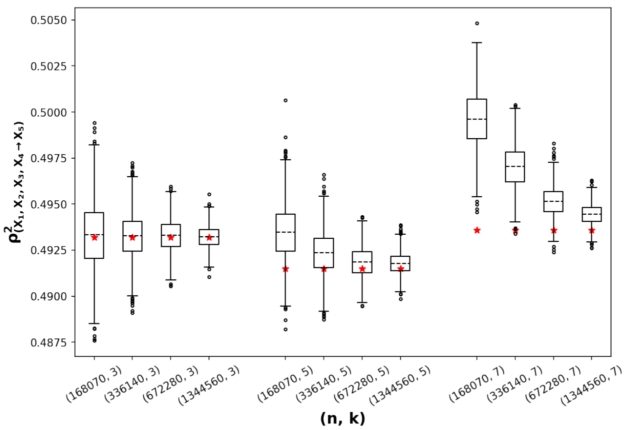
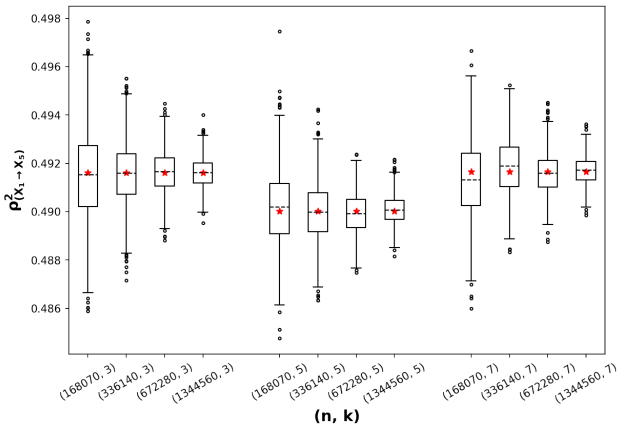
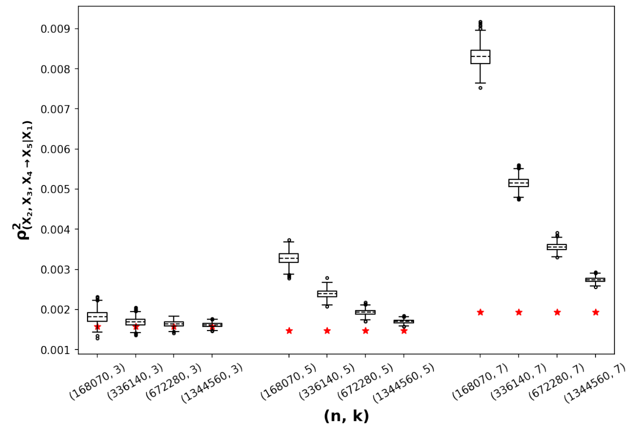


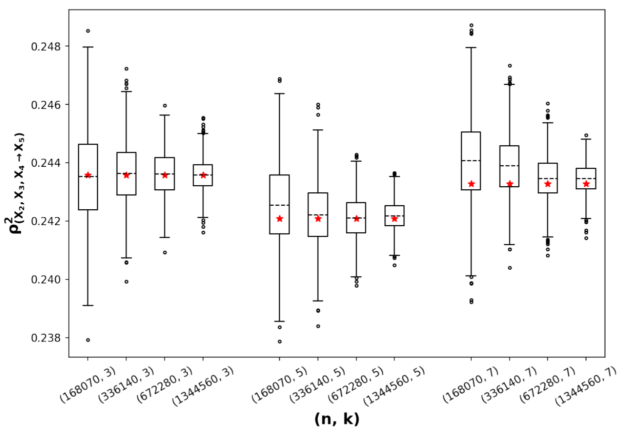
Figure K.19: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case



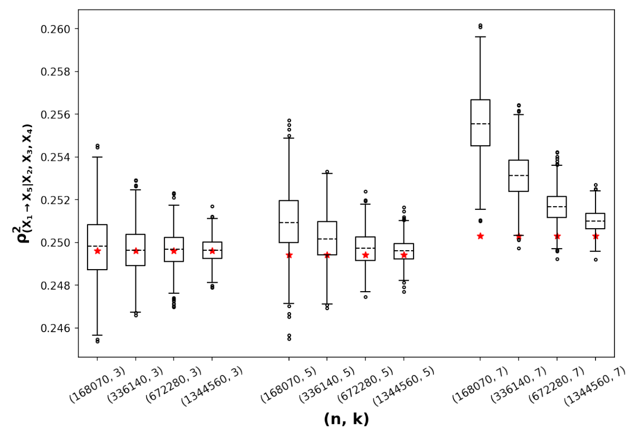
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.20: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case

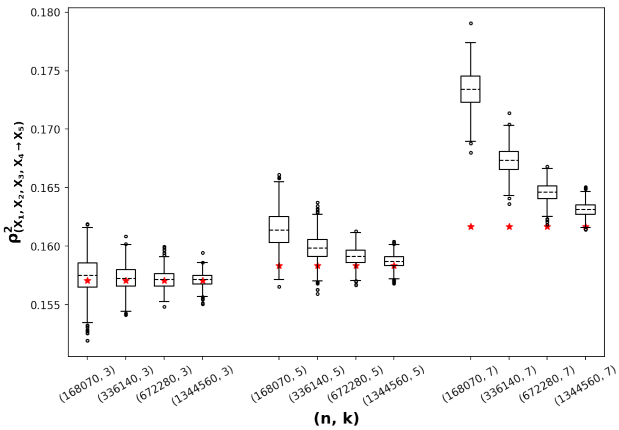
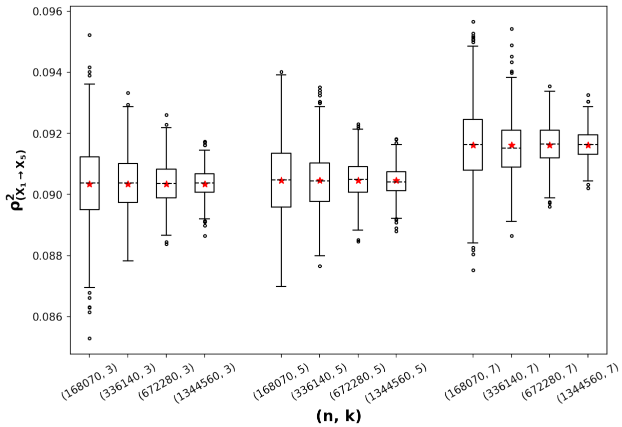
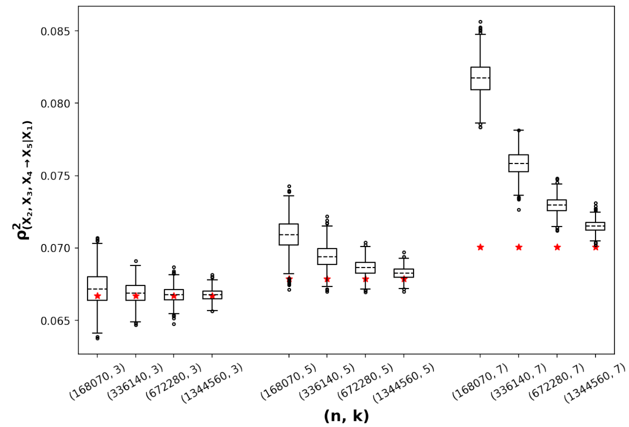


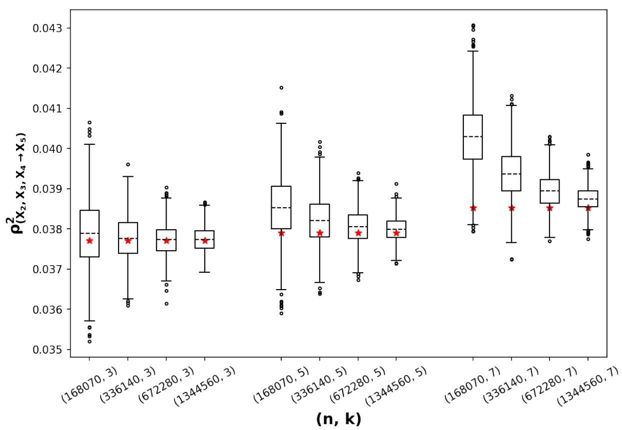
Figure K.21: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



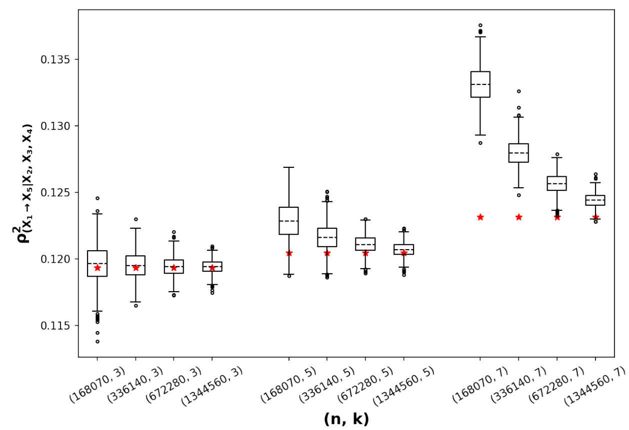
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.22: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

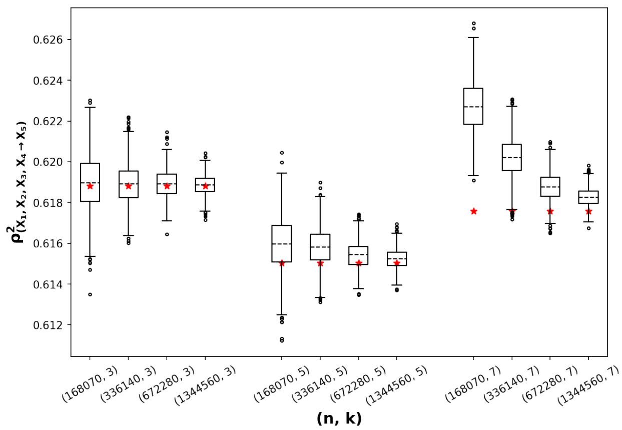
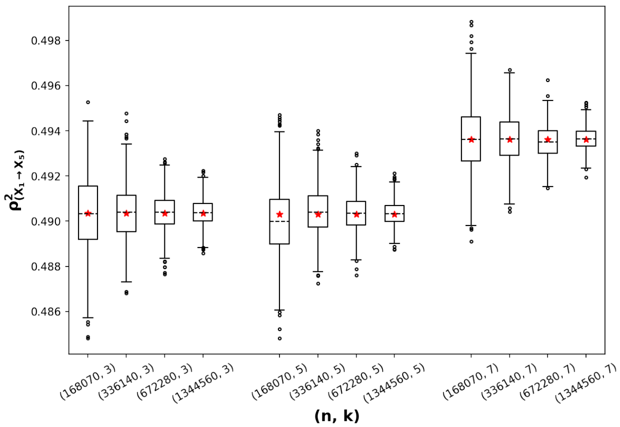
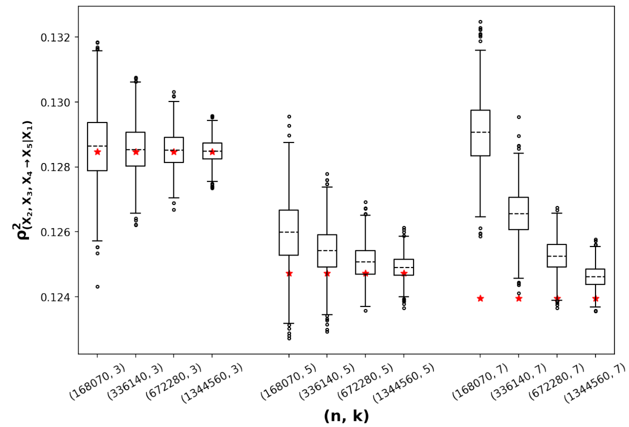


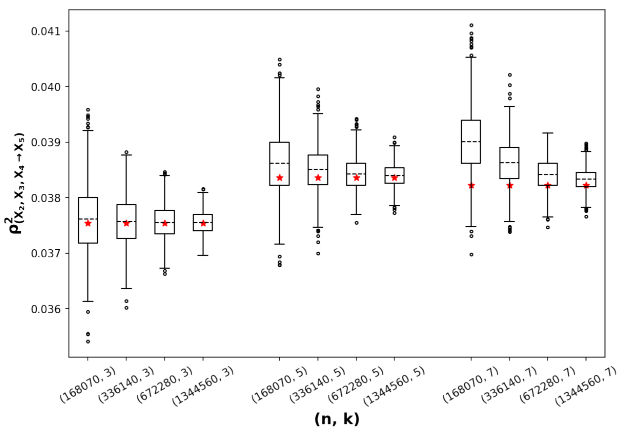
Figure K.23: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



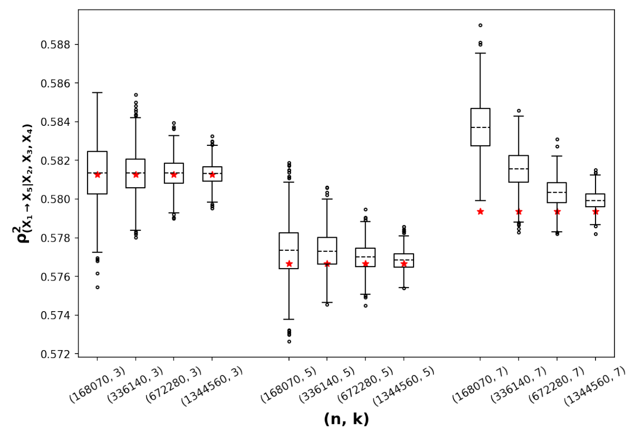
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.24: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

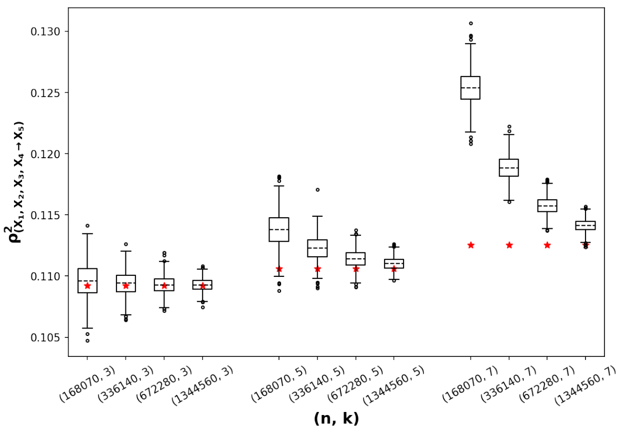
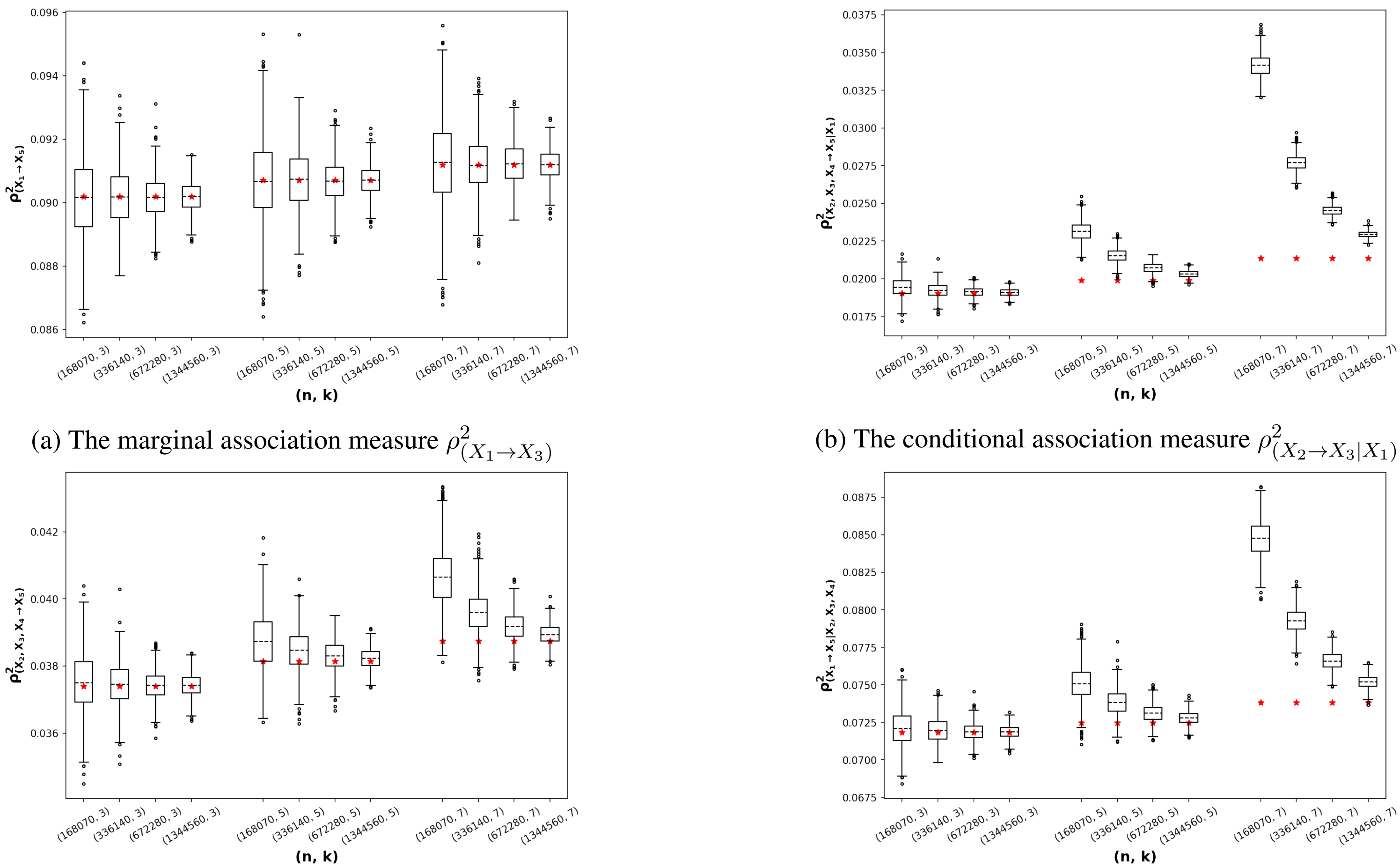


Figure K.25: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$

(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$

(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$

(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.26: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

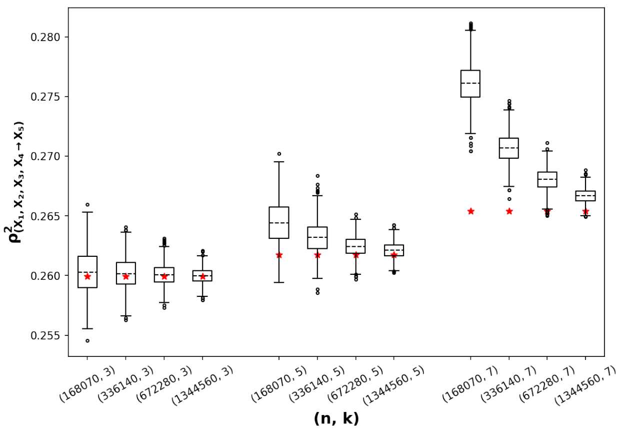
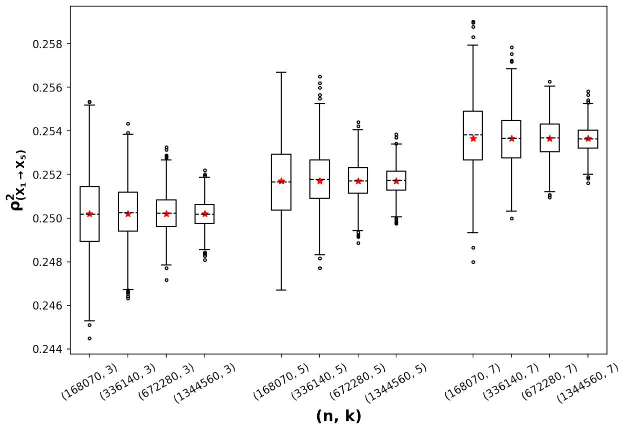
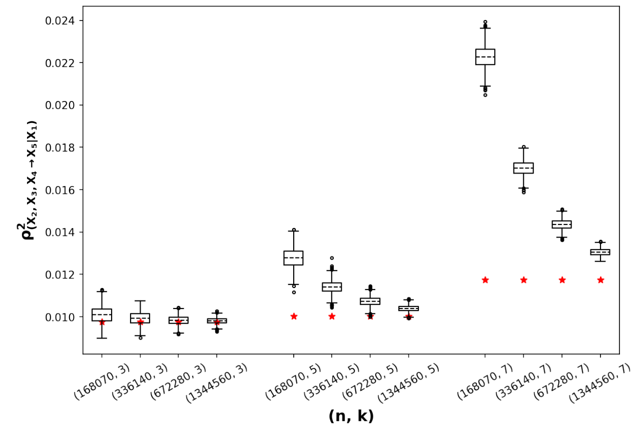


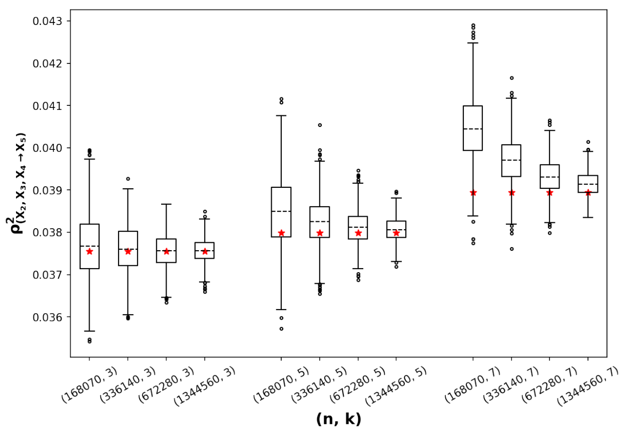
Figure K.27: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case



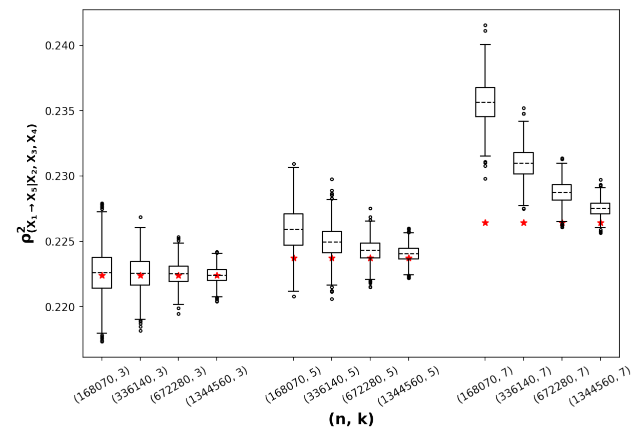
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.28: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case

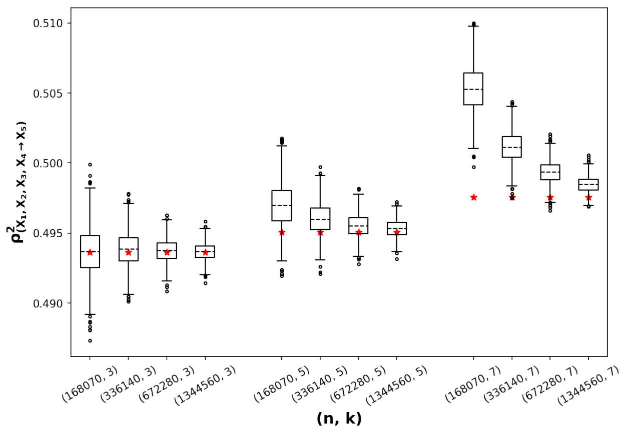
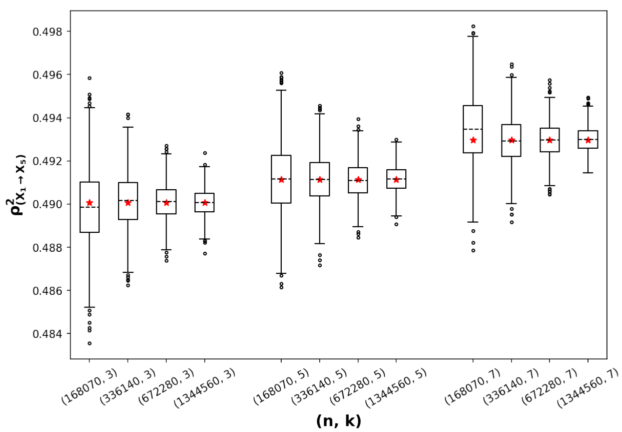
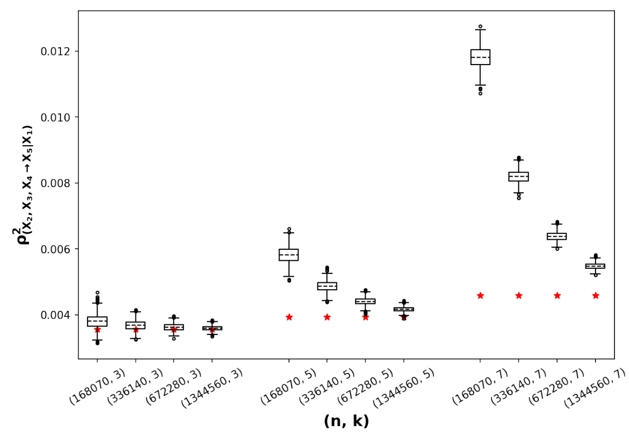


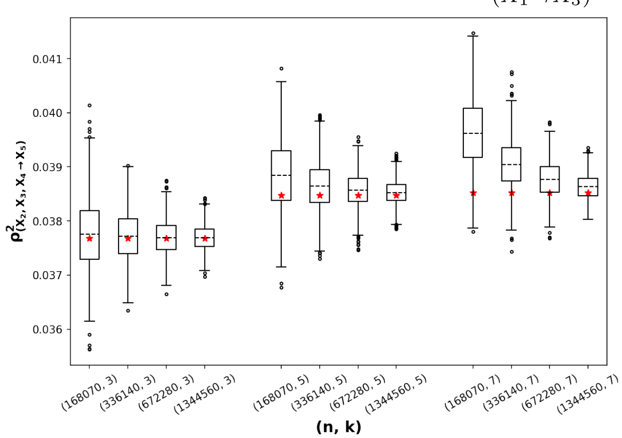
Figure K.29: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



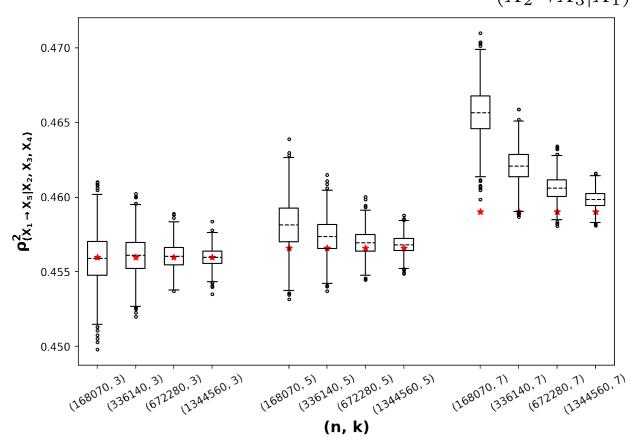
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.30: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

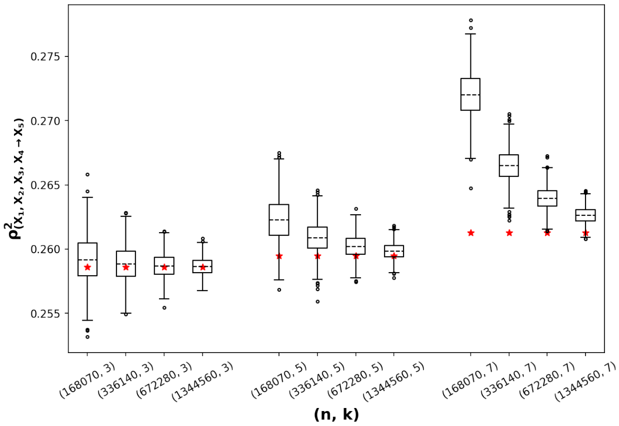
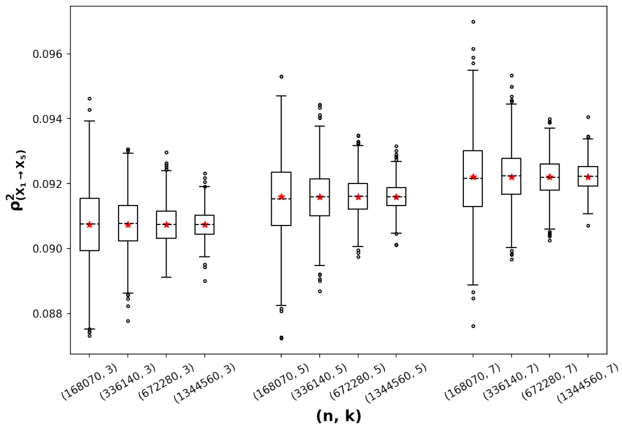
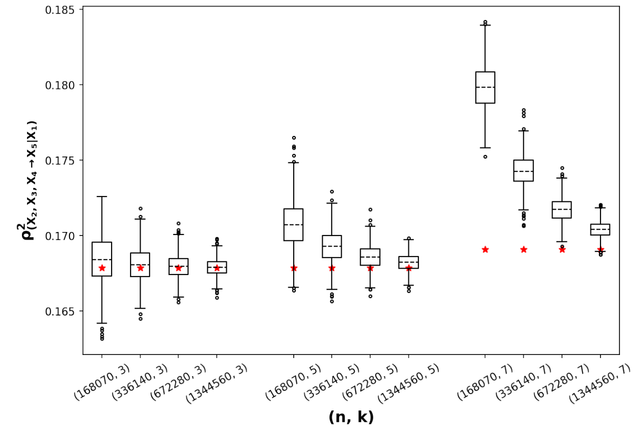


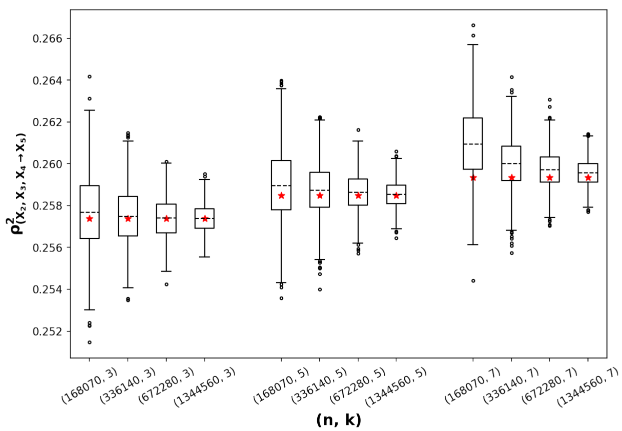
Figure K.31: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



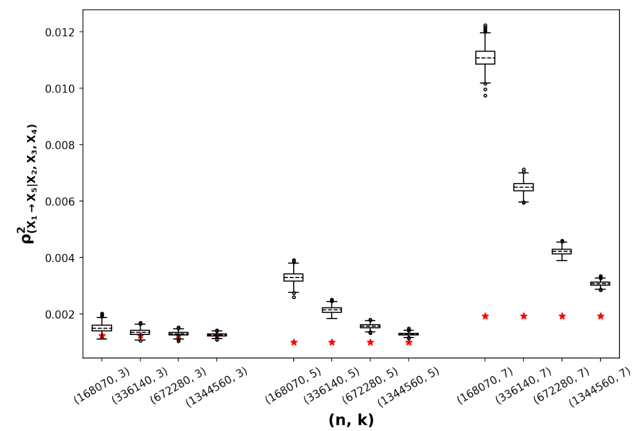
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.32: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

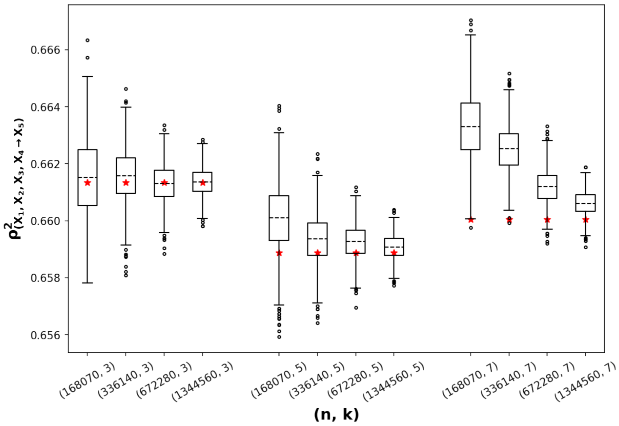
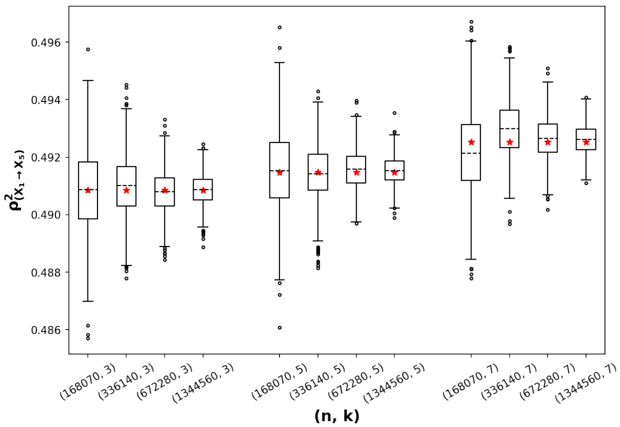
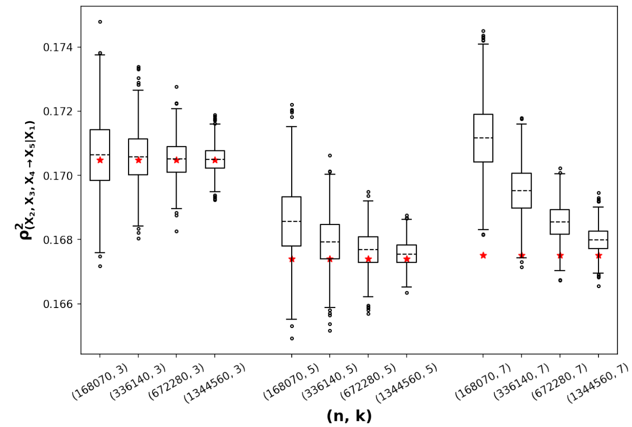


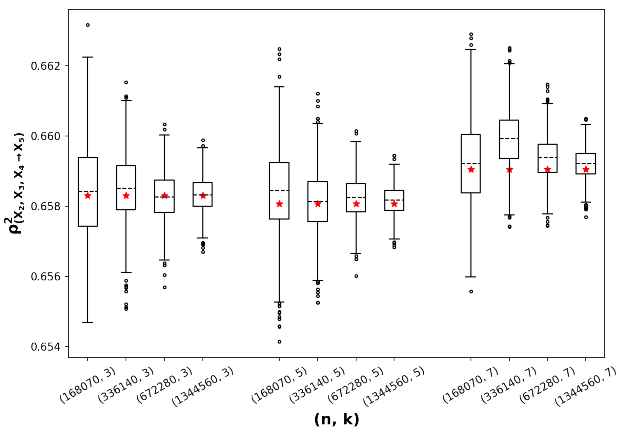
Figure K.33: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



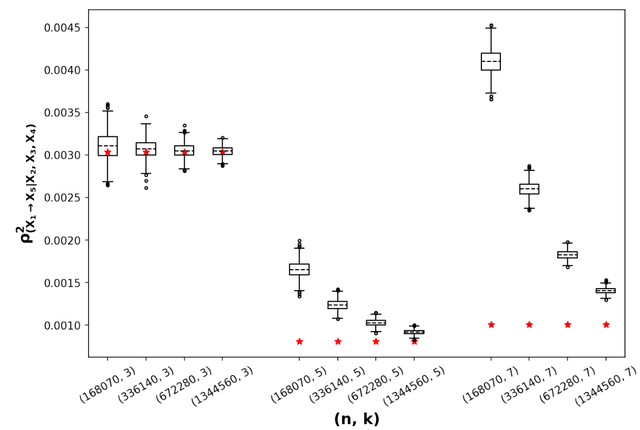
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.34: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

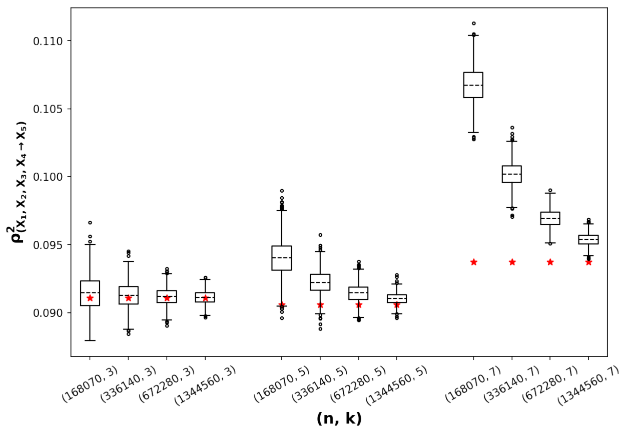
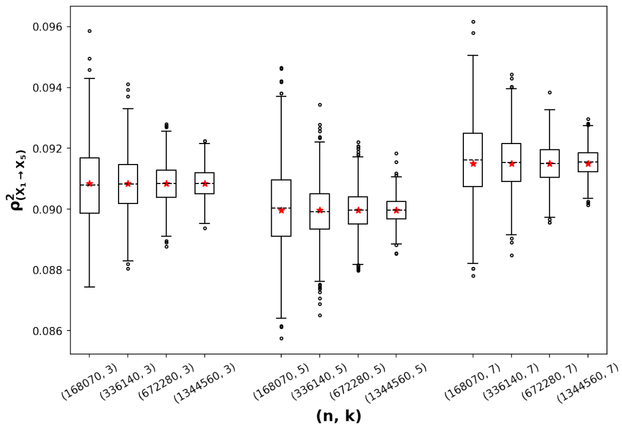
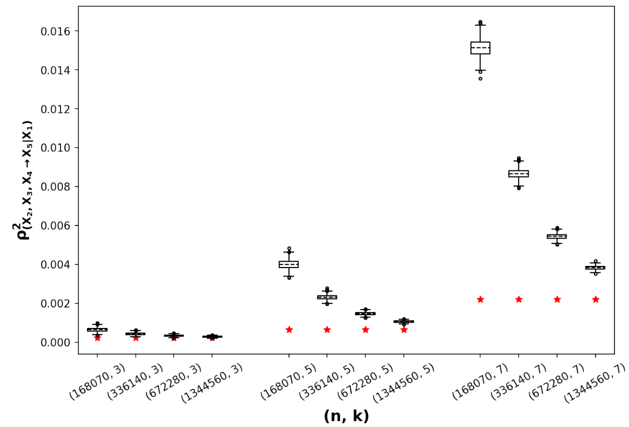


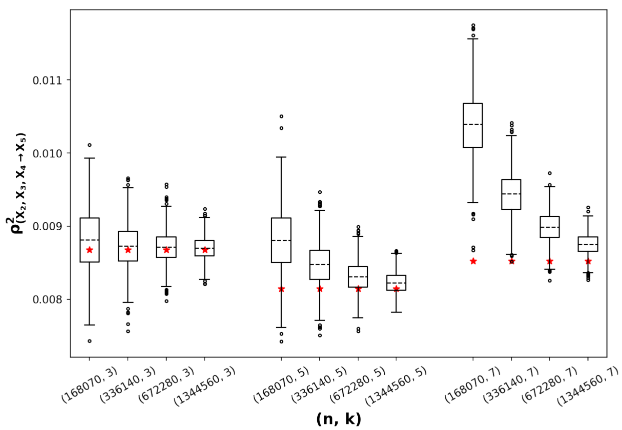
Figure K.35: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case



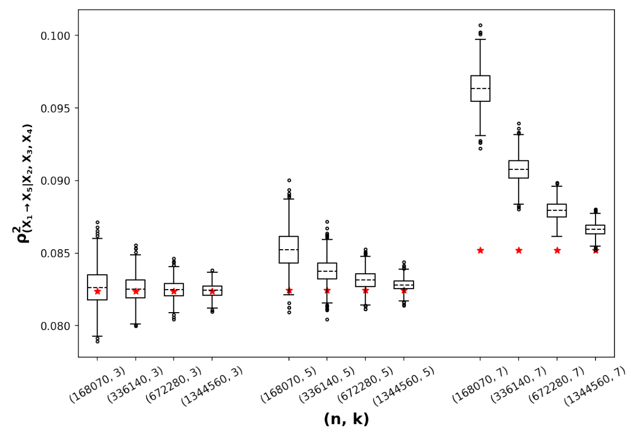
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.36: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case

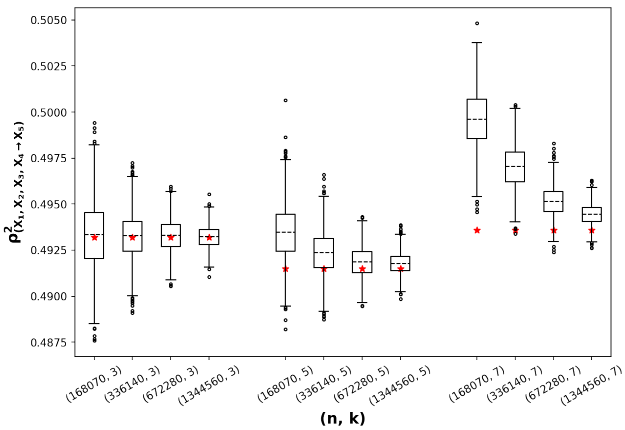


Figure K.37: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case

K.3 Scenario 4

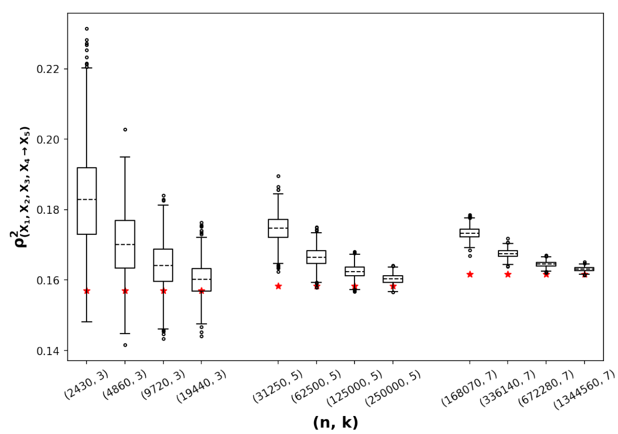
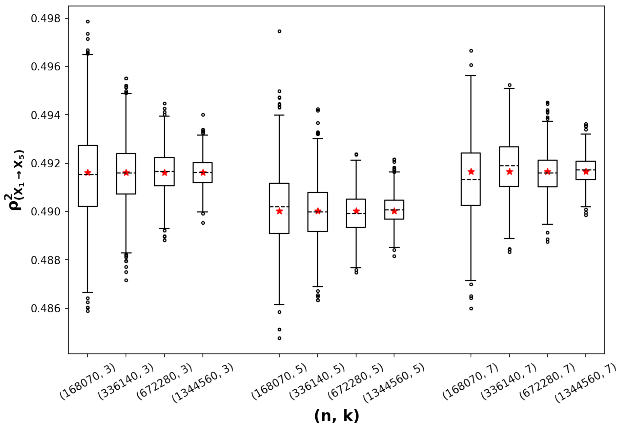
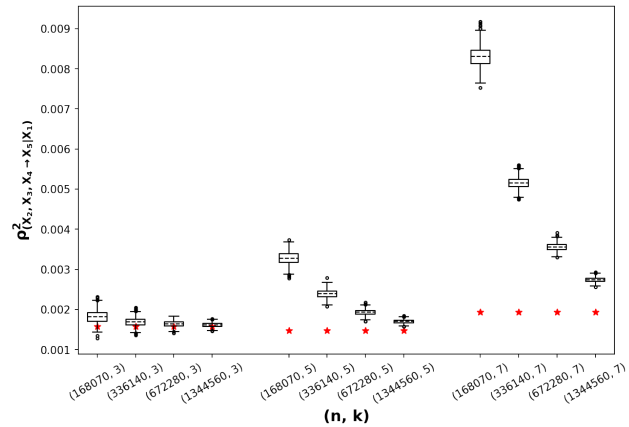


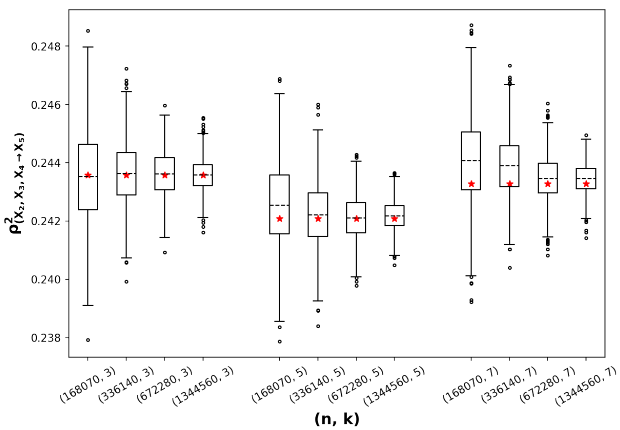
Figure K.39: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



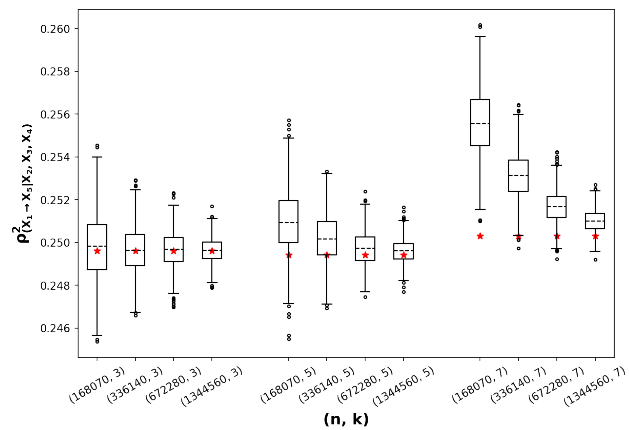
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$

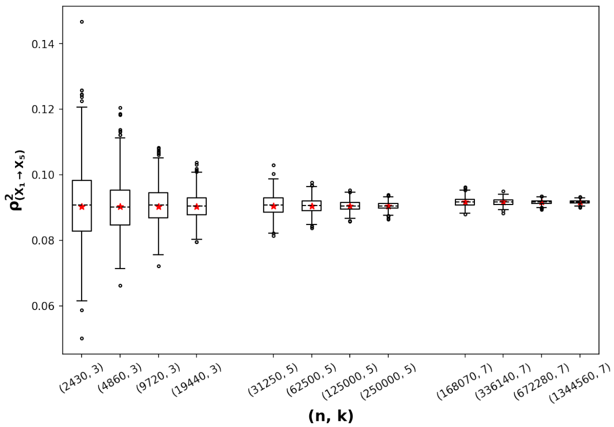


(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$

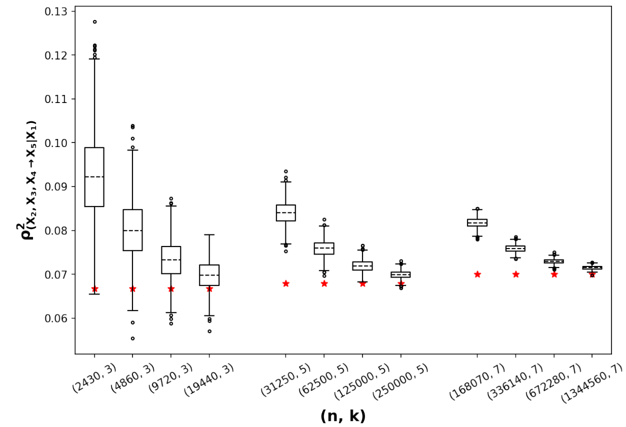


(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

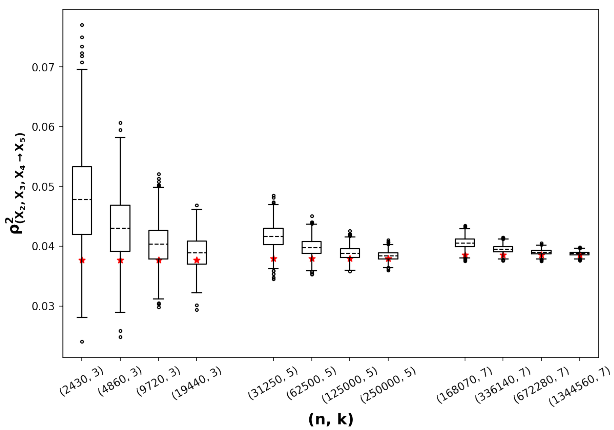
Figure K.38: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case



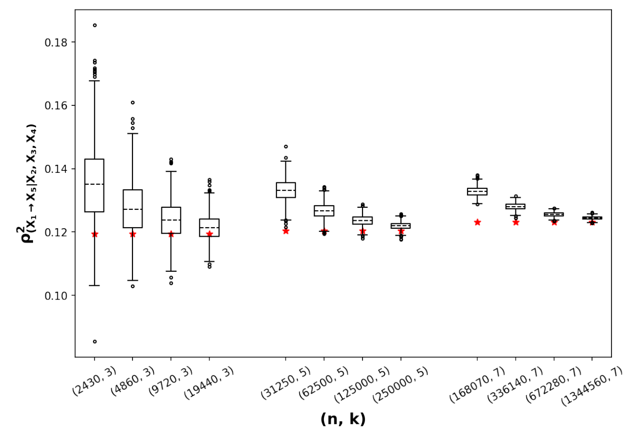
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.40: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

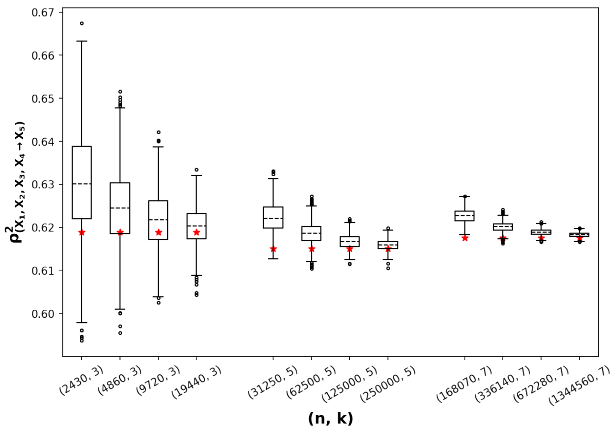
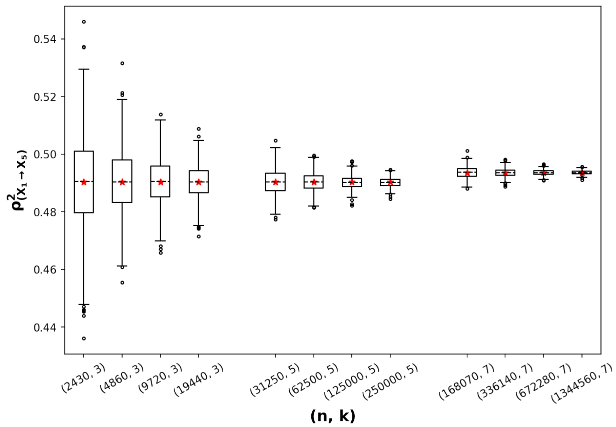
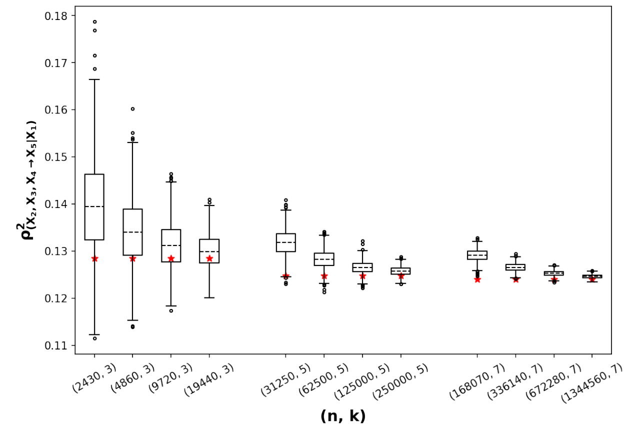


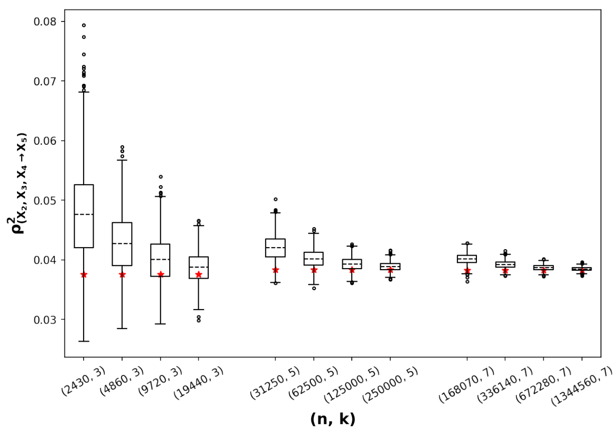
Figure K.41: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



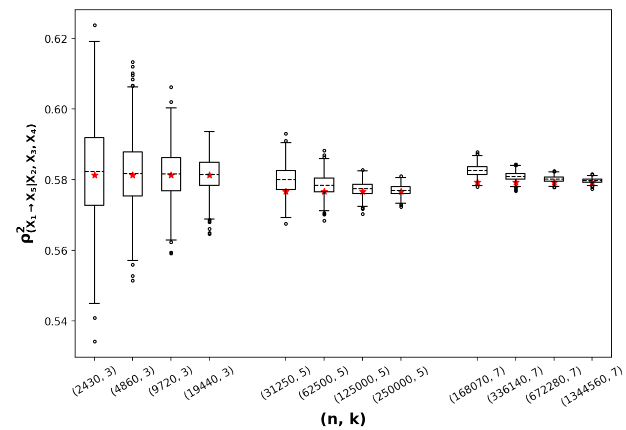
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.42: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) > 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

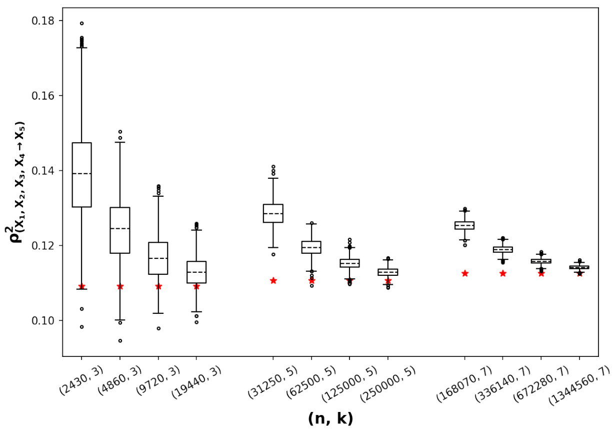
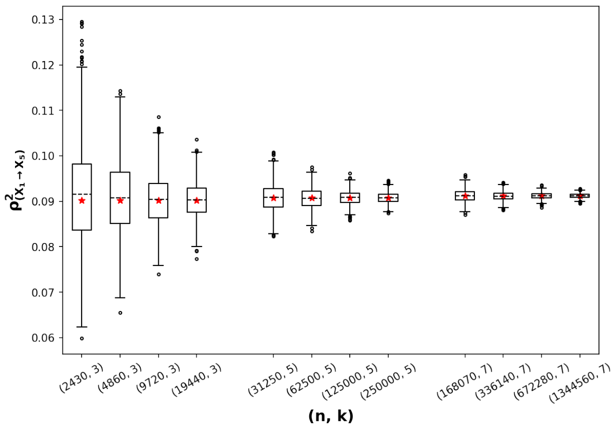
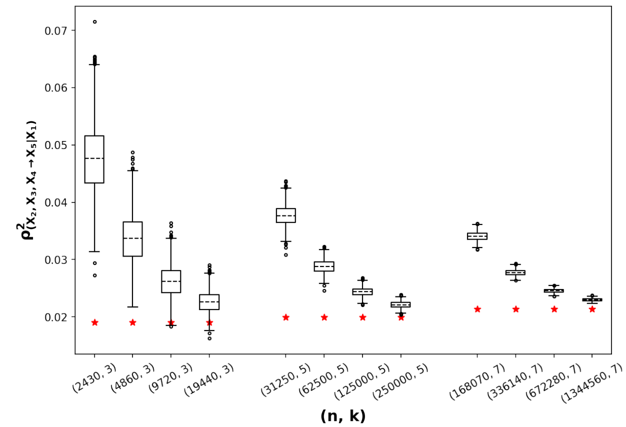


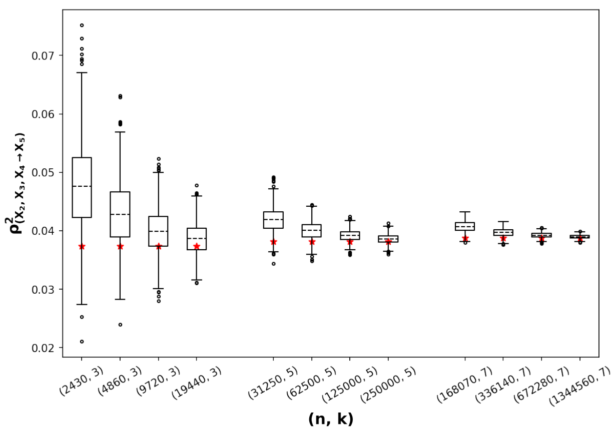
Figure K.43: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



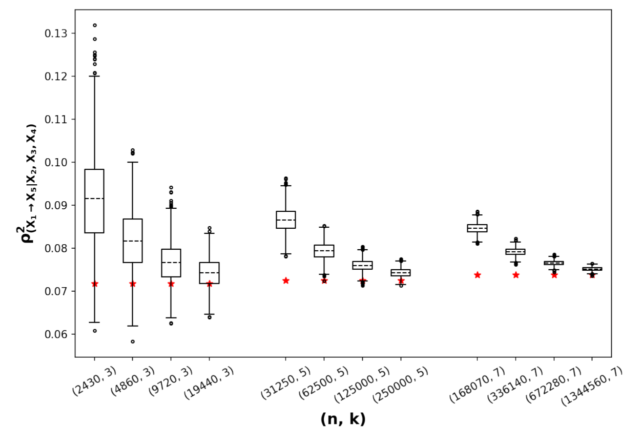
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.44: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

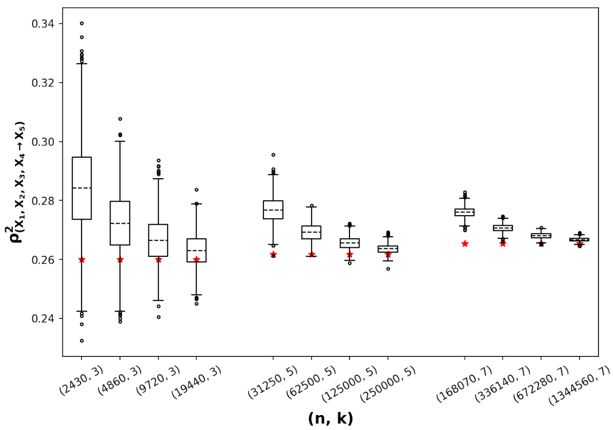
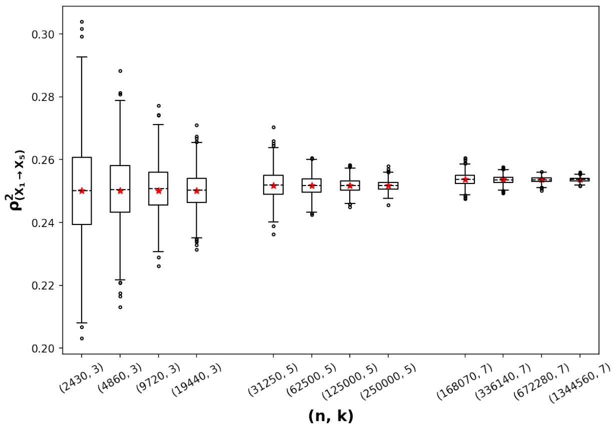
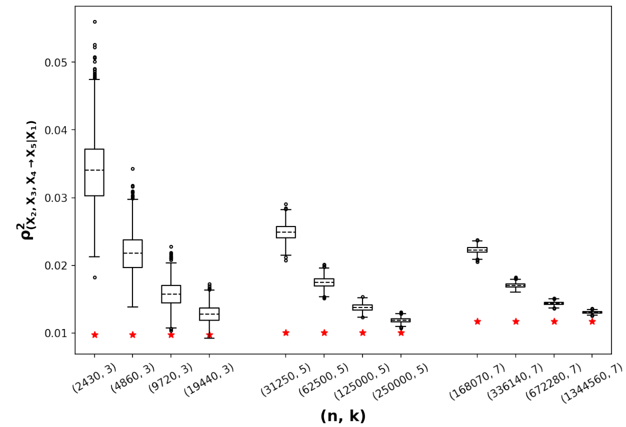


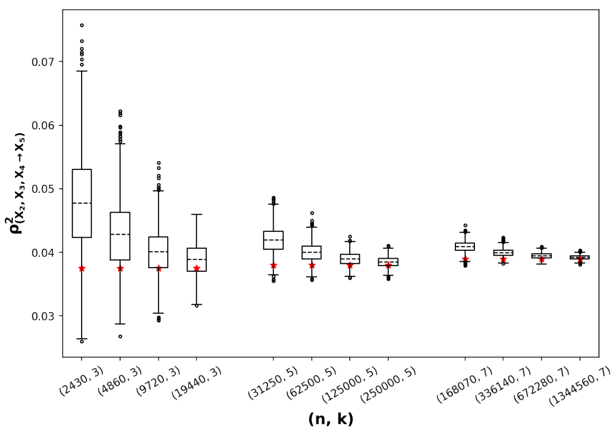
Figure K.45: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case



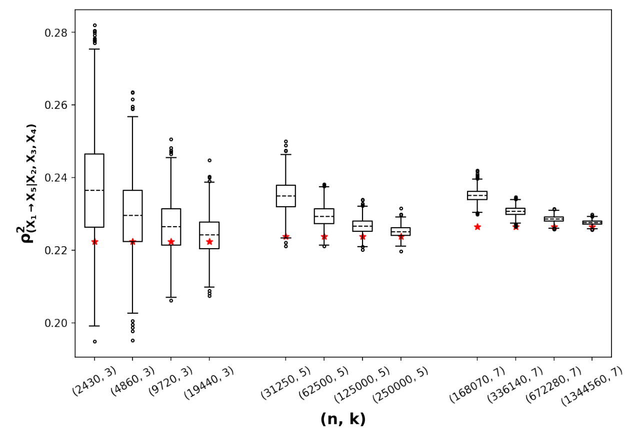
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.46: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) < 0$ and moderate association $|corr(X_5, X_1)| = 0.5$ in five-dimensional case

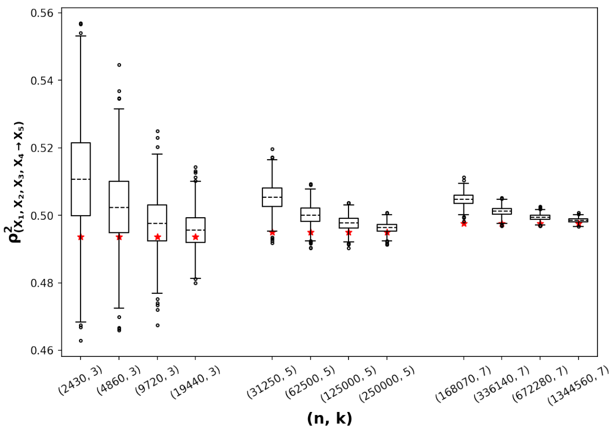
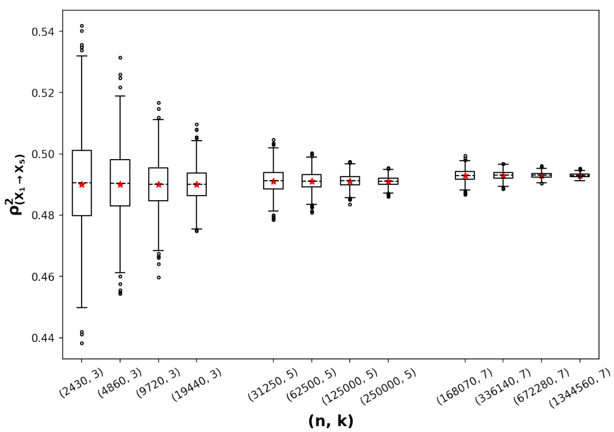
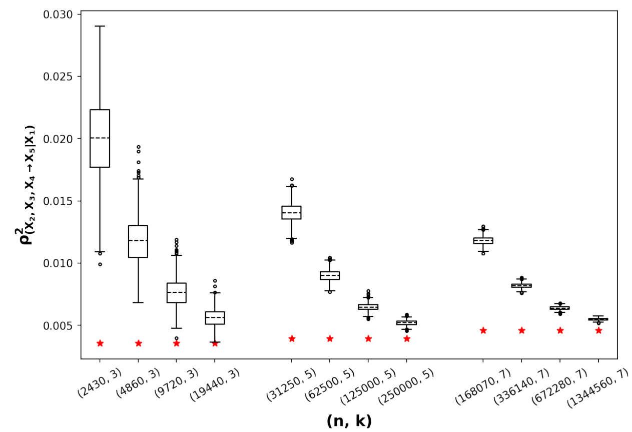


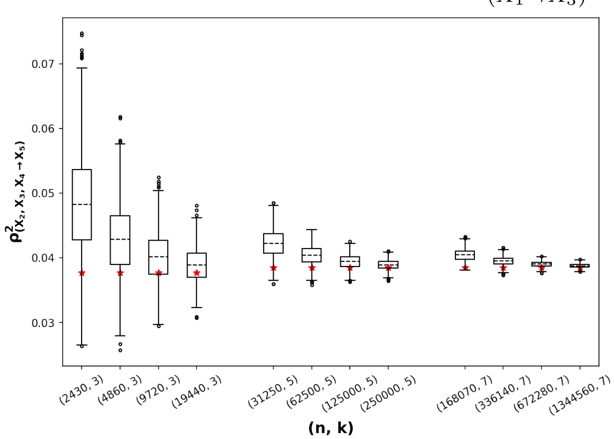
Figure K.47: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



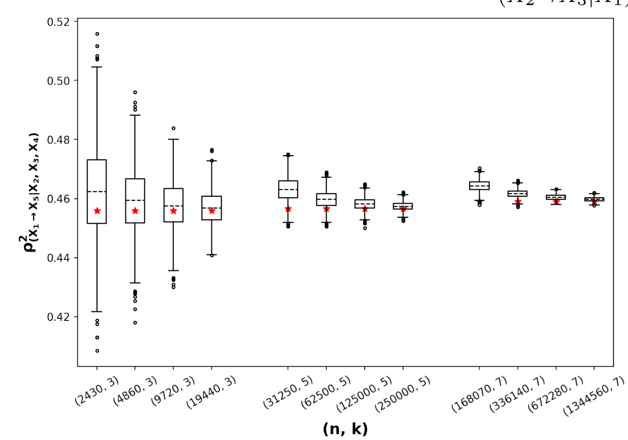
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.48: The marginal and conditional association measures for $pcorr(X_5, X_1 | X_2, X_3, X_4) < 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

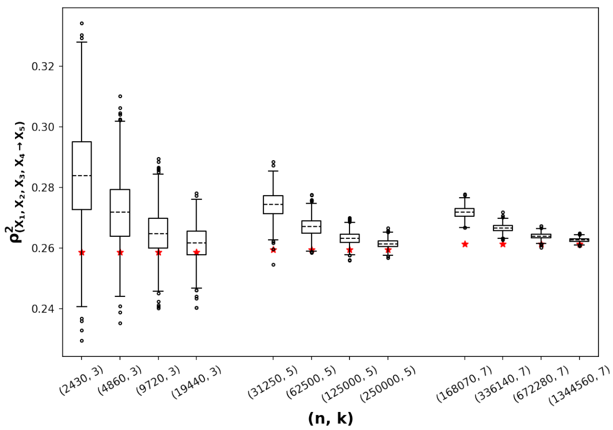
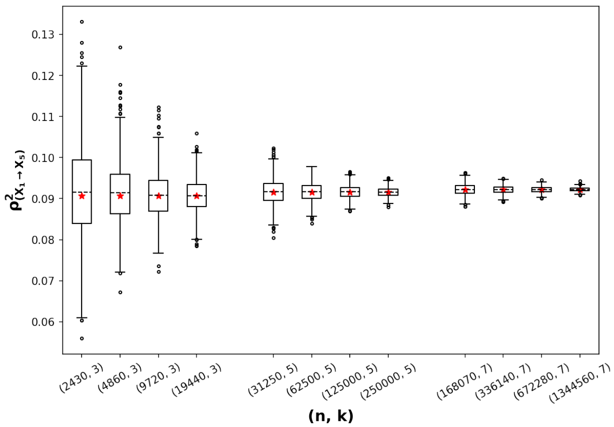
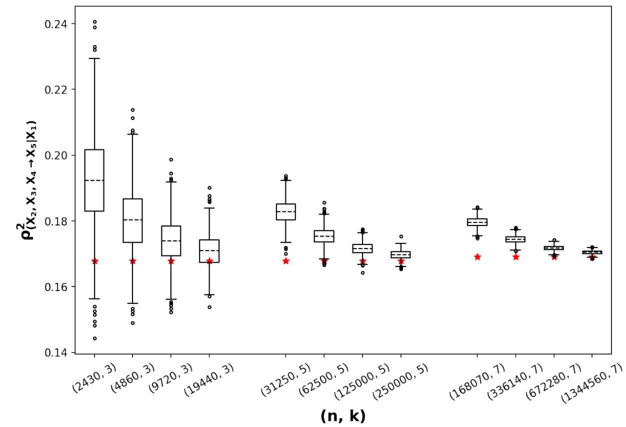


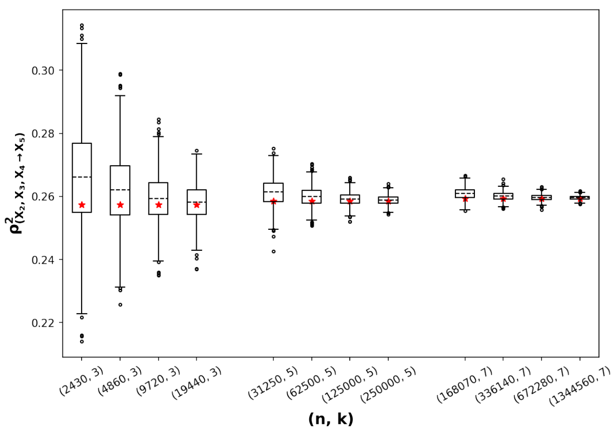
Figure K.49: the overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case



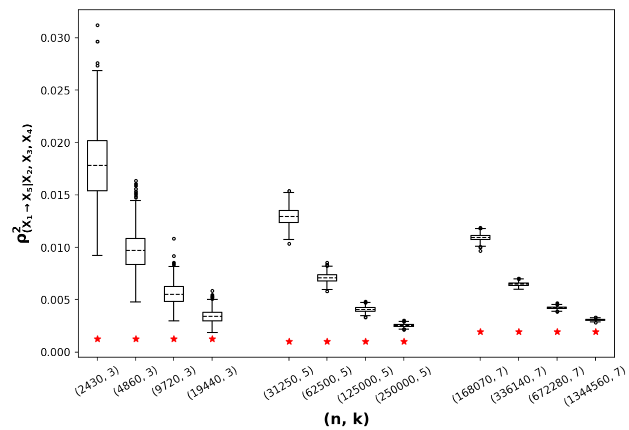
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.50: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and weak association $|corr(X_5, X_1)| = 0.3$ in five-dimensional case

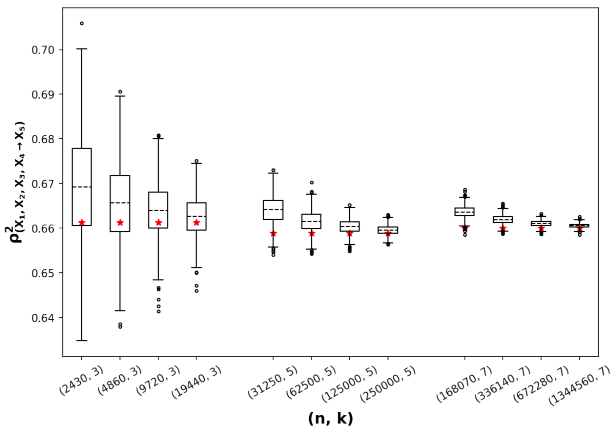
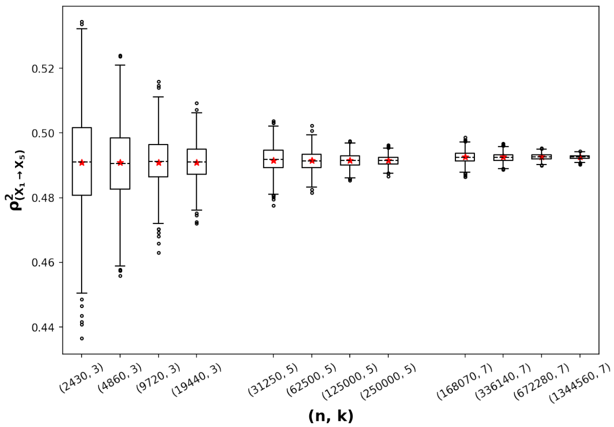
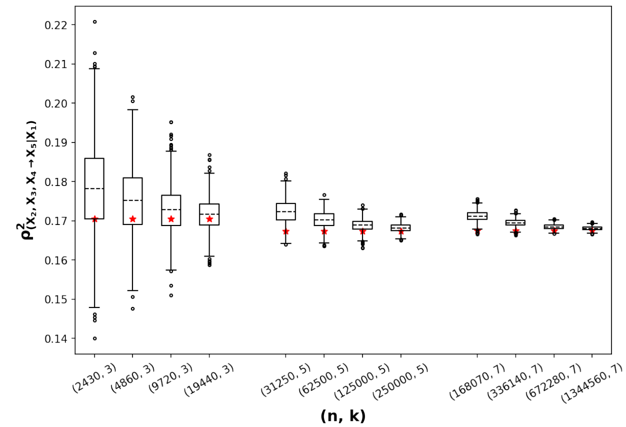


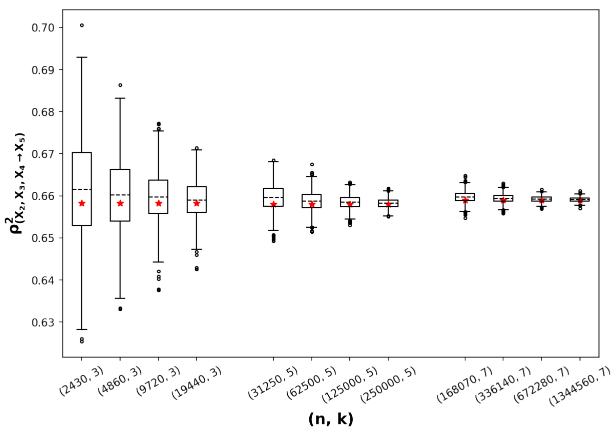
Figure K.51: The overall association measure for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case



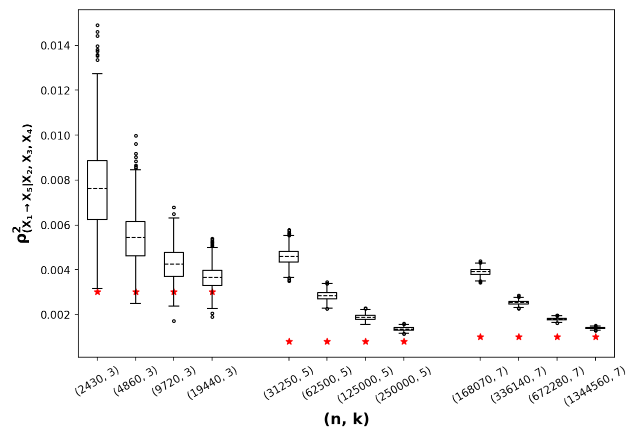
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.52: The marginal and conditional association measures for $pcorr(X_5, X_1|X_2, X_3, X_4) = 0$ and strong association $|corr(X_5, X_1)| = 0.7$ in five-dimensional case

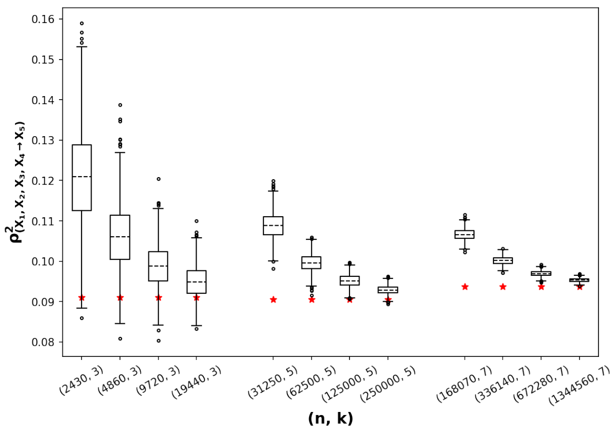
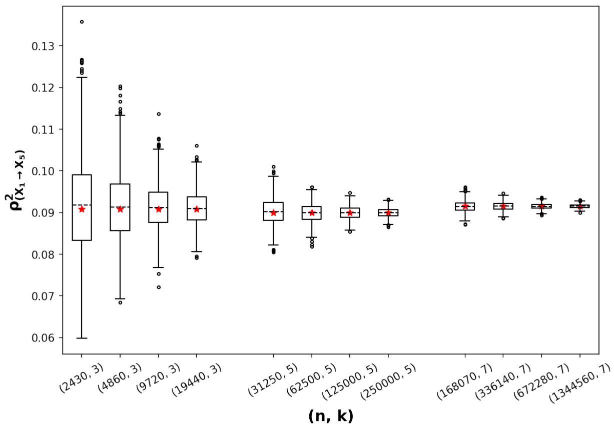
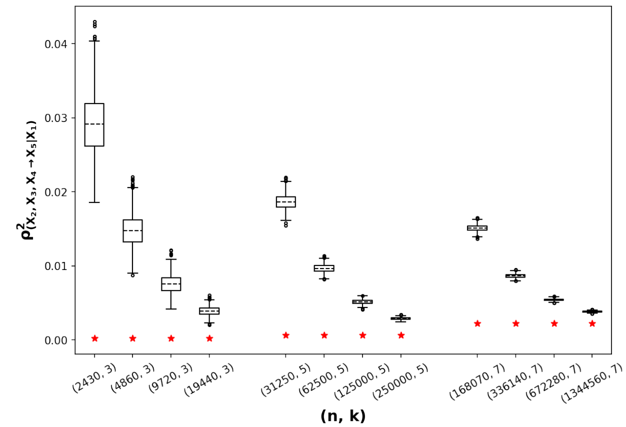


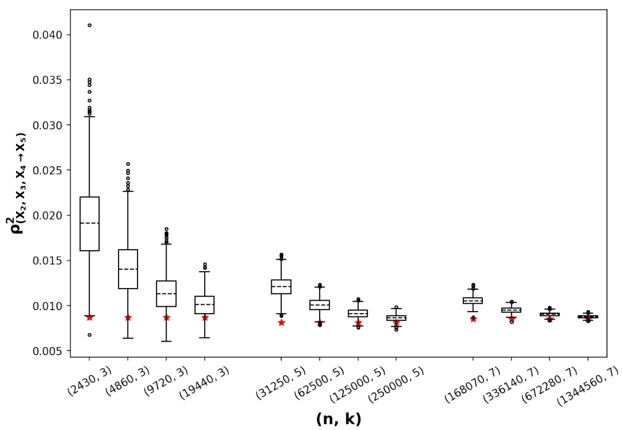
Figure K.53: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case



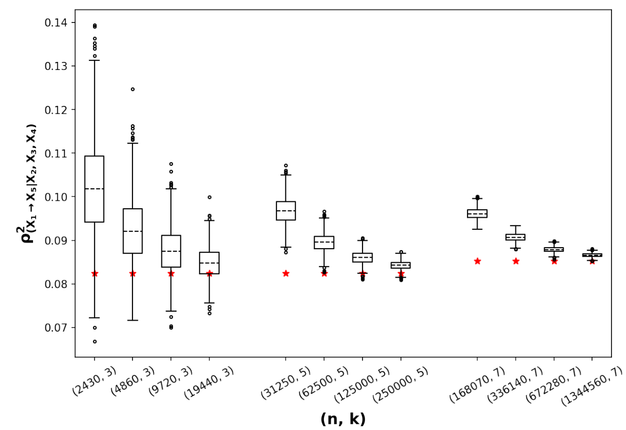
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3|X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3|X_2)}$

Figure K.54: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and weak association $|\phi| = |corr(X_5, X_1)| = 0.3$ in five-dimensional case

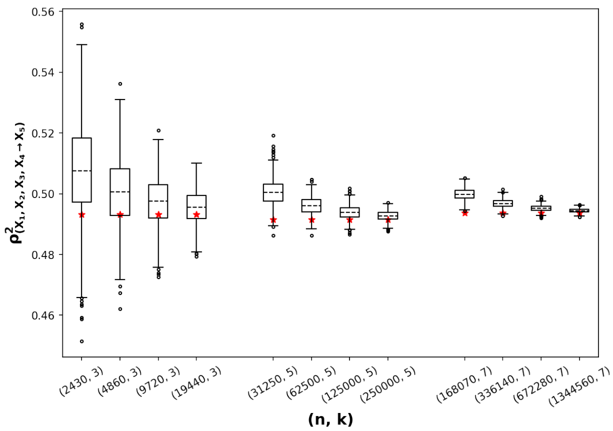
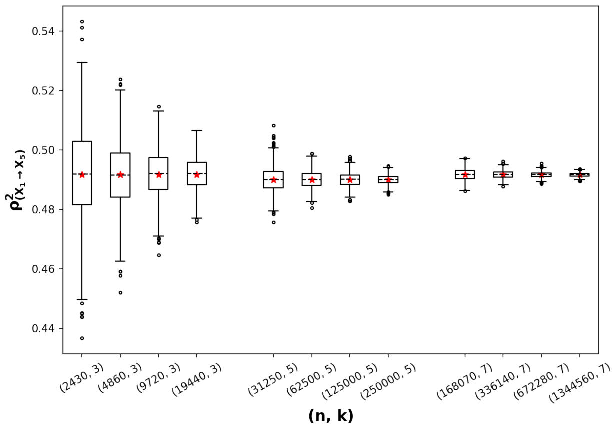
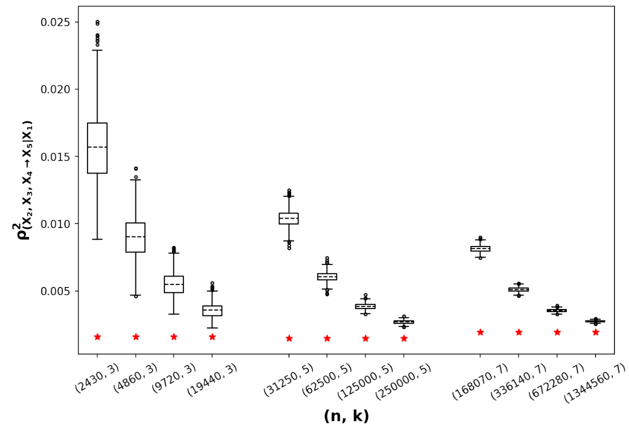


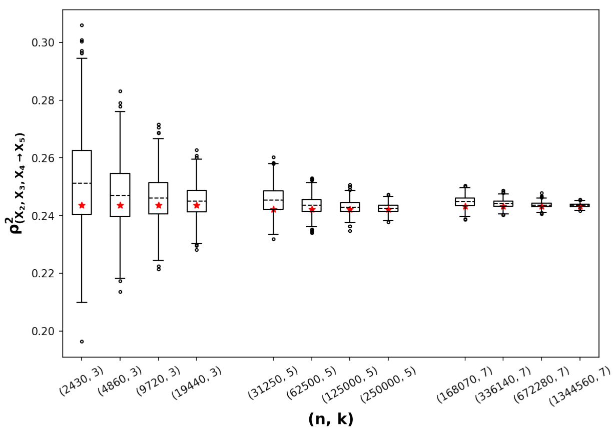
Figure K.55: The overall association measure for $auto1corr(X_i, X_j)$ where $i, j = 1, \ldots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case



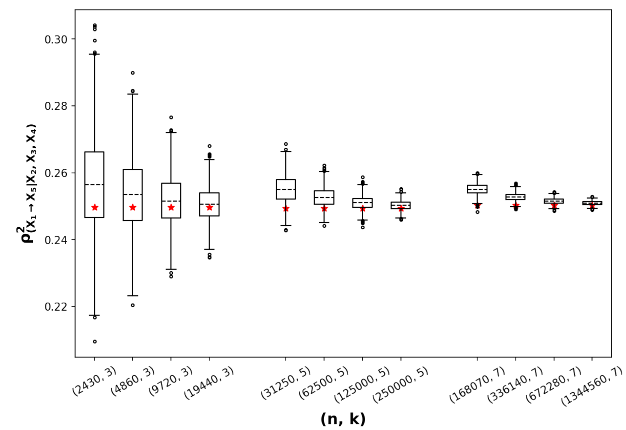
(a) The marginal association measure $\rho^2_{(X_1 \to X_3)}$



(b) The conditional association measure $\rho^2_{(X_2 \to X_3 | X_1)}$



(c) The marginal association measure $\rho^2_{(X_2 \to X_3)}$



(d) The conditional association measure $\rho^2_{(X_1 \to X_3 | X_2)}$

Figure K.56: The marginal and conditional association measures for $auto1corr(X_i, X_j)$ where $i, j = 1, \dots, 5$ and strong association $|\phi| = |corr(X_5, X_1)| = 0.7$ in five-dimensional case

BIBLIOGRAPHY

Agresti, A. (2002), *Categorical data analysis*, John Wiley & Sons.

Agresti, A. (2010), *Analysis of ordinal categorical data*, Vol. 656, John Wiley & Sons.

Anderson, R. J. and Landis, J. R. (1982), Catanova for multidimensional contingency tables: ordinal-scale resfqnse, *Communications in Statistics-Theory and Methods*, 11(3), 257–270.

Baglivo, J., Olivier, D. and Pagano, M. (1988), Methods for the analysis of contingency tables with large and small cell counts, *Journal of the American Statistical Association*, 83(404), 1006–1013.

Barbiero, A. and Ferrari, P. A. (2015), *GenOrd: Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions*. R package version 1.4.0.

Beh, E. J. and Lombardo, R. (2014), *Correspondence analysis: theory, practice and new strategies*, John Wiley & Sons.

Beh, E. J., Simonetti, B. and D'Ambra, L. (2007), Partitioning a non-symmetric measure of association for three-way contingency tables, *Journal of Multivariate Analysis*, 98(7), 1391–1411.

Benjamini, Y. and Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289–300.

Benjamini, Y. and Yekutieli, D. (2001), The control of the false discovery rate in multiple testing under dependency, *Annals of statistics*, pp. 1165–1188.

Berry, K. J., Johnston, J. E. and Mielke, J. P. W. (2018), *The measurement of association: a permutation statistical approach*, Springer.

Berry, K. J., Johnston, J. E. and Mielke, J. P. W. (2019), *A primer of permutation statistical methods*, Springer.

Berry, K. J., Johnston, J. E. and Mielke Jr, P. W. (2011), Permutation methods, *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(6), 527–542.

Bonferroni, C. (1936), Teoria statistica delle classi e calcolo delle probabilita, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8, 3–62.

Borgoni, R. (2004), A two-step smoothing procedure for the analysis of sparse contingency tables with ordered categories, *Statistica*, 64(4), 751–766.

Box, G. and Jones, S. (1986), Testing in industrial experiments with ordered categorical data: Discussion, *Technometrics*, 28(4), 295–301.

Brockett, P. L. (1981), A note on the numerical assignment of scores t0 ranked categorical data, *Journal of Mathematical Sociology*, 8, 91–101.

Bross, I. D. (1958), How to use ridit analysis, *Biometrics*, pp. 18–38.

Budihardjo, A., Grzymala-Busse, J. and Woolery, L. (1991), Program lers-lb 2.5 as a tool for knowledge acquisition in nursing, *in* 'Proceedings of the 4th Int. Conference on Industrial & Engineering Applications of AI & Expert Systems', pp. 735–740.

Carlier, A. and Kroonenberg, P. M. (1996), Decompositions and biplots in three-way correspondence analysis, *Psychometrika*, 61(2), 355–373.

Chen, H.-C. and Wang, N.-S. (2014), The assignment of scores procedure for ordinal categorical data, *The Scientific World Journal*, 2014.

Coull, B. A. and Agresti, A. (2003), Generalized log-linear models with random effects, with application to smoothing contingency tables, *Statistical Modelling*, 3(4), 251–271.

Cressie, N. and Read, T. R. (1984), Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3), 440–464.

Dash, S. R. and Dehuri, S. (2013), Comparative study of different classification techniques for post operative patient dataset, *International Journal of Innovative Research in Computer and Communication Engineering*, 1(5), 1101–1108.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, Vol. 1, Cambridge university press.

Doran, G., Muandet, K., Zhang, K. and Schölkopf, B. (2014), A permutation-based kernel conditional independence test., *in* 'Proccedings of Uncertainty in Artificial Intelligence 2014', pp. 132–141.

Efron, B. (1981), Nonparametric standard errors and confidence intervals, *canadian Journal of Statistics*, 9(2), 139–158.

Efron, B. (1987), Better bootstrap confidence intervals, *Journal of the American statistical Association*, 82(397), 171–185.

Efron, B. and Tibshirani, R. J. (1994), *An introduction to the bootstrap*, CRC press.

Ferrari, P. A. and Barbiero, A. (2012), Simulating ordinal data, *Multivariate Behavioral Research*, 47(4), 566–589.

Fielding, A. (1993), Scoring functions for ordered classifications in statistical analysis, *Quality and Quantity*, 27(1), 1–17.

Fienberg, S. E. (2007), *The analysis of cross-classified categorical data*, Springer Science & Business Media.

Fienberg, S. E., Bromet, E. J., Follmann, D., Lambert, D. and May, S. M. (1985), Longitudinal analysis of categorical epidemiological data: a study of three mile island., *Environmental Health Perspectives*, 63, 241–248.

Fisher, R. A. (1992), Statistical methods for research workers, *in* 'Breakthroughs in statistics', Springer, pp. 66–70.

Fox, J. and Weisberg, S. (2018), *An R companion to applied regression*, Sage publications.

Freeman, M. F. and Tukey, J. W. (1950), Transformations related to the angular and the square root, *The Annals of Mathematical Statistics*, pp. 607–611.

García-Pérez, M. A. and Núñez-Antón, V. (2009), Accuracy of power-divergence statistics for testing independence and homogeneity in two-way contingency tables, *Communications in Statistics—Simulation and Computation®*, 38(3), 503–512.

Genest, C. and Nešlehová, J. (2007), A primer on copulas for count data, *ASTIN Bulletin: The Journal of the IAA*, 37(2), 475–515.

Good, I. (1975), The number of hypotheses of independence for a random vector or for a multidimensional contingency table, and the bell numbers, *Iranian Journal of Science and Technology*, 4, 77–83.

Goodman, L. and Kruskal, W. H. (1954), Measures of association for cross classifications, *Journal of the American Statistical Association*, 49(268), 732–764.

Gray, L. N. and Williams, J. S. (1981), Goodman and kruskal's tau b: multiple and partial analogs, *Sociological Methods & Research*, 10(1), 50–62.

Grzymala-Busse, J. W. (1997), A new version of the rule induction system lers, *Fundamenta Informaticae*, 31(1), 27–39.

Guo, S. W. and Thompson, E. A. (1989), Analysis of sparse contingency tables: Monte carlo estimation of exact p-values, *Technical Report*, 187.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009), The weka data mining software: an update, *ACM SIGKDD explorations newsletter*, 11(1), 10–18.

Harrell, F. E. (2015), Ordinal logistic regression, *in* 'Regression modeling strategies', Springer, pp. 311–325.

Holm, S. (1979), A simple sequentially rejective multiple test procedure, *Scandinavian journal of statistics*, pp. 65–70.

Kateri, M. (2014), Contingency table analysis, *Statistics for Industry and Technology*, 525.

Kendall, M. G. (1938), A new measure of rank correlation, *Biometrika*, 30(1/2), 81–93.

Kendall, M. G. (1945), The treatment of ties in ranking problems, *Biometrika*, pp. 239–251.

Kendall, M. G. (1948), Rank correlation methods.

Kullback, S. (1997), *Information theory and statistics*, Courier Corporation.

Liebetrau, A. M. (1983), *Measures of association*, Vol. 32, Sage.

Light, R. J. and Margolin, B. H. (1971), An analysis of variance for categorical data, *Journal of the American Statistical Association*, 66(335), 534–544.

Likert, R. (1932), A technique for the measurement of attitudes, *Archives of psychology*, 140, 55.

Liu, I. and Agresti, A. (2005), The analysis of ordered categorical data: An overview and a survey of recent developments, *Test*, 14(1), 1–73.

Lombardo, R. and Camminatiello, I. (2010), Catanova for two-way cross classified categorical data, *Statistics*, 44(1), 57–71.

Mantel, N. (1963), Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure, *Journal of the American Statistical Association*, 58(303), 690–700.

Marcotorchino, F. (1984*a*), Utilisation des comparaisons par paires en statistique des contingencies: Partie i, *Report# F069*, 69.

Marcotorchino, F. (1984*b*), Utilisation des comparaisons par paires en statistique des contingencies: Partie ii, *Report# F070*, 69.

Marcotorchino, F. (1984*c*), Utilisation des comparaisons par paires en statistique des contingencies: Partie iii, *Report# F071*, 69.

Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science & Business Media.

Nešlehová, J. (2007), On rank correlation measures for non-continuous random variables, *Journal of Multivariate Analysis*, 98(3), 544–567.

Neyman, J. (1949), Contribution to the theory of the x2 test, *in* 'Proceedings of the Berkeley symposium on mathematical statistics and probability', Vol. 1, University of California Press Berkeley, pp. 239–273.

Neyman, J. and Pearson, E. S. (1928*a*), On the use and interpretation of certain test criteria for purposes of statistical inference: Part i, *Biometrika*, pp. 175–240.

Neyman, J. and Pearson, E. S. (1928*b*), On the use and interpretation of certain test criteria for purposes of statistical inference: Part ii, *Biometrika*, pp. 263–294.

Pearson, K. (1895), Notes on regression and inheritance in the case of two parents proceedings of the royal society of london, 58, 240-242.

Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157–175.

Pitman, E. J. (1937*a*), Significance tests which may be applied to samples from any populations, *Supplement to the Journal of the Royal Statistical Society*, 4(1), 119–130.

Pitman, E. J. G. (1937*b*), Significance tests which may be applied to samples from any populations. ii. the correlation coefficient test, *Supplement to the Journal of the Royal Statistical Society*, 4(2), 225–232.

Pitman, E. J. G. (1938), Significance tests which may be applied to samples from any populations: Iii. the analysis of variance test, *Biometrika*, 29(3/4), 322–335.

Roy, S. N. and Mitra, S. K. (1956), An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis, *Biometrika*, 43(3-4), 361–376.

Schweizer, B. et al. (1974), Operations on distribution functions not derivable from operations on random variables., *Studia Mathematica*, 52(1), 43–52.

Shmueli, G. et al. (2010), To explain or to predict?, *Statistical science*, 25(3), 289–310.

Sklar, M. (1959), Fonctions de repartition an dimensions et leurs marges, *Publ. inst. statist. univ. Paris*, 8, 229–231.

Somers, R. H. (1962), A new asymmetric measure of association for ordinal variables, *American sociological review*, pp. 799–811.

Spearman, C. (1904), The proof and measurement of association between two things, *The American journal of psychology*, 15, 72–101.

The Ice Cream Study at Penn State (2012), `https://online.stat.psu.edu/stat504/node/187/`.

Theil, H. (1970), On the estimation of relationships involving qualitative variables, *American Journal of Sociology*, 76(1), 103–154.

Thomas, G. and Kiwanga, S. (1993), Use of ranking and scoring methods in the analysis of ordered categorical data from factorial experiments, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 42(1), 55–67.

Tsamardinos, I. and Borboudakis, G. (2010), Permutation testing improves bayesian network learning, *in* 'Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases', Springer, pp. 322–337.

Tutz, G. (2011), *Regression for categorical data*, Vol. 34, Cambridge University Press.

Vandenhende, F. and Lambert, P. (2000), Modeling repeated ordered categorical data using copulas. Discussion Paper No.00-25: www.stat.ucl.ac.be/pub/papers/dp/dp00/dpO025.ps Institut de statistique, Uni versit.

Wei, Z. and Kim, D. (2017), Subcopula-based measure of asymmetric association for contingency tables, *Statistics in medicine*, 36(24), 3875–3894.

Wei, Z. and Kim, D. (2021), On exploratory analytic method for multi-way contingency tables with an ordinal response. Manuscript submitted for publication (copy on file with auther).

Wilks, S. (1935), The likelihood test of independence in contingency tables, *The annals of mathematical statistics*, 6(4), 190–196.

Woolery, L., Grzymala-Busse, J., Summers, S. and Budihardjo, A. (1991), The use of machine learning program lers-lb 2.5 in knowledge acquisition for expert system development in nursing, *Computers in nursing*, 9(6), 227–234.