Testing Independence in High Dimensions
&
Identifiability of Graphical Models

Dennis Leung

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Mathias Drton, Chair

Thomas Richardson

Jon Wakefield

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Testing Independence in High Dimensions
&
Identifiability of Graphical Models

Dennis Leung

Chair of the Supervisory Committee:
Professor Mathias Drton
Department of Statistics

In this thesis two problems in multivariate statistics will be studied.

In the first chaper, we treat the problem of testing independence between m continuous observations when m can be larger than the available sample size n. We consider three types of test statistics that are constructed as sums of many pairwise rank correlation signals. In the asymptotic regime where both m and n converge to ∞, a martingale central limit theorem is applied to show that the null distributions of these statistics converge to Gaussian limits, which are valid with no specific distributional or moment assumptions on the data. Using the framework of U-statistics, our result covers a variety of rank correlations including Kendall's tau and a dominating term of Spearman's rank correlation coefficient (rho), but also degenerate U-statistics such as Hoeffding's D, or the τ^* of Bergsma and Dassios (2014). Like the classical theory for U-statistics, the test statistics need to be scaled differently when the rank correlations used to construct them are degenerate U-statistics. The power of the considered tests is explored in rate-optimality theory under a Gaussian equicorrelation alternative as well as in numerical experiments for specific cases of more general alternatives.

In the second chapter, we study parameter identifiability of directed Gaussian graphical models with one latent variable. In the scenario we consider, the latent variable is a confounder that forms a source node of the graph and is a parent to all other nodes, which correspond to the observed variables. We give a graphical condition that is sufficient for the Jacobian matrix of the parametrization map to be full rank, which entails that the parametrization is generically finite-to-one, a fact that is sometimes also referred to as local identifiability. We also derive a graphical condition that is necessary for such identifiability. Finally, we give a condition under which generic parameter identifiability can be determined from identifiability of a model associated with a subgraph. The power of these criteria is assessed via an exhaustive algebraic computational study for small models with 4, 5, and 6 observable variables, and a simulation study for large models with 25 or 35 observable variables.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

ACKNOWLEDGMENTS

The author would like to give thanks to his advisor Mathias Drton for both his mentorship and friendship. The author has learned more than statistics from him and will continue to seek his advices in future. He would also like to acknowledge all other senior faculty in the Department of Statistics at the University of Washington who have helped him make his way through the PhD program, in particular Thomas Richardson, Jon Wellner and Jon Wakefield.

Lastly, he would like to express great gratitude for the love of his mother over the years.

Chapter 1

NON-PARAMETRIC TESTING OF INDEPENDENCE IN HIGH DIMENSIONS

1.1 Introduction

This chapter is concerned with nonparametric tests of independence between the coordinates of a continuous random vector $\mathbf{X} = (X^{(1)}, \ldots, X^{(m)})$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d. sample, with each $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(m)})$ following the same distribution as $\mathbf{X}$. We then wish to test the null hypothesis

$$H_0 : \; X^{(1)}, \ldots, X^{(m)} \text{ are independent.} \tag{1.1.1}$$

The natural approach is to form a test statistic that measures the dependence among $X^{(1)}, \ldots, X^{(m)}$ based on the sample, and reject H_0 when its value is too large, where the critical value of rejection is calibrated by the asymptotic distribution of the test statistic under the null. Our focus is on the use of rank correlations in problems where the dimension m can be of a larger order than the sample size n. Specifically, our testing procedures will be studied under the asymptotic regime where $m = m(n)$ grows as a function of n such that m also tends to infinity. This regime will be denoted by $m, n \longrightarrow \infty$ throughout our chapter.

In the literature, one line of work uses the *maximum* of many pairwise dependency measures to test for (1.1.1). For $p = 1, \ldots, m$, let $\mathbf{X}^{(p)} = (X_1^{(p)}, \ldots, X_n^{(p)})$ be the sample of observations for the p-th variable. For $1 \le p \ne q \le m$, let $r^{(pq)}$ denote the sample Pearson (product-moment) correlation of $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$. Jiang (2004) proved that, after a squaring transformation and suitable renormalization, the null distribution of the statistic

$$\max_{1 \le p < q \le m} \left| r^{(pq)} \right| \tag{1.1.2}$$

converges to an extreme value distribution of type 1 when m/n converges to a constant

$\gamma \in (0, \infty)$ as $m, n \longrightarrow \infty$ (abbreviated as $m/n \longrightarrow \gamma \in (0, \infty)$ below). He assumed higher-order moment conditions that were weakened in subsequent work (Li et al., 2010, 2012, Liu et al., 2008, Zhou, 2007). Cai and Jiang (2011) derived a similar asymptotic distribution for the statistic from (1.1.2), allowing for subexponential growth in the dimension m. Further weakening distributional assumptions, the recent work of Han and Liu (2014) treated maxima of rank correlations, that is, the sample Pearson correlation in (1.1.2) is replaced by a rank correlation measure such as Kendall's tau. This maximum was shown to have a similar extreme value type null distribution. Statistics such as (1.1.2) are of obvious appeal when strong dependence is expected between some variables.

Our treatment, however, aligns with a more classical approach that is appealing when moderate dependence is expected between many variables. Such works base their tests on estimates of the *sum* of many pairwise dependency signals. Let $\Sigma = (\sigma^{(pq)})$ and $R = (\rho^{(pq)})$ be, respectively, the population covariance and correlation matrix of the random vector $\mathbf{X}$. Under Gaussian assumption for $\mathbf{X}$, Schott (2005) proposed the use of the "plug-in" estimate

$$S_r := \sum_{1 \leq p < q \leq m} (r^{(pq)})^2 \tag{1.1.3}$$

for the overall dependency signal $\sum_{1 \leq p < q \leq m} (\rho^{(pq)})^2$ and showed its asymptotic normality when the null hypothesis (1.1.1) is true, for the regime $m/n \longrightarrow \gamma \in (0, \infty)$. Often not mentioned in the literature, S_r is in fact the Rao score statistic in a multivariate normal setting; see Appendix A.1. Mao (2014) suggested a related statistic, namely, the sum of $f(r^{(pq)})$ for $f(x) = x^2/(1-x^2)$, and again the null distribution is shown to be asymptotically normal. For the two related problems of testing the equality and the proportionality of Σ to the identity matrix, similar "plug-in" statistics have been studied (John, 1972, Ledoit and Wolf, 2002, Nagao, 1973). Motivated by this approach, we construct our first class of test statistics by plugging in rank correlations to obtain nonparametric tests for (1.1.1). We will illustrate it for Kendall's tau. For $1 \leq p \neq q \leq m$, let

$$\tau^{(pq)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \operatorname{sgn}\left(X_i^{(p)} - X_j^{(p)}\right) \operatorname{sgn}\left(X_i^{(q)} - X_j^{(q)}\right) \tag{1.1.4}$$

be the sample Kendall's tau correlation coefficient for $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$. A natural test is then to reject H_0 for large values of the statistic

$$\sum_{1\leq p<q\leq m} (\tau^{(pq)})^2. \tag{1.1.5}$$

As an estimator of the dependency signal

$$\sum_{1\leq p<q\leq m} \left(\mathbb{E}\left[\tau^{(pq)}\right]\right)^2, \tag{1.1.6}$$

the "plug-in" statistic (1.1.5) is biased and hence needs to be recentered to obtain a mean 0 asymptotic null distribution under our considered regime $m, n \longrightarrow \infty$. As an alternative we may attempt to form an unbiased estimator of (1.1.6) to serve as a test statistic instead. As shown in Section 1.3, such an unbiased estimator is given by

$$\begin{aligned}\frac{1}{4!\binom{n}{4}} \sum \operatorname{sgn}\left(X^{(p)}_{i_{\pi(1)}} - X^{(p)}_{i_{\pi(2)}}\right) \operatorname{sgn}\left(X^{(p)}_{i_{\pi(3)}} - X^{(p)}_{i_{\pi(4)}}\right)& \\ \times \operatorname{sgn}\left(X^{(q)}_{i_{\pi(1)}} - X^{(q)}_{i_{\pi(2)}}\right) \operatorname{sgn}\left(X^{(q)}_{i_{\pi(3)}} - X^{(q)}_{i_{\pi(4)}}\right)&, \end{aligned} \tag{1.1.7}$$

where the summation is over all variable pairs $1 \leq p < q \leq m$, ordered 4-tuple of indices $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$ and permutations π on four elements. This type of test statistics is motivated by the work of Chen et al. (2010) and Cai and Ma (2013) who tested the equality of Σ to the identity based on unbiased estimates of the squared Frobenius norm $\|\Sigma - I_m\|_F^2$, where I_m is the m-by-m identity matrix. Under a Gaussian assumption for $\mathbf{X}$, Cai and Ma (2013) showed their resulting test to be asymptotically minimax rate optimal. As a last variant, when one is only interested in testing for *positive* associations between the m variables, it may be of interest to consider the statistic

$$\sum_{1\leq p<q\leq m} \tau^{(pq)} \tag{1.1.8}$$

formed by a simple sum of all pairwise sample Kendall's tau correlations to give a "one-sided" test. This way of constructing a statistic also provides a "two-sided" test for H_0 when rank correlations like τ^* of Bergsma and Dassios (2014) are used instead, as seen below. In

Section 1.4, we will show that all the statistics introduced above are asymptotically normal under suitable recentering and rescaling.

Kendall's tau is an example of a U-statistic whose values only depend on the data via ranks (van der Vaart, 1998, Example 12.5). Indeed, the values of (1.1.4), (1.1.7) and (1.1.8) remain unchanged if each observation $X_i^{(p)}$ is replaced with its rank $R_i^{(p)}$. To be specific, for each $p = 1, \ldots, m$, the rank $R_i^{(p)}$ is the rank of $X_i^{(p)}$ among $X_1^{(p)}, \ldots, X_n^{(p)}$. Other examples of measures of association that are both U-statistics and rank correlations are the D of Hoeffding (1948a) and the τ^* of Bergsma and Dassios (2014) mentioned. We note that for a pair of continuous random variables both of these statistics lead to consistent tests of independence, that is, their expectations are zero if and only if the two random variables are independent. Another classical example is Spearman's rho, which is not a U-statistic but can be approximated by a rank-based U-statistic.

The above examples of U-statistics are reviewed in Section 1.2, which also introduces a general framework of rank-based U-statistics that we adopt for a unified theory. In Section 1.3 we will construct our classes of test statistics for the null hypothesis H_0 (1.1.1). Their asymptotic null distributions when $m, n \longrightarrow \infty$ are derived in Section 1.4. Our arguments make use of a central limit theorem for martingale arrays and U-statistic theory. We emphasize that all our statistics admit a normal limit after appropriate rescaling, but just as in the classical theory for U-statistics, the scaling factors have a different order when degenerate U-statistics are considered. In Section 1.5, we will explore aspects of power of our tests from a minimax point of view. Simulation experiments will be presented in Section 1.6, which also discusses computational considerations in the implementation of the tests. Throughout, for our null distributional theory, we make no distributional or moment assumption on $(X^{(1)}, \ldots, X^{(m)})$ other than that it is a continuous random vector. This assumption is needed to avoid ties in observations and ranks. We conclude with a brief discussion in Section 1.7.

1.1.1 Notational convention

For $p \in \{1,\dots,m\}$, we let $\mathbf{R}^{(p)} := (R_1^{(p)},\dots,R_n^{(p)})$ be the vector of ranks of $\mathbf{X}^{(p)} = (X_1^{(p)},\dots,X_n^{(p)})$. The symmetric group of order l is denoted by $\mathfrak{S}_l$. Depending on the context, its elements will either be treated as permutation functions or ordered tuples of the set $\{1,\dots,l\}$. For $k \le n$, $\mathcal{P}(n,k)$ denotes the set of k-tuples $\mathbf{i} = (i_1,\dots,i_k)$ with $1 \le i_1 < \cdots < i_k \le n$, and we will also identify the tuple $\mathbf{i}$ with its set of elements $\{i_1,\dots,i_k\}$. Hence, for any two elements $\mathbf{i},\mathbf{j} \in \mathcal{P}(n,k)$, the operations $\mathbf{i}\cup\mathbf{j}$, $\mathbf{i}\cap\mathbf{j}$, and $\mathbf{i}\setminus\mathbf{j}$ give the tuples with increasing components that, as sets, equal the union, intersection and difference of $\mathbf{i}$ and $\mathbf{j}$ respectively. For $\mathbf{i} \in \mathcal{P}(n,k)$, we let $\mathbf{X}_{\mathbf{i}}^{(p)} := (X_{i_1}^{(p)},\dots,X_{i_k}^{(p)})$, and define the rank vector

$$\mathbf{R}_{\mathbf{i}}^{(p)} := \left(R_{\mathbf{i},1}^{(p)},\dots,R_{\mathbf{i},k}^{(p)}\right),$$

where $R_{\mathbf{i},c}^{(p)}$ is the rank of $X_{i_c}^{(p)}$ among $X_{i_1}^{(p)},\dots,X_{i_k}^{(p)}$.

Let $p \neq q$ index two distinct variables. Then $\mathbf{X}_c^{(pq)}$ and $\mathbf{R}_c^{(pq)}$ denotes the pairs $(X_c^{(p)},X_c^{(q)})$ and $(R_c^{(p)},R_c^{(q)})$, respectively, for $c = 1,\dots,n$. Similarly, given $\mathbf{i} = (i_1,\dots,i_k) \in \mathcal{P}(n,k)$, we let $\mathbf{X}_{\mathbf{i},c}^{(pq)} := (X_{i_c}^{(p)},X_{i_c}^{(q)})$ and $\mathbf{R}_{\mathbf{i},c}^{(pq)} := (R_{\mathbf{i},c}^{(p)},R_{\mathbf{i},c}^{(q)})$ for $c \in \{1,\dots,k\}$. We then define the observation and rank vectors of pairs

$$\mathbf{R}_{\mathbf{i}}^{(pq)} := \left(\mathbf{R}_{\mathbf{i},1}^{(pq)},\dots,\mathbf{R}_{\mathbf{i},k}^{(pq)}\right) \text{ and } \mathbf{X}_{\mathbf{i}}^{(pq)} := \left(\mathbf{X}_{\mathbf{i},1}^{(pq)},\dots,\mathbf{X}_{\mathbf{i},k}^{(pq)}\right).$$

When taking expectations under the null hypothesis H_0, we write $\mathbb{E}_0[\cdot]$, whereas $\mathbb{E}[\cdot]$ is the general expectation operator, possibly under alternative hypotheses. Similarly, we write $P_0[\cdot]$, $P[\cdot]$, $\mathrm{Var}_0[\cdot]$, $\mathrm{Var}[\cdot]$, $\mathrm{Cov}_0[\cdot]$ and $\mathrm{Cov}[\cdot]$ for the probability, variance and covariance operator under H_0 and possibly alternatives respectively. Finally, $\|\cdot\|_{\max}$ and $\|\cdot\|_2$ are the max norm and Euclidean norm for vectors, respectively, and the Froebenius norm of a matrix is written $\|\cdot\|_F$. For two sequences (a_n) and (b_n), $a_n \asymp b_n$ is used to indicate that there exist universal constants $c, C > 0$ such that $ca_n \le b_n \le Ca_n$.

1.2 Rank correlations as U-statistics

This section lays out a rank-based U-statistic framework that encompasses all rank correlations we will use to construct specific test statistics for H_0 in Section 1.3. Let

$$h : \left(\mathbb{R}^2\right)^k \longrightarrow \mathbb{R}$$

be a symmetric function of $k \geq 2$ arguments in $\mathbb{R}^2$, i.e., for all choices of $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})' \in \mathbb{R}^2$, $i = 1, \ldots, k$, and any permutation $\pi \in \mathfrak{S}_k$, it holds that $h\left(\mathbf{x}_1, \ldots, \mathbf{x}_k\right) = h\left(\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(k)}\right)$. For any pair of variable indices $p, q \in \{1, \ldots, m\}$, the function h yields a *U-statistic*

$$U_h^{(pq)} = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h\left(\mathbf{X}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{X}_{\mathbf{i},k}^{(pq)}\right) = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h\left(\mathbf{X}_{\mathbf{i}}^{(pq)}\right). \tag{1.2.1}$$

In this context, h is termed the *kernel* of the U-statistics and is said to be of *degree* k.

We will always assume that the kernel h and the induced U-statistics from (1.2.1) are *rank-based*, that is, the kernel has the property that $h(\mathbf{x}_1, \ldots, \mathbf{x}_k) = h(\mathbf{r}_1, \ldots, \mathbf{r}_k)$ for all arguments $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^2$. Here, for each argument $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})' \in \mathbb{R}^2$, we let $\mathbf{r}_i = (r_i^{(1)}, r_i^{(2)})'$ with $r_i^{(j)}$ being the rank of $x_i^{(j)}$ among $x_1^{(j)}, \ldots, x_k^{(j)}$ for $j = 1, 2$. If $U_h^{(pq)}$ from (1.2.1) is rank-based, then

$$U_h^{(pq)} = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h\left(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},k}^{(pq)}\right) = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h\left(\mathbf{R}_{\mathbf{i}}^{(pq)}\right). \tag{1.2.2}$$

Note that $(\mathbf{R}_1^{(pq)}, \ldots, \mathbf{R}_n^{(pq)})$ uniquely determines all k-tuples $(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},k}^{(pq)})$.

The following lemma lists elementary properties of $U_h^{(pq)}$ under H_0. It relies on the fact that under H_0 the distribution of $h(\mathbf{R}_{\mathbf{i}}^{(pq)})$ does not depend on the choice of $\mathbf{i}$, p and q because the rank vectors $\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(m)}$ are i.i.d. according to a uniform distribution on the symmetric group $\mathfrak{S}_n$; recall that we assume the original observations to be continuous random vectors such that ties among the ranks have probability zero.

Lemma 1.2.1. *Suppose $g(\cdot)$ is a real-valued function defined on $\left(\mathbb{R}^2\right)^n$, and for $1 \leq p \neq q \leq m$,*

$$g^{(pq)} := g\left(\mathbf{R}_1^{(pq)}, \ldots, \mathbf{R}_n^{(pq)}\right)$$

is symmetric in the n arguments $\mathbf{R}_1^{(pq)}, \ldots, \mathbf{R}_n^{(pq)}$. The random variables $g^{(pq)}$ satisfy the following properties under H_0:

i. If $p \neq q$, then $g^{(pq)}$ has the same distribution as $g^{(12)}$.

ii. If $p \neq q$, then $g^{(pq)}$ is independent of $\mathbf{X}^{(p)}$ (and also independent of $\mathbf{X}^{(q)}$).

iii. For any fixed $1 \leq l \leq m$, the $m-1$ random variables $g^{(pl)}$, $p \neq l$, are mutually independent.

iv. If $p \neq q$, $r \neq s$ and $\{p, q\} \neq \{r, s\}$, then $g^{(pq)}$ and $g^{(rs)}$ are independent.

Proof of Lemma 1.2.1. Claim (i) holds because the independence of $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}$ implies that the rank vectors $\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(m)}$ are i.i.d. For assertion (ii), note that, by the permutation symmetry of g in its n arguments, $g^{(pq)}$ is a function of the antirank of $\mathbf{X}^{(q)}$ in relation to $\mathbf{X}^{(p)}$ (Hájek et al., 1999, p. 63). These antiranks, which we denote by $\mathbf{R}^{(q)|(p)}$, are uniformly distributed on $\mathfrak{S}_n$ for any fixed choice of $\mathbf{X}^{(p)}$, which yields the independence of $g^{(pq)}$ and $\mathbf{X}^{(p)}$. Similarly, $g^{(pq)}$ is independent $\mathbf{X}^{(q)}$. (Of course, $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$ together determine $g^{(pq)}$.) Claim (iii) holds since the independence of $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}$ implies that the $m-1$ vectors of antiranks $\mathbf{R}^{(l)|(p)}$ for $p \neq l$ are mutually independent. Finally, the pairwise independence stated in (iv) is implied by the independence of $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}$ and (iii). $\square$

For simplicity we assume all kernel functions h to be *bounded*. Since h can be recentered if needed, without loss of generality, we will further assume that $\mathbb{E}_0[h(\mathbf{R}_{\mathbf{i}}^{(pq)})] = 0$, a property exhibited by all the examples below.

Example 1.2.1 (Kendall's tau). If we take h in (1.2.2) to be the kernel of degree $k = 2$ given by

$$h_\tau(\mathbf{r}_1, \mathbf{r}_2) = \operatorname{sgn}\left(\left(r_1^{(1)} - r_2^{(1)}\right)\left(r_1^{(2)} - r_2^{(2)}\right)\right),$$

then $\tau^{(pq)} := U_{h_\tau}^{(pq)}$ is Kendall's tau, which measures the association of $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$ by counting concordant versus disconcordant pairs of points.

Example 1.2.2 (Spearman's rho). Let

$$\rho_s^{(pq)} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} \left(R_i^{(p)} - R_i^{(q)} \right)^2 . \tag{1.2.3}$$

be the Spearman's rank correlation coefficient (rho) between $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(p)}$. Define $\hat{\rho}_s^{(pq)} := U_{h_{\hat{\rho}_s}}$, where $h_{\hat{\rho}_s}$ is the kernel function of degree 3 given by

$$h_{\hat{\rho}_s}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{1}{2} \sum_{\pi \in \mathfrak{S}_3} \operatorname{sgn}\left(r_{\pi_1}^{(1)} - r_{\pi_2}^{(1)} \right) \operatorname{sgn}\left(r_{\pi_1}^{(2)} - r_{\pi_3}^{(2)} \right) . \tag{1.2.4}$$

Hoeffding (1948b, p.318) showed that

$$\rho_s^{(pq)} = \frac{n-2}{n+1} \hat{\rho}_s^{(pq)} + \frac{3}{n+1} \tau^{(pq)}. \tag{1.2.5}$$

Hence, the dominating term $\hat{\rho}_s$ of Spearman's rho is a U-statistic.

Example 1.2.3 (Hoeffding's D statistic). Let

$$h_D(\mathbf{r}_1, \cdots, \mathbf{r}_5) = \frac{1}{5!} \sum_{\pi \in \mathfrak{S}_5} \frac{\phi\left(r_{\pi_1}^{(1)}, \ldots, r_{\pi_5}^{(1)} \right) \phi\left(r_{\pi_1}^{(2)}, \ldots, r_{\pi_5}^{(2)} \right)}{4},$$

where

$$\phi(r_1, \ldots, r_5) = \left(I(r_1 \geq r_2) - I(r_1 \geq r_3) \right) \left(I(r_1 \geq r_4) - I(r_1 \geq r_5) \right)$$

and $I(\cdot)$ is the indicator function. Hoeffding (1948a) suggested the statistic $D^{(pq)} := U_{h_D}^{(pq)}$ to measure association between the vectors $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$. When the joint distribution of $(X^{(p)}, X^{(q)})$ has continuous joint and marginal densities, the kernel expectation

$$\mathbb{E}\left[h_D(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \cdots, \mathbf{R}_{\mathbf{i},5}^{(pq)}) \right]$$

equals zero if and only if $X^{(p)}$ and $X^{(q)}$ are independent (Hoeffding, 1948a, Theorem 3.1).

Example 1.2.4 (Bergsma and Dassios' t^*). In a recent paper, Bergsma and Dassios (2014) introduced $t^{*(pq)} := U_{h_{t^*}}^{(pq)}$, a U-statistic of degree 4 with the kernel

$$h_{t^*}(\mathbf{r}_1, \cdots, \mathbf{r}_4) = \frac{1}{4!} \sum_{\pi \in \mathfrak{S}_4} \phi\left(r_{\pi(1)}^{(1)}, \ldots, r_{\pi(4)}^{(1)} \right) \phi\left(r_{\pi(1)}^{(2)}, \ldots, r_{\pi(4)}^{(2)} \right),$$

where now

$$\phi(r_1,\ldots,r_4) = I(r_1,r_3<r_2,r_4) + I(r_1,r_3>r_2,r_4) - I(r_1,r_2<r_3,r_4) - I(r_1,r_2>r_3,r_4).$$

According to Theorem 1 in Bergsma and Dassios (2014), t^* is an improvement over Hoeffding's D in the sense that the vanishing of $\mathbb{E}[h_{t^*}(\mathbf{R}^{(pq)}_{\mathbf{i},1},\cdots,\mathbf{R}^{(pq)}_{\mathbf{i},4})]$ characterizes the independence of $X^{(p)}$ and $X^{(q)}$ under the weaker assumption that $(X^{(p)},X^{(q)})$ has a bivariate distribution that is discrete or (absolutely) continuous, or a mixture of both.

Returning to our general setup, the variance and also the large-sample behavior of the statistic $U_h^{(pq)}$ is determined by the covariance quantities

$$\zeta_c^h \;:=\; \mathrm{Cov}\left[h\left(\mathbf{R}^{(pq)}_{\mathbf{i}}\right)h\left(\mathbf{R}^{(pq)}_{\mathbf{j}}\right)\right], \quad c=0,\ldots,k, \tag{1.2.6}$$

where $\mathbf{i},\mathbf{j}\in\mathcal{P}(n,k)$ are such that $|\mathbf{i}\cap\mathbf{j}|=c$. When H_0 is true,

$$\zeta_c^h \;=\; \mathbb{E}_0\left[h\left(\mathbf{R}^{(pq)}_{\mathbf{i}}\right)h\left(\mathbf{R}^{(pq)}_{\mathbf{j}}\right)\right] \tag{1.2.7}$$

as we are assuming that $\mathbb{E}_0[h(\mathbf{R}^{(pq)}_{\mathbf{i}})]=0$. Furthermore, the value of ζ_c^h does not depend on the choice of $(\mathbf{i},p,q)$ under H_0. In the sequel it will be clear from the context whether ζ_c^h is defined under H_0 or an alternative hypothesis.

It is well known that $0=\zeta_0^h\le\zeta_1^h,\ldots,\le\zeta_k^h$, and the kernel h is said to have order of degeneracy d if $\zeta_0^h=\zeta_1^h=\cdots=\zeta_{d-1}^h=0$ and $\zeta_d^h>0$ (Serfling, 1980, chapter 5). If $d\ge 2$, the kernel and the U-statistic it defines are referred to as degenerate. For any $c=1,\ldots,k$, it holds under H_0 that

$$\zeta_c^h=0 \iff \mathbb{E}_0\left[h\left(\mathbf{R}^{(pq)}_{\mathbf{i},1},\ldots,\mathbf{R}^{(pq)}_{\mathbf{i},k}\right)\middle|\mathbf{X}^{(pq)}_{\mathbf{i}'}\right]=0, \quad \text{almost surely}, \tag{1.2.8}$$

as a function of $\mathbf{X}^{(pq)}_{\mathbf{i}'}$, where $\mathbf{i}'\subset\mathbf{i}$ may be any subset with $|\mathbf{i}'|=c$. In particular, for the kernels h_D and h_{t^*}, the right-hand side of (1.2.8) holds with $c\le 1$.

Similar to the classical distribution theory of U-statistics, ζ_d^h will play a role in our asymptotic results for the test statistics we construct from rank-based U-statistics whose kernels have order of degeneracy $d=1$ or $d=2$ under H_0 in the next section. However, in the latter case, in addition to ζ_2^h, we will also need another quantity to describe our asymptotic results. For a symmetric kernel $h : (\mathbb{R}^2)^k \longrightarrow \mathbb{R}$ with order of degeneracy $d = 2$ under H_0, we define

$$\eta^h \ := \ \mathbb{E}_0 \left[h\left(\mathbf{R}_{\mathbf{i}^1}^{(pq)}\right) h\left(\mathbf{R}_{\mathbf{i}^2}^{(pq)}\right) h\left(\mathbf{R}_{\mathbf{i}^3}^{(pq)}\right) h\left(\mathbf{R}_{\mathbf{i}^4}^{(pq)}\right) \right], \tag{1.2.9}$$

where $\mathbf{i}^1, \ldots \mathbf{i}^4 \in \mathcal{P}(n, k)$ are any four tuples such that

i. $| \cup_{\omega=1}^4 \mathbf{i}^\omega | = 4k - 4$,

ii. $|\mathbf{i}^1 \cap \mathbf{i}^2| = |\mathbf{i}^2 \cap \mathbf{i}^3| = |\mathbf{i}^3 \cap \mathbf{i}^4| = |\mathbf{i}^4 \cap \mathbf{i}^1| = 1$, and

iii. no index $i \in \cup_{\omega=1}^4 \mathbf{i}^\omega$ is an element of more than two of the sets $\mathbf{i}^1, \ldots \mathbf{i}^4$.

For our purpose we only need to define η^h under H_0, and it is also easy to see that the choice of p, q, $\mathbf{i}^\omega$, $\omega = 1, \ldots, 4$, does not matter in its definition. Table 1.1 collects the order of degeneracy d under H_0, and the quantities ζ_d^h and η^h for the kernels in Example 1.2.1–1.2.4.

Finally, it is easy to check that all the kernels in Example 1.2.1–1.2.4 satisfy the property below that will be assumed for our null asymptotic results.

Assumption 1.2.2. *Let $h : (\mathbb{R}^2)^k \longrightarrow \mathbb{R}$ be a symmetric kernel with order of degeneracy $d \geq 1$ under H_0. Then given $\mathbf{i} = (i_1, \ldots, i_k) \in \mathcal{P}(n, k)$ and $1 \leq p \neq q \leq m$,*

$$\mathbb{E}_0 \left[h\left(\mathbf{R}_{\mathbf{i}}^{(pq)}\right) \middle| \mathbf{X}_{\mathbf{j}}^{(p)}, \mathbf{X}_{\mathbf{j}'}^{(q)} \right] = 0$$

for all $\mathbf{j}, \mathbf{j}' \subset \mathbf{i}$ such that $\min(|\mathbf{j}|, |\mathbf{j}'|) < d$.

1.3 Test statistics

We now proceed to construct test statistics for the independence hypothesis H_0 (1.1.1). Building on the pairwise rank correlations from Section 1.2, we introduce general classes of statistics and derive their respective asymptotic null distributions when $m, n \longrightarrow \infty$.

Table 1.1: Degree k, order of degeneracy d, covariance ζ_d^h and fourth moment η^h for the kernel functions in Example 1.2.1–1.2.4 when independence holds. The moment quantities are found in Hoeffding (1948a,b), and by our own calculations.

Kernel	h_τ	$h_{\hat{\rho}_s}$	h_D	h_{t^*}
k	2	3	5	4
d	1	1	2	2
ζ_d^h	1/9	1/9	1/810000	1/225
η^h	–	–	$(7/864000)^2$	$(2/525)^2$

1.3.1 Sum of squared sample rank correlations

Let $U_h^{(pq)}$ be a rank-based U-statistic as defined in (1.2.2), with mean 0 when $X^{(p)}$ and $X^{(q)}$ are independent. Suppose further that large absolute values of $U_h^{(pq)}$ indicate strong association (positive or negative) between $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$. Following the approach of Schott (2005), it is then natural to reject H_0 for large values of the centered quantity

$$S_h := \sum_{1 \le p < q \le m} \left(U_h^{(pq)}\right)^2 - \binom{m}{2}\mu_h. \tag{1.3.1}$$

Here, $\mu_h := \mathbb{E}_0[(U_h^{(pq)})^2]$. Note that, as indicated by our notation, this expectation does not depend on the choice of p and q by Lemma 1.2.1(i). The following lemma specifies μ_h and gives a result on other moments of $U_h^{(pq)}$ that will be used later.

Lemma 1.3.1. *Let $n \ge 2k \ge 2$, and suppose that $U_h^{(pq)}$ from* (1.2.2) *has a kernel h with order of degeneracy d under H_0. Then given $1 \le p < q \le n$ and under H_0,*

i.

$$\mu_h = \binom{n}{k}^{-1} \sum_{c=1}^{k} \binom{k}{c}\binom{n-k}{k-c}\zeta_c = \binom{k}{d}^2 \frac{d!\zeta_d}{n^d} + O\left(n^{-d-1}\right)$$

ii. and for any $r > 2$,

$$\mathbb{E}_0\left[\left(U_h^{(pq)}\right)^r\right] = O\left(n^{-\lfloor (rd+1)/2 \rfloor}\right),$$

where $\lfloor\cdot\rfloor$ denotes the floor function.

iii. Moreover,

$$\mathbb{E}_0\left[\left(U_h^{(pq)}\right)^4\right] = \begin{cases} \dfrac{3k^4(\zeta_1^h)^2}{n^2} + O\left(n^{-3}\right) & \textit{if} \quad d = 1, \\ \dbinom{k}{2}^4 \dfrac{12}{n^4}\left((\zeta_2^h)^2 + 4\eta^h\right) + O\left(n^{-5}\right) & \textit{if} \quad d = 2. \end{cases}$$

For Lemma 1.3.1(i) and (ii), see Lemma 5.2.1A and 5.2.2B in Serfling (1980). The last claim about the leading term of the fourth moment is proven in Appendix A.3. Let μ_τ, $\mu_{\hat{\rho}_s}$, μ_D and μ_{t^*} be the values of μ_h when h is equal to h_τ, $h_{\hat{\rho}_s}$, h_D and h_{t^*} respectively. Then

$$\mu_\tau = \frac{2(2n+5)}{9n(n-1)}, \qquad \mu_{\hat{\rho}_s} = \frac{(n^2-3)}{n(n-1)(n-2)},$$
$$\mu_D = \frac{2(n^2+5n-32)}{9n(n-1)(n-3)(n-4)}, \qquad \mu_{t^*} = \frac{8}{75}\frac{3n^2+5n-18}{n(n-1)(n-2)(n-3)}.$$

The first three quantities can be found in Hoeffding (1948a,b). The stated value of μ_{t^*} is based on our own calculations.

1.3.2 Unbiased estimator of the sum of squared population correlations

The kernel function h is central to the role of $U_h^{(pq)}$ as a measure of association between the vectors of observations $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$. At the population level the association (positive or negative) is captured by the expectation of $U_h^{(pq)}$, which is also equal to

$$\mathbb{E}\left[h\left(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},k}^{(pq)}\right)\right], \tag{1.3.2}$$

where $\mathbf{j}$ may be any element in $\mathcal{P}(n,k)$. Hence,

$$\sum_{1\leq p<q\leq m}\left(\mathbb{E}\left[h\left(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},k}^{(pq)}\right)\right]\right)^2 \tag{1.3.3}$$

is a population measure of overall dependency in the joint distribution of $X^{(1)}, \ldots, X^{(m)}$. As an alternative approach to Section 1.3.1, we now construct an unbiased estimator of (1.3.3), targeting more directly the problem of global (in-)dependence.

Recall that given $\mathbf{i} \in \mathcal{P}(n, 2k)$ and $\mathbf{j} \in \mathcal{P}(n, k)$ such that $\mathbf{j} \subset \mathbf{i}$ as sets, $\mathbf{i} \setminus \mathbf{j}$ is the k-tuple in $\mathcal{P}(n, k)$ that is given by their set difference. The function

$$h^W\left(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},2k}^{(pq)}\right) := \binom{2k}{k}^{-1} \sum_{\substack{\mathbf{j} \subset \mathbf{i} \\ |\mathbf{j}|=k}} h\left(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},k}^{(pq)}\right) h\left(\mathbf{R}_{\mathbf{i}\setminus\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i}\setminus\mathbf{j},k}^{(pq)}\right) \tag{1.3.4}$$

can be seen to be symmetric in its $2k$ arguments $\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},2k}^{(pq)}$, due to the symmetry of h and the summation over all possible tuples $\mathbf{j} \in \mathcal{P}(n, k)$ on the right hand side of (1.3.4). Moreover, h^W is an unbiased estimator of the square of the expectation in (1.3.2), since each summand on the right hand side of (1.3.4) is a product of two independent unbiased estimators of $\mathbb{E}[h(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},k}^{(pq)})]$. Therefore, defining the U-statistic

$$W_h^{(pq)} = W_h^{(pq)}\left(\mathbf{R}_1^{(pq)}, \ldots, \mathbf{R}_n^{(pq)}\right) = \binom{n}{2k}^{-1} \sum_{\mathbf{i} \in \mathcal{P}(n,2k)} h^W\left(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},2k}^{(pq)}\right), \tag{1.3.5}$$

we have that the sum

$$T_h := \sum_{1 \le p < q \le m} W_h^{(pq)} \tag{1.3.6}$$

is an unbiased estimator of (1.3.3). The statistic T_h is a U-statistic itself and serves as a natural test statistic for H_0. Large values of T_h indicate departures from H_0. When $h = h_\tau$, i.e., the case of Kendall's tau, T_h equals the statistic displayed in (1.1.7) in the introduction.

Clearly, $W_h^{(pq)}$ is a rank-based U-statistic with the kernel h^W of degree $2k$. The following lemma summarizes the degeneracy properties of h^W under H_0.

Lemma 1.3.2. *Suppose $h : (\mathbb{R}^2)^k \longrightarrow \mathbb{R}$ is a symmetric kernel function of degree k with order of degeneracy $d \in \{1, 2\}$ under H_0. So, $\zeta_d^h > 0$. Then, under H_0, the induced symmetric kernel function h^W defined in (1.3.4) has order of degeneracy $2d$ and*

$$\begin{aligned}
\zeta_{2d}^{h^W} &:= \mathbb{E}_0\left[h^W\left(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{i},2k}^{(pq)}\right) h^W\left(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},2k}^{(pq)}\right)\right] \\
&= \begin{cases} 4\binom{2k-2}{k-1}^2 \binom{2k}{k}^{-2} \left(\zeta_d^h\right)^2 & \text{if} \quad d = 1, \\ 12\binom{2k-4}{k-2}^2 \binom{2k}{k}^{-2} \left\{(\zeta_d^h)^2 + 2\eta^h\right\} & \text{if} \quad d = 2, \end{cases}
\end{aligned}$$

where $\mathbf{i}, \mathbf{j} \in \mathcal{P}(n, 2k)$ and $|\mathbf{i} \cap \mathbf{j}| = 2d$.

The proof of the lemma is deferred to Appendix A.3.

1.3.3 Simple sum of sample rank correlations

It is also instructive to consider the simple sum

$$Z_h := \sum_{1 \leq p < q \leq m} U_h^{(pq)} \tag{1.3.7}$$

for testing the null hypothesis H_0. For the kernels $h_{\hat{\rho}_s}$ and h_τ, (1.3.7) makes sense as a "one-sided" test when one is only interested in testing H_0 against the alternative of positive associations among $X^{(1)}, \dots, X^{(m)}$. For the kernels $h_{\tau*}$ and h_D, (1.3.7) also forms a legitimate "two-sided" test since, as noted in Bergsma and Dassios (2014), $\mathbb{E}[h_{\tau^*}(\mathbf{R}_\mathbf{i}^{(pq)})]$ and $\mathbb{E}[h_D(\mathbf{R}_\mathbf{i}^{(pq)})]$ are always non-negative and equal zero if and only if $X^{(p)}$ and $X^{(q)}$ are independent, under weak distributional assumptions.

1.4 Asymptotic null distributions

We are now ready to state our results on the asymptotic distributions for the test statistics introduced in Section 1.3. As mentioned in Section 1.2, we will focus on rank-based U-statistics with a kernel h satisfying Assumption 1.2.2 and order of degeneracy $d \in \{1, 2\}$ under H_0.

Theorem 1.4.1. *Suppose the null hypothesis H_0 from (1.1.1) is true. Let h be a symmetric bounded kernel function of degree k satisfying Assumption 1.2.2, and consider the asymptotic regime $m, n \longrightarrow \infty$. If $d = 1$, after suitable rescaling, S_h, T_h and Z_h are asymptotically normal, namely,*

$$\frac{nS_h}{k^2 m \zeta_1^h}, \ \frac{nT_h}{k^2 m \zeta_1^h}, \ \frac{\sqrt{2n} Z_h}{km\sqrt{\zeta_1^h}} \Longrightarrow \mathcal{N}(0,1).$$

If $d = 2$, then

$$\frac{n^2 \binom{k}{2}^{-2} S_h}{2m\sqrt{(\zeta_2^h)^2 + 6\eta^h}}, \ \frac{n^2 \binom{k}{2}^{-2} T_h}{2m\sqrt{(\zeta_2^h)^2 + 2\eta^h}}, \ \frac{n \binom{k}{2}^{-1} Z_h}{m\sqrt{\zeta_2^h}} \Longrightarrow \mathcal{N}(0,1).$$

The theorem covers in particular the rank correlations from Examples 1.2.1–1.2.4. Upon rescaling with appropriate multiplicative factors, the statistics S_h, T_h and Z_h for these four choices of the kernel h converge to standard normal limits as $m, n \longrightarrow \infty$ under H_0. A critical value for an approximate α-size test can thus be calibrated based on normal quantiles. Just like in the classical theory for U-statistics, these rescaling factors for the non-degenerate and degenerate cases differ in order; for instance, we have to multiply S_h with a factor of order $O(n/m)$ when h has order of degeneracy $d = 1$, and with a factor of order $O(n^2/m)$ when h has order of degeneracy $d = 2$. The ingredients needed to compute these rescaling factors were given in Table 1.1. In slight abbreviation, we write S_τ, $S_{\hat{\rho}_s}$, S_D and S_{t^*} for the four versions of the statistic S_h from (1.3.1) with the different kernels reviewed in Section 1.2, and analogously, T_τ, $T_{\hat{\rho}_s}$, T_D, T_{t^*} and Z_τ, $Z_{\hat{\rho}_s}$, Z_D, Z_{t^*} for the versions of T_h and Z_h from (1.3.6) and (1.3.7). Hence S_τ, T_τ and Z_τ are the statistics in the displays (1.1.5), (1.1.7) and (1.1.8), respectively.

We remark that while the classical Spearman's rho is not a U-statistic one may of course consider the centered test statistic

$$S_{\rho_s} := \sum_{1 \le p < q \le m} \left(\rho_s^{(pq)}\right)^2 - \binom{m}{2} \mu_{\rho_s}, \tag{1.4.1}$$

where $\mu_{\rho_s} := \mathbb{E}_0[(\rho_s^{(pq)})^2] = 1/(n-1)$; see Hoeffding (1948b, p.321). The convergence of $\frac{n}{m} S_{\hat{\rho}_s}$ to a standard normal distribution, as suggested by Theorem 1.4.1, implies the following distributional convergence for S_{ρ_s}. Its proof, given in Appendix A.4, makes use of the decomposition from (1.2.5). The same result has been obtained by Zhou (2007) and Wang et al. (2013) via different methods.

Corollary 1.4.2. *Under H_0, $\frac{n}{m} S_{\rho_s} \Longrightarrow \mathcal{N}(0,1)$ as $m, n \longrightarrow \infty$.*

Our proof of Theorem 1.4.1 is based on a central limit theorem for martingale arrays (Hall and Heyde, 1980, Corollary 3.1) that was also applied by Schott (2005). We outline the approach here, postponing computations verifying the conditions of the martingale CLT to Appendix A.4.

Proof of Theorem 1.4.1. Fix a sample size n. For $q = 1, \ldots, m$, let $\mathcal{F}_{nq}$ be the σ-algebra generated by $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(q)}$ (or for our purposes, equivalently, $\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(q)}$) under H_0. For convenience we will use the shorthand $\bar{U}_h^{(pq)} := \left(U_h^{(pq)}\right)^2 - \mu_h$ for $1 \leq p < q \leq m$. Let

$$D_{nq}^S := \sum_{p=1}^{q-1} \bar{U}_h^{(pq)}, \quad D_{nq}^T := \sum_{p=1}^{q-1} W_h^{(pq)} \text{and} \quad D_{nq}^Z := \sum_{p=1}^{q-1} U_h^{(pq)} \tag{1.4.2}$$

and set $D_{n1}^S = D_{n1}^T = D_{n1}^Z = 0$. Writing $S_{nq} = \sum_{l=1}^q D_{nl}^S$, $T_{nq} = \sum_{l=1}^q D_{nl}^T$ and $Z_{nq} = \sum_{l=1}^q D_{nl}^Z$, we have that $S_h = S_{nm}$, $T_h = T_{nm}$ and $Z_h = Z_{nm}$.

We claim that, for each n, the sequences

$$\{S_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\}, \quad \{T_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\} \quad \text{and} \quad \{Z_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\} \tag{1.4.3}$$

are martingales, i.e., $\mathbb{E}_0\left[S_{nq} | \mathcal{F}_{n,q-1}\right] = S_{n,q-1}$, $\mathbb{E}_0\left[T_{nq} | \mathcal{F}_{n,q-1}\right] = T_{n,q-1}$ and $\mathbb{E}_0\left[Z_{nq} | \mathcal{F}_{n,q-1}\right] = Z_{n,q-1}$ for $q = 2, \ldots, m$. Given the way S_{nq}, T_{nq} and Z_{nq} are defined as sums, it suffices to show that

$$\mathbb{E}_0\left[\bar{U}_h^{(pq)} \middle| \mathcal{F}_{n,q-1}\right] = \mathbb{E}_0\left[W_h^{(pq)} \middle| \mathcal{F}_{n,q-1}\right] = \mathbb{E}_0\left[U_h^{(pq)} \middle| \mathcal{F}_{n,q-1}\right] = 0 \tag{1.4.4}$$

for all $1 \leq p < q \leq m$. Since $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}$ are independent under H_0, conditioning on $\mathcal{F}_{n,q-1}$ is the same as conditioning on $X^{(p)}$ alone in (1.4.4). As $\bar{U}_h^{(pq)}$, $W_h^{(pq)}$ and $U_h^{(pq)}$are all symmetric functions of the n arguments $\mathbf{R}_1^{(pq)}, \ldots, \mathbf{R}_n^{(pq)}$, (1.4.4) follows from Lemma 1.2.1(i) and (ii).

By the boundedness of our kernel h, each of the martingales in (1.4.3) is trivially square-integrable. As such, the central limit theorem for martingale arrays from Corollary 3.1 in Hall and Heyde (1980) implies the assertion of Theorem 1.4.1 if we can show that the squares of the martingale differences D_{nl}^S, D_{nl}^T and D_{nl}^Z each satisfy the following two conditions. The first condition requires that as $m, n \longrightarrow \infty$,

$$\frac{n^2}{m^2} \sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^S)^2 \middle| \mathcal{F}_{n,l-1}\right], \quad \frac{n^2}{m^2} \sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^T)^2 \middle| \mathcal{F}_{n,l-1}\right] \quad \underset{p}{\longrightarrow} \quad k^4 (\zeta_1^h)^2, \tag{1.4.5}$$

$$\frac{n}{m^2} \sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^Z)^2 \middle| \mathcal{F}_{n,l-1}\right] \quad \underset{p}{\longrightarrow} \quad 2^{-1} k^2 \zeta_1^h, \tag{1.4.6}$$

for $d = 1$, and

$$\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^S)^2\middle|\mathcal{F}_{n,l-1}\right] \quad \underset{p}{\longrightarrow} \quad 4\binom{k}{2}^4\left\{(\zeta_2^h)^2+6\eta^h\right\}\ , \tag{1.4.7}$$

$$\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^T)^2\middle|\mathcal{F}_{n,l-1}\right] \quad \underset{p}{\longrightarrow} \quad 4\binom{k}{2}^4\left\{(\zeta_2^h)^2+2\eta^h\right\} \tag{1.4.8}$$

$$\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^2\middle|\mathcal{F}_{n,l-1}\right] \quad \underset{p}{\longrightarrow} \quad \binom{k}{2}^2\zeta_2^h \tag{1.4.9}$$

for $d = 2$, where the convergence symbol stands for convergence in probability. The second condition is a Lindeberg condition. In Lemma A.4.1 in the Appendix A.4, we show that, in fact, (1.4.5) – (1.4.9) also hold in the stronger sense of L^2 (or quadratic mean). Lemma A.4.2 proves a Lyapunov condition that implies the Lindeberg condition, which completes the proof of Theorem 1.4.1. $\square$

1.5 Preliminary power analysis

In this section, we will conduct preliminary investigation into the power of our tests from an asymptotic minimax perspective. While our null distributional results in Section 1.3 are valid under the more general asymptotic regime $m, n \longrightarrow \infty$, in this section we will focus on the particular regime $\frac{m}{n} \longrightarrow \gamma \in (0, \infty)$ considered in works such as Ledoit and Wolf (2002). For $1 \le p < q \le m$, define

$$\theta^{(pq)} \ := \ \mathbb{E}[h(\mathbf{R}_{\mathbf{j},1}^{(pq)}, \ldots, \mathbf{R}_{\mathbf{j},k}^{(pq)})],$$

which will generally be nonzero when $X^{(p)}$ and $X^{(q)}$ are dependent, and also let $\Theta = (\theta^{(pq)})_{1\le p<q\le m}$ be the $\binom{m}{2}$-vector comprising all these pairwise measures of association. Our power analysis will focus on the two classes of statistics S_h and T_h since for a general chioce of h, they are motivated as estimates of the square of the undirectional dependency signal $\|\Theta\|_2$ and thus define proper "two-sided" tests for the null hypothesis H_0. For further simplicity, we will focus on the statistics S_τ and T_τ constructed with the Kendall's tau kernel $h = h_\tau$. To indicate this restriction in our notation, we define $\theta_\tau^{(pq)} := \mathbb{E}[h_\tau(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \mathbf{R}_{\mathbf{i},2}^{(pq)})]$ and

$\Theta_\tau = (\theta_\tau^{(pq)})_{1\le p<q\le m}$. In light of our non-parametric approach to extend beyond the Gaussian assumption, let $\mathcal{D}_m$ be a family of continuous joint distributions on $\mathbb{R}^m$ containing all m-variate Gaussian distributions, to be considered as joint distributions for $(X^{(1)}, \dots, X^{(m)})$. For a given significance level $\alpha \in (0,1)$, We will aim at studying what sequences of lower bound ϵ_n on the dependency signal $\|\Theta_\tau\|_2$ allow our tests to uniformly achieve a fixed power $\beta > \alpha$ over the subset of alternative distributions

$$\mathcal{D}_m(\|\Theta_\tau\|_2 \ge \epsilon_n) := \Big\{ D \in \mathcal{D}_m : \|\Theta_\tau\|_2 \ge \epsilon_n \Big\}. \tag{1.5.1}$$

As usual we take a test ϕ to be a function mapping the data into the unit interval $[0,1]$. Given a test statistic $S = S(\mathbf{X}_1, \dots, \mathbf{X}_n)$, we write $\phi_\alpha(S)$ for the test that rejects for large values of S and has (asymptotic) size α.

It is intuitive to use our statistics S_τ or T_τ when one is interested in detecting the alternatives described by (1.5.1) because these statistics are natural estimates for the squared Euclidean norm signal $\|\Theta_\tau\|_2^2$. The following theorem gives a rough lower bound on the signal size $\|\Theta_\tau\|_2$ for detectability.

Theorem 1.5.1. *Let $0 < \alpha < \beta < 1$. Under the asymptotic regime $m/n \longrightarrow \gamma \in (0,\infty)$, there exist constants $C_i = C_i(\alpha,\beta,\gamma) > 0$ for $i = 1,2$, such that*

i.

$$\liminf_{n\longrightarrow\infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \ge \epsilon_n)} \mathbb{E}[\phi_\alpha(S_\tau)] > \beta \quad \textit{for} \quad \epsilon_n = C_1\sqrt{n} \quad \textit{and}$$

ii.

$$\liminf_{n\longrightarrow\infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \ge \epsilon_n)} \mathbb{E}[\phi_\alpha(T_\tau)] > \beta \quad \textit{for} \quad \epsilon_n = C_2\sqrt{n}.$$

We would like to emphasize that our proof of Theorem 1.5.1 is based on rather general concentration bounds. It should be possible to sharpen the analysis and show asymptotic power for $\phi_\alpha(S_\tau)$ and $\phi_\alpha(T_\tau)$ under smaller signal strength. Indeed, based on the result from the next subsection we *conjecture* that a test based on T_τ can asymptotically attain uniform power β when the signal size $\|\Theta_\tau\|_2$ is of constant order $O(1)$. This conjecture is partially supported by Theorem 1.5.2 below, whose proof makes use of Theorem 1.5.1.

1.5.1 Rate-optimality under equicorrelation

When the joint distribution of $X^{(1)}, \dots, X^{(m)}$ is a regular Gaussian distribution, then H_0 is equivalent to $R - I_m = 0$, where $R = (\rho^{(pq)})$ is the population Pearson correlation matrix of $(X^{(1)}, \dots, X^{(m)})$, and I_m is the m-by-m identity matrix. For any $\epsilon > 0$, define the alternative

$$\mathcal{N}_m(\|R - I_m\|_F \geq \epsilon) \tag{1.5.2}$$

as the family of regular m-variate Gaussian distributions whose correlation matrix R satisfies $\|R - I_m\|_F \geq \epsilon$. A result of Cai and Ma (2013, Remark 1(a)) implies that in the regime $m/n \longrightarrow \gamma$, for given $0 < \alpha < \beta < 1$, there exists a sufficiently small constant $c = c(\alpha, \beta, \gamma) > 0$ such that

$$\limsup_{n\to\infty} \inf_{\mathcal{N}_m(\|R-I_m\|_F \geq c)} \mathbb{E}[\phi] < \beta$$

for any α-level test ϕ. In other words, asymptotically, no α-level test can uniformly achieve the desired power against the alternative (1.5.2) when the signal size $\|R - I_m\|_F$ is allowed to be as small as c. An immediate consequence of this in our nonparametric setup is that there also exists a constant $\tilde{c} = \tilde{c}(\alpha, \beta, \gamma) > 0$ such that

$$\limsup_{n\longrightarrow\infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 > \tilde{c})} \mathbb{E}[\phi] < \beta$$

for any α-level test ϕ. This is true because the nonparametric class $\mathcal{D}_m$ contains all m-variate Gaussian distributions, and because $\theta_\tau^{(pq)} \asymp \rho^{(pq)}$ when $X^{(p)}$ and $X^{(q)}$ are jointly Gaussian. The latter fact follows from $\rho^{(pq)} = \sin\left(\frac{\pi}{2}\theta_\tau^{(pq)}\right)$; see Kruskal (1958, p.823).

Given the observation just made, a α-level test ϕ that satisfies

$$\liminf_{n\longrightarrow\infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \geq \tilde{C})} \mathbb{E}[\phi] > \beta \tag{1.5.3}$$

for a large enough constant $\tilde{C} = \tilde{C}(\alpha, \beta, \gamma) > 0$ would be rate-optimal. If the signal $\|\Theta_\tau\|_2$ is large, being an unbiased estimator of $\|\Theta_\tau\|_2^2$ our statistic T_τ always centers around the the square of the same large value regardless of the true underlying distribution of $\mathbf{X}$. It is hence natural to *conjecture* that the optimality condition (1.5.3) is satisfied by the test $\phi_\alpha(T_\tau)$, for a reasonable class of distributions $\mathcal{D}_m$ that extends beyond the Gaussians. Such a choice for $\mathcal{D}_m$ could be a wide regular class of elliptical distributions (Anderson, 2003, p.47), which still satisfy the property that $\theta_\tau^{(pq)} \asymp \rho^{(pq)}$; see Lindskog et al. (2003). Our next result supports the conjecture.

Let $\mathcal{N}_m^{\text{equi}}(\|\Theta_\tau\|_2 \geq \tilde{C})$ be the set of m-variate Gaussian distributions that have all pairwise (Pearson and thus also Kendall) correlations equal to a common value such that $\|\Theta_\tau\|_2 \geq \tilde{C}$. If $\theta_\tau^{(pq)} = \theta$ for all $1 \leq p \neq q \leq m$, then $\|\Theta_\tau\|_2^2 = \theta^2 \binom{m}{2}$.

Theorem 1.5.2. *As $\frac{m}{n} \longrightarrow \gamma$, there exists a constant $\tilde{C} = \tilde{C}(\alpha, \beta, \gamma) > 0$ such that*

$$\liminf_{n \longrightarrow \infty} \inf_{\mathcal{N}_m^{equi}(\|\Theta_\tau\|_2 \geq \tilde{C})} \mathbb{E}[\phi_\alpha(T_\tau)] > \beta.$$

The theorem is proven in Appendix A.5. Empirically, our simulation experiments on power in the next section corroborate the conjecture made above. At this point we do not yet have general results on asymptotic minimax optimality for the class of statistics T_h with h being a general kernel we considered. The result we have for T_τ in Theorem 1.5.2 is more of a proof of concept with the simplest case of a non-degenerate kernel of degree 1 and a particularly simple alternative.

1.6 Implementation and simulation experiments

We now compare several tests of H_0 based on specific versions of the test statistics introduced above. In our simulations we will explore the accuracy of the normal distribution approximation, by exploring the sizes of the tests. We then compare their power. Before turning to the simulations, however, we will discuss the computation of the involved test statistics.

1.6.1 Implementation

Given a kernel function h, to compute the statistics S_h from (1.3.1) and Z_h from (1.3.7) for the m variables, one has to make $\binom{m}{2}$ evaluations of the U-statistics $U_h^{(pq)}$. In general, for a U-statistic of degree k, a naïve calculation following the definition from (1.2.2) requires $O(n^k)$ operations. Fortunately, more efficient algorithms have been developed for the specific examples we have covered. For instance, Spearman's $\rho_s^{(pq)}$ from Example 1.2.2 can be computed in $O(n)$ operations. Kendall's $\tau^{(pq)}$ from Example 1.2.1 has kernel h_τ of degree $k = 2$ but can again be computed in $O(n \log n)$ operations (Christensen, 2005). Similarly, Weihs et al. (2015) devised a related algorithm to compute the Bergsma-Dassios sign covariance $t^{*(pq)}$ in $O(n^2 \log n)$ operations despite the fact that its kernel has degree $k = 4$, as reviewed in Example 1.2.4.

The situation with the class of statistics T_h from (1.3.6) is more complicated. Since a given kernel h of degree k gives rise to an induced kernel h^W of degree $2k$, the number of operations equals $O(n^{2k})$ if we compute $W_h^{(pq)}$ by naïvely following its definition. This would lead to a total of $\binom{m}{2}O(n^{2k})$ operations to find all $W_h^{(pq)}$, $1 \le p < q \le m$. A more efficient way to compute each $W_h^{(pq)}$ in $O(n^k)$ time proceeds as follows. Using (1.3.4) and (1.3.5), $W_h^{(pq)}$ can be seen to be equal to

$$\frac{1}{\binom{n}{k}\binom{n-k}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h_{\mathbf{i}} h_{\bar{\mathbf{i}}}, \tag{1.6.1}$$

where for each $\mathbf{i} \in \mathcal{P}(n,k)$ and suppressing the dependence on the pair (p,q), we define

$$h_{\mathbf{i}} := h(R_{\mathbf{i}}^{(pq)}) \quad \text{and} \quad h_{\bar{\mathbf{i}}} := \sum_{\mathbf{j} \in \mathcal{P}(n,k): \mathbf{j} \cap \mathbf{i} = \emptyset} h_{\mathbf{j}}.$$

Hence, it suffices to calculate i. $h_{\mathbf{i}}$ for all $\mathbf{i} \in \mathcal{P}(n,k)$, ii. $h_{\bar{\mathbf{i}}}$ for all $\mathbf{i} \in \mathcal{P}(n,k)$ and iii. the summation in (1.6.1), in that order. Evidently, step (i) involves $O(n^k)$ operations. By the inclusion-exclusion principle,

$$h_{\bar{\mathbf{i}}} = \sum_{\mathbf{j} \in \mathcal{P}(n,k)} h_{\mathbf{j}} + \sum_{1 \le \ell \le k} (-1)^\ell \sum_{\substack{\mathbf{j}' \in \mathcal{P}(n,\ell): \\ \mathbf{j}' \subset \mathbf{i}}} h_{\mathbf{j}'}, \tag{1.6.2}$$

where $h_{\mathbf{j}'} := \sum_{\mathbf{j} \subset \mathcal{P}(n,k): \mathbf{j}' \subset \mathbf{j}} h_{\mathbf{j}}$ for each $1 \le \ell < k$ and $\mathbf{j}' \subset \mathcal{P}(n,\ell)$. Note that there are $O(n^\ell)$ many $\mathbf{j}' \in \mathcal{P}(n,\ell)$, and each $h_{\mathbf{j}'}$ is a sum of $O(n^{k-\ell})$ many terms. Finding $h_{\mathbf{j}'}$ for all $\mathbf{j}' \in \mathcal{P}(n,\ell)$ and $1 \le \ell < k$ thus requires $O(n^k)$ operations, and with these as ingredients, by

(1.6.2), one can compute each $h_{\bar{\mathbf{i}}}$ in $O(1)$ operations if $\sum_{\mathbf{j}\in\mathcal{P}(n,k)} h_{\mathbf{j}}$ is already known. But the quantity $\sum_{\mathbf{j}\in\mathcal{P}(n,k)} h_{\mathbf{j}}$ only has to be computed once, with another $O(n^k)$ computations. Consequently, step (ii) involves $O(n^k)$ operations, and so does the final summation in step (iii).

1.6.2 Simulations

We first consider the sizes of tests based on our statistics S_τ, S_{ρ_s},S_{t^*}, T_τ, $T_{\hat{\rho}_s}$ and Z_{t^*} defined in Section 1.4. For comparison, we also consider the sum of squared Pearson correlations S_r from Schott (2005); recall (1.1.3). Each test proceeds by comparing its rescaled test statistic to the limiting standard normal distribution as in Theorem 1.4.1 and Corollary 1.4.2. Here, we would like to remark that by a proof similar to that of Theorem 1.4.1, one can show that

$$\frac{n}{m} S_r \Longrightarrow \mathcal{N}(0,1)$$

under H_0 as $m, n \longrightarrow \infty$, whereas Schott (2005)'s original asymptotic result is based on the more restrictive regime $\frac{m}{n} \longrightarrow \gamma \in (0, \infty)$. Targeting a size of 5%, the null hypothesis H_0 is rejected if the value of the rescaled statistic exceeds the 95th percentile of the standard normal distribution. The finite-sample sizes are listed in Table 1.2. The data underlying the table are i.i.d. noncentral t with $\nu = 3$ degrees of freedom and noncentrality parameter $\mu = 2$. For each combination of m and n, the sample sizes of the tests are calculated from $5,000$ independently generated data sets. As expected, the sample sizes corresponding to our rank-based statistics all get closer to 0.05 when m and n increase, but the test based on Schott (2005)'s S_r rejects too often, reflecting the fact that his limit theorem involves a Gaussian assumption.

Turning to an empirical study of the tests' power, our results provide evidence for the power analysis we did in Section 1.5. For different combinations of (m, n), we generate data as n independent draws from three different m-variate elliptical distributions. These are

i. the m-variate normal distribution: $\mathcal{N}_m(0, \Sigma)$,

ii. the m-variate t distribution: $t_{\nu=20,m}(\mu = 2 \cdot \mathbf{1}_m, \Sigma)$ and

iii. the m-variate power exponential distribution: $PE(\mu = 0, \Sigma, \nu = 20)$.

Here, $\mathbf{1}_m$ is the m-vector with all entries equal to 1, and the parameter specifications of these distributions are in accordance with Oja (2010, pp. 8–10). For each distribution, the scatter matrix $\Sigma = (\sigma_{ij})$ is a matrix with 1's on the diagonal and equal values for the off-diagonal entries, picked so that Σ gives rise to the signal strengths $\|\Theta_\tau\|_2^2 = 0.1$, 0.3, and 0.7. We refer again to Lindskog et al. (2003) for the relationship between Σ and $\|\Theta_\tau\|_2^2$. In the case of multivariate normal distribution Σ is simply the covariance matrix of $\mathbf{X}$. The empirical power, computed based on 500 repetitions of experiments, for tests based on S_τ, T_τ, and Han and Liu (2014)'s $S_\tau^{\max}$ are compared in Tables 1.3 and 1.4. $S_\tau^{\max}$ is, as mentioned in Section 1.1, defined as the maximum of the absolute values of all pairwise sample Kendall's tau correlation,

$$\max_{1 \le p < q \le m} |\tau^{(pq)}|.$$

A test based on $S_\tau^{\max}$ is powerful in detecting alternatives that have strong dependence but only among a few of the random variables $X^{(1)}, \ldots, X^{(m)}$, since $S_\tau^{\max}$ attempts to estimate the largest pairwise (population) rank correlation. In fact Han and Liu (2014) showed that $S_\tau^{\max}$ is rate-optimal in detecting such alternatives that are measured in the max norm signal $\|\Theta_\tau\|_\infty$. However, for alternatives that have small but non-zero dependence among many pairs of random variables as in our simulation setups, the statistics S_τ and T_τ should be more powerful since they estimate the squared Euclidean norm $\|\Theta_\tau\|_2^2$. As expected, since Han and Liu (2014)'s $S_\tau^{\max}$ is not well-adapted for detecting the alternatives we generated, its power can hardly match the other statistics in almost all setups. On one hand, for each (m, n) combination and a given value of $\|\Theta_\tau\|_2^2$, the power of the test based on T_τ is similar across different data-generating distributions. On the other hand, S_τ tend to have more power for t-distributed data, and less power for data with power exponential distribution. The stability of the power rendered by T_τ points to our conjecture in Section 1.5 on the minimax optimality of T_τ over a wide class of elliptical distributions.

Our empirical power results deserve special comments when the data are generated from multivariate normal distributions, in which case the power of two other statistics are also compared in Table 1.3. Schott (2005)'s S_r (1.1.3) is a legitimate test in this case since it has the right asymptotic size when the data are normally distributed. As seen in Table 1.3, the three statistics, S_τ, T_τ and Schott (2005)'s S_r, have comparable power for different combinations of (m, n) and signal strength $\|\Theta_\tau\|_2^2$. We have also experimented with the statistic proposed in Cai and Ma (2013) which demonstrated similar power. We note that as proved in Cai and Ma (2013), their statistic is minimax rate optimal in detecting the Frobenius norm signal $\|\Sigma - I_m\|_2$, but only for testing the different hypothesis of whether the covariance matrix Σ is equal to the identitiy under Gaussian assumption on $\mathbf{X}$. Hence, in principle, Cai and Ma (2013)'s statistic should not be treated as a benchmark on the best achievable power for testing the null hypothesis of independence in (1.1.1), which in the multivariate normal setup is equivalent to the hypothesis that $R = I_m$. However, the comparable empirical power of Cai and Ma (2013)'s statistic still gives certain indication that the three statistics S_τ, T_τ and S_r are all powerful in detecting the signal $\|R - I_m\|_2 \asymp \|\Theta_\tau\|_2$, since our data-generating choice of Σ has 1's on the diagonal which means $\|\Sigma - I_m\|_2 = \|R - I_m\|_2$. Lastly, we speculate that S_r is minimax optimal in detecting the signal $\|R - I_m\|_2$ for the null hypothesis (1.1.1) *under Gaussian assumption* on $\mathbf{X}$, although to our knowledge there hasn't been theoretical result on this in the literature; see also the last section of Cai and Ma (2013) for other related open problems.

1.6.3 Choice of statistics in practice

As per our numerical study, if normality is a reasonable assumption for our data, the three statistics S_r, S_τ, T_τ are all comparable choices in terms of test power, but Schott (2005)'s S_r has its appeal since the computation of each $r^{(pq)}$ is linear in the sample size n. When the Gaussian assumption is dubious, tests with our classes of rank-based statistics should be employed to guaranteed reasonable test sizes. Generally speaking, the classes of statistics S_h and T_h are preferable to Z_h since the former provide proper "two-sided" tests for the null hypothesis (1.1.1), unless one forms Z_h with kernels such as h_{t*} and h_D which always have non-negative expectation but are computationally more costly due to the higher degree of these kernels. For a given kernel h, T_h is generally more costly to compute than S_h, but it has more robust power regardless of the distribution of $\mathbf{X}$ since it unbiasedly estimates the dependency signal strength $\|\Theta\|_2^2$.

1.7 Conclusion

This chapter treats nonparametric tests of independence using pairwise rank correlations or, more precisely, rank correlations that are also U-statistics. As reviewed in Section 1.2, the motivating examples are Kendall's tau and Spearman's rho but also Hoeffding's D and Bergsma and Dassios' sign covariance t^*. The latter two correlations allow for consistent assessment of pairwise independence but form degenerate U-statistics. With a view towards alternatives in which pairwise dependence is "spread out over many coordinates", we proposed statistics that are formed as sums of many pairwise dependency signals as explained in Section 1.3. In a high-dimensional regime in which both the number of variables m and the sample size n tend to infinity, we derived normal limits for the null distribution of these statistics (Section 1.4). Our framework allows for U-statistic degeneracy of order up to two. Finally, we explored aspects of power theoretically and empirically (Sections 1.5 and 1.6).

Under the null hypothesis of independence, the m rank vectors are independent, each following a uniform distribution on the symmetric group $\mathfrak{S}_n$. In small to moderate size problems, we may thus implement exact tests using Monte Carlo simulation to compute critical values. However, for large-scale problems and/or when using the computationally more involved t^* or D, the asymptotic normal distributions we derived furnish accurate approximations and allow for great computational savings.

Our study of power has focused on the case of Kendall's tau. In a minimax paradigm and for Gaussian equicorrelation alternatives we showed rate-optimality for the test based on T_τ, the unbiased estimator of the signal strength defined via (1.3.6) with kernel $h = h_\tau$. It would be an interesting problem for future work to prove such rate-optimality more broadly, for more general alternatives as well as other kernels. In particular, for the kernel associated to Kendall's tau, we conjectured in Section 1.5.2 that rate-optimality holds for alternatives from a wide class of elliptical distributions.

Table 1.2: Simulated size of tests when $X^{(1)}, \ldots, X^{(m)}$ are i.i.d. $t_{3,2}$ data. For each combination of (m, n) and each test, the empirical sizes are computed from 5000 independently generated datasets.

Statistics	$n \backslash m$	4	8	16	32	64	128	256
S_r	32	0.066	0.078	0.076	0.081	0.076	0.089	0.079
S_τ		0.059	0.069	0.067	0.077	0.073	0.071	0.070
T_τ		0.064	0.078	0.075	0.087	0.081	0.082	0.080
S_ρ		0.047	0.054	0.052	0.061	0.056	0.053	0.056
$T_{\hat{\rho}_s}$		0.062	0.075	0.072	0.082	0.080	0.079	0.072
S_{t^*}		0.056	0.081	0.085	0.090	0.088	0.078	0.087
Z_{t^*}		0.062	0.069	0.067	0.081	0.077	0.077	0.079
S_r	64	0.073	0.083	0.095	0.095	0.102	0.097	0.096
S_τ		0.057	0.061	0.062	0.065	0.058	0.058	0.065
T_τ		0.058	0.064	0.066	0.069	0.061	0.064	0.067
S_ρ		0.048	0.053	0.055	0.055	0.050	0.052	0.057
$T_{\hat{\rho}_s}$		0.057	0.061	0.065	0.067	0.060	0.064	0.059
S_{t^*}		0.045	0.074	0.064	0.070	0.068	0.070	0.069
Z_{t^*}		0.054	0.061	0.058	0.064	0.065	0.062	0.063
S_r	128	0.072	0.089	0.107	0.112	0.101	0.109	0.110
S_τ		0.047	0.061	0.053	0.061	0.052	0.056	0.053
T_τ		0.049	0.063	0.053	0.064	0.054	0.060	0.054
S_ρ		0.043	0.059	0.049	0.056	0.048	0.052	0.048
$T_{\hat{\rho}_s}$		0.048	0.062	0.052	0.060	0.055	0.057	0.058
S_{t^*}		0.041	0.066	0.070	0.071	0.060	0.058	0.052
Z_{t^*}		0.050	0.055	0.058	0.062	0.053	0.056	0.055

Table 1.3: Simulated power of different test statistics for data generated from the multivariate normal distribution (MVN) with three different values for the dependency signal $\|\Theta_\tau\|_2^2$. All pairwise (population) Kendall's tau correlations $\theta_\tau^{(pq)}, 1 \leq p < q \leq m$ are equal to the same value θ so that $\|\Theta_\tau\|_2^2 = \binom{m}{2}\theta^2$. For each combination of (m, n) and each test, the empirical power is calculated from 500 independently generated datasets.

		$\|\Theta_\tau\|_2^2 = 0.1$			$\|\Theta_\tau\|_2^2 = 0.3$			$\|\Theta_\tau\|_2^2 = 0.7$		
Statistic	$n\backslash m$	64	128	256	64	128	256	64	128	256
MVN										
S_τ	64	0.094	0.054	0.070	0.182	0.108	0.092	0.424	0.218	0.114
T_τ	64	0.100	0.068	0.078	0.194	0.110	0.090	0.426	0.228	0.134
$S_\tau^{\max}$	64	0.046	0.046	0.020	0.040	0.058	0.046	0.056	0.054	0.058
S_r	64	0.070	0.058	0.070	0.178	0.114	0.080	0.448	0.222	0.110
Cai & Ma	64	0.076	0.076	0.060	0.190	0.116	0.086	0.456	0.278	0.130
S_τ	128	0.130	0.086	0.056	0.342	0.164	0.080	0.794	0.444	0.176
T_τ	128	0.132	0.088	0.058	0.352	0.174	0.084	0.806	0.446	0.186
$S_\tau^{\max}$	128	0.062	0.064	0.052	0.046	0.058	0.060	0.094	0.058	0.060
S_r	128	0.142	0.072	0.066	0.378	0.172	0.084	0.832	0.514	0.198
Cai & Ma	128	0.134	0.064	0.068	0.386	0.172	0.096	0.834	0.520	0.204
S_τ	256	0.256	0.108	0.096	0.780	0.358	0.198	0.992	0.838	0.476
T_τ	256	0.262	0.114	0.094	0.782	0.364	0.200	0.992	0.830	0.470
$S_\tau^{\max}$	256	0.048	0.050	0.046	0.064	0.056	0.058	0.124	0.082	0.052
S_r	256	0.282	0.126	0.094	0.816	0.420	0.224	1.000	0.880	0.502
Cai & Ma	256	0.282	0.124	0.110	0.812	0.422	0.234	1.000	0.882	0.494

Table 1.4: Simulated power of different test statistics for data generated from the multivariate t (MVT) and multivariate power exponential (MVPE) distributions with three different values for the dependency signal $\|\Theta_\tau\|_2^2$. All pairwise (population) Kendall's tau correlations $\theta_\tau^{(pq)}, 1 \le p < q \le m$ are equal to the same value θ so that $\|\Theta_\tau\|_2^2 = \binom{m}{2}\theta^2$. For each combination of (m, n) and each test, the empirical power is calculated from 500 independently generated datasets.

		$\|\Theta_\tau\|_2^2 = 0.1$			$\|\Theta_\tau\|_2^2 = 0.3$			$\|\Theta_\tau\|_2^2 = 0.7$		
Statistic	$n\backslash m$	64	128	256	64	128	256	64	128	256
MVT										
S_τ	64	0.506	0.866	0.998	0.628	0.896	0.998	0.802	0.926	0.998
T_τ		0.130	0.080	0.078	0.232	0.128	0.096	0.488	0.234	0.114
$S_\tau^{\max}$		0.080	0.066	0.060	0.086	0.074	0.060	0.110	0.074	0.068
S_τ	128	0.554	0.912	0.998	0.806	0.948	1.000	0.962	0.990	1.000
T_τ		0.130	0.102	0.094	0.384	0.210	0.114	0.796	0.494	0.244
$S_\tau^{\max}$		0.064	0.060	0.054	0.080	0.064	0.066	0.114	0.074	0.076
S_τ	256	0.694	0.924	1.000	0.972	0.992	1.000	1.000	1.000	1.000
T_τ		0.268	0.130	0.084	0.740	0.348	0.188	0.998	0.832	0.456
$S_\tau^{\max}$		0.076	0.062	0.072	0.110	0.066	0.076	0.186	0.102	0.078
MVPE										
S_τ	64	0.052	0.042	0.022	0.128	0.056	0.044	0.358	0.122	0.060
T_τ		0.114	0.076	0.076	0.222	0.110	0.082	0.462	0.216	0.134
$S_\tau^{\max}$		0.056	0.050	0.032	0.046	0.050	0.034	0.062	0.054	0.036
S_τ	128	0.074	0.038	0.028	0.274	0.094	0.036	0.744	0.314	0.112
T_τ		0.128	0.084	0.056	0.398	0.174	0.096	0.836	0.454	0.214
$S_\tau^{\max}$		0.038	0.054	0.050	0.050	0.056	0.044	0.084	0.060	0.046
S_τ	256	0.134	0.066	0.050	0.638	0.256	0.102	0.992	0.794	0.306
T_τ		0.232	0.152	0.100	0.768	0.370	0.184	0.998	0.862	0.450
$S_\tau^{\max}$		0.052	0.036	0.060	0.074	0.040	0.060	0.120	0.064	0.062

Chapter 2

IDENTIFIABILITY OF DIRECTED GAUSSIAN GRAPHICAL MODELS WITH ONE LATENT SOURCE VARIABLE

2.1 Introduction

In this chapter we study parameter identifiability in directed Gaussian graphical models with a latent variable. Our work falls in a line of work where the graphical representation of causally interpretable latent variable models is used to give tractable criteria to decide whether parameters can be uniquely recovered from the joint distribution of the observed variables (Pearl, 2009). Some examples of prior work in this context are Chen et al. (2014), Drton et al. (2011), Foygel et al. (2012), Grzebyk et al. (2004), Kuroki and Miyakawa (2004), Kuroki and Pearl (2014), Stanghellini and Wermuth (2005), Tian (2005), and Tian (2009).

The setup we consider has a single latent variable appearing as a source node in the directed graph defining the Gaussian model. The resulting models can be described as follows. Let $X_1, \dots, X_m$ be observable variables, and let L be a hidden variable, and suppose the variables are related by linear equations as

$$X_v = \sum_{w \neq v} \lambda_{wv} X_w + \delta_v L + \epsilon_v, \qquad v = 1, \dots, m,$$

where λ_{wv}, δ_v are real coefficients quantifying linear relationships, and the ϵ_v are independent mean zero Gaussian noise terms with variances $\omega_v > 0$. The latent variable L is assumed to be standard normal and independent of the noise terms ϵ_v. Letting $X = (X_1, \dots, X_m)^T$, $\epsilon = (\epsilon_1, \dots, \epsilon_m)^T$ and $\delta = (\delta_1, \dots, \delta_m)^T$, we may present the model in the vectorized form

$$X = \Lambda^T X + \delta L + \epsilon, \tag{2.1.1}$$

where Λ is the matrix (λ_{wv}) with $\lambda_{vv} = 0$ for all $v = 1, \dots, m$. We are then interested in specific models, in which for certain pairs of nodes $w \neq v$ the coefficient λ_{wv} is constrained to zero. In particular, we are interested in recursive models, that is, models in which the matrix Λ can be brought into strictly upper triangular form by permuting the indices of the variables (and thus the rows and columns of Λ). This implies that $I_m - \Lambda$ is invertible, where I_m is the $m \times m$ identity matrix. It follows that the observable variate vector X has a m-variate normal distribution $N_m(0, \Sigma)$ with covariance matrix

$$\Sigma = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}, \tag{2.1.2}$$

where Ω is the diagonal matrix with $\Omega_{vv} = \omega_v$. For additional background on graphical models we refer the reader to Lauritzen (1996) and Pearl (2009). We note that the models we consider also belong to the class of linear structural equation models (Bollen, 1989).

A Gaussian latent variable model postulating recursive zero structure in the matrix Λ from (2.1.1) can be thought of as associated with a graph $G = (V, E)$ whose vertex set $V = \{1, \dots, m\}$ is the index set for the observable variables $X_1, \dots, X_m$. For two distinct nodes $w, v \in V$, the edge set E includes the directed edge (w, v), denoted as $w \to v$ if and only if the model includes λ_{wv} as a free parameter. When the model is recursive, the directed graph G is acyclic and following common terminology we refer to G as a DAG (for directed acyclic graph). In this chapter, we will then always assume that the nodes are labeled in topological order, that is, we have $V = \{1, \dots, m\}$ and $w \to v \in E$ only if $w < v$.

To emphasize the presence of the latent variable L, one could equivalently represent the model by an extended DAG $\overline{G} = (\overline{V}, \overline{E})$ on $m+1$ nodes enumerated as $\overline{V} := \{0, 1, \dots, m\}$, where the node 0 corresponds to the latent variable L, and if $G = (V, E)$ is the graph on m nodes representing the model as in the preceding paragraph, then $\overline{E} = E \cup \{0 \to v : v \in \{1, \dots, m\}\}$. The edges $0 \to v$ correspond to the coefficients δ_v.

For the DAG $G = (V, E)$, let

$$\mathbb{R}_E := \{\Lambda = (\lambda_{wv}) \in \mathbb{R}^{m \times m} : w \to v \notin E \Rightarrow \lambda_{wv} = 0\}$$

be the linear space of coefficient matrices, and let diag_m^+ be the set of all $m \times m$ diagonal matrices with a positive diagonal.

Definition 2.1.1. The *Gaussian one latent source model* associated with a given DAG $G = (V, E)$, denoted as $\mathcal{N}_*(G)$, is the family of all m-variate normal distributions $N_m(0, \Sigma)$ with a covariance matrix of the form

$$\Sigma = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1},$$

for $\Lambda \in \mathbb{R}_E$, $\Omega \in \mathrm{diag}_m^+$ and $\delta \in \mathbb{R}^m$.

The model $\mathcal{N}_*(G)$ has the parametrization map

$$\phi_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{2.1.3}$$

defined on the set $\Theta := \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$, which we may also view as an open subset of $\mathbb{R}^{2m+|E|}$, where $|E|$ is the cardinality of the directed edge set E. Clearly, the image of ϕ_G is in PD_m, the cone of positive definite $m \times m$ matrices. Note that since G is acyclic, we have $(I_m - \Lambda)^{-1} = I_m + \Lambda + \Lambda^2 + \cdots + \Lambda^{m-1}$ and thus the covariance parametrization ϕ_G is a polynomial map.

In this chapter we will derive graphical conditions on G that are sufficient/necessary for identifiability of the model $\mathcal{N}_*(G)$. We begin by clarifying what precisely we mean by identifiability. The most stringent notion, namely that of global identifiability, requires ϕ_G to be injective on all of Θ. While this notion is important (Drton et al., 2011), it is too stringent for the setting we consider here. Indeed, for any triple $(\Lambda, \Omega, \delta) \in \Theta$, $\phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda, \Omega, -\delta)$, which implies that the *fiber*

$$\{(\Lambda', \Omega', \delta') \in \Theta : \phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda', \Omega', \delta')\}$$

always has cardinality ≥ 2. We may account for this symmetry by requiring ϕ_G to be 2-to-1 on all of Θ but this is not enough as there are always some fibers that are infinite. For instance, it is easy to show that the fiber in the above display is infinite when $\delta = 0$. As such, it is natural to consider notions of generic identifiability. Specifically, our contributions will pertain to the notion of generic finite identifiability, as defined below, that only requires finite identification of parameters away from a fixed null set in Θ; here a null set is a set of

Lebesgue measure zero. This notion is also referred to as local identifiability in other related work such as Anderson and Rubin (1956).

Null sets appearing in our work are algebraic sets, where an algebraic set $A \subset \mathbb{R}^n$ is the set of common zeros of a collection of multivariate polynomials, i.e.,

$$A = \{a \in \mathbb{R}^n : f_i(a) = 0,\ i = 1, \ldots, k\},$$

for $f_i \in \mathbb{R}[x_1, \ldots, x_n]$, where $\mathbb{R}[x_1, \ldots, x_n]$ is the ring of polynomials in n variables with coefficients in $\mathbb{R}$. If all polynomials f_i are the zero polynomial then $A = \mathbb{R}^n$. Otherwise, A is a proper subset, $A \subsetneq \mathbb{R}^n$, and its dimension is then less than n. In particular, a proper algebraic subset of $\mathbb{R}^n$ has measure zero.

Definition 2.1.2. Let S be an open subset of $\mathbb{R}^n$, and let f be a map defined on S. Then f is said to be *generically finite-to-one* if there exists a proper algebraic set $\tilde{S} \subset \mathbb{R}^n$ such that the *fiber* of s, i.e. the set $\{s' \in S : f(s') = f(s)\}$, is finite for all $s \in S \setminus \tilde{S}$. Otherwise, f is said to be generically infinite-to-one.

Definition 2.1.3. The model $\mathcal{N}_*(G)$ of a given DAG $G = (V, E)$ is said to be *generically finitely identifiable* if its parametrization ϕ_G defined on Θ is generically finite-to-one. We also say the DAG G is generically finitely identifiable for short.

Hereafter for any map f defined on an open domain $S \subset \mathbb{R}^n$, we will use

$$\mathcal{F}_f(s) := \{s' \in S : f(s') = f(s)\} \tag{2.1.4}$$

to denote the fiber of a point $s \in \mathcal{S}$. If T is a subset of S, we will use $f|_T$ to denote the restriction of f to T, in which case for any $t \in T$, we have the fiber

$$\mathcal{F}_{f|_T}(t) = \{t' \in T : f(t') = f(t)\}.$$

The term "generic point" will refer to any point in the domain S that lies outside a fixed proper algebraic subset $\tilde{S}$, and a property is said to hold generically if it holds everywhere on $S \setminus \tilde{S}$. The following well-known lemma is a main tool in this chapter, and its proof will be included in Appendix B.1 for completeness. It gives as an immediate corollary a trivial necessary condition for generic finite identifiability.

Lemma 2.1.1. *Suppose $f : S \to \mathbb{R}^d$ is a polynomial map defined on an open set $S \subset \mathbb{R}^n$. The following statements are equivalent:*

i. f is generically finite-to-one.

ii. There exists a proper algebraic subset $\tilde{S} \subset \mathbb{R}^n$ such that the fibers of the restricted map $f|_{S \setminus \tilde{S}}$ are all finite, i.e. $|\mathcal{F}_{f|_{S\setminus\tilde{S}}}(s)| < \infty$ for all $s \in S \setminus \tilde{S}$.

iii. The Jacobian matrix of f is generically of full column rank.

Corollary 2.1.2. *Given a DAG $G = (V, E)$, a necessary condition for generic finite identifiability of its associated model $\mathcal{N}_*(G)$ is that $\binom{m+1}{2} - 2m \geq |E|$.*

Proof of Corollary 2.1.2. The Jacobian matrix of ϕ_G is of size $\binom{m+1}{2} \times (|E| + 2m)$, and it is necessary that $\binom{m+1}{2} \geq |E| + 2m$ for it to have full column rank. □

Property (ii) is seemingly weaker than (i) in Lemma 2.1.1. It is useful in proving our results in Section 2.5. In light of Corollary 2.1.2, for the rest of this chapter we will restrict our attention to DAGs $G = (V, E)$ with $\binom{m+1}{2} - 2m \geq |E|$, in which case m must be at least 3.

One of our contributions is a sufficient graphical condition stated in Theorem 2.1.3 below. For $v \neq w \in V$, we will use $v — w$ or $w — v$ to denote the edge $(v, w) = (w, v)$ of an undirected graph on V. With slight abuse of notation, we may also use $v — w$ or $w — v$ to denote an edge $v \to w \in E$ when the directionality of edges in a DAG $G = (V, E)$ is to be ignored. For any directed/undirected graph $G = (V, E)$, the complement of G, denoted as $G^c = (V, E^c)$, is the undirected graph on V with the edge set $E^c = \{v — w : (v, w) \notin E \text{ and } (w, v) \notin E\}$.

Figure 2.1: A DAG G that satisfies the sufficient condition in Theorem 2.1.3; its undirected complement G^c is shown on the right.

Theorem 2.1.3 (Sufficient condition for generic finite identifiability)**.** *The model $\mathcal{N}_*(G)$ given by a DAG $G = (V, E)$ is generically finitely identifiable if every connected component of G^c contains an odd cycle.*

Figure 2.1 shows a DAG G that satisfies the sufficient condition in Theorem 2.1.3; its undirected complement G^c is shown on the right of the figure. We will revisit this example in Section 2.4, where we report on algebraic computations that show that for this graph G the fibers of ϕ_G are generically of size 2 or 4.

Our approach to proving Theorem 2.1.3 also yields a necessary condition for generic finite identifiability. This condition can be stated in terms of two undirected graphs on the node set V, denoted $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$, where $E_{|L,cov}$ captures the dependency of variable pairs after conditioning on the latent variable L, and E_{con} captures the dependency of variable pairs after conditioning on all other variables. From (2.1.1) it can be seen that $\Sigma_{|L} := (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$ is the covariance matrix of X conditioning on L, hence $v - w \in E_{|L,cov}$ if and only if $(\Sigma_{|L})_{vw} \not\equiv 0$, and analogously $v - w \in E_{con}$ if and only if $(\Sigma_{|L}^{-1})_{vw} \not\equiv 0$. It is well known that these two undirected graphs can be obtained by using the d-separation criterion applied to the extended DAG $\overline{G}$; see Drton et al. (2009, p. 73) for example.

Theorem 2.1.4 (Necessary condition for generic finite identifiability)**.** *Given a DAG $G = (V, E)$, for the model $\mathcal{N}_*(G)$ to be generically finitely identifiable, it is necessary that the following two conditions both hold:*

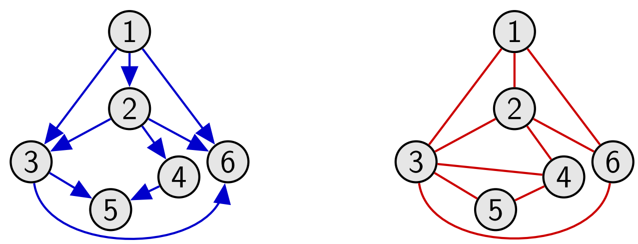

Figure 2.2: A graph G (left), G_{con} (middle) and G^c_{con} (right). Since $|E_{con}| - |E| = 1 < 2 = d_{con}$, the necessary condition in Thm. 2.1.4 does not hold.

i. $|E_{con}| - |E| \geq d_{con}$*, where* d_{con} *is the number of connected components in the graph* $(G_{con})^c$ *that do not contain any odd cycle;*

ii. $|E_{|L,cov}| - |E| \geq d_{cov}$*, where* d_{cov} *is the number of connected components in the graph* $(G_{|L,cov})^c$ *that do not contain any odd cycle.*

Figure 2.2 gives an example of a DAG that fails to satisfy our necessary condition, specifically, condition (*ii*).

In addition to the closely related work of Stanghellini (1997) and Vicard (2000), identifiability of directed Gaussian models with one latent variable has been studied by Stanghellini and Wermuth (2005). The models we treat here are special cases with the latent node being a common parent of all the observable nodes. As we review in more detail in Section 2.2, we can readily adapt the sufficient graphical criteria given in Stanghellini and Wermuth (2005) for certifying that the model $\mathcal{N}_*(G)$ of a given DAG G is generically finitely identifiable with respect to Definition 2.1.3. Our own sufficient condition stated in Theorem 2.1.3 is stronger, in the sense that every DAG G satisfying the sufficient conditions in Stanghellini and Wermuth (2005) necessarily satisfies the condition in Theorem 2.1.3. However, when it applies the result of Stanghellini and Wermuth (2005) yields a stronger conclusion than our generic finiteness result. Indeed as we also emphasize in the discussion in Section 2.6, their conditions imply that the parmetrization is generically 2-to-1.

We will prove the above stated Theorems 2.1.3 and 2.1.4 in Section 2.3. Since the parametrization map in (2.1.3) is polynomial, the generic finite identifiability of a given model is decidable by algebraic techniques that involve Gröbner basis computations. In Section 2.4, we will study the applicability of our graphical criteria via such algebraic computations for all models $\mathcal{N}_*(G)$ of DAGs G with $m = 4, 5, 6$ nodes. Section 2.4 also contains simulation experiments for larger graphs and a discussion of the computational complexity of checking the graphical conditions. Section 2.5 will give results on situations where we can determine generic finite identifiability of a model $\mathcal{N}_*(G)$ based on knowledge about the generic finite identifiability of a model $\mathcal{N}_*(G')$, where G' is an induced subgraph of G.

Before ending this introduction, however, we comment on the role that Markov equivalence plays in our problem. Recall that two DAGs defined on the same set of nodes are Markov equivalent if they have the same d-separation relations. The following theorem, which will be proved in Appendix B.1, says that generic finite identifiability is a property of Markov equivalence classes of DAGs.

Theorem 2.1.5. *Suppose $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are two Markov equivalent DAGs on the same set of nodes V. Then the model $\mathcal{N}_*(G_1)$ is generically finitely identifiable if and only if the same is true for $\mathcal{N}_*(G_2)$.*

2.2 Prior work

Stanghellini and Wermuth (2005) give sufficient graphical conditions for identifiability of directed Gaussian graphical models with one latent variable that can be any node in the DAG. We revisit their result in the context of the models from Definition 2.1.1 and formulate it in terms of generic finite identifiability. (As was mentioned in the Introduction, their result yields in fact the stronger conclusion of a generically 2-to-1 parametrization.) We begin by stating a well-known fact about DAG models without latent variables.

Lemma 2.2.1. *For any DAG $G = (V, E)$ with $m = |V|$ nodes, the map*

$$(\Lambda, \Omega) \;\mapsto\; (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$$

is injective on the domain $\mathbb{R}_E \times \mathrm{diag}_m^+$ *and has a rational inverse.*

Proof. For any $(\Lambda, \Omega) \in \mathbb{R}_E \times \mathrm{diag}_m^+$, let $\Sigma = (\sigma_{vw}) = (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$. Let $pa(v) = \{w : w \to v \in E\}$ be the parent set of the node v. Then one can show, by induction on m and considering a topological ordering of V, that

$$\Lambda_{pa(v),v} \;=\; \left(\Sigma_{pa(v),pa(v)}\right)^{-1} \Sigma_{pa(v),v}$$

and

$$\Omega_{vv} \;=\; \sigma_{vv} - \Sigma_{v,pa(v)} \left(\Sigma_{pa(v),pa(v)}\right)^{-1} \Sigma_{pa(v),v};$$

compare, for instance, Richardson and Spirtes (2002, §8). □

Let the random vector X and the latent variable L have their joint distribution specified via the equation system from (2.1.1). Write $\Sigma_{|L}$ for the conditional covariance matrix of X given L. Then it holds that

$$\Sigma_{|L} = (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}. \tag{2.2.1}$$

Hence, by Lemma 2.2.1, when knowing $\Sigma_{|L}$ we can uniquely solve for the pair (Λ, Ω), which are rational functions of $\Sigma_{|L}$. Writing Σ for the (unconditional) covariance matrix of X, we have from (2.1.2) that

$$\Sigma_{|L} = \Sigma - (I_m - \Lambda^T)^{-1}\delta\delta^T(I_m - \Lambda)^{-1}.$$

Consequently, (Λ, Ω) can be recovered uniquely from Σ and $(I_m - \Lambda^T)^{-1}\delta$. The results of Stanghellini and Wermuth (2005) then address identification of the vector $(I_m - \Lambda^T)^{-1}\delta$, which holds the covariances between each coordinate of X and the latent variable L. We obtain the following observation.

Proposition 2.2.2 (Adapted from Stanghellini and Wermuth, 2005)**.** *Let* $G = (V, E)$ *be a DAG. The model* $\mathcal{N}_*(G)$ *is generically finitely identifiable if*

i. every connected component of $G^c_{|L,cov} = (V, E^c_{|L,cov})$ *has an odd cycle, or*

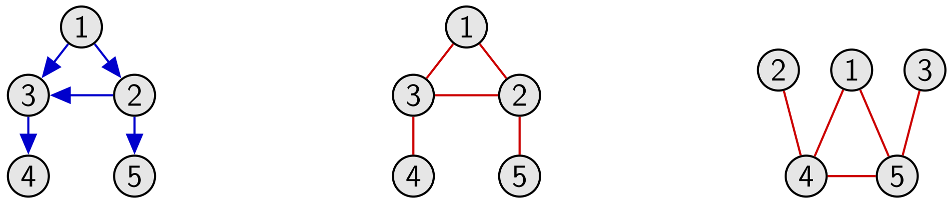

Figure 2.3: A graph G (left) satisfying the sufficient condition in Proposition 2.2.2, with G_{con} (middle) and G^c_{con} (right).

ii. every connected component of $G^c_{con} = (V, E^c_{con})$ *has an odd cycle.*

Proof. Theorem 1 in Stanghellini and Wermuth (2005) gives (i) or (ii) as a sufficient condition for identifying, up to sign, the m-vector $(I_m - \Lambda^T)^{-1}\delta$ when $\Sigma = \phi_G(\Lambda, \Omega, \delta)$ for a generic point $(\Lambda, \Omega, \delta)$ in Θ. In this case, we can uniquely recover the conditional covariance matrix $\Sigma_{|L}$ from (2.2.1) and also the pair (Λ, Ω) by Lemma 2.2.1. After identifying Λ, δ can be solved for, up to sign, by the previous knowledge of $(I_m - \Lambda^T)^{-1}\delta$. Hence, (i) or (ii) is in fact a sufficient condition for generic finite identifiability of $\mathcal{N}_*(G)$. $\square$

Figure 2.3 shows a DAG with $m = 5$ nodes that satisfies the condition of Proposition 2.2.2(ii).

We conclude this review of prior work by pointing out that any model $\mathcal{N}_*(G)$ that can be determined to be generically finitely identifiable using Proposition 2.2.2 can also be found to have this property using our new Theorem 2.1.3.

Proposition 2.2.3. *A DAG* $G = (V, E)$ *satisfying either one of the conditions in Proposition 2.2.2 necessarily satisfies the condition in Theorem 2.1.3.*

Proof. Let $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$. An edge $v \to w \in E$ also presents itself as an undirected edge in both $E_{|L,cov}$ and E_{con}. Hence, when ignoring the directionality of its edges, G is a subgraph of both $G_{|L,cov}$ and G_{con} and, thus, G^c is a supergraph of both $G^c_{|L,cov}$ and G^c_{con}. As such, if every connected component of $G^c_{|L,cov}$, or of G^c_{con}, contains an odd cycle, the same is true of G^c. $\square$

2.3 Criteria based on the Jacobian of parametrization maps

In this section, we prove Theorems 2.1.3 and 2.1.4. Let $G = (V, E)$ be a fixed DAG with $m = |V|$ nodes, and let $\Theta := \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$ denote again the domain of the parametrization

$$\phi_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}$$

of the covariance matrix of the distributions in model $\mathcal{N}_*(G)$. We begin by introducing other mappings that are generically finite-to-one if and only if ϕ_G is generically finite-to-one.

First, it will be helpful to study the map

$$\tilde{\phi}_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1} + \delta\delta^T, \tag{2.3.1}$$

defined on Θ. Second, focusing on concentration instead of covariance matrices, we will also consider the maps

$$\varphi_G : (\Lambda, \Psi, \gamma) \longmapsto (I_m - \Lambda)(\Psi - \gamma\gamma^T)(I_m - \Lambda^T), \tag{2.3.2}$$

$$\tilde{\varphi}_G : (\Lambda, \Psi, \gamma) \longmapsto (I_m - \Lambda)\Psi(I_m - \Lambda^T) - \gamma\gamma^T. \tag{2.3.3}$$

Lemma 2.3.1. *The parametrization ϕ_G is generically finite-to-one if and only if any one of the maps $\tilde{\phi}_G$, φ_G and $\tilde{\varphi}_G$ is generically finite-to-one.*

Proof. Consider first the map $\tilde{\phi}_G$ for which it holds that $\phi_G = \tilde{\phi}_G \circ g$, where

$$g : (\Lambda, \Omega, \delta) \longmapsto (\Lambda, \Omega, (I_m - \Lambda^T)^{-1}\delta)$$

is a diffeomorphism that maps Θ to itself. By the chain rule, the Jacobian of ϕ_G at $(\Lambda, \Omega, \delta)$ is the product of the Jacobian of $\tilde{\phi}_G$ at $g(\Lambda, \Omega, \delta)$ and the Jacobian of g at $(\Lambda, \Omega, \delta)$. Now the latter matrix is invertible on all of Θ since g is a diffeomorphism. It follows that there exists a point in Θ at which the Jacobian of ϕ_G has full column rank if and only if the same is true for $\tilde{\phi}_G$. For the Jacobian of a polynomial map such as ϕ_G and $\tilde{\phi}_G$, full column rank at a single point implies generically full column rank; use the subdeterminants that characterize a drop in rank to define a proper algebraic subset of exceptions, see also Geiger et al. (2001, Lemma 9). The claim about ϕ_G and $\tilde{\phi}_G$ follows from Lemma 2.1.1.

Let $h : (\Lambda, \Psi, \gamma) \longmapsto (\Lambda, \Psi, (I_m - \Lambda)\gamma)$. Since $\varphi_G = \tilde{\varphi}_G \circ h$, by the same argument as above it also holds that φ_G is generically finite-to-one if and only if $\tilde{\phi}_G$ has this property.

In order to complete the proof of the lemma it suffices to show that ϕ_G is generically finite-to-one if and only if the same holds for φ_G. Define another diffeomorphism from Θ to itself as

$$\rho : (\Lambda, \Omega, \delta) \longmapsto \left(\Lambda, \Omega^{-1}, (1 + \delta^T \Omega^{-1} \delta)^{-1/2} \Omega^{-1} \delta\right).$$

Writing *inv* for matrix inversion, we then have that

$$inv \circ \phi_G = \varphi_G \circ \rho \tag{2.3.4}$$

because of the identity $(\Omega + \delta\delta^T)^{-1} = (\Psi - \gamma\gamma^T)$ with $\Psi = \Omega^{-1}$ and $\gamma = k^{-1/2}\Psi\delta$, where $k = 1 + \delta^T \Psi \delta > 0$; see e.g. Rao (1973, p. 33). Using (2.3.4), the equivalence of being generically finite-to-one for ϕ_G and φ_G may be argued similarly as for the maps considered earlier. □

Let $J(\tilde{\varphi}_G)$ be the Jacobian matrix of the map $\tilde{\varphi}_G$ from (2.3.3). It will be examined to prove Theorem 2.1.3. In light of Lemmas 2.1.1 and 2.3.1, we will show that if G satisfies the condition in Theorem 2.1.3, then $J(\tilde{\varphi}_G)$ is generically of full column rank, implying that ϕ_G is generically finite-to-one. Our arguments will make use of the following lemma that rests on observations made in Vicard (2000).

Lemma 2.3.2. *Let $G = (V, E)$ be an undirected graph, and let $f_G : \mathbb{R}^V \to \mathbb{R}^E$ be the map with coordinate functions*

$$f_{G,vw}(x) \;=\; x_v x_w, \quad v - w \in E.$$

Then the Jacobian of f_G has generic rank $m - d$, where $m = |V|$ is the number of nodes and d is the number of connected components of G that do not contain an odd cycle.

Proof. For simpler notation, let $f := f_G$. Let J_f be the Jacobian matrix of the polynomial map f, and let $\ker(J_f)$ be its kernel. By the rank theorem (Rudin, 1976, p. 229), the dimension of $\ker(J_f)$ is generically equal to the dimension of the fiber $\mathcal{F}_f$; recall (2.1.4). Since $\operatorname{rank}(J_f) = m - \dim(\ker(J_f))$, it suffices to show that $\mathcal{F}_f$ has generic dimension d.

Since the claim is about a generic property, we may restrict the domain of f to the open set $\mathcal{X} := (\mathbb{R}\backslash\{0\})^m$. This assumption is made so that Lemma 1 in Vicard (2000) is applicable later without difficulty. Now, fix a point $y \in f(\mathcal{X}) \subset \mathbb{R}^E$. The elements of the fiber $\mathcal{F}_f(y)$ are the vectors $x \in \mathbb{R}^m$, or equivalently, $x \in \mathcal{X}$, that are solutions to the system of equations

$$y_{vw} = x_v x_w, \quad v — w \in E. \tag{2.3.5}$$

Let $G_1 = (V_1, E_1), \ldots, G_k = (V_k, E_k)$ be the connected components of G, so that $V_1, \ldots, V_k$ form a partition of V and $E_1, \ldots, E_k$ partition E. Let $k' \leq k$ be the number of connected components containing two nodes at least. Without loss of generality, assume $G_{k'+1}, \ldots, G_k$ are all the connected components with only a single node. Then the equations listed in (2.3.5) can be arranged to form k' disjoint subsystems indexed by $i = 1, \ldots, k'$. The i-th subsystem has the form

$$y_{vw} = x_v x_w, \quad v — w \in E_i \tag{2.3.6}$$

and exclusively involves the variables $\{x_v : v \in V_i\}$. By Lemma 1 in Vicard (2000) and also the relevant discussion in the proof of Theorem 1 in the same paper, the solution set to (2.3.6) either contains two points or can be parametrized by a single free variable in $\mathbb{R}$. The former case arises if and only if G_i contains an odd cycle. It follows that the dimension of the solution set of (2.3.6) is zero when G_i contains an odd cycle, and it has dimension one if G_i does not contain an odd cycle. In addition, each singleton component $G_i = (V_i, \emptyset)$ for $i = k'+1, \ldots, k$ provides one additional dimension to the fiber $\mathcal{F}_f(y)$, since the corresponding variables in x are not restricted by any equations. We conclude that the dimension of $\mathcal{F}_f(y)$ equals the number of connected components G_i that do not contain an odd cycle. $\square$

We return to the object of study, namely, the map $\tilde{\varphi}_G$ which sends the $(2m + |E|)$-dimensional set $\Theta = \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$ to the $\binom{m+1}{2}$-dimensional space of symmetric $m \times m$

matrices. The Jacobian $J(\tilde{\varphi}_G)$ is of size $\binom{m+1}{2} \times (2m + |E|)$, and we index its rows by pairs (v, w) with $1 \leq v < w \leq m$, whereas in Section 2.1 we assume the vertex set $V = \{1, \ldots, m\}$ to be topologically ordered. We now describe a particular way of arranging the rows and columns of $J(\tilde{\varphi}_G)$.

Define the set of "non-edges" as $N := \{(v, w) : v < w \text{ and } (v, w) \notin E\}$; we will also write $v \not\to w$ to express that $(v, w) \in N$. Also, define $D := \{(v, v) : v \in V\}$, so that $D \cup E \cup N$ indexes all entries in the upper triangular half of an $m \times m$ symmetric matrix. The rows of $J(\tilde{\varphi}_G)$ are now arranged in the order D, E and N. The columns of $J(\tilde{\varphi}_G)$ are indexed such that partial derivatives with respect to the free input variables in the triple (Λ, Ψ, γ) appear from left to right, in the order Ψ, Λ and γ. In other words, we partition $J(\tilde{\varphi}_G)$ into 9 blocks as follows:

$$J(\tilde{\varphi}_G) = \begin{array}{c c} & \begin{array}{ccc} \Psi & \Lambda & \gamma \end{array} \\ \begin{array}{c} D \\ E \\ N \end{array} & \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \end{array}. \tag{2.3.7}$$

The following lemma is obtained by inspection of the partial derivatives of $\tilde{\varphi}_G$. Its proof appears in Appendix B.2.

Lemma 2.3.3. *The Jacobian matrix $J(\tilde{\varphi}_G)$ is generically of full column rank provided that the submatrix $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is so.*

We now give the proof of Theorem 2.1.3.

Proof of Theorem 2.1.3. By Lemmas 2.1.1 and 2.3.3 , it suffices to show that $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank. For each $v \not\to w \in N$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = \left[(I_m - \Lambda)\Psi(I_m - \Lambda^T)\right]_{vw} - \gamma_v \gamma_w. \tag{2.3.8}$$

Note that only the right most term in (2.3.8) contributes to the partial derivatives of $\tilde{\varphi}_G$ with respect to $\gamma = (\gamma_v)_{v \in \{1,\ldots,m\}}$.

Ignoring the directionality of non-edges in N, define the undirected graph $H = (V, N)$ to which we associate a map f_H as in Lemma 2.3.2. Then

$$[J(\tilde{\varphi}_G)]_{N,\gamma} = -J_{f_H}.$$

But J_{f_H} has generically full column rank by Lemma 2.3.2 because, in fact, H is equal to the complementary graph G^c for which we assume that all connected components contain an odd cycle. □

We remark that Theorem 2.1.3 can also be proved by studying the Jacobian of the map $\tilde{\phi}_G$ from (2.3.1). We chose to work with $\tilde{\varphi}_G$ above since this allowed us to avoid consideration of the inverse of the matrix $I_m - \Lambda$. For Theorem 2.1.4, however, we consider both $\tilde{\varphi}_G$ and $\tilde{\phi}_G$.

Proof of Theorem 2.1.4. We first prove the necessity of condition (i) by showing that if $|E_{con}| - |E| < d_{con}$, then the Jacobian matrix $J(\tilde{\varphi}_G)$ always has row rank less than $2m + |E|$. This implies that it cannot be of full column rank which implies the failure of generic finite identifiability by Lemma 2.1.1.

As in the proof of Theorem 2.1.3, we consider the set of non-edges N, which we now partition as $N = N_1 \dot{\cup} N_2$, where $N_1 = \{v \not\to w \in E : \; v - w \in E_{con}\}$, and $N_2 = N \setminus N_1$. Accordingly, we can partition the submatrix $[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$ into two block of rows indexed by N_1 and N_2 as

$$[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}} = \begin{array}{c} \\ N_1 \\ N_2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} \Psi & \Lambda & \gamma \end{array} \\ \begin{bmatrix} \cdots & \cdots & \cdots \\ 0 & 0 & \cdots \end{bmatrix} \end{array}. \tag{2.3.9}$$

To see that the submatrix $[J(\tilde{\varphi}_G)]_{N_2,\{\Psi,\Lambda\}} = 0$, observe first that an entry of $(I-\Lambda)\Psi(I-\Lambda^T)$ is the zero polynomial if and only if the same is true for $\Sigma_{|L}^{-1}$, where $\Sigma_{|L}$ is the matrix from (2.2.1). Second, by definition of E_{con} and N_2, if $(v, w) \in N_2$ then $(\Sigma_{|L}^{-1})_{vw} = 0$.

Next, observe that to prove the necessity of condition (i) it suffices to show that the rank of $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ cannot be larger than $m - d_{con}$. Indeed, if this is true, then there exists a subset $N_2' \subset N_2$ with $|N_2'| = m - d_{con}$, such that the submatrix

$$[J(\tilde{\varphi}_G)]_{\{D,E,N_1,N_2'\},\{\Psi,\Lambda,\gamma\}} \tag{2.3.10}$$

has the same rank as the original Jacobian matrix $J(\tilde{\varphi}_G)$. However, the submatrix (2.3.10) has $2m + |E_{con}| - d_{con}$ rows, and thus its rank is less than $2m + |E|$ because under condition (i) we have $|E_{con}| - |E| < d_{con}$. As a result, $J(\tilde{\varphi}_G)$ cannot be of full column rank.

It now remains to show that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ has rank at most $m - d_{con}$. Observe that the undirected graph (V, N_2) is equal to the complementary graph $(G_{con})^c$. Moreover, $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ is equal to the negative Jacobian of the map $f_{(G_{con})^c}$ that we get by applying the construction from Lemma 2.3.2 to $(G_{con})^c$; recall the proof of Theorem 2.1.3. Applying Lemma 2.3.2, we find that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ has generic rank $m - d_{con}$, which is also the maximal rank that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ may have.

The proof of (ii) follows the exact same argument as that of (i), by replacing (a) G_{con} with $G_{|L,cov}$, (b) d_{con} with d_{cov}, (c) $\tilde{\varphi}_G$ with $\tilde{\phi}_G$, (d) Ψ with Ω, (e) γ with δ and (f) $J(\tilde{\varphi}_G)$ with $J(\tilde{\phi}_G)$, where $J(\tilde{\phi}_G)$ is partitioned as

$$J(\tilde{\phi}_G) = \begin{array}{c} \\ D \\ E \\ N \end{array} \begin{array}{c} \begin{array}{ccc} \Omega & \Lambda & \delta \end{array} \\ \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \end{array}, \tag{2.3.11}$$

similarly to (2.3.7). □

2.4 Computations and simulation experiments

According to Theorems 2.1.3 and 2.1.4 generic (non-)identifiability can be verified using algorithms that determine the existence of odd cycles for each connected component of a given graph. Efficient algorithms based on depth-first search techniques exist for that purpose: Finding all connected components of a graph $G = (V, E)$, directed or not, requires $O(|V| + |E|)$ operations. If G is connected, determining the oddness of its cycles is equivalent to determining the oddness of the *fundamental cycles* with respect to a spanning tree of G, which requires $O(|V| + |E| + l)$ operations, where l is the sum of the lengths of these fundamental cycles. The relevant algorithms are discussed in Reingold et al. (1977, Chapter 8).

To certify generic finite identifiability of a model $\mathcal{N}_*(G)$ for a DAG G based on Theorem 2.1.3, its complement G^c has to be submitted as an input to the graph algorithms. When G is *sparse*, one may want to avoid handling the dense complement G^c. The following corollary gives two simple criteria.

Corollary 2.4.1. *The model $\mathcal{N}_*(G)$ for a DAG $G = (V, E)$ is generically finitely identifiable if*

i. G has two connected components, at least one of which is incomplete, or

ii. G has three connected components.

Proof. We will prove (i) here. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be the two connected components of G such that $V = V_1 \cup V_2$ and $E = E_1 \cup E_2$. Clearly, any two nodes $v \in V_1$ and $w \in V_2$ are adjacent in the complement $G^c = (V, E^c)$. Moreover, two nodes $v, w \in V_1$ are connected in G^c as we may pick any $u \in V_2$ to have $v - u$, $u - w \in E^c$. Similarly, any two nodes in V_2 are connected in G^c. Hence, G^c is a connected graph. The claim is then proven if we can show that G^c contains an odd cycle. By assumption of incompleteness, without loss of generality, there are $v_1, w_1 \in V_1$ such that $v_1 \to w_1, w_1 \to v_1 \notin E_1$. This implies that G^c contains the odd cycle $v_1 - u$, $u - w_1$, $w_1 - v_1$ for arbitrary $u \in V_2$.

The proof of (ii) is similar and simpler, and we omit it. □

2.4.1 Algebraic computations for small graphs

As explained in Drton (2006, §3) and Garcia-Puente et al. (2010), identifiability properties of a model such as $\mathcal{N}_*(G)$ can be decided using Gröbner basis techniques from computational algebraic geometry (Cox et al., 2007). While these techniques are tractable only for small to moderate size problems, we were able to perform an exhaustive algebraic study of all DAGs $G = (V, E)$ with $m \leq 6$ nodes. Beyond a mere decision on whether the parametrization map ϕ_G is generically 1-to-1, the algebraic methods also provide information about the generic cardinality of the fibers of ϕ_G as a map defined on complex space.

Definition 2.4.1. For a DAG $G = (V, E)$, let $\phi_G^{\mathbb{C}}$ be the map obtained by extending ϕ_G to the complex domain $\mathbb{C}^{2m+|E|}$. If the (complex) fibers of $\phi_G^{\mathbb{C}}$ are generically of cardinality k, then we say that $\phi_G^{\mathbb{C}}$ is generically k-to-one.

The language of Definition 2.4.1 allows us to give a refined classification of DAGs G in terms of the identifiability properties of the parametrization of model $\mathcal{N}_*(G)$. Indeed, $\mathcal{N}_*(G)$ is generically finitely identifiable if and only if $\phi_G^{\mathbb{C}}$ is generically k-to-one for some $k < \infty$.

Remark 2.4.1. The generic size of the fibers of $\phi_G^{\mathbb{C}}$ equals the generic size of the fibers of the complex extensions of the three maps from Lemma 2.3.1. The map $\tilde{\varphi}_G$ has low degree coordinates and tends to be the easiest to work with in algebraic computation. Another approach that can be useful is to adapt the algorithm described in Section 8 of the supplementary material for Foygel et al. (2012). To do this, note that for $\Lambda \in \mathbb{C}^{|E|}$ there exist complex choices of Ω and δ such that $\phi_G(\Lambda, \Omega, \delta) = \Sigma$ if and only if $(I - \Lambda^T)\Sigma(I - \Lambda)$ is a matrix that is the sum of a diagonal matrix, namely, Ω, and a symmetric matrix of rank 1, namely, $\delta\delta^T$. Whether a matrix is of the latter type can be tested using tetrads, that is, 2×2 subdeterminants involving only off-diagonal entries of the matrix; see also (2.5.4) below. The tetrads of a matrix form a Gröbner basis (de Loera et al., 1995, Drton et al., 2007).

Table 2.1 lists out the counts of DAGs $G = (V, E)$, with $4 \leq m \leq 6$ nodes, that have $\phi_G^{\mathbb{C}}$ generically k-to-one, for all possible values of k. The table also gives the the counts of DAGs satisfying the conditions in Theorems 2.1.3 and 2.1.4 as well as Proposition 2.2.2. DAGs with $\binom{m+1}{2} - 2m < |E|$, which trivially give generically ∞-to-one maps $\phi_G^{\mathbb{C}}$ in view of Corollary 2.1.2, are excluded. We emphasize that the counts are with respect to unlabeled

Table 2.1: Counts of unlabeled DAGs G with m nodes, at most $\binom{m+1}{2} - 2m$ edges, and complex parametrization $\phi_G^{\mathbb{C}}$ generically k-to-one. Counts are also given for DAGs that satisfy the sufficient conditions from Thm. 2.1.3 and Prop. 2.2.2, and DAGs that fail to satisfy the necessary condition from Thm. 2.1.4.

m	4	5	6
$k < \infty$	5	95	3344
$k = 2$	5	87	2961
$k = 4$	0	8	345
$k = 6$	0	0	24
$k = 8$	0	0	14
Prop. 2.2.2	5	49	985
Thm. 2.1.3	5	88	2957
$k = \infty$	1	20	552
Thm. 2.1.4	1	20	361
Total # of DAGs	6	115	3896

DAGs, that is, all DAGs that are isomorphic with respect to relabeling of nodes are counted as one unlabeled graph.

In the considered settings the condition in Theorem 2.1.3 is very successful in certifying DAGs with a generically finitely identifiable model. For instance, when $m = 6$, it is able to correctly identify 2957 out of 3344 such graphs. The previously known sufficient condition of Stanghellini and Wermuth (2005) identifies 985 of them. Our necessary condition in Theorem 2.1.4 is also useful in assessing graphs that give generically infinite-to-one models. For instance, when $m = 6$, we find that 361 of 552 such graphs violate the condition; recall the example from Figure 2.2.

While, by Proposition 2.2.3, our sufficient condition in Theorem 2.1.3 is stronger than that in Proposition 2.2.2 for generic finite identifiability, the latter condition, due to Stanghellini and Wermuth (2005), in fact implies that $\phi_G^{\mathbb{C}}$ is generically 2-to-one. For $m = 5$, there are 6 DAGs that satisfy the condition in Theorem 2.1.3 but give generically 4-to-one maps $\phi_G^{\mathbb{C}}$. The graph from Figure 2.1 is an example. We note that for this DAG G the fibers of $\phi_G^{\mathbb{C}}$ intersect the statistically relevant set Θ in either 2 or 4 points, and both possibilities do occur.

2.4.2 Simulation study for larger graphs

For DAGs with a large number of nodes m, using exhaustive algebraic computations to determine their identifiability is not feasible. We instead assess the power of our graphical conditions by simulations. For $m = 25$ and $m = 35$, we randomly generate 5000 *labeled* DAGs with k edges, where k ranges from 226 to 275 when $m = 25$ and from 461 to 560 when $m = 35$. Note that for each m, the maximum number of edges in the DAGs we draw equals $\binom{m+1}{2} - 2m$; over that limit the DAGs must give generically non-identifiable models by Corollary 2.1.2. The number of graphs satisfying the conditions in Theorems 2.1.3 and Theorems 2.1.4 are plotted in Figure 2.4. As the number of edges increases, our sufficient condition in Theorems 2.1.3 certifies fewer of the randomly generated graphs to be identifiable. This agrees with the intuition that it is less likely to be a generically finitely identifiable model as the number of edges increases, since there are more free parameters in the coefficient matrix Λ. Conversely, the necessary condition in Theorem 2.1.4 is not very helpful in telling apart generically infinite-to-one models since almost all random graphs satisfy it.

2.5 Subgraph extension

This section concerns results on how we can extend knowledge about identifiability of an induced subgraph to that of the original DAG. We recall standard terminology in graphical modeling. For a given DAG $G = (V, E)$, we write $pa(v) = \{w : w \to v \in E\}$ for the parent set of the node v, and $ch(v) = \{w : v \to w \in E\}$ for the child set of v. If for some node $s \in V$ there does not exist a node $s' \in V$ with $s \to s' \in E$, then s is a sink node. If there

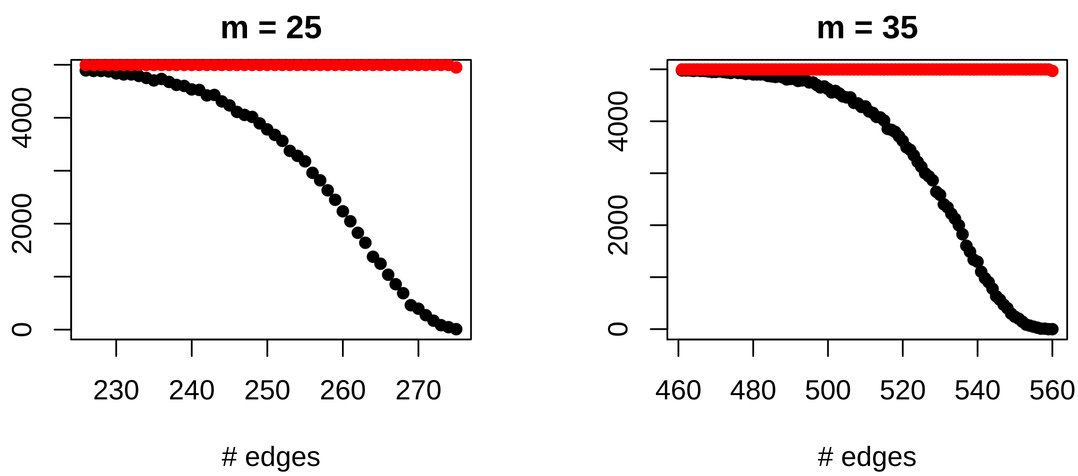

Figure 2.4: For $m = 25$ and 35 and each fixed number of edges, the counts of 5000 randomly drawn labeled DAGs satisfying the sufficient condition in Theorem 2.1.3 are plotted in black, and the counts of labeled DAGs satisfying the necessary condition in Theorem 2.1.4 are plotted in red.

is no other node $s' \in V$ with $s' \to s \in E$, then s is a source node. The following theorem is the main result of this section.

Theorem 2.5.1. *Given a DAG $G = (V, E)$, if there exists*

i. a sink node $s \in V$ such that $pa(s) \neq V \setminus \{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V \setminus \{s\}$ is generically finitely identifiable, or

ii. a source node $s \in V$ such that $ch(s) \neq V \setminus \{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V \setminus \{s\}$ is generically finitely identifiable,

then the model $\mathcal{N}_*(G)$ is generically finitely identifiable.

Recall that in Table 2.1 there are $3344 - 2957 = 387$ DAGs with $m = 6$ nodes that are generically finitely identifiable but do not satisfy our sufficient condition from Theorem 2.1.3. The above Theorem 2.5.1 provides a way to certify identifiability of models falling within this "gap", provided that we have knowledge of which DAGs on $m = 5$ nodes are generically finitely identifiable. For instance, from our algebraic computations we know that there are $95 - 88 = 7$ DAGs that are generically finitely identifiable but cannot be proven to be so by Theorem 2.1.3. Of the 387 aforementioned DAGs on 6 nodes, 194 can be proven to be generically finitely identifiable by using the knowledge about the 7 graphs on $m = 5$ nodes and applying Theorem 2.5.1. We remark that if a DAG satisfies the condition in Theorem 2.1.3, the resulting supergraph obtained by augmenting a sink (source) node that does not have every other node as its parent (child) must also satisfy the condition in Theorem 2.1.3. Hence, given current state-of-the-art, Theorem 2.5.1 is useful primarily as a tool to reduce the identifiability problem to smaller subgraphs that may then be tackled by algebraic methods.

Theorem 2.5.1 is obtained by studying the maps ϕ_G and φ_G in (2.1.3) and (2.3.2). First consider (2.1.3). In light of Lemma 2.1.1(ii), we can show that ϕ_G is generically finite-to-one if there exists a proper algebraic subset $\Xi \subset \mathbb{R}^{2m+|E|}$ such that $|\mathcal{F}_{\phi_G|_{\Theta\setminus\Xi}}(\theta_0)| < \infty$ for all $\theta_0 = (\Lambda_0, \Omega_0, \delta_0) \in \Theta \setminus \Xi$, or equivalently,

$$(I_m - \Lambda^T)\phi_G(\theta_0)(I_m - \Lambda) = \Omega + \delta\delta^T, \tag{2.5.1}$$

has finitely many solutions for $(\Lambda, \Omega, \delta)$ in $\Theta \setminus \Xi$. *Throughout* this section, Ξ is taken so that all points $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ have $\delta_i \neq 0$ for all $i = 1, \ldots, m$. As such, the matrix $\Omega + \delta\delta^T$ on the right hand side of (2.5.1) has all entries nonzero and is known as a Spearman matrix.

Definition 2.5.1. A symmetric matrix $\Upsilon \in \mathbb{R}^{m\times m}$ of size $m \geq 3$ is a *Spearman matrix* if $\Upsilon = \Omega + \delta\delta^T$ for a diagonal matrix Ω with positive diagonal and a vector δ with no zero elements.

Any Spearman matrix Υ is positive definite, and it is not difficult to show that if $\Upsilon = \Omega + \delta\delta^T$ is Spearman with $m \geq 3$ then the two summands Ω and $\delta\delta^T$ are uniquely determined as rational functions of Υ. Moreover, $\delta\delta^T$ determines δ up to sign change. For these facts see, for instance, Theorem 5.5 in Anderson and Rubin (1956). We term Ω the *diagonal component* of Υ, and $\delta\delta'$ the *rank-1 component*. The following theorem gives an implicit characterization of Spearman matrices of size $m \geq 4$.

Theorem 2.5.2. *A positive definite symmetric matrix $\Upsilon = (\upsilon_{ij}) \in \mathbb{R}^{m\times m}$ of size $m \geq 4$ is a Spearman matrix if and only if, after sign changes of rows and corresponding columns, all its elements are positive and such that*

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} \;=\; 0 \tag{2.5.2}$$

for $i < j < k < l$, and

$$\upsilon_{ii}\upsilon_{jk} - \upsilon_{ik}\upsilon_{ji} > 0 \tag{2.5.3}$$

for $i \neq j \neq k$.

This is essentially the same as Theorem 1 in Bekker and de Leeuw (1987), which the reader is referred to for a proof. Unlike Bekker and de Leeuw (1987), we have a strict inequality in (2.5.3) since in Definition 2.5.1 we require the diagonal component of a Spearman matrix to be strictly positive.

The three polynomial expressions in (2.5.2) are the 2×2 off-diagonal minors of the matrix Υ, which are also known as tetrads in the literature. We call the quadruple $i < j < k < l$ the indices of the tetrad they define. Note that

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} = (\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl}) - (\upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl})$$

so that the three tetrads in (2.5.2) are algebraically dependent. In general, a symmetric $m \times m$ matrix Υ has $2\binom{m}{4}$ algebraically independent tetrads and we write TETRADS(Υ) to denote a column vector comprising a choice of $2\binom{m}{4}$ algebraically independent tetrads.

For each triple $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ that solves (2.5.1), it must be true that

$$\text{TETRADS}\left((I_m - \Lambda^T)\phi_G(\theta_0)(I_m - \Lambda)\right) = 0. \tag{2.5.4}$$

Together with the uniqueness of the diagonal and rank-1 components for a Spearman matrix, if we can show only finitely many Λ's solve the system (2.5.4), then we have shown that the model $\mathcal{N}_*(G)$ is generically finitely identifiable. Our proof for Theorem 2.5.1(i) follows this approach.

Alternatively, based on Lemma 2.3.1, we can also prove generic finite identifiability by considering the map φ_G from (2.3.2). We then need to show that there exists a proper algebraic subset $\Xi \subset \mathbb{R}^{2m+|E|}$ so that $|\mathcal{F}_{\varphi_G|_{\Theta\setminus\Xi}}(\theta_0)| < \infty$ for all $\theta_0 = (\Lambda_0, \Psi_0, \gamma_0) \in \Theta \setminus \Xi$, or equivalently,

$$(I_m - \Lambda)^{-1}\varphi_G\,(\theta_0)\,(I_m - \Lambda^T)^{-1} = \Psi - \gamma\gamma^T \tag{2.5.5}$$

has finitely many solutions for (Λ, Ψ, γ) in $\Theta \setminus \Xi$. Again we assume that Ξ is defined to avoid issues due to zeros, that is, every triple $(\Lambda, \Psi, \gamma) \in \Theta \setminus \Xi$ has $\gamma_i \neq 0$ for all $i = 1, \dots, m$. We introduce the term coSpearman matrix to describe the matrix on the right hand side of (2.5.5).

Definition 2.5.2. A symmetric matrix $\Upsilon \in \mathbb{R}^{m\times m}$ of size $m \geq 3$ is a *coSpearman matrix* if $\Upsilon = \Psi - \gamma\gamma^T$ for a diagonal matrix Ψ with positive diagonal and a vector γ with no zero elements.

Again, the diagonal component Ψ and the rank-1 component $\gamma\gamma^T$ are uniquely determined by Υ; compare Stanghellini (1997, p. 243). The following theorem is analogous to Theorem 2.5.2.

Theorem 2.5.3. *A positive definite symmetric matrix $\Upsilon = (\upsilon_{ij}) \in \mathbb{R}^{m\times m}$ of size $m \geq 4$ is a coSpearman matrix if and only if, after sign changes of rows and corresponding columns, all its non-diagonal elements are negative and such that*

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} \;=\; 0 \tag{2.5.6}$$

for $i < j < k < l$, and

$$\upsilon_{ii}\upsilon_{jk} - \upsilon_{ik}\upsilon_{ji} < 0 \tag{2.5.7}$$

for $i \neq j \neq k$.

Using the tetrad characterizations (2.5.6) and the uniqueness of diagonal and rank-1 components, one can now demonstrate that the restricted map $\varphi_G|_{\Theta\setminus\Xi}$ has finite fibers by showing that the system of tetrad equations

$$\text{TETRADS}\big((I_m - \Lambda)^{-1}\varphi_G\,(\theta_0)\,(I_m - \Lambda^T)^{-1}\big) = 0 \tag{2.5.8}$$

admits only finitely many solutions for Λ when $\theta_0 \in \Theta \setminus \Xi$.

The finiteness of solutions in Λ for the system (2.5.4), or (2.5.8), is a sufficient condition for the generic finite identifiability of $\mathcal{N}_*(G)$. It is, however, not obvious that these two systems necessarily have finitely many solutions when $\mathcal{N}_*(G)$ is generically finitely identifiable. The following lemma states that such a converse does hold for the following two types of DAGs, whose generic finite identifiability can be easily checked by Theorem 2.1.3. Recall that the notation " $\subsetneq$ " means "being a proper subset of".

Lemma 2.5.4. *Let $G = (V, E)$ be a DAG with vertex set $V = \{1, \ldots, m\}$.*

i. If $E \subsetneq \{(k, m) : k \leq m - 1\}$, then there exists a proper algebraic subset Ξ such that for all $\theta_0 = (\Lambda_0, \Omega_0, \delta_0) \in \Theta \setminus \Xi$, the system

$$\mathit{TETRADS}\big((I_m - \Lambda^T)\phi_G\,(\theta_0)\,(I_m - \Lambda)\big) = 0$$

is linear in the variable $\Lambda \in \mathbb{R}_E$ and is solved uniquely by $\Lambda = \Lambda_0$.

ii. If $E \subsetneq \{(1, k) : k \geq 2\}$, then there exists a proper algebraic subset Ξ such that for all $\theta_0 = (\Lambda_0, \Psi_0, \gamma_0) \in \Theta \setminus \Xi$, the system

$$\mathit{TETRADS}\big((I_m - \Lambda)^{-1}\varphi_G\,(\theta_0)\,(I_m - \Lambda^T)^{-1}\big) = 0$$

is linear in the variable $\Lambda \in \mathbb{R}_E$ and is solved uniquely by $\Lambda = \Lambda_0$.

The proof of Lemma 2.5.4 is deferred to Appendix B.3.

Proof of Theorem 2.5.1. We will first prove (i), which uses Lemma 2.5.4(i). The proof of (ii) will follow from similar reasoning using Lemma 2.5.4(ii).

Without loss of generality, assume that the sink node $s = m$, by giving the nodes a new topological order if necessary. Define two DAGs as follows. First, let $G_1 = (V_1, E_1)$ be the subgraph of G induced by the set $V_1 = V \setminus \{m\} = [m-1]$, where we adopt the shorthand $[k] := \{1, \dots, k\}$, $k \in \mathbb{N}$. Second, let $G_2 = (V, E \setminus E_1)$ be the graph on V obtained from G by removing all edges that do not have the sink node m as their head. As before, let $\Theta := \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$. We will construct a proper algebraic subset Ξ, such that for any $\theta \in \Theta \backslash \Xi$, the fiber $\mathcal{F}_{\phi_G|_{\Theta\backslash\Xi}}(\theta)$ is finite. Then Lemma 2.1.1(ii) applies and yields the assertion of Theorem 2.5.1(i).

Let $\Theta_1 := \mathbb{R}_{E_1} \times \mathrm{diag}_{m-1}^+ \times \mathbb{R}^{m-1}$ be the open set on which the parametrization ϕ_{G_1} of model $\mathcal{N}_*(G_1)$ is defined. By assumption, there exists a proper algebraic subset $\Xi_1' \subset \mathbb{R}^{2(m-1)+|E_1|}$ such that the restricted map $\phi_{G_1}|_{\Theta_1 \backslash \Xi_1'}$ has finite fibers, by Lemma 2.1.1(ii). Extend Ξ_1' to a proper algebraic subset of $\mathbb{R}^{2m+|E|}$ by defining

$$\Xi_1 := \Xi_1' \times \mathbb{R}^{E \backslash E_1} \times \mathbb{R}^2,$$

where $\mathbb{R}^{E\backslash E_1}$ accommodates the additional free variables λ_{vm} with $v \in pa(m)$, and $\mathbb{R}^2$ accommodates the two variables $\Omega_{mm} = \omega_m$ and δ_m.

Next, recall that for a given point $\theta' = (\Lambda', \Omega', \delta') \in \Theta$, any $(\Lambda, \Omega, \delta) \in \mathcal{F}_{\phi_G}(\theta')$ must satisfy the tetrad equations

$$\text{TETRADS}\left((I_m - \Lambda^T)\phi_G(\theta')(I_m - \Lambda)\right) = 0. \tag{2.5.9}$$

Let $\lambda_{E_1} := (\lambda_{vw})^T_{(v,w)\in E_1}$. Then any tetrad in (2.5.9) with indices $i < j < k < m$ has the form

$$\sum_{m' \in pa(m)} a_{m'}\left(\lambda_{E_1}, \phi_G(\theta')\right)\lambda_{m'm} - b\left(\lambda_{E_1}, \phi_G(\theta')\right),$$

where the $a_{m'}$ as well as b are polynomials with the entries of λ_{E_1} and the entries of a symmetric $m \times m$ matrix being their variables. Let $\lambda_{pa(m),m}$ be the vector with entries λ_{vm} for $v \in pa(m)$. Then the part of the system (2.5.9) involving the variables $\lambda_{v,m}$, $v \in pa(m)$, has the form

$$C\left(\lambda_{E_1}, \phi_G(\theta')\right)\lambda_{pa(m),m} = c\left(\lambda_{E_1}, \phi_G(\theta')\right), \tag{2.5.10}$$

where C is a matrix of size $2\binom{m-1}{3} \times |pa(m)|$, and c is a vector of length $2\binom{m-1}{3}$. Both C and c are filled with polynomials in the entries of λ_{E_1} and a symmetric $m \times m$ matrix. Since

$(\Lambda, \Omega, \delta) \in \mathcal{F}_{\phi_G}(\theta')$, we have $\phi_G(\theta') = \phi_G(\Lambda, \Omega, \delta)$ and, thus,

$$C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right) \lambda_{pa(m),m} = c\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right). \tag{2.5.11}$$

As θ' was an arbitrary point in Θ, (2.5.11) holds for all $(\Lambda, \Omega, \delta) \in \Theta$. We claim that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ is of full rank for generic choices of $(\Lambda, \Omega, \delta)$. To see this, note that if λ_{E_1} is set to 0, then (2.5.11) becomes the system of tetrad equations for the graph G_2. Using Lemma 2.5.4(i) and the assumption that $pa(m) \subsetneq V \setminus \{s\}$, we see that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ achieves full rank for $\lambda_{E_1} = 0$ and a generic choice of $(\lambda_{pa(m),m}, \Omega, \delta)$. We deduce that the rank is full generically.

Let Ξ_2 be a proper algebraic subset such that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ is of full rank for any $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi_2$. Let Ξ_3 be the (algebraic) set comprising all triples $(\Lambda, \Omega, \delta)$ with at least one coordinate $\delta_i = 0$, and define $\Xi := \Xi_1 \cup \Xi_2 \cup \Xi_3$. Clearly, Ξ is a proper algebraic subset of $\mathbb{R}^{2m+|E|}$. Take $(\Lambda_0, \Omega_0, \delta_0)$ to be a point in $\Theta \setminus \Xi$ and define $\Sigma_0 := \phi_G(\Lambda_0, \Omega_0, \delta_0)$. It remains to show that the equation system

$$\Sigma_0 \;=\; \phi_G(\Lambda, \Omega, \delta) = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{2.5.12}$$

has only finitely many solutions in $(\Lambda, \Omega, \delta)$ over the set $\Theta \setminus \Xi$.

We begin by observing that because $s = m$ is a sink node, by taking submatrices in (2.5.12), we obtain the equation system

$$\begin{aligned}
(\Sigma_0)_{[m-1]} &= [(I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}]_{[m-1]} \\
&= (I_{m-1} - \Lambda^T_{[m-1]})^{-1}(\Omega_{[m-1]} + \delta_{[m-1]}\delta^T_{[m-1]})(I_{m-1} - \Lambda_{[m-1]})^{-1} \\
&= \phi_{G_1}\left(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]}\right).
\end{aligned}$$

Here, for an index set $W \subset [m]$, we write x_W to denote the subvector $x_W = (x_v : v \in W)$ of vector $x = (x_1, \ldots, x_m)^T$, and we similarly write A_W for the $W \times W$ principal submatrix of a matrix A. Let $\mathcal{S} \subset \Theta_1$ be the projection of the set of all triples $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ that solve (2.5.12) onto their triple of submatrices/subvector $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$. By choice of Ξ, we have that $\mathcal{S} \subset \Theta_1 \setminus \Xi_1'$ and, since $\phi_{G_1}|_{\Theta_1 \setminus \Xi_1'}$ has finite fibers, we know that $\mathcal{S}$ is finite. However, a triple $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]}) \in \mathcal{S}$ determines the matrix C and the vector c in (2.5.11) and, by choice of Ξ, we may deduce that $\lambda_{pa(m),m}$ is uniquely determined by $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$. It follows that the solutions to (2.5.12) that are in $\Theta \setminus \Xi$ have their Λ part equal to one of $|\mathcal{S}|/2$ many choices; recall that if $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$ is in $\mathcal{S}$ then so is $(\Lambda_{[m-1]}, \Omega_{[m-1]}, -\delta_{[m-1]})$. The proof is now complete because Λ determines the Spearman matrix

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta^T,$$

for which the diagonal component Ω and the rank-1 component $\delta\delta^T$ are uniquely determined. Given the fact that $\delta\delta^T$ determines δ only up to sign, (2.5.12) has $|\mathcal{S}| < \infty$ solutions over $\Theta \setminus \Xi$, which concludes the proof of (i).

The proof of (ii) is analogous, and we only give a sketch. Instead of considering ϕ_G we turn to φ_G, which also has domain Θ. Without loss of generality, we let the source node be $s = 1$. We then define $G_1 = (V_1, E_1)$ to be the subgraph of G that is induced by $V_1 = \{2, \ldots, m\}$, and we let $G_2 = (V, E \setminus E_1)$. We consider the parametrization φ_{G_1} with domain $\Theta_1 = \mathbb{R}_{E_1} \times \mathrm{diag}^+_{m-1} \times \mathbb{R}^{m-1}$. By assumption, $\mathcal{N}_*(G_1)$ is generically finitely identifiable, so there exists a proper algebraic subset Ξ'_1 such that $\varphi_{G_1}|_{\Theta\setminus\Xi'_1}$ has finite fibers, by Lemma 2.1.1(ii).

On the other hand, for any $(\Lambda, \Psi, \gamma) \in \Theta$, we have

$$\mathrm{TETRADS}\left((I_m - \Lambda)^{-1}\varphi_G(\Lambda, \Psi, \gamma)(I_m - \Lambda^T)^{-1}\right) = 0.$$

Let $\lambda_{E_1} := (\lambda_{vw})^T_{(v,w)\in E_1}$ and $\lambda_{1,ch(1)} := (\lambda_{1v})^T_{v\in ch(1)}$. Then the tetrad equations with one index equal to $s = 1$ yield the equation system

$$C\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right)\lambda_{1,ch(1)} = c\left(\lambda_{E_1}, \varphi_G\left(\Lambda, \Psi, \gamma\right)\right),$$

where part (ii) of Lemma 2.5.4 can be applied to show that $C\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right)$ is of full rank outside some proper algebraic subset Ξ_2. We may then define a set Ξ as in the proof of part (i) and use arguments similar to the ones above for a proof of part (ii) of our theorem. $\square$

2.6 Discussion

In this chapter we studied identifiability of directed Gaussian graphical models with one latent variable that is a common cause of all observed variables. To our knowledge, previously the best criteria to decide on identifiability of such models are those given by Stanghellini and Wermuth (2005) who consider a more general setup of Gaussian graphical models with one latent variable. Their results provide a sufficient condition for the strictest notion of identifiability that is meaningful is this context, namely, whether the parametrization map is generically 2-to-one. Recall that the coefficients associated with the edges pointing from the latent variable to the observables can only be recovered up to a common sign change.

In our work, we take a different approach and study the Jacobian matrix of the parametrization, which leads to graphical criteria to check whether the parametrization is finite-to-one. Our sufficient condition covers all graphs that can be shown to have a 2-to-one parametrization by the conditions of Stanghellini and Wermuth (2005). However, our sufficient condition, which is stated as Theorem 2.1.3, covers far more graphs as was shown in the computational experiments in Section 2.4. Our Theorem 2.1.4 describes a complementary necessary condition.

By studying tetrad equations, we also give a criterion that allows one to deduce identifiability of certain graphs from identifiability of subgraphs (Theorem 2.5.1). This result is stated for generic finite identifiability but as is clear from the proof, the result would also confirm that the parametrization of a graph is generically 2-to-one provided the involved subgraph has a generically 2-to-one parametrization.

The extension result from Theorem 2.5.1 can be used in conjunction with the results obtained by the algebraic computations in Section 2.4. These computations solve the identifiability problem for graphs with up to 6 nodes. In particular, we confirm that the sufficient conditions of Stanghellini and Wermuth (2005) are not necessary for the parametrization map to be generically 2-to-one. We also provide examples of graphs that yield a generically finite but not 2-to-one parametrization.

As mentioned above, we studied models with one latent source 0 that is connected to all nodes that represent observed variables. However, the graphical criteria in Theorems 2.1.3 and 2.1.4 can be readily extended to models with some of these factor loading edges missing. Given the previously used notation, we describe such models as follows. Let $G = (V, E)$ be a DAG with vertex set of size $m = |V|$; these vertices index the observed variables. Let $V' \subset V$ be the nodes representing observed variables that do not directly depend on the latent variable. Then only the edges $0 \to v$ with $v \in V \setminus V'$ are added when forming the extended DAG $\overline{G}$. The parametrization of the Gaussian graphical model determined by G and V' is the restriction of ϕ_G from (2.1.3) to the domain

$$\Theta(V') := \{(\Lambda, \Omega, \delta) \in \Theta : \delta_v = 0 \text{ for all } v \in V'\}\,.$$

When the parametrization maps $\tilde{\phi}_G$, φ_G and $\tilde{\varphi}_G$ are restricted to the same domain, the assertion of Lemma 2.3.1 still holds. The corresponding identifiability results, which are in the spirit of Corollary 1 in Grzebyk et al. (2004), are stated below. A brief outline of their proofs is given in Appendix B.4.

Theorem 2.6.1 (Sufficient condition)**.** *Let $G = (V, E)$ be a DAG, and let $V' \subset V$. If every connected component of $(G^c)_{V\setminus V'}$, the subgraph of G^c induced by $V \setminus V'$, contains an odd cycle, then the parametrization map ϕ_G is generically finite-to-one when restricted to the domain $\Theta(V')$.*

The necessary condition given next makes references to the graphs G_{con} and $G_{|L,cov}$ that were defined in the introduction.

Theorem 2.6.2 (Necessary condition)**.** *Let $G = (V, E)$ be a DAG, and let $V' \subset V$. In order for the restriction of ϕ_G to the domain $\Theta(V')$ to be generically finite-to-one, it is necessary that the following two conditions both hold:*

i. Let $\widetilde{G}^c_{con} = (V \setminus V', \widetilde{E}_{con})$ be the subgraph of G^c_{con} induced by $V \setminus V'$. If d_{con} is the number of connected components in the graph $\widetilde{G}^c_{con}$ that do not contain any odd cycle, then $|\widetilde{E}_{con}| - |E| \geq d_{con}$.

ii. Let $\widetilde{G}^c_{|L,cov} = (V \setminus V', \widetilde{E}_{|L,cov})$ *be the subgraph of* G^c_{con} *induced by* $V \setminus V'$*. If* d_{cov} *is the number of connected components in the graph* $\widetilde{G}^c_{|L,cov}$ *that do not contain any odd cycle, then* $|\widetilde{E}_{|L,cov}| - |E| \geq d_{cov}$*.*

While Theorems 2.6.1 and 2.6.2 may be useful in some contexts, models in which latent variables are parents to only some of the observables deserve a more in-depth treatment in future work. In particular, it would be natural to seek ways to combine the results of Stanghellini and Wermuth (2005) and the present chapter with the work of Foygel et al. (2012) and Drton and Weihs (2015).

BIBLIOGRAPHY

T. W. Anderson. *An introduction to multivariate statistical analysis.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

T. W. Anderson and Herman Rubin. Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. V*, pages 111–150. University of California Press, Berkeley and Los Angeles, 1956.

Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics.* Springer-Verlag, Berlin, second edition, 2006.

Paul A. Bekker and Jan de Leeuw. The rank of reduced dispersion matrices. *Psychometrika*, 52(1):125–135, 1987.

Wicher Bergsma and Angelos Dassios. A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20(2):1006–1028, 2014.

Kenneth A. Bollen. *Structural equations with latent variables.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1989. A Wiley-Interscience Publication.

Yu. V. Borovskikh. *U-statistics in Banach spaces.* VSP, Utrecht, 1996.

T. Tony Cai and Tiefeng Jiang. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.*, 39(3):1496–1525, 2011.

T. Tony Cai and Zongming Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388, 2013.

Bryant Chen, Jin Tian, and Judea Pearl. Testable implications of linear structural equations models. In Carla E. Brodley and Peter Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 2424–2430. AAAI Press, 2014.

Song Xi Chen, Li-Xin Zhang, and Ping-Shou Zhong. Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 105(490):810–819, 2010.

David Christensen. Fast algorithms for the calculation of Kendall's τ. *Comput. Statist.*, 20 (1):51–62, 2005.

David Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, New York, third edition, 2007. An introduction to computational algebraic geometry and commutative algebra.

Jesús A. de Loera, Bernd Sturmfels, and Rekha R. Thomas. Gröbner bases and triangulations of the second hypersimplex. *Combinatorica*, 15(3):409–424, 1995.

M. Drton. Algebraic techniques for Gaussian models. In M. Hušková and M. Janžura, editors, *Prague Stochastics*, pages 81–90, Charles University Prague, 2006. Matfyzpress.

Mathias Drton and Luca Weihs. Generic identifiability of linear structural equation models by ancestor decomposition. *ArXiv e-prints*, 2015. 1504.02992.

Mathias Drton, Bernd Sturmfels, and Seth Sullivant. Algebraic factor analysis: tetrads, pentads and beyond. *Probab. Theory Related Fields*, 138(3-4):463–493, 2007.

Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel, 2009.

Mathias Drton, Rina Foygel, and Seth Sullivant. Global identifiability of linear structural equation models. *Ann. Statist.*, 39(2):865–886, 2011.

Rina Foygel, Jan Draisma, and Mathias Drton. Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.*, 40(3):1682–1713, 2012.

Luis D. Garcia-Puente, Sarah Spielvogel, and Seth Sullivant. Identifying causal effects with computer algebra. In Peter Grünwald and Peter Spirtes, editors, *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press, 2010.

Dan Geiger, David Heckerman, Henry King, and Christopher Meek. Stratified exponential families: graphical models and model selection. *Ann. Statist.*, 29(2):505–529, 2001.

Michel Grzebyk, Pascal Wild, and Dominique Chouanière. On identification of multi-factor models with correlated residuals. *Biometrika*, 91(1):141–151, 2004.

Jaroslav Hájek, Zbyněk Šidák, and Pranab K. Sen. *Theory of rank tests*. Probability and Mathematical Statistics. Academic Press, Inc., San Diego, CA, second edition, 1999.

P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

Fang Han and Han Liu. Distribution-free tests of independence with applications to testing more structures. arXiv:1410.4179, October 2014.

Wassily Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, pages 546–557, 1948a.

Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948b.

Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

Tiefeng Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.*, 14(2):865–880, 2004.

S. John. The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, 59:169–173, 1972.

William H. Kruskal. Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53:814–861, 1958.

Manabu Kuroki and Masami Miyakawa. Graphical identifiability criteria for total effects in studies with an unobserved response variable. *Behaviormetrika*, 31(1):13–28, 2004.

Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014.

Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.

Olivier Ledoit and Michael Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102, 2002.

E. L. Lehmann and George Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.

Deli Li, Wei-Dong Liu, and Andrew Rosalsky. Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields*, 148(1-2):5–35, 2010.

Deli Li, Yongcheng Qi, and Andrew Rosalsky. On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.*, 111:256–270, 2012.

Filip Lindskog, Alexander Mcneil, and Uwe Schmock. *Kendalls tau for elliptical distributions*. Springer, 2003.

Wei-Dong Liu, Zhengyan Lin, and Qi-Man Shao. The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.*, 18(6):2337–2366, 2008.

Guangyu Mao. A new test of independence for high-dimensional data. *Statist. Probab. Lett.*, 93:14–18, 2014.

Hisao Nagao. On some test criteria for covariance matrix. *Ann. Statist.*, 1:700–709, 1973.

Hannu Oja. *Multivariate nonparametric methods with R*, volume 199 of *Lecture Notes in Statistics*. Springer, New York, 2010. An approach based on spatial signs and ranks.

Judea Pearl. *Causality*. Cambridge University Press, Cambridge, second edition, 2009. Models, reasoning, and inference.

C. Radhakrishna Rao. *Linear statistical inference and its applications*. John Wiley & Sons, New York-London-Sydney, second edition, 1973. Wiley Series in Probability and Mathematical Statistics.

Edward M. Reingold, Jurg Nievergelt, and Narsingh Deo. *Combinatorial algorithms: theory and practice*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1977.

Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *Ann. Statist.*, 30 (4):962–1030, 2002.

Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition, 1976. International Series in Pure and Applied Mathematics.

James R. Schott. Testing for complete independence in high dimensions. *Biometrika*, 92(4): 951–956, 2005.

Robert J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York, 1980. Wiley Series in Probability and Mathematical Statistics.

Elena Stanghellini. Identification of a single-factor model using graphical gaussian rules. *Biometrika*, 84(1):241–244, 1997.

Elena Stanghellini and Nanny Wermuth. On the identification of path analysis models with one hidden variable. *Biometrika*, 92(2):337–350, 2005.

Jin Tian. Identifying direct causal effects in linear models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 346–352. AAAI Press/The MIT Press, 2005.

Jin Tian. Parameter identication in a class of linear structural equation models. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1970–1975. AAAI Press, 2009.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

Paola Vicard. On the identification of a single-factor model with correlated residuals. *Biometrika*, 87(1):199–205, 2000.

Guanghui Wang, Changliang Zou, and Zhaojun Wang. A necessary test for complete independence in high dimensions using rank-correlations. *J. Multivariate Anal.*, 121:224–232, 2013.

Luca Weihs, Mathias Drton, and Dennis Leung. Efficient computation of the bergsma–dassios sign covariance. *Computational Statistics*, pages 1–14, 2015.

Wang Zhou. Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.*, 359(11):5345–5363, 2007.

Appendix A
PROOFS FOR CHAPTER 1

A.1 Motivation of Schott's statistic as a Rao score

We show that, up to a rescaling by the squared sample size, the statistic S_r (1.1.3) is Rao's score statistic in the multivariate normal setting. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. m-variate normal random vectors with mean vector $\boldsymbol{\mu}$ and *precision* matrix $K = (K^{(pq)})$, and define $\bar{\mathbf{X}} := n^{-1} \sum_i \mathbf{X}_i$, $W = (W^{(pq)}) := n^{-1} \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$. Recall that the score vector is the first derivative of their joint likelihood

$$l_n = \frac{1}{2}\left(n \log |K| - \sum_i (\mathbf{X}_i - \boldsymbol{\mu})' K (\mathbf{X}_i - \boldsymbol{\mu})\right)$$

with respect to $\boldsymbol{\theta} = (\boldsymbol{\mu}, K)$ evaluated at the maximum likelihood estimate $(\hat{\boldsymbol{\mu}}_{H_0}, \hat{K}_{H_0}) = (\bar{X}, \mathrm{diag}((W^{(11)})^{-1}, \ldots, (W^{(mm)})^{-1}))$ of $\boldsymbol{\theta}$ under H_0 (1.1.1). By routine calculations, it can be shown that

$$\left.\frac{\partial l_n}{\partial \mu}\right|_{\boldsymbol{\theta}=(\hat{\boldsymbol{\mu}}_{H_0}, \hat{K}_{H_0})} = 0, \quad \left.\frac{\partial l_n}{\partial K^{(pq)}}\right|_{\boldsymbol{\theta}=(\hat{\boldsymbol{\mu}}_{H_0}, \hat{K}_{H_0})} = \begin{cases} 0 & \text{if} \quad p = q \\ -nW^{(pq)} & \text{if} \quad p < q \end{cases}. \tag{A.1.1}$$

Let $I(\boldsymbol{\theta})$ be the Fisher Information of $\mathbf{X}$, and by the classical theory of Rao score test, one reject H_0 when

$$\left.\left(\frac{\partial l_n}{\partial \boldsymbol{\theta}}\right)' I(\boldsymbol{\theta})^{-1} \left(\frac{\partial l_n}{\partial \boldsymbol{\theta}}\right)\right|_{\boldsymbol{\theta}=(\hat{\boldsymbol{\mu}}_{H_0}, \hat{K}_{H_0})} \tag{A.1.2}$$

is large. For $p < q$ and $p' < q'$,

$$
\begin{aligned}
&[I(\hat{\boldsymbol{\mu}}_{H_0}, \hat{K}_{H_0})]_{K^{(pq)},K^{(p'q')}} \\
&= \mathbb{E}_{\mathcal{N}(\hat{\boldsymbol{\mu}}_{H_0},\hat{K}_{H_0})}[(X^{(p)} - \mu^{(p)})(X^{(q)} - \mu^{(q)})(X^{(p')} - \mu^{(p')})(X^{(q')} - \mu^{(q')})] \\
&= \begin{cases} ([\hat{K}_{H_0}]_{pp}[\hat{K}_{H_0}]_{qq})^{-1} & \text{if} \quad (p,q) = (p',q') \\ 0 & \text{if} \quad (p,q) \neq (p',q') \end{cases},
\end{aligned} \tag{A.1.3}
$$

where $\mathbb{E}_{\mathcal{N}(\hat{\boldsymbol{\mu}}_{H_0},\hat{K}_{H_0})}$ means taking expectation with respect to $\mathbf{X}$ having a multivariate normal distribution with mean $\hat{\boldsymbol{\mu}}_{H_0}$ and precision matrix $\hat{K}_{H_0}$. In light of (A.1.1) and (A.1.3), one obtains that the statistic from (A.1.2) is equal to n^2 times Schott (2005)'s statistic S_r from (1.1.3)

A.2 Technical lemmas

The following lemma will be used to prove both Lemmas A.2.2 and A.2.3 below, as well as Lemmas 1.3.1 and 1.3.2. We make use of the following notion of multisets. For $1 \leq k \leq n$, if $\mathbf{i}^1, \ldots, \mathbf{i}^r$ are tuples in $\mathcal{P}(n,k)$, let the duple $(\cup_{\omega=1}^r \mathbf{i}^\omega, f_m)$ be the multiset associated with $\cup_{\omega=1}^r \mathbf{i}^\omega$, where $f_m : \cup_{\omega=1}^r \mathbf{i}^\omega \longrightarrow \mathbb{N}$ is the multiplicity function such that $f_m(i)$ is the number of occurrences of index i in the sets $\mathbf{i}^1, \ldots, \mathbf{i}^r$.

Lemma A.2.1. *Let $h : (\mathbb{N}^2)^k \longrightarrow \mathbb{R}$ be a kernel that is symmetric in its k arguments and has order of degeneracy d under H_0.*

i. Suppose $\mathbf{i}^1, \ldots, \mathbf{i}^4 \in \mathcal{P}(n,k)$. If $|\cup_{\omega=1}^4 \mathbf{i}^\omega| > 4k - 2d$, then

$$\mathbb{E}_0 \left[\prod_{\omega=1}^4 h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)}\right) \right] = 0$$

for all $1 \leq p^\omega \neq q^\omega \leq m$, $\omega = 1, \ldots, 4$. If $|\cup_{\omega=1}^4 \mathbf{i}^\omega| = 4k - 2d$, then $\mathbb{E}_0[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)})]$ is nonzero only if $|\mathbf{i}^\omega \cap (\cup_{\omega' \neq \omega} \mathbf{i}^{\omega'})| = d$ for all $\omega = 1, \ldots, 4$, and in this case the multiplicity function f_m of the multiset $(\cup_{\omega=1}^4 \mathbf{i}^\omega, f_m)$ takes value either 1 or 2.

ii. Suppose $\mathbf{i}^1, \ldots, \mathbf{i}^8 \in \mathcal{P}(n, k)$. *If* $|\cup_{\omega=1}^8 \mathbf{i}^\omega| > 8k - 4d$, *then*

$$\mathbb{E}_0\left[\prod_{\omega=1}^{8} h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)}\right)\right] = 0$$

for all $1 \leq p^\omega \neq q^\omega \leq m$, $\omega = 1, \ldots, 8$. *If* $|\cup_{\omega=1}^8 \mathbf{i}^\omega| = 8k - 4d$, *then* $\mathbb{E}_0[\prod_{\omega=1}^8 h(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)})]$ *is nonzero only if* $|\mathbf{i}^\omega \cap (\cup_{\omega' \neq \omega} \mathbf{i}^{\omega'})| = d$ *for all* $\omega = 1, \ldots, 8$, *and in this case the multiplicity function* f_m *of the multiset* $(\cup_{\omega=1}^8 \mathbf{i}^\omega, f_m)$ *takes value either* 1 *or* 2.

Proof. We consider the first claim (i). Since $\mathbf{i}^1, \ldots, \mathbf{i}^4$ are tuples in $\mathcal{P}(n, k)$, the multiplicity function f_m of the multiset $(\cup_{\omega=1}^4 \mathbf{i}^\omega, f_m)$ is such that $\sum_{i \in \cup_{\omega=1}^4 \mathbf{i}^\omega} f_m(i) = 4k$. If $|\cup_{w=1}^4 \mathbf{i}^\omega| > 4k - 2d$, the cardinality of the set $\{i \in \cup_{w=1}^8 \mathbf{i}^\omega : f_m(i) = 1\}$ must be greater than $4k - 4d$, in which case there exists an ω' so that $c := |\mathbf{i}^{\omega'} \cap (\cup_{\omega \neq \omega'} \mathbf{i}^\omega)| < d$. By symmetry, we may assume $w' = 1$ without loss of generality.

Let $\mathbf{j} = (j_1, \ldots, j_c) = \mathbf{i}^1 \cap (\cup_{w \neq 1} \mathbf{i}^\omega)$ as sets. Then, conditional on $\mathbf{X}_{\mathbf{j}}^{(p^1 q^1)}$, we have that $h(\mathbf{R}_{\mathbf{i}^1}^{(p^1 q^1)})$ is independent of all other factors $h(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)})$ for $\omega = 2, \ldots, 4$. Since h has order of degeneracy d under H_0, by the equivalence relation in (1.2.8), $\mathbb{E}_0[h(\mathbf{R}_{\mathbf{i}^1}^{(p^1 q^1)}) | \mathbf{X}_{\mathbf{j}}^{(p^1 q^1)}] = 0$, and therefore by the aforementioned conditional independence

$$\mathbb{E}_0\left[\prod_{\omega=1}^{4} h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)}\right) \middle| \mathbf{X}_{\mathbf{j}}^{(p^1 q^1)}\right] \equiv 0$$

as a function of $\mathbf{X}_{\mathbf{j}}^{(p^1 q^1)}$, which in turn implies that $\mathbb{E}_0[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)})] = 0$.

The necessary condition for $\mathbb{E}_0[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(p^\omega q^\omega)})]$ to be nonzero when $|\cup_{\omega=1}^4 \mathbf{i}^\omega| = 4k - 2d$ can be argued similarly, and we omit the details.

The proof of (ii) is analogous to that of (i). Again, we omit the details. $\square$

The following three lemmas will be used to prove Lemma A.4.1. Recall the notational shorthand $\bar{U}_h^{(pq)} := \left(U_h^{(pq)}\right)^2 - \mu_h$ for $1 \leq p < q \leq m$, defined in the proof of Theorem 1.4.1.

Lemma A.2.2. *Suppose* $1 \leq p, q, l, u \leq m$ *are four distinct indices, and* h *is a kernel of order of degeneracy* d *satisfying Assumption 1.2.2 under* H_0. *Then*

$$\mathbb{E}_0\left[\bar{U}_h^{(pl)} \bar{U}_h^{(ql)} \bar{U}_h^{(pu)} \bar{U}_h^{(qu)}\right] = O(n^{-4d-1}).$$

Proof. Without loss of generality, we prove the result for $(p,q,l,u)=(1,2,3,4)$. Note that for any four distinct indices $1\le p_1,p_2,p_3,p_4\le m$, the antiranks $\mathbf{R}^{(p_1)|(p_2)}$, $\mathbf{R}^{(p_2)|(p_3)}$, $\mathbf{R}^{(p_3)|(p_4)}$ are independent. Since $\bar U^{(13)}$, $\bar U^{(23)}$, $\bar U^{(14)}$, $\bar U^{(24)}$ are functions of $\mathbf{R}^{(1)|(3)}$, $\mathbf{R}^{(2)|(3)}$, $\mathbf{R}^{(1)|(4)}$, $\mathbf{R}^{(2)|(4)}$, respectively, on expansion,

$$\begin{aligned}\mathbb{E}_0\left[\bar U^{(13)}\bar U^{(23)}\bar U^{(14)}\bar U^{(24)}\right] &= \mathbb{E}_0\left[\left(U_h^{(13)}\right)^2\left(U_h^{(23)}\right)^2\left(U_h^{(14)}\right)^2\left(U_h^{(24)}\right)^2\right]-\mu_h^4\\ &= \mathbb{E}_0\left[\left(U_h^{(13)}\right)^2\left(U_h^{(23)}\right)^2\left(U_h^{(14)}\right)^2\left(U_h^{(24)}\right)^2\right]-\binom{k}{d}^8\left(\frac{d!\zeta_d}{n^d}\right)^4+O\left(n^{-4d-1}\right),\end{aligned}$$

where the last equality follows from Lemma 1.3.1(i). The proof is completed if we are able to show that

$$\mathbb{E}_0\left[\left(U_h^{(13)}\right)^2\left(U_h^{(23)}\right)^2\left(U_h^{(14)}\right)^2\left(U_h^{(24)}\right)^2\right]=\binom{k}{d}^8\left(\frac{d!\zeta_d}{n^d}\right)^4+O\left(n^{-4d-1}\right). \tag{A.2.1}$$

For $\mathbf{i}^\omega\in\mathcal{P}(n,k)$, $\omega=1,\dots,8$, we define

$$P(\mathbf{i}^1,\dots,\mathbf{i}^8) \;=\; \left(\prod_{\omega=1}^{2}h\left(\mathbf{R}_{\mathbf{i}^w}^{(13)}\right)\right)\left(\prod_{\omega=3}^{4}h\left(\mathbf{R}_{\mathbf{i}^w}^{(23)}\right)\right)\left(\prod_{\omega=5}^{6}h\left(\mathbf{R}_{\mathbf{i}^w}^{(14)}\right)\right)\left(\prod_{\omega=7}^{8}h\left(\mathbf{R}_{\mathbf{i}^w}^{(24)}\right)\right). \tag{A.2.2}$$

Then on expansion,

$$\mathbb{E}_0\left[\left(U_h^{(13)}\right)^2\left(U_h^{(23)}\right)^2\left(U_h^{(14)}\right)^2\left(U_h^{(24)}\right)^2\right]=\binom{n}{k}^{-8}\sum_{\substack{\mathbf{i}^\omega\in\mathcal{P}(n,k)\\1\le\omega\le 8}}\mathbb{E}_0\left[P(\mathbf{i}^1,\dots,\mathbf{i}^8)\right]. \tag{A.2.3}$$

Each summand $\mathbb{E}_0[P(\mathbf{i}^1,\dots,\mathbf{i}^8)]$ on the right hand side of (A.2.3) depends on the multiset $(\cup_{w=1}^8\mathbf{i}^w, f_m)$. If $|\cup_{w=1}^8\mathbf{i}^\omega|>8k-4d$, by Lemma A.2.1$(ii)$, $\mathbb{E}_0[P(\mathbf{i}^1,\dots,\mathbf{i}^8)]=0$.

If $|\cup_{w=1}^8\mathbf{i}^\omega|=8k-4d$, by Lemma A.2.1$(ii)$, for $\mathbb{E}_0[P(\mathbf{i}^1,\dots,\mathbf{i}^8)]$ to be non-zero it is necessary that $|\mathbf{i}^{\omega'}\cap(\cup_{\omega\neq\omega'}\mathbf{i}^\omega)|=d$ for all $\omega'=1,\dots,8$, in which case f_m takes the value 1 or 2. Suppose this is true. Under H_0, conditioning on $\mathbf{X}^{(1)}_{\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^5\cup\mathbf{i}^6)}$ and $\mathbf{X}^{(3)}_{\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^3\cup\mathbf{i}^4)}$, $h(\mathbf{R}^{(1,3)}_{\mathbf{i}^1})$ is independent of all other multiplicative factors on the right hand side of (A.2.2). If $\mathbf{i}^1$ intersects with the set $\cup_{\omega=3}^8\mathbf{i}^\omega\setminus\mathbf{i}^2$, at least one of $\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^5\cup\mathbf{i}^6)$ and $\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^3\cup\mathbf{i}^4)$ has cardinality less than d given that $f_m\le 2$, and by Assumption 1.2.2

$$\mathbb{E}_0\left[h\left(\mathbf{R}_{\mathbf{i}^1}^{(13)}\right)\middle|\mathbf{X}^{(1)}_{\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^5\cup\mathbf{i}^6)},\mathbf{X}^{(2)}_{\mathbf{i}^1\cap(\mathbf{i}^2\cup\mathbf{i}^3\cup\mathbf{i}^4)}\right]=0,$$

Hence, $\mathbb{E}_0[P(\mathbf{i}^1,\ldots,\mathbf{i}^8)] = 0$ by the aforementioned conditional independence. Similarly, $\mathbf{i}^3,\mathbf{i}^5,\mathbf{i}^7$ can only intersect with $\mathbf{i}^4,\mathbf{i}^6,\mathbf{i}^8$, respectively, to ensure that $\mathbb{E}_0[P(\mathbf{i}^1,\ldots,\mathbf{i}^8)]$ does not equal zero. When this is the case, $|\mathbf{i}^\omega \cap \mathbf{i}^{w+1}| = d$ for $w = 1,3,5,7$, then the four sets $\mathbf{i}^1\cap\mathbf{i}^2, \mathbf{i}^3\cap\mathbf{i}^4, \mathbf{i}^5\cap\mathbf{i}^6, \mathbf{i}^7\cap\mathbf{i}^8$ are disjoint and $\mathbb{E}_0[P(\mathbf{i}^1,\ldots,\mathbf{i}^8)] = (\zeta_d^h)^4$.

As a result, when $|\cup_{w=1}^8 \mathbf{i}^\omega| = 8k-4d$, $\mathbb{E}_0[P(\mathbf{i}^1,\ldots,\mathbf{i}^8)]$ is only nonzero with value $(\zeta_d^h)^4$ for

$$\binom{n}{8k-4d}\binom{8k-4d}{2k-d,2k-d,2k-d,2k-d}\binom{2k-d}{d}^4\binom{2k-2d}{k-d}^4 = \frac{n!}{(n-8k+4d)!((k-d)!)^8(d!)^4}$$

choices of $(\mathbf{i}^1,\ldots,\mathbf{i}^8)$, which can be seen as follows. First, pick $8k-4d$ indices from the set $\{1,\ldots,n\}$, and note that there are $\binom{8k-4d}{2k-d,2k-d,2k-d,2k-d}$ ways of partitioning the $8k-4d$ indices into the four sets $\mathbf{i}^1\cap\mathbf{i}^2, \mathbf{i}^3\cap\mathbf{i}^4, \mathbf{i}^5\cap\mathbf{i}^6, \mathbf{i}^7\cap\mathbf{i}^8$. For each $w \in 1,3,5,7$, there are $\binom{2k-d}{d}$ choices for the d shared common index in $\mathbf{i}^w\cap\mathbf{i}^{w+1}$, and there are $\binom{2k-2d}{k-d}$ ways of distributing the remaining $2k-2d$ indices to $\mathbf{i}^\omega$ and $\mathbf{i}^{w+1}$. Since the count of the summands $\mathbb{E}_0[P(\mathbf{i}^1,\ldots,\mathbf{i}^8)]$ with $|\cup_{w=1}^8 \mathbf{i}^\omega| < 8k-4d$ is of the order $O(n^{8k-4d-1})$, we find from (A.2.3) that

$$\begin{aligned}
&\mathbb{E}_0\left[\left(U_h^{(13)}\right)^2\left(U_h^{(23)}\right)^2\left(U_h^{(14)}\right)^2\left(U_h^{(24)}\right)^2\right] \\
&= \binom{n}{k}^{-8}\left(\frac{(\zeta_d^h)^4 n!}{(n-8k+4d)!((k-d)!)^8(d!)^4} + O\left(n^{8k-4d-1}\right)\right) \\
&= \binom{k}{d}^8 \frac{(d!\zeta_d^h)^4}{n^{4d}} + O\left(n^{-4d-1}\right),
\end{aligned}$$

and we are done proving (A.2.1). □

Lemma A.2.3. *Suppose $1 \le p,q,l,u \le m$ are four distinct indices, and h is a kernel of order of degeneracy d satisfying Assumption 1.2.2 under H_0. Then*

$$\mathbb{E}_0\left[W_h^{(pl)}W_h^{(ql)}W_h^{(pu)}W_h^{(qu)}\right] = O(n^{-4d-1}).$$

Proof. Again, without loss of generality, we prove the result for $(p,q,l,u)=(1,2,3,4)$. Given $\mathbf{i}^\omega \in \mathcal{P}(n,2k)$, $\omega=1,\ldots,4$, we define

$$
\begin{aligned}
Q(\mathbf{i}^1,\ldots,\mathbf{i}^4) &= h^W\left(\mathbf{R}^{(13)}_{\mathbf{i}^1}\right) h^W\left(\mathbf{R}^{(23)}_{\mathbf{i}^2}\right) h^W\left(\mathbf{R}^{(14)}_{\mathbf{i}^3}\right) h^W\left(\mathbf{R}^{(24)}_{\mathbf{i}^4}\right) \\
&= \binom{2k}{k}^{-4} \sum_{\substack{\tilde{\mathbf{i}}^\omega \subset \mathbf{i}^\omega \\ |\tilde{\mathbf{i}}^\omega|=k}} h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4},
\end{aligned}
\tag{A.2.4}
$$

where $h^{(pq)}_{\mathbf{i}^\omega,\tilde{\mathbf{i}}^\omega} := h\left(\mathbf{R}^{(pq)}_{\tilde{\mathbf{i}}^\omega}\right) h\left(\mathbf{R}^{(pq)}_{\mathbf{i}^\omega\setminus\tilde{\mathbf{i}}^\omega}\right)$. By the definition from (1.3.5), on expansion,

$$
\begin{aligned}
\mathbb{E}_0\left[W_h^{(13)}W_h^{(23)}W_h^{(14)}W_h^{(24)}\right] &= \frac{1}{\binom{n}{2k}^4} \sum_{\substack{\mathbf{i}^w \in \mathcal{P}(n,2k), \\ 1\le w\le 4}} \mathbb{E}_0\left[Q\left(\mathbf{i}^1,\mathbf{i}^2,\mathbf{i}^3,\mathbf{i}^4\right)\right] \\
&= \frac{1}{\left(\binom{n}{2k}\binom{2k}{k}\right)^4} \sum_{\substack{\mathbf{i}^w \in \mathcal{P}(n,2k), \\ \omega=1,\ldots,4}} \sum_{\substack{\tilde{\mathbf{i}}^\omega \subset \mathbf{i}^\omega \\ |\tilde{\mathbf{i}}^\omega|=k}} \mathbb{E}_0\left[h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4}\right].
\end{aligned}
\tag{A.2.5}
$$

It now suffices to show that

$$
\mathbb{E}_0\left[h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4}\right] = 0 \tag{A.2.6}
$$

whenever $|\cup_{\omega=1}^4 \mathbf{i}^\omega| \geq 8k-4d$, because then the right hand side of (A.2.5) is of the order $\binom{n}{2k}^{-4}\binom{n}{8k-4d-1} = O(n^{-4d-1})$.

The value of a term

$$
h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4} = h\left(\mathbf{R}^{(13)}_{\tilde{\mathbf{i}}^1}\right) h\left(\mathbf{R}^{(13)}_{\mathbf{i}^1\setminus\tilde{\mathbf{i}}^1}\right) \cdots\cdots h\left(\mathbf{R}^{(24)}_{\tilde{\mathbf{i}}^4}\right) h\left(\mathbf{R}^{(24)}_{\mathbf{i}^4\setminus\tilde{\mathbf{i}}^4}\right), \tag{A.2.7}
$$

depends on the multi set $(\cup_{\omega=1}^4\mathbf{i}^\omega, f_m)$, where $f_m : \cup_{\omega=1}^4\mathbf{i}^\omega \longrightarrow \mathbb{N}$ is the multiplicity function with $f(i)$ equal to the occurrences of i among the eight tuples

$$
\tilde{\mathbf{i}}^1, \mathbf{i}^1\setminus\tilde{\mathbf{i}}^1, \ldots, \tilde{\mathbf{i}}^4, \mathbf{i}^4\setminus\tilde{\mathbf{i}}^4 \in \mathcal{P}(n,k) \tag{A.2.8}
$$

and $\sum_{i\in\cup_{\omega=1}^4\mathbf{i}^\omega} f(i) = 8k$. If $|\cup_{\omega=1}^4\mathbf{i}^\omega| = |\cup_{\omega=1}^4(\tilde{\mathbf{i}}^\omega)\cup(\mathbf{i}^\omega\setminus\tilde{\mathbf{i}}^\omega)| > 8k-4d$, by Lemma A.2.1$(ii)$, $\mathbb{E}_0[h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4}] = 0$. We are left with the case $|\cup_{\omega=1}^4\mathbf{i}^\omega| = 8k-4d$.

If $|\cup_{\omega=1}^4 \mathbf{i}^\omega| = 8k-4d$, by Lemma A.2.1($ii$) for $\mathbb{E}_0[h^{(13)}_{\mathbf{i}^1,\tilde{\mathbf{i}}^1} \cdot h^{(23)}_{\mathbf{i}^2,\tilde{\mathbf{i}}^2} \cdot h^{(14)}_{\mathbf{i}^3,\tilde{\mathbf{i}}^3} \cdot h^{(24)}_{\mathbf{i}^4,\tilde{\mathbf{i}}^4}]$ to be non-zero, it is necessary (but not sufficient, as seen below) that each of the eight tuples in (A.2.8) intersects with the union of the other seven at exactly d elements, with $f_m(i) \leq 2$ for all $i \in \cup_{\omega=1}^4 \mathbf{i}^\omega$. In particular, since $\tilde{\mathbf{i}}^1$ is disjoint from $\mathbf{i}^1 \setminus \tilde{\mathbf{i}}^1$, it is the case that

$$|\tilde{\mathbf{i}}^1 \cap (\cup_{\omega=2}^4 \mathbf{i}^\omega)| = d. \tag{A.2.9}$$

When conditioning on $\mathbf{X}^{(3)}_{\tilde{\mathbf{i}}^1 \cap \mathbf{i}^2}$ and $\mathbf{X}^{(1)}_{\tilde{\mathbf{i}}^1 \cap \mathbf{i}^3}$, it is seen that $h(\mathbf{R}^{(13)}_{\tilde{\mathbf{i}}^1})$ is independent of the other multiplicative factors on the right hand side of (A.2.7). Note that since f_m is always less than or equal to 2, by (A.2.9) one of $\tilde{\mathbf{i}}^1 \cap \mathbf{i}^2$ and $\tilde{\mathbf{i}}^1 \cap \mathbf{i}^3$ must have cardinality less than d. Hence, by Assumption 1.2.2 we have that

$$\mathbb{E}_0\left[h(\mathbf{R}^{(13)}_{\tilde{\mathbf{i}}^1}) \middle| \mathbf{X}^{(3)}_{\tilde{\mathbf{i}}^1 \cap \mathbf{i}^2}, \mathbf{X}^{(1)}_{\tilde{\mathbf{i}}^1 \cap \mathbf{i}^3}\right] = 0,$$

and the aforementioned conditional independence yields the claim from (A.2.6). $\square$

Lemma A.2.4. *Suppose* $1 \leq p, q, l, u \leq m$ *are four distinct indices, and* h *is a kernel of order of degeneracy* d *satisfying Assumption 1.2.2 under* H_0*. Then*

$$\mathbb{E}_0\left[U_h^{(pl)} U_h^{(ql)} U_h^{(pu)} U_h^{(qu)}\right] = O(n^{-2d-1}).$$

Proof. The proof uses similar counting techniques to that of Lemmas A.2.2 and A.2.3 and is only simpler. We sketch it here and leave the reader to fill in the details. Likewise, let $(p,q,l,u) = (1,2,3,4)$ without loss of generality. On expansion, by defining $B(\mathbf{i}^1, \dots, \mathbf{i}^4) := h\left(\mathbf{R}^{(13)}_{\mathbf{i}^1}\right) h\left(\mathbf{R}^{(23)}_{\mathbf{i}^2}\right) h\left(\mathbf{R}^{(14)}_{\mathbf{i}^3}\right) h\left(\mathbf{R}^{(24)}_{\mathbf{i}^4}\right)$,

$$\mathbb{E}_0\left[U_h^{(13)} U_h^{(23)} U_h^{(14)} U_h^{(24)}\right] = \binom{n}{k}^{-4} \sum_{\substack{\mathbf{i}^\omega \in \mathcal{P}(n,k) \\ 1 \leq \omega \leq 4}} \mathbb{E}_0\left[B(\mathbf{i}^1, \dots, \mathbf{i}^4)\right]. \tag{A.2.10}$$

By Lemma A.2.1(i), $\mathbb{E}_0\left[B(\mathbf{i}^1, \dots, \mathbf{i}^4)\right] = 0$ if $|\cup_{\omega=1}^4 \mathbf{i}^\omega| > 4k-2d$. When $|\cup_{\omega=1}^4 \mathbf{i}^\omega| = 4k-2d$, one can also show $\mathbb{E}_0\left[P(\mathbf{i}^1, \dots, \mathbf{i}^4)\right] = 0$ by using Lemma A.2.1(i) and the property of the kernel given by Assumption 1.2.2. Hence, there are only at most $O(n^{4k-2d-1})$ summands on the right hand side of (A.2.10) and hence $\mathbb{E}_0\left[U_h^{(13)} U_h^{(23)} U_h^{(14)} U_h^{(24)}\right] = \binom{n}{k}^{-4} O(n^{4k-2d-1}) = O(n^{-2d-1})$. $\square$

A.3 Proofs for Section 1.3

Proof of Lemma 1.3.1. It remains to prove claim (iii) about the fourth moment of $U_h^{(pq)}$ when the kernel h has its order of degeneracy d equal to 1 or 2 under H_0. Without loss of generality, we can assume $(p, q) = (1, 2)$. The fourth moment can be written as

$$\mathbb{E}_0\left[\left(U_h^{(12)}\right)^4\right] = \binom{n}{k}^{-4} \sum_{\mathbf{i}^1,\mathbf{i}^2,\mathbf{i}^3,\mathbf{i}^4 \in \mathcal{P}(n,k)} \mathbb{E}_0\left[\prod_{\omega=1}^4 h\left(\mathbf{R}_{\mathbf{i}^\omega,1}^{(12)}, \ldots, \mathbf{R}_{\mathbf{i}^\omega,k}^{(12)}\right)\right]. \tag{A.3.1}$$

The value of each summand $\mathbb{E}_0\big[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(12)})\big]$ in (A.3.1) depends on the multiset $(\cup_{\omega=1}^4 \mathbf{i}^\omega, f_m)$ with

$$\sum_{i \in \cup_{\omega=1}^4 \mathbf{i}^\omega} f_m(i) = 4k; \tag{A.3.2}$$

we use the multiset notation introduced in the first paragraph of Appendix A.2.

By Lemma A.2.1(i), we have $\mathbb{E}_0\big[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(12)})\big] = 0$ if $|\cup_{\omega=1}^4 \mathbf{i}^\omega| > 4k - 2d$. If $|\cup_{\omega=1}^4 \mathbf{i}^\omega| < 4k - 2d$, there are at most $\binom{n}{4k-2d-1}$ choices for the set $\cup_{\omega=1}^4 \mathbf{i}^\omega$. Since h is bounded, it thus holds that

$$\binom{n}{k}^{-4} \sum_{\substack{\mathbf{i}^1,\mathbf{i}^2,\mathbf{i}^3,\mathbf{i}^4 \in \mathcal{P}(n,k) \\ |\cup_{\omega=1}^4 \mathbf{i}^\omega| < 4k-2d}} \mathbb{E}_0\left[\prod_{\omega=1}^4 h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(12)}\right)\right] = O(n^{-2d-1}).$$

Therefore, to complete the proof, it suffices to show that

$$\binom{n}{k}^{-4} \sum_{\substack{\mathbf{i}^1,\mathbf{i}^2,\mathbf{i}^3,\mathbf{i}^4 \in \mathcal{P}(n,k) \\ |\cup_{\omega=1}^4 \mathbf{i}^\omega| = 4k-2d}} \mathbb{E}_0\left[\prod_{\omega=1}^4 h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(12)}\right)\right] = \begin{cases} \frac{3k^4(\zeta_1^h)^2}{n^2} + O(n^{-3}) & \text{if} \quad d = 1, \\ \binom{k}{2}^4 \frac{12}{n^4}\left((\zeta_2^h)^2 + 4\eta^h\right) + O(n^{-5}) & \text{if} \quad d = 2. \end{cases} \tag{A.3.3}$$

By Lemma A.2.1(i), when $|\cup_{\omega=1}^4 \mathbf{i}^\omega| = 4k - 2d$, a summand $\mathbb{E}_0\big[\prod_{\omega=1}^4 h(\mathbf{R}_{\mathbf{i}^\omega}^{(12)})\big]$ on the left hand side of (A.3.3) is non-zero only if

$$|\mathbf{i}^\omega \cap (\cup_{\omega' \neq \omega} \mathbf{i}^{\omega'})| = d \text{ for all } \omega = 1, \ldots, 4. \tag{A.3.4}$$

For both $d = 1$ and $d = 2$, (A.3.4) is true when the set $\{1, 2, 3, 4\}$ can be partitioned into two disjoint sets Ω_1 and Ω_2 such that

$$|\Omega_1| = |\Omega_2| = 2 \quad \text{and} \quad |\cap_{\omega \in \Omega_1} \mathbf{i}^\omega| = |\cap_{\omega \in \Omega_2} \mathbf{i}^\omega| = d, \tag{A.3.5}$$

in which case $(\cup_{\omega\in\Omega_1}\mathbf{i}^\omega) \cap (\cup_{\omega\in\Omega_2}\mathbf{i}^\omega) = \emptyset$ and, by independence,

$$\mathbb{E}_0\left[\prod_{\omega=1}^{4}\left(h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(12)}\right)\right)\right] = \prod_{j=1}^{2}\mathbb{E}_0\left[\prod_{\omega\in\Omega_j}\left(h\left(\mathbf{R}_{\mathbf{i}^\omega}^{(12)}\right)\right)\right] = (\zeta_d^h)^2. \tag{A.3.6}$$

Next, we count how many summands on the left hand side of (A.3.3) have their indices $\mathbf{i}^1,\ldots,\mathbf{i}^4$ satisfying the constellation in (A.3.5). There are $\binom{n}{4k-2d}$ choices for the set $\cup_{\omega=1}^4\mathbf{i}^\omega$. Then there are $\frac{1}{2}\binom{4k-2d}{2k-d}$ partitions of $\cup_{\omega=1}^4\mathbf{i}^\omega$ into two subsets of equal cardinality. Each of these subsets with cardinality $2k-d$ is to be split into two subsets that have d elements in common. We have $\binom{2k-d}{d}$ choices for this common element, and there are $\frac{1}{2}\binom{2k-2d}{k-d}$ ways of partitioning the remaining elements to form the two subsets. In the above counting process, no ordering is taken into account. Hence, the number of summands in (A.3.1) whose indices $\mathbf{i}^1,\ldots,\mathbf{i}^4$ satisfy (A.3.5) is

$$4!\binom{n}{4k-2d}\frac{1}{2}\binom{4k-2d}{2k-d}\left[\binom{2k-d}{d}\frac{1}{2}\binom{2k-2d}{k-d}\right]^2 = \frac{3n!}{(n-4k+2d)!\left[d!\left((k-d)!\right)^2\right]^2}. \tag{A.3.7}$$

When $d=1$, for any four tuples $\mathbf{i}^1,\ldots,\mathbf{i}^4\in\mathcal{P}(n,k)$ with $|\cup_{\omega=1}^4\mathbf{i}^\omega| = 4k-2d = 4k-2$, (A.3.4) is only satisfied when they can be described by the constellation in (A.3.5). Since

$$\binom{n}{k}^{-4}\frac{3n!}{(n-4k+2d)!\left[d!\left((k-d)!\right)^2\right]^2} = \binom{k}{d}^4\frac{3(d!)^2}{n^{2d}} + O\left(n^{-2d-1}\right), \tag{A.3.8}$$

by (A.3.6) and (A.3.7), we have proved the equality in (A.3.3) for $d=1$.

When $d=2$, in addition to (A.3.5), there is another constellation for $\mathbf{i}^1,\ldots,\mathbf{i}^1\in\mathcal{P}(n,k)$ that satisfies the condition in (A.3.4) subject to $|\cup_{\omega=1}^4\mathbf{i}^\omega| = 4k-2d = 4k-4$. If, up to relabeling of superscripts $\{1,\ldots,4\}$ for $\mathbf{i}^1,\ldots,\mathbf{i}^4$, the multiset $(\cup_{\omega=1}^4\mathbf{i}^\omega, f_m)$ is such that

$$|\mathbf{i}^1\cap\mathbf{i}^2| = |\mathbf{i}^2\cap\mathbf{i}^3| = |\mathbf{i}^3\cap\mathbf{i}^4| = |\mathbf{i}^4\cap\mathbf{i}^1| = 1 \quad \text{and} \tag{A.3.9}$$

$$f_m(i) = \begin{cases} 2 & \text{if } i \text{ belongs to any one of } \mathbf{i}^1\cap\mathbf{i}^2, \mathbf{i}^2\cap\mathbf{i}^3, \mathbf{i}^3\cap\mathbf{i}^4 \text{ or } \mathbf{i}^4\cap\mathbf{i}^1, \\ 1 & \text{otherwise,} \end{cases} \tag{A.3.10}$$

then (A.3.4) is satisfied with

$$\mathbb{E}_0 \left[\prod_{\omega=1}^{4} \left(h \left(\mathbf{R}^{(12)}_{\mathbf{i}^\omega} \right) \right) \right] = \eta^h. \tag{A.3.11}$$

We will conclude the proof of (A.3.3) for $d = 2$ by showing there are

$$3 \cdot 4! \cdot \binom{n}{4k-4} \binom{4k-4}{4} \binom{4k-8}{k-2, k-2, k-2, k-2} = \frac{3n!}{(n-4k+4)!((k-2)!)^4} \tag{A.3.12}$$

choices of $\mathbf{i}^1, \ldots, \mathbf{i}^4$ that satisfy (A.3.9) and (A.3.10), possibly after relabeling of their superscripts. If so, since $\binom{n}{4}^{-4} \frac{3n!}{(n-4k+4)!((k-2)!)^4} = \binom{k}{2}^4 \frac{48}{n^4} + O(n^{-5})$, combining (A.3.8) with the summand values (A.3.6) and (A.3.11), we have shown that for $d = 2$, the left hand side of (A.3.3) equals

$$\binom{k}{2}^4 \frac{3(2!)^2}{n^4} (\zeta_d^h)^2 + \binom{k}{2}^4 \frac{48}{n^4} \eta^h + O(n^{-5}) = \binom{k}{2}^4 \frac{12}{n^4} \left\{ (\zeta_2^h)^2 + 4\eta^h \right\} + O(n^{-5}).$$

It remains to show the count in (A.3.12). First, we count how many such constellations there are *without* any relabeling of superscripts. Given each of the $\binom{n}{4k-4}$ choice for the set $\cup_{\omega=1}^4 \mathbf{i}^\omega$, there are $4!\binom{4k-4}{4}$ ways of picking the disjoint singleton sets $(\mathbf{i}^1 \cap \mathbf{i}^2)$, $(\mathbf{i}^2 \cap \mathbf{i}^3)$, $(\mathbf{i}^3 \cap \mathbf{i}^4)$ and $(\mathbf{i}^4 \cap \mathbf{i}^1)$. Now there are $\binom{4k-8}{k-2,k-2,k-2,k-2}$ ways to partition the remaining $4k-8$ elements of the set $\cup_{\omega=1}^4 \mathbf{i}^\omega$ into the four sets $\mathbf{i}^1 \setminus (\mathbf{i}^2 \cup \mathbf{i}^4)$, $\mathbf{i}^2 \setminus (\mathbf{i}^1 \cup \mathbf{i}^3)$, $\mathbf{i}^3 \setminus (\mathbf{i}^2 \cup \mathbf{i}^4)$ and $\mathbf{i}^4 \setminus (\mathbf{i}^1 \cup \mathbf{i}^3)$. Hence, there are

$$4 \cdot \binom{n}{4k-4} \binom{4k-4}{4} \binom{4k-8}{k-2, k-2, k-2, k-2}$$

choices of $\mathbf{i}^1, \ldots, \mathbf{i}^4$ that satisfy (A.3.9) and (A.3.10) without having to relabel their superscripts. To obtain the factor of 3 in (A.3.12), we note that the constellation of $\mathbf{i}^1, \ldots, \mathbf{i}^4$ described by (A.3.9) and (A.3.10) is such that $\mathbf{i}^1$ intersects with $\mathbf{i}^2$ and $\mathbf{i}^4$. Alternatively, $\mathbf{i}^1$ can intersect with $\mathbf{i}^3$ and $\mathbf{i}^4$, or $\mathbf{i}^2$ and $\mathbf{i}^3$, to give a constellation satisfying (A.3.9) and (A.3.10) after relabeling of index superscripts. □

Proof of Lemma 1.3.2. As in the proof of Lemma 1.3.1, without loss of generality, we assume

$(p,q) = (1,2)$. For any given $\mathbf{i}, \mathbf{j} \in \mathcal{P}(n, 2k)$,

$$\begin{aligned}\mathbb{E}_0 &\left[h^W\left(\mathbf{R}_{\mathbf{i}}^{(12)}\right) h^W\left(\mathbf{R}_{\mathbf{j}}^{(12)}\right)\right] \\ &= \binom{2k}{k}^{-2} \sum_{\substack{\mathbf{i}^1 \subset \mathbf{i} \\ |\mathbf{i}^1| = k}} \sum_{\substack{\mathbf{j}^1 \subset \mathbf{j} \\ |\mathbf{j}^1| = k}} \mathbb{E}_0\left[h\left(\mathbf{R}_{\mathbf{i}^1}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{i}\setminus\mathbf{i}^1}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{j}^1}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{j}\setminus\mathbf{j}^1}^{(12)}\right)\right], \end{aligned} \tag{A.3.13}$$

Since $\mathbf{i}^1$, $\mathbf{i} \setminus \mathbf{i}^1$, $\mathbf{j}^1$ and $\mathbf{j} \setminus \mathbf{j}^1$ are tuples in $\mathcal{P}(n, k)$, if $|\mathbf{i} \cap \mathbf{j}| < 2d$, or equivalently $|\mathbf{i} \cup \mathbf{j}| > 4k - 2d$, by Lemma A.2.1($i$), all summands on the right hand side of (A.3.13) equal zero, and thus $\mathbb{E}_0\left[h^W\left(\mathbf{R}_{\mathbf{i}}^{(pq)}\right) h^W\left(\mathbf{R}_{\mathbf{j}}^{(pq)}\right)\right] = 0$.

Suppose $|\mathbf{i} \cap \mathbf{j}| = 2d$. If $\mathbf{i}^1, \mathbf{j}^1 \in \mathcal{P}(n,k)$ are such that $\mathbf{i}^1 \subset \mathbf{i}$ and $\mathbf{j}^1 \subset \mathbf{j}$, we define $\mathbf{i}^2 = \mathbf{i} \setminus \mathbf{i}^1$ and $\mathbf{j}^2 = \mathbf{j} \setminus \mathbf{j}^1$ to simplify notation. If

$$|\mathbf{i}^1 \cap \mathbf{j}^1| = d \quad \text{and} \quad |\mathbf{i}^2 \cap \mathbf{j}^2| = d, \tag{A.3.14}$$

then the necessary condition in Lemma A.2.1(i) is satisfied. Since $\mathbf{i}^1 \cup \mathbf{j}^1$ and $\mathbf{i}^2 \cup \mathbf{j}^2$ are disjoint, independence gives

$$\mathbb{E}_0\left[h\left(\mathbf{R}_{\mathbf{i}^1}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{i}^2}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{j}^1}^{(12)}\right) h\left(\mathbf{R}_{\mathbf{j}^2}^{(12)}\right)\right] = (\zeta_d^h)^2. \tag{A.3.15}$$

Similarly, if

$$|\mathbf{i}^1 \cap \mathbf{j}^2| = d \quad \text{and} \quad |\mathbf{i}^2 \cap \mathbf{j}^1| = d, \tag{A.3.16}$$

then (A.3.15) holds too.

Now we give the count for how many combinations of $\mathbf{i}^1$ and $\mathbf{j}^1$ satisfy (A.3.14). Since $|\mathbf{i} \cap \mathbf{j}| = 2d$, there are $\binom{2d}{d}$ choices for the set $\mathbf{i}^1 \cap \mathbf{j}^1$, which determines $\mathbf{i}^2 \cap \mathbf{j}^2$. For each such choice, there are then $\binom{2k-2d}{k-d}$ choices for each of $\mathbf{i}^1 \setminus (\mathbf{i}^1 \cap \mathbf{j}^1)$ and $\mathbf{j}^2 \setminus (\mathbf{i}^2 \cap \mathbf{j}^2)$, which determine $\mathbf{i}^2$ and $\mathbf{j}^2$. Hence, there are $\binom{2d}{d}\binom{2k-2d}{k-d}^2$ choices of $(\mathbf{i}^1, \mathbf{j}^1)$ satisfying (A.3.14). Analogously, there are also $\binom{2d}{d}\binom{2k-2d}{k-d}^2$ choices of $(\mathbf{i}^1, \mathbf{j}^1)$ satisfying (A.3.16). In total, there are

$$2\binom{2d}{d}\binom{2k-2d}{k-d}^2 \tag{A.3.17}$$

summands in (A.3.13) with the value $(\zeta_d^h)^2$.

If $d = 1$, then no constellations for $\mathbf{i}^1$ and $\mathbf{i}^2$ other than the ones given by (A.3.14) and (A.3.16) yield a non-zero value for $\mathbb{E}_0[h(\mathbf{R}_{\mathbf{i}^1}^{(12)})h(\mathbf{R}_{\mathbf{i}^2}^{(12)})h(\mathbf{R}_{\mathbf{j}^1}^{(12)})h(\mathbf{R}_{\mathbf{j}^2}^{(12)})]$. Therefore, we deduce from (A.3.13) that, for $d = 1$,

$$\zeta_{2d}^{h^W} = 2\binom{2d}{d}\binom{2k-2d}{k-d}^2\binom{2k}{k}^{-2}(\zeta_1^h)^2 = 4\binom{2k-2}{k-1}^2\binom{2k}{k}^{-2}(\zeta_1^h)^2.$$

It remains to prove the formula for $\zeta_{2d}^{h^W}$ when $d = 2$. In this case, besides (A.3.14) and (A.3.16), there is one other constellation for $\mathbf{i}^1$, $\mathbf{i}^2$, $\mathbf{j}^1$, $\mathbf{j}^2$ so that the necessary condition in Lemma A.2.1(i) is satisfied. If the multiset $(\mathbf{i}^1 \cup \mathbf{j}^1 \cup \mathbf{i}^2 \cup \mathbf{j}^2, f_m)$ is such that

$$|\mathbf{i}^1 \cap \mathbf{j}^1| = |\mathbf{j}^1 \cap \mathbf{i}^2| = |\mathbf{i}^2 \cap \mathbf{j}^2| = |\mathbf{j}^2 \cap \mathbf{i}^1| = 1 \quad \text{and} \tag{A.3.18}$$

$$f_m(i) = \begin{cases} 2 & \text{if } i \text{ belongs to any one of } \mathbf{i}^1 \cap \mathbf{j}^1, \mathbf{j}^1 \cap \mathbf{i}^2, \mathbf{i}^2 \cap \mathbf{j}^2 \text{ or } \mathbf{j}^2 \cap \mathbf{i}^1, \\ 1 & \text{otherwise,} \end{cases} \tag{A.3.19}$$

then

$$\mathbb{E}_0\left[h\left(\mathbf{R}_{\mathbf{i}^1}^{(12)}\right)h\left(\mathbf{R}_{\mathbf{i}^2}^{(12)}\right)h\left(\mathbf{R}_{\mathbf{j}^1}^{(12)}\right)h\left(\mathbf{R}_{\mathbf{j}^2}^{(12)}\right)\right] = \eta^h. \tag{A.3.20}$$

Now we count: For a fixed pair $(\mathbf{i}, \mathbf{j})$ such that $|\mathbf{i} \cap \mathbf{j}| = 4$, there are $4!$ choices for the singletons $\mathbf{i}^1 \cap \mathbf{j}^1$, $\mathbf{j}^1 \cap \mathbf{i}^2$, $\mathbf{i}^2 \cap \mathbf{j}^2$ and $\mathbf{j}^2 \cap \mathbf{i}^1$. Given each such choice for these singletons, there are $\binom{2k-4}{k-2}$ choices for each one of $\mathbf{i}^1$ and $\mathbf{j}^1$, hence there are

$$4!\binom{2k-4}{k-2}^2$$

summands on the right hand side of (A.3.13) with the value η^h. Combining with the count (A.3.17) for summands with the value $(\zeta_d^h)^2$, we conclude that if $d = 2$ then

$$\begin{aligned} \zeta_{2d}^{h^W} &= \binom{2k}{k}^{-2}\left\{2\binom{2d}{d}\binom{2k-2d}{k-d}^2(\zeta_d^h)^2 + 4!\binom{2k-4}{k-2}^2\eta^h\right\} \\ &= 12\binom{2k-4}{k-2}^2\binom{2k}{k}^{-2}\left[(\zeta_2^h)^2 + 2\eta^h\right]. \quad \square \end{aligned}$$

A.4 Proofs for Section 1.4

Here, we prove Lemmas A.4.1 and A.4.2 that were used in the proof of Theorem 1.4.1.

Lemma A.4.1. *The martingale differences from (1.4.2) satisfy the L^2 convergences*

$$\mathbb{E}_0\left[\left(\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^S)^2\middle|\mathcal{F}_{n,l-1}\right]-k^4(\zeta_1^h)^2\right)^2\right]\longrightarrow 0\,, \tag{A.4.1}$$

$$\mathbb{E}_0\left[\left(\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^T)^2\middle|\mathcal{F}_{n,l-1}\right]-k^4(\zeta_1^h)^2\right)^2\right]\longrightarrow 0\,, \tag{A.4.2}$$

$$\mathbb{E}_0\left[\left(\frac{n}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^2\middle|\mathcal{F}_{n,l-1}\right]-\frac{k^2\zeta_1^h}{2}\right)^2\right]\longrightarrow 0 \tag{A.4.3}$$

when $d=1$, and the L^2 convergences

$$\mathbb{E}_0\left[\left(\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^S)^2\middle|\mathcal{F}_{n,l-1}\right]-4\binom{k}{2}^4\left\{(\zeta_2^h)^2+6\eta^h\right\}\right)^2\right]\longrightarrow 0\,, \tag{A.4.4}$$

$$\mathbb{E}_0\left[\left(\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^T)^2\middle|\mathcal{F}_{n,l-1}\right]-4\binom{k}{2}^4\left\{(\zeta_2^h)^2+2\eta^h\right\}\right)^2\right]\longrightarrow 0\,, \tag{A.4.5}$$

$$\mathbb{E}_0\left[\left(\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^2\middle|\mathcal{F}_{n,l-1}\right]-\binom{k}{2}^2\zeta_2^h\right)^2\right]\longrightarrow 0 \tag{A.4.6}$$

when $d=2$.

Proof. When $d=1$, for the L^2 convergences in (A.4.1), (A.4.2) and (A.4.3), it is sufficient to show that, as $m,n\longrightarrow\infty$,

$$\begin{aligned}&\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2],\quad \frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^T)^2]\longrightarrow k^4(\zeta_1^h)^2,\\ &\frac{n}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^2\right]\longrightarrow\frac{k^2\zeta_1^h}{2},\end{aligned}\qquad \textit{and} \tag{A.4.7}$$

$$\begin{aligned}\mathrm{Var}_0\left[\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2|\mathcal{F}_{n,l-1}]\right]&\longrightarrow 0,\\ \mathrm{Var}_0\left[\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^T)^2|\mathcal{F}_{n,l-1}]\right]&\longrightarrow 0,\\ \mathrm{Var}_0\left[\frac{n}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^Z)^2|\mathcal{F}_{n,l-1}]\right]&\longrightarrow 0.\end{aligned} \tag{A.4.8}$$

When $d = 2$, for the L^2 convergences in (A.4.4) and (A.4.5), it suffices to show that, as $m, n \longrightarrow \infty$,

$$\begin{aligned}\sum_{l=2}^{m} \frac{n^4}{m^2}\mathbb{E}_0[(D_{nl}^S)^2] &\longrightarrow 4\binom{k}{2}^4 \left\{(\zeta_2^h)^2 + 6\eta^h\right\}, \\ \sum_{l=2}^{m} \frac{n^4}{m^2}\mathbb{E}_0[(D_{nl}^T)^2] &\longrightarrow 4\binom{k}{2}^4 \left\{(\zeta_2^h)^2 + 2\eta^h\right\}, \qquad \textit{and} \\ \sum_{l=2}^{m} \frac{n^2}{m^2}\mathbb{E}_0\left[(D_{nl}^Z)^2\right] &\longrightarrow \binom{k}{2}^2 \zeta_2^h,\end{aligned} \tag{A.4.9}$$

$$\begin{aligned}\mathrm{Var}_0\left[\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2|\mathcal{F}_{n,l-1}]\right] &\longrightarrow 0, \\ \mathrm{Var}_0\left[\frac{n^4}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^T)^2|\mathcal{F}_{n,l-1}]\right] &\longrightarrow 0, \\ \mathrm{Var}_0\left[\frac{n^2}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^Z)^2|\mathcal{F}_{n,l-1}]\right] &\longrightarrow 0.\end{aligned} \tag{A.4.10}$$

We will first show the convergences of expectations in (A.4.7) and (A.4.9). Suppose $d = 1$ or 2 is the order of degeneracy of h under H_0. By Lemma 1.2.1(i) and (iii), the terms $\bar{U}_h^{(pl)}$ that are summed to form D_{nl}^S are i.i.d. such that

$$\frac{n^{2d}}{m^2}\mathbb{E}_0[(D_{nl}^S)^2] = \frac{n^{2d}}{m^2}\sum_{p=1}^{l-1}\mathrm{Var}_0\left[\bar{U}_h^{(pl)}\right] = \frac{n^{2d}}{m^2}(l-1)\mathrm{Var}_0\left[\bar{U}_h^{(12)}\right].$$

It follows that

$$\frac{n^{2d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2] = \frac{(m-1)n^{2d}}{2m}\mathrm{Var}_0\left[\bar{U}_h^{(12)}\right]. \tag{A.4.11}$$

Similarly, by Lemma 1.2.1(i) and (iii), we have that

$$\frac{n^{2d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^T)^2] = \frac{(m-1)n^{2d}}{2m}\mathrm{Var}_0\left[W_h^{(12)}\right] \quad \text{and} \tag{A.4.12}$$

$$\frac{n^{d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^2\right] = \frac{(m-1)n^{d}}{2m}\mathrm{Var}_0\left[U_h^{(12)}\right]. \tag{A.4.13}$$

By Lemma 1.3.1(i) and (iii),

$$\begin{aligned}\mathrm{Var}_0\left[\bar{U}_h^{(12)}\right] &= \mathbb{E}_0\left[\left(U_h^{(12)}\right)^4\right] - \mu_h^2 \\ &= \begin{cases} \frac{2k^4(\zeta_1^h)^2}{n^2} + O\left(n^{-3}\right) & \text{if} \quad d=1, \\ \frac{8}{n^4}\binom{k}{2}^4\{(\zeta_2^h)^2 + 6\eta^h\} + O\left(n^{-5}\right) & \text{if} \quad d=2. \end{cases}\end{aligned} \tag{A.4.14}$$

Since $W_h^{(12)}$ is a rank-based U-statistic with the induced kernel function h^W of degree $2k$, via Lemma 1.3.2, Lemma 1.3.1(i) applies to give

$$\begin{aligned}\mathrm{Var}_0\left[W_h^{(12)}\right] = \mathbb{E}_0\left[\left(W_h^{(12)}\right)^2\right] &= \binom{2k}{2d}^2 \frac{(2d)!}{n^{2d}}\zeta_{2d}^{h^W} + O(n^{-2d-1}) \\ &= \begin{cases} \frac{2k^4(\zeta_1^h)^2}{n^2} + O\left(n^{-3}\right) & \text{if} \quad d=1, \\ \frac{8}{n^4}\binom{k}{2}^4\{(\zeta_2^h)^2 + 2\eta^h\} + O(n^{-5}) & \text{if} \quad d=2. \end{cases}\end{aligned} \tag{A.4.15}$$

Moreover, Lemma 1.3.1(i) yields that

$$\mathrm{Var}_0\left[\left(U_h^{(12)}\right)^2\right] = \mathbb{E}_0\left[\left(U_h^{(12)}\right)^2\right] = \begin{cases} \frac{k^2\zeta_1^h}{n} + O\left(n^{-2}\right) & \text{if} \quad d=1, \\ \frac{2\zeta_2^h}{n^2}\binom{k}{2}^2 + O(n^{-3}) & \text{if} \quad d=2. \end{cases} \tag{A.4.16}$$

Plugging (A.4.14), (A.4.15) and (A.4.16) into (A.4.11) , (A.4.12) and (A.4.13) for $d=1$ and $d=2$, respectively, and taking the limit, we obtain the convergences in (A.4.7) and (A.4.9).

Next, we show that the variances in (A.4.8) and (A.4.10) converges to zero. For $d \in \{1,2\}$, write

$$\begin{aligned}&\frac{n^{2d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0\left[(D^S)_{nl}^2|\mathcal{F}_{n,l-1}\right] \\ &= \frac{n^{2d}}{m^2}\left\{\sum_{l=2}^{m}\sum_{p=1}^{l-1}\mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^2\middle|\mathcal{F}_{n,l-1}\right] + 2\sum_{l=3}^{m}\sum_{1\le p<q<l}\mathbb{E}_0\left[\bar{U}_h^{(pl)}\bar{U}_h^{(ql)}\middle|\mathcal{F}_{n,l-1}\right]\right\},\end{aligned}$$

and notice that the first sum on the right-hand side is a constant because, by Lemma 1.2.1(ii),

$$\mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^2\middle|\mathcal{F}_{n,l-1}\right] = \mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^2\middle|\mathbf{X}^{(p)}\right] = \mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^2\right].$$

We observe that in order to show $\mathrm{Var}_0\left[\frac{n^{2d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2|\mathcal{F}_{n,l-1}]\right] \longrightarrow 0$, it suffices to show

$$\frac{n^{4d}}{m^4}\mathrm{Var}_0\left[\sum_{l=3}^{m}\sum_{1\le p<q<l}\mathbb{E}_0\left[\bar{U}_h^{(pl)}\bar{U}_h^{(ql)}\middle|\mathcal{F}_{n,l-1}\right]\right] \longrightarrow 0. \tag{A.4.17}$$

By exactly analogous arguments, it suffices to show

$$\frac{n^{4d}}{m^4}\mathrm{Var}_0\left[\sum_{l=3}^{m}\sum_{1\le p<q<l}\mathbb{E}_0\left[W_h^{(pl)}W_h^{(ql)}\middle|\mathcal{F}_{n,l-1}\right]\right] \longrightarrow 0 \quad \text{and} \tag{A.4.18}$$

$$\frac{n^{2d}}{m^4}\mathrm{Var}_0\left[\sum_{l=3}^{m}\sum_{1\le p<q<l}\mathbb{E}_0\left[U_h^{(pl)}U_h^{(ql)}\middle|\mathcal{F}_{n,l-1}\right]\right] \longrightarrow 0 \tag{A.4.19}$$

to prove $\mathrm{Var}_0\left[\frac{n^{2d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^2|\mathcal{F}_{n,l-1}]\right],\ \ \mathrm{Var}_0\left[\frac{n^{d}}{m^2}\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^Z)^2|\mathcal{F}_{n,l-1}]\right] \longrightarrow 0$.

We first prove (A.4.17). For $p<q<l$, consider

$$C^{(pq)} := \mathbb{E}_0\left[\bar{U}_h^{(pl)}\bar{U}_h^{(ql)}\middle|\ \mathcal{F}_{n,l-1}\right] \ = \ \mathbb{E}_0\left[\bar{U}_h^{(pl)}\bar{U}_h^{(ql)}\middle|\mathbf{X}^{(p)},\mathbf{X}^{(q)}\right],$$

which is a function of $\mathbf{X}^{(p)}$ and $\mathbf{X}^{(q)}$ alone. Since

$$\bar{U}_h^{(pl)}\bar{U}_h^{(ql)} = f(\mathbf{R}_1^{(pl)},\dots,\mathbf{R}_k^{(pl)})f(\mathbf{R}_1^{(ql)},\dots,\mathbf{R}_k^{(ql)})$$

for a function $f:(\mathbb{R}^2)^k \longrightarrow \mathbb{R}$ that is permutation symmetric in its k arguments, and since the rank vectors $\mathbf{R}^{(p)}$, $\mathbf{R}^{(q)}$, $\mathbf{R}^{(l)}$ are independent and uniformly distributed on $\mathfrak{S}_n$ under H_0, the conditional expectation $C^{(pq)}$ is in fact a function of the tuple $(\mathbf{R}_1^{(pq)},\dots,\mathbf{R}_n^{(pq)})$ that is symmetric in its n arguments. Therefore, Lemma 1.2.1 applies to the collection of $C^{(pq)}$, $1\le p\ne q\le m$. The variance in (A.4.17) is thus

$$\begin{aligned}\mathrm{Var}_0\left[\sum_{l=3}^{m}\sum_{1\le p<q<l}C^{(pq)}\right] &= \sum_{1\le p<q\le m-1}(m-q)^2\mathrm{Var}_0\left[C^{(pq)}\right]\\ &= \frac{1}{12}m(m-2)(m-1)^2\mathrm{Var}_0\left[C^{(12)}\right].\end{aligned}$$

Now under the asymptotic regime $m,n\longrightarrow\infty$, (A.4.17) holds if $\mathrm{Var}_0\left[C^{(12)}\right]$ is of order $O(n^{-4d-1})$.

Suppose $2 < l < u \leq m$, then by definition

$$C^{(12)} = \mathbb{E}_0\left[\bar{U}_h^{(1l)}\bar{U}_h^{(2l)}\middle|\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right] = \mathbb{E}_0\left[\bar{U}_h^{(1u)}\bar{U}_h^{(2u)}\middle|\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right],$$

from this it follows that

$$\begin{aligned}\mathbb{E}_0\left[\bar{U}_h^{(1l)}\bar{U}_h^{(2l)}\bar{U}_h^{(1u)}\bar{U}_h^{(2u)}\right] &= \mathbb{E}_0\left[\mathbb{E}_0\left[\bar{U}_h^{(1l)}\bar{U}_h^{(2l)}\bar{U}_h^{(1u)}\bar{U}_h^{(2u)}\middle|\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right]\right]\\ &= \mathbb{E}_0\left[\mathbb{E}_0\left[\bar{U}_h^{(1l)}\bar{U}_h^{(2l)}\middle|\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right]\mathbb{E}_0\left[\bar{U}_h^{(1u)}\bar{U}_h^{(2u)}\middle|\mathbf{X}^{(1)},\mathbf{X}^{(2)}\right]\right] \qquad \text{(A.4.20)}\\ &= \mathbb{E}_0\left[\left(C^{(12)}\right)^2\right],\end{aligned}$$

where (A.4.20) follows from independence of $\mathbf{X}^{(l)}$ and $\mathbf{X}^{(u)}$. Applying Lemma A.2.2, we deduce that $\mathbb{E}_0[(C^{(12)})^2]$ is of order $O(n^{-4d-1})$. This concludes the proof as an application of Lemma 1.2.1(*iii*) shows that $C^{(12)}$ has mean zero, and thus $\mathrm{Var}_0[C^{(12)}] = \mathbb{E}_0[(C^{(12)})^2]$.

The proof of (A.4.18) and (A.4.19) proceeds line by line as the proof of (A.4.17), where for all $1 \leq p \neq q \leq m$ we replace $\bar{U}_h^{(pq)}$ by $W_h^{(pq)}$ or $U_h^{(pq)}$, define $C^{(pq)}$ alternatively as

$$C^{(pq)} := \mathbb{E}_0\left[W_h^{(pl)}W_h^{(ql)}\middle|\ \mathcal{F}_{n,l-1}\right] \quad \text{or} \quad C^{(pq)} := \mathbb{E}_0\left[U_h^{(pl)}U_h^{(ql)}\middle|\ \mathcal{F}_{n,l-1}\right],$$

and apply Lemma A.2.3 or Lemma A.2.4. We omit the details. □

Lemma A.4.2. *For $d = 1$ or 2, the martingale differences from (1.4.2) satisfy the Lyapunov conditions*

$$\frac{n^{4d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^S)^4\middle|\mathcal{F}_{n,l-1}\right], \quad \frac{n^{4d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^T)^4\middle|\mathcal{F}_{n,l-1}\right] \underset{p}{\longrightarrow} 0 \quad \textit{and} \tag{A.4.21}$$

$$\frac{n^{2d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^4\middle|\mathcal{F}_{n,l-1}\right] \underset{p}{\longrightarrow} 0 \tag{A.4.22}$$

as $m, n \longrightarrow \infty$.

Proof. Since $\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^S)^4|\mathcal{F}_{n,l-1}]$, $\sum_{l=2}^{m}\mathbb{E}_0[(D_{nl}^T)^4|\mathcal{F}_{n,l-1}]$ and $\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^4\middle|\mathcal{F}_{n,l-1}\right]$ are nonnegative random variables, it suffices to show that all three expectations converge to zero, that is,

$$\frac{n^{4d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^S)^4\right], \quad \frac{n^{4d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^T)^4\right], \quad \frac{n^{2d}}{m^4}\sum_{l=2}^{m}\mathbb{E}_0\left[(D_{nl}^Z)^4\middle|\mathcal{F}_{n,l-1}\right] \longrightarrow 0.$$

We first show it for $\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^S)^4\right]$. By Lemma 1.2.1($i$) and ($iii$), D_{nl}^S is a sum of $l-1$ centered i.i.d. random variables. On expansion, we have that

$$\begin{aligned}\mathbb{E}_0\left[\left(D_{nl}^S\right)^4\right] &= \sum_{p=1}^{l-1}\mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^4\right] + 6\sum_{1\le p<q<l}\mathbb{E}_0\left[\left(\bar{U}_h^{(pl)}\right)^2\right]\mathbb{E}_0\left[\left(\bar{U}_h^{(ql)}\right)^2\right] \\ &= (l-1)\mathbb{E}_0\left[\left(\bar{U}_h^{(12)}\right)^4\right] + 6\binom{l-1}{2}\left(\mathrm{Var}_0\left[\bar{U}_h^{(12)}\right]\right)^2.\end{aligned}$$

It follows that

$$\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[\left(D_{nl}^S\right)^4\right] = \frac{n^{4d}}{m^4}\left\{\binom{m}{2}\mathbb{E}_0\left[\left(\bar{U}_h^{(12)}\right)^4\right] + 6\binom{m}{3}\left(\mathrm{Var}_0\left[\bar{U}_h^{(12)}\right]\right)^2\right\}. \quad \text{(A.4.23)}$$

Now recall from (A.4.14) that the variance of $\bar{U}_h^{(12)}$ is of order $O(n^{-2d})$. Furthermore,

$$\begin{aligned}\mathbb{E}_0\left[\left(\bar{U}_h^{(12)}\right)^4\right] &= \mathbb{E}_0\left[\left(\left(U_h^{(12)}\right)^2 - \mu_h\right)^4\right] \\ &= \mathbb{E}_0\left[\left(U_h^{(12)}\right)^8 - 4\mu_h\left(U_h^{(12)}\right)^6 + 6\mu_h^2\left(U_h^{(12)}\right)^4 - 4\mu_h^3\left(U_h^{(12)}\right)^2 + \mu_h^4\right]\end{aligned}$$

is of order $O(n^{-4d})$ by Lemma 1.3.1(ii). Substituting these into (A.4.23) we conclude that

$$\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^S)^4\right] = O(m^{-1}) \longrightarrow 0 \quad \text{as} \quad m,n\longrightarrow\infty.$$

The proof for $\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^T)^4\right]$ and $\frac{n^{2d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[(D_{nl}^Z)^4\right]$ is similar. On expansion, we have

$$\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[\left(D_{nl}^T\right)^4\right] = \frac{n^{4d}}{m^4}\left\{\binom{m}{2}\mathbb{E}_0\left[\left(W_h^{(12)}\right)^4\right] + 6\binom{m}{3}\left(\mathbb{E}_0\left[\left(W_h^{(12)}\right)^2\right]\right)^2\right\}, \quad \text{(A.4.24)}$$

$$\frac{n^{2d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[\left(D_{nl}^Z\right)^4\right] = \frac{n^{2d}}{m^4}\left\{\binom{m}{2}\mathbb{E}_0\left[\left(U_h^{(12)}\right)^4\right] + 6\binom{m}{3}\left(\mathbb{E}_0\left[\left(U_h^{(12)}\right)^2\right]\right)^2\right\} \quad \text{(A.4.25)}$$

by Lemma 1.2.1(i) and (iii). By Lemmas 1.3.1(ii) and 1.3.2, since h^W has order of degeneracy $2d$, $\mathbb{E}_0\left[(W_h^{(12)})^4\right]$ and $\mathbb{E}_0\left[(W_h^{(12)})^2\right]$ are of order $O(n^{-4d})$ and $O(n^{-2d})$ respectively. Another application of Lemma 1.3.1(ii) gives that $\mathbb{E}_0\left[(U_h^{(12)})^4\right] = O(n^{-2d})$ and $\mathbb{E}_0\left[(U_h^{(12)})^2\right] = O(n^{-d})$. On substituting these into (A.4.24) and (A.4.25) we get that both $\frac{n^{4d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[\left(D_{nl}^T\right)^4\right]$ and $\frac{n^{2d}}{m^4}\sum_{l=2}^m \mathbb{E}_0\left[\left(D_{nl}^Z\right)^4\right]$ are of order $O(m^{-1})$ and converge to 0 as $m,n\longrightarrow\infty$. □

Proof of Corollary 1.4.2. It suffices to show that $\frac{n}{m}(S_\rho - S_{\hat{\rho}}) = o_p(1)$, in which case the corollary is implied by the fact that $\frac{n}{m}S_{\hat{\rho}} \longrightarrow N(0,1)$ as given in Theorem 1.4.1 and the value of $\zeta_1^{h_{\hat{\rho}_s}}$ in Table 1.1. By the decomposition in (1.2.5), the statistic S_ρ from (1.4.1) may be written as

$$S_\rho = \sum_{1 \le p < q \le m} \left(\frac{n-2}{n+1} \hat{\rho}^{(pq)} + \frac{3}{n+1} \tau^{(pq)} \right)^2 - \binom{m}{2} \mu_{\rho^2}.$$

Expanding the square in the summands on the right-hand side, we obtain that

$$\begin{aligned} S_\rho = \left(\frac{n-2}{n+1}\right)^2 S_{\hat{\rho}} + \frac{9}{(n+1)^2} S_\tau + \frac{6(n-2)}{(n+1)^2} \sum_{1 \le p < q \le m} \hat{\rho}^{(pq)} \tau^{(pq)} \\ + \binom{m}{2} \left[\left(\frac{n-2}{n+1}\right)^2 \mu_{\hat{\rho}^2} + \frac{9}{(n+1)^2} \mu_{\tau^2} - \mu_{\rho^2} \right]; \end{aligned} \quad \text{(A.4.26)}$$

recall the definition of S_τ and $S_{\hat{\rho}}$. Note that since S_ρ, S_τ and $S_{\hat{\rho}}$ have mean zero, it holds that

$$\mu_{\hat{\rho}\tau} := \mathbb{E}_0 \left[\hat{\rho}^{(pq)} \tau^{(pq)} \right] = \frac{(n+1)^2}{6(n-2)} \left[\mu_{\rho^2} - \left(\frac{n-2}{n+1}\right)^2 \mu_{\hat{\rho}^2} - \frac{9}{(n+1)^2} \mu_{\tau^2} \right],$$

and hence (A.4.26) can be rewritten as

$$S_\rho = \left(\frac{n-2}{n+1}\right)^2 S_{\hat{\rho}} + \frac{9}{(n+1)^2} S_\tau + \frac{6(n-2)}{(n+1)^2} \left[\sum_{1 \le p < q \le m} \hat{\rho}^{(pq)} \tau^{(pq)} - \binom{m}{2} \mu_{\hat{\rho}\tau} \right].$$

Since $\frac{9n}{m(n+1)^2} S_\tau = o_p(1)$ by Theorem 1.4.1, in order to prove the assertion that $\frac{n}{m}(S_\rho - S_{\hat{\rho}}) = o_p(1)$, it thus suffices to show that

$$\frac{6n(n-2)}{m(n+1)^2} \left[\sum_{1 \le p < q \le m} \hat{\rho}^{(pq)} \tau^{(pq)} - \binom{m}{2} \mu_{\hat{\rho}\tau} \right] \xrightarrow[p]{} 0.$$

We show this by proving convergence to zero in L^2, for which we need to argue that

$$\frac{36n^2(n-2)^2}{m^2(n+1)^4} \mathbb{E}_0 \left[\left\{ \sum_{1 \le p < q \le m} \hat{\rho}^{(pq)} \tau^{(pq)} - \binom{m}{2} \mu_{\hat{\rho}\tau} \right\}^2 \right] \longrightarrow 0. \quad \text{(A.4.27)}$$

Note that Lemma 1.2.1 applies to the collection of statistics $\hat{\rho}^{(pq)}\tau^{(pq)}$. By Lemma 1.2.1($i$) and ($iv$), the term in (A.4.27) equals

$$\frac{18n^2(n-2)^2(m-1)}{(n+1)^4 m} \left\{ \mathbb{E}_0 \left[\left(\hat{\rho}^{(12)} \tau^{(12)} \right)^2 \right] - \mu_{\hat{\rho}\tau}^2 \right\}. \quad \text{(A.4.28)}$$

Since $\frac{18n^2(n-2)^2(m-1)}{(n+1)^4 m} = O(1)$ as $m, n \longrightarrow \infty$, for the convergence from (A.4.27) it remains to show that

$$\mathrm{Var}_0\left[\hat{\rho}^{(12)}\tau^{(12)}\right] = \mathbb{E}_0\left[\left(\hat{\rho}^{(12)}\tau^{(12)}\right)^2\right] - \mu^2_{\hat{\rho}\tau} \ \longrightarrow \ 0.$$

However, using the inequality $2xy \le (x^2 + y^2)$, we see that

$$0 \le \mathrm{Var}_0\left[\hat{\rho}^{(12)}\tau^{(12)}\right] \le \mathbb{E}_0\left[\left(\hat{\rho}^{(12)}\tau^{(12)}\right)^2\right] \le \frac{1}{2}\mathbb{E}_0\left[\left(\hat{\rho}^{(12)}\right)^4\right] + \frac{1}{2}\mathbb{E}_0\left[\left(\tau^{(12)}\right)^4\right],$$

which is of order $O(n^{-2})$ by Lemma 1.3.1(ii). $\square$

A.5 Proofs for Section 1.5

Unlike in other sections, here all the rank-based U-statistics will be treated as functions of the original data $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}$ in our presentation.

Proof of Theorem 1.5.1. In this proof, all operators $\mathbb{E}[\cdot]$, $\mathrm{Cov}[\cdot]$, $\mathrm{Var}(\cdot)$, $P(\cdot)$ are with respect to a general distribution in $\mathcal{D}_m$.

(i): Let $\mathbf{U}_\tau$ be the $\binom{m}{2}$-vector $(U^{(pq)}_{h_\tau})_{1\le p<q\le m}$. Then $\mathbf{U}_\tau$ is a U-statistic taking values in $\mathbb{R}^{\binom{m}{2}}$, with the $\binom{m}{2}$-dimensional vector-valued kernel

$$\mathbf{h}_\tau(\mathbf{X}_i, \mathbf{X}_j) = \Big(h_\tau(\mathbf{X}^{(pq)}_i, \mathbf{X}^{(pq)}_j)\Big)_{1\le p<q\le m}$$

of degree $k = 2$. Here, $i \ne j$ index any pair of samples. Note that $S_\tau = \|\mathbf{U}_\tau\|_2^2 - \binom{m}{2}\mu_{h_\tau}$, and under the regime $\frac{m}{n} \longrightarrow \gamma$, using Theorem 1.4.1 $\phi_\alpha(S_\tau)$ rejects H_0 when $\|\mathbf{U}_\tau\|_2 \ge \sqrt{\binom{m}{2}\mu_\tau + \frac{4m}{9n} z_{1-\alpha}} = O(\sqrt{n})$; recall $\mu_\tau = \frac{2(2n+5)}{9n(n-1)}$ and the value of $\zeta^{h_\tau}_1$ in Table 1.1. By the triangle inequality

$$\|\mathbf{U}_\tau\|_2 \ge \|\Theta_\tau\|_2 - \|\mathbf{U}_\tau - \Theta_\tau\|_2,$$

it suffices to show that as $n \longrightarrow \infty$, uniformly over $\mathcal{D}_m$,

$$P(\|\mathbf{U}_\tau - \Theta_\tau\|_2 \ge C\sqrt{n}) \le 1 - \beta$$

for some constant $C > 0$ that only depends on β and γ. For any pair $i \neq j$, let $\mathbf{h}_{\tau,1}(\mathbf{X}_i) = \mathbb{E}[\mathbf{h}_\tau(\mathbf{X}_i, \mathbf{X}_j)|\mathbf{X}_i]$ and define the canonical functions (Borovskikh, 1996, p.8)

$$\mathbf{g}_1(\mathbf{X}_i) := \mathbf{h}_{\tau,1}(\mathbf{X}_i) - \Theta, \tag{A.5.1}$$

$$\mathbf{g}_2(\mathbf{X}_i, \mathbf{X}_j) := \mathbf{h}_\tau(\mathbf{X}_i, \mathbf{X}_j) - \mathbf{h}_{\tau,1}(\mathbf{X}_i) - \mathbf{h}_{\tau,1}(\mathbf{X}_j) + \Theta. \tag{A.5.2}$$

Since the Kendall kernel h_τ is bounded, $\|\mathbf{g}_1\|_2^2$ and $\|\mathbf{g}_2\|_2^2$ are both less than $\binom{m}{2}M$ for a certain constant $M > 0$ that does not depend on n and m. Suppose $d \in \{1, 2\}$ is the order of degeneracy for the kernel $\mathbf{h}_\tau$. By Borovskikh (1996, Corollary 8.1.7), we have that for any $t > 0$,

$$P(\|\mathbf{U}_\tau - \Theta_\tau\|_2 > t) \ \leq \ C_1 \exp\left\{-C_2 n \left(\frac{t^2}{\lambda^2}\right)^{1/d}\right\},$$

where $C_1, C_2 > 0$ are universal constants and $\lambda^2 = M\binom{m}{2}\sum_{c=0}^{2-d} n^{-c} = M\binom{m}{2}\frac{1-n^{d-3}}{1-n^{-1}}$. Using the fact that $\frac{1-n^{d-3}}{1-n^{-1}} \leq \frac{1}{1-n^{-1}}$ and letting $t = C\sqrt{n}$ for some $C > 0$, we get

$$P(\|\mathbf{U}_\tau - \Theta_\tau\|_2 > C\sqrt{n}) \ \leq \ C_1 \exp\left\{-C_2 \left(\frac{2n(n-1)C^2}{Mm(m-1)}\right)\right\}, \tag{A.5.3}$$

for large enough n as $\frac{m}{n} \longrightarrow \gamma$. The proof for (i) is completed by picking C large so that the right hand side of (A.5.3) is less than $1 - \beta$ as $\frac{m}{n} \longrightarrow \gamma \in (0, \infty)$.

(ii): Recall that $\mathbb{E}[T_\tau] = \|\Theta\|_2^2$, and the test $\phi(T_\tau)$ rejects H_0 when $T_\tau \geq \frac{4m}{9n} z_{1-\alpha}$. In what follows we let $\|\Theta\|_2 = C\sqrt{n}$ for an arbitrary fixed constant $C > 0$. By Chebyshev's inequality, for large enough n under the regime $\frac{m}{n} \longrightarrow \gamma$,

$$\begin{aligned} 1 - \mathbb{E}[\phi(T_\tau)] &= P\left(T_\tau - \|\Theta_\tau\|_2^2 \leq \frac{4m}{9n} z_{1-\alpha} - \|\Theta_\tau\|_2^2\right) \\ &\leq P\left(\left|T_\tau - \|\Theta_\tau\|_2^2\right| \geq \left|\frac{4m}{9n} z_{1-\alpha} - \|\Theta_\tau\|_2^2\right|\right) \leq \frac{\mathrm{Var}(T_\tau)}{(\frac{4m}{9n} z_{1-\alpha} - \|\Theta_\tau\|_2^2)^2}, \end{aligned} \tag{A.5.4}$$

where the first inequality is true when C is taken large enough. We will finish the proof by showing that as $\frac{m}{n} \longrightarrow \gamma$, the rightmost term of (A.5.4) is less than $1 - \beta$ when C is chosen large enough. To that end we will study the variance of the statistic T_τ. Note that

$$T_\tau = \frac{1}{\binom{n}{4}} \sum_{1 \leq i < j < k < l \leq n} h_\tau^T(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k, \mathbf{X}_l)$$

is a U-statistic with the kernel of degree 4

$$h_\tau^T(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k, \mathbf{X}_l) := \sum_{1 \le p < q \le m} h_\tau^W(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}, \mathbf{X}_k^{(pq)}, \mathbf{X}_l^{(pq)}),$$

where h_τ^W is the function h^W defined in (1.3.4) when h is the Kendall kernel h_τ. Here it is important to note that the kernel h_τ^T also depends on the number of variables m since it is a sum of $\binom{m}{2}$ terms. By Lemma 5.2.1A in Serfling (1980), the variance of T_τ satisfies

$$\mathrm{Var}(T_\tau) := \binom{n}{4}^{-1} \sum_{c=1}^{4} \binom{4}{c}\binom{n-4}{4-c} \zeta_c^\tau \le \frac{16\zeta_1^{h_\tau^T}}{n} + \frac{\widetilde{C}}{n^2}\left(\zeta_2^{h_\tau^T} + \zeta_3^{h_\tau^T} + \zeta_4^{h_\tau^T}\right) \tag{A.5.5}$$

for a constant $\widetilde{C} > 0$ that does not depend on C; recall definition (1.2.6) for the kernel $h = h_\tau^T$.

Claim. $\zeta_1^{h_\tau^T} \le C^2 nm(m-1)$

Proof of the claim. For seven distinct sample indices $i_1, \ldots, i_7 \in \{1, \ldots, n\}$,

$$\begin{aligned}
\zeta_1^{h_\tau^T} &= \mathbb{E}[h_\tau^T(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_4}) h_\tau^T(\mathbf{X}_{i_4}, \ldots, \mathbf{X}_{i_7})] - \|\Theta_\tau\|_2^4 \\
&= \sum_{\substack{1 \le p < q \le m \\ 1 \le p' < q' \le m}} \mathbb{E}[h_\tau^W(\mathbf{X}_{i_1}^{(pq)}, \ldots, \mathbf{X}_{i_4}^{(pq)}) h_\tau^W(\mathbf{X}_{i_4}^{(p'q')}, \ldots, \mathbf{X}_{i_7}^{(p'q')})] - \|\Theta_\tau\|_2^4 \\
&= \sum_{\substack{1 \le p < q \le m \\ 1 \le p' < q' \le m}} \theta_\tau^{(pq)} \theta_\tau^{(p'q')} \zeta_1^{h_\tau},
\end{aligned}$$

where the last equality is true by the defnition of h_τ^W and independence. Since $|h_\tau| \le 1$, it is true that $\zeta_1^{h_\tau} = |\zeta_1^{h_\tau}| \le 2$. This in turns implies that $\zeta_1^{h_\tau^T}$ is less than the quadratic form $2\Theta_\tau^T \mathbf{J}_{\binom{m}{2}} \Theta_\tau$, where $\mathbf{J}_{\binom{m}{2}}$ is the $\binom{m}{2}$-by-$\binom{m}{2}$ semi-positive definite matrix with all 1's. Since the largest eigenvalue of $\mathbf{J}_{\binom{m}{2}}$ is $\binom{m}{2}$, given that $\|\Theta_\tau\|_2 = C\sqrt{n}$,

$$2\Theta_\tau^T \mathbf{J}_{\binom{m}{2}} \Theta_\tau \le C^2 nm(m-1),$$

and the claim is proved. $\square$

Returning to the other quantities in (A.5.5), since $|h_\tau^W| \leq 1$, it is easy to show that each of $\zeta_2^{h_\tau^T}$, $\zeta_3^{h_\tau^T}$ and $\zeta_4^{h_\tau^T}$ is bounded by $2\binom{m}{2}^2$. Hence, under the regime $\frac{m}{n} \longrightarrow \gamma$, together with the claim above, (A.5.5) gives that for all large n,

$$\mathrm{Var}(T_\tau) \leq m^2(16C^2 + 3\gamma^2\widetilde{C}). \tag{A.5.6}$$

Recalling that $\|\Theta_\tau\|^2 = C\sqrt{n}$, and applying (A.5.6) to (A.5.4), we get that

$$1 - \mathbb{E}[\phi(T_\tau)] \leq \frac{m^2(16C^2 + 3\gamma^2\widetilde{C})}{C^4n^2 - C^2\frac{8}{9}mz_{1-\alpha} + \frac{16m^2}{81n^2}z_{1-\alpha}^2} \tag{A.5.7}$$

for all large n. Since C is arbitrary, by choosing it large enough the right hand side of (A.5.9) can be made less than $1-\beta$ as $\frac{m}{n} \longrightarrow \gamma$. $\square$

The following lemma is needed for the proof of Theorem 1.5.2.

Lemma A.5.1. *Let $I = [0, 1-\epsilon] \subset \mathbb{R}$ for some small fixed $\epsilon > 0$. For fixed positive integers $c_1, \ldots, c_b$ such that $\sum_{i=1}^b c_i = c$, suppose $\mathbf{X} = (X^{(1)}, \ldots, X^{(c)})' \sim N(0, \Sigma)$ is a c-variate normal random vector with an invertible block diagonal covariance matrix*

$$\Sigma = \Sigma(\rho) = \begin{bmatrix} B_1(\rho) & & \\ & \ddots & \\ & & B_b(\rho) \end{bmatrix},$$

where each $B_i(\rho)$ is a c_i-by-c_i matrix with 1's on the diagonal and all off-diagonal entries equal to some $\rho \in I$. If $H : \mathbb{R}^c \longrightarrow \mathbb{R}$ is a bounded function such that $\mathbb{E}[H(\mathbf{X})] = 0$ when $\rho = 0$, then there exists a constant $C = C(H, \epsilon) > 0$ such that $|\mathbb{E}[H(\mathbf{X})]| \leq C\rho$ for all $\rho \in I$.

Proof. For all $\rho \in I$, the matrix $\Sigma(\rho)$ is invertible and the precision matrix $\Sigma^{-1}(\rho)$ is a smooth function of ρ. Hence, the set of distributions $N(0, \Sigma(\rho))$ forms a curved exponential family. By standard results on exponential families (Lehmann and Casella, 1998, Theorem 5.8), the expectation $\mathbb{E}[H(\mathbf{X})]$ is a continuous function of ρ that is differentiable on $(0, 1-\epsilon)$. The lemma is thus implied by the mean value theorem and the compactness of $[0, 1-\epsilon]$. $\square$

Proof of Theorem 1.5.2. The value of T_τ depends only on the rank vectors $\mathbf{R}^{(1)},\ldots,\mathbf{R}^{(m)}$. Without loss of generality, we may thus assume that each $X^{(p)}$ is centered with unit variance, i.e., $(X^{(1)},\ldots,X^{(m)})' \sim N(0,R)$, where $R=(\rho^{(pq)})$ is a correlation matrix, with 1's on the diagonal.

It suffices to prove the result under the restriction that θ can only take values in a closed interval $[0,1-\epsilon]$, for some fixed small $\epsilon>0$. In other words, in the statement of the theorem, replace the set of distributions $\mathcal{N}_m(\|\Theta_\tau\|_2 \geq \tilde{C}; \theta_\tau^{(pq)}=\theta)$ under the infimum by the subset

$$\{N \in \mathcal{N}_m(\|\Theta_\tau\|_2 \geq \tilde{C}; \theta_\tau^{(pq)}=\theta) : \theta \in [0,1-\epsilon]\}. \tag{A.5.8}$$

To see that this restriction can be made, note that $\theta > 1-\epsilon$ implies that $\|\Theta_\tau\|_2 > \sqrt{\binom{m}{2}}(1-\epsilon) = O(m)$. Since $O(m) > O(\sqrt{n})$ asymptotically under the regime $\frac{m}{n} \longrightarrow \gamma$, by Theorem 1.5.1$(ii)$, nothing is lost by ignoring the normal distributions in $\mathcal{N}_m(\|\Theta_\tau\|_2 \geq \tilde{C}; \theta_\tau^{(pq)}=\theta)$ with $\theta > 1-\epsilon$. In addition, for all $p \neq q$, by we have the classical result (Kruskal, 1958, p.823),

$$\rho^{(pq)} = \rho = \sin\left(\frac{\pi\theta}{2}\right)$$

when $\theta_\tau^{(pq)}=\theta$. As a consequence, for the covariance matrix R to be positive definite it must be that $\theta > -\frac{2}{\pi}\arcsin[\frac{1}{m-1}]$ (Horn and Johnson, 2013, Theorem 7.2.5). Hence, as n and m grow, it can be seen that $\|\Theta_\tau\|_2 < 1/\sqrt{2}$ when θ lies in the interval $(-\frac{2}{\pi}\arcsin[\frac{1}{m-1}],0)$. As such, by taking the constant $\tilde{C}$ to be larger than $1/\sqrt{2}$ when necessary, it suffices to consider the subset of distributions (A.5.8) under the infimum.

In what follows, the operators $\mathbb{E}[\cdot]$, $\mathrm{Var}[\cdot]$ and $\mathrm{Cov}[\cdot]$ are all with respect to an m-variate normal distribution for $(X^{(1)},\ldots,X^{(m)})'$ in (A.5.8). Recall from (A.5.5) that

$$\mathrm{Var}(T_\tau) := \binom{n}{4}^{-1}\sum_{c=1}^{4}\binom{4}{c}\binom{n-4}{4-c}\zeta_c^{h_\tau^T}.$$

Our proof now begins with the Chebyshev's inequality from (A.5.4):

$$1-\mathbb{E}[\phi_\alpha(T_\tau)] \leq \frac{\mathrm{Var}(T_\tau)}{(\frac{4m}{9n}z_{1-\alpha} - \|\Theta_\tau\|_2^2)^2} \leq \frac{\zeta_c^{h_\tau^T} B \sum_{c=1}^{4} n^{-c}}{(\frac{4m}{9n}z_{1-\alpha})^2 - \frac{8m}{9n}z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4}, \tag{A.5.9}$$

where the last inequality is true since $\binom{n}{4}^{-1}\binom{4}{c}\binom{n-4}{4-c} \leq Bn^{-c}$ for a constant $B > 0$. To finish the proof, it suffices to show that for each $c = 1, \ldots, 4$, a constant $\tilde{C}_c(\alpha, \beta, \gamma) > 0$ exists such that for large enough n (depending on $\tilde{C}_c$),

$$\frac{B\zeta_c^{h_\tau^T} n^{-c}}{(\frac{4m}{9n} z_{1-\alpha})^2 - \frac{8m}{9n} z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4} < \frac{1-\beta}{4} \tag{A.5.10}$$

whenever $\|\Theta_\tau\|_2 > \tilde{C}_c$. We may then take $\tilde{C} = \max_{c=1,\ldots,4} \tilde{C}_c$.

For notational convenience, we define

$$f_{\mathbf{i},\mathbf{j}} := \sum_{\substack{1\leq p<q\leq m \\ 1\leq p'<q'\leq m}} \mathbb{E}[h_\tau^W(\mathbf{X}_{i_1}^{(pq)}, \ldots, \mathbf{X}_{i_4}^{(pq)}) h_\tau^W(\mathbf{X}_{j_1}^{(p'q')}, \ldots, \mathbf{X}_{j_4}^{(p'q')})] \geq 0,$$

for any tuples $\mathbf{i} = (i_1, \ldots, i_4)$, $\mathbf{j} = (j_1, \ldots, j_4) \in \mathcal{P}(n, 4)$ such that $|\mathbf{i} \cap \mathbf{j}| = c$. Then

$$\zeta_c^{h_\tau^T} = f_{\mathbf{i},\mathbf{j}} - \|\Theta_\tau\|_2^4. \tag{A.5.11}$$

Since the ratio

$$\frac{B\|\Theta_\tau\|_2^4}{(\frac{4m}{9n} z_{1-\alpha})^2 - \frac{8m}{9n} z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4}$$

is bounded for all values of $\|\Theta_\tau\|_2$, for each $c = 1, \ldots, 4$,

$$\frac{B\|\Theta_\tau\|_2^4 n^{-c}}{(\frac{4m}{9n} z_{1-\alpha})^2 - \frac{8m}{9n} z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4} \longrightarrow 0$$

as $\frac{m}{n} \longrightarrow \gamma$. Upon substituting (A.5.11) into (A.5.10), we see that the proof is finished if the below claim is shown to be true. $\square$

Claim. Under $\theta_\tau^{(pq)} = \theta$, there exists for each $c = 1, \ldots, 4$, a constant $\tilde{C}_c(\alpha, \beta, \gamma) > 0$ such that for large enough n (depending on $\tilde{C}_c$),

$$\frac{B f_{\mathbf{i},\mathbf{j}} n^{-c}}{(\frac{4m}{9n} z_{1-\alpha})^2 - \frac{8m}{9n} z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4} < \frac{1-\beta}{5}. \tag{A.5.12}$$

whenever $\|\Theta_\tau\|_2 = \theta\sqrt{\binom{m}{2}} > \tilde{C}_c$.

Proof of the claim when $c = 1$. Using independence, we find that for any four distinct indices $1 \leq i, j, k \leq n$,

$$f_{\mathbf{i},\mathbf{j}} = \underbrace{\sum_{\substack{1 \leq p < q \leq m \\ 1 \leq p' < q' \leq m \\ |\{p,q\} \cap \{p',q'\}| \geq 1}} \theta^2 \mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}) h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})]}_{(1)} + \underbrace{\sum_{\substack{1 \leq p < q \leq m \\ 1 \leq p' < q' \leq m \\ |\{p,q\} \cap \{p',q'\}| = 0}} \theta^2 \mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}) h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})]}_{(2)} . \quad \text{(A.5.13)}$$

Since $|h_\tau| \leq 1$, the term (1) is bounded in absolute value by $[\binom{m}{2}^2 - \binom{m}{2}\binom{m-2}{2}]\theta^2 = O(m)\|\Theta_\tau\|_2^2$. To bound (2), note that when $|\{p, q\} \cap \{p', q'\}| = 0$, the expectation term

$$\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}) h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})] \quad \text{(A.5.14)}$$

equals 0 when $\theta = 0$ due to the independence of $\{\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}\}$ and $\{\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')}\}$. Moreover, for $\theta \neq 0$, the pairs $\{\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)}, \mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')}\}$ jointly follow a 8-variate normal distribution with block diagonal covariance matrix, where each block has 1's on the diagonal and all its off-diagonal entries equal to $\rho = \sin[\pi\theta/2]$. By Lemma A.5.1, the expectation (A.5.14) is bounded in absolute value, up to a multiplying constant, by θ, and hence (2) bounded by $O(m^4)\theta^3 = O(m)\|\Theta_\tau\|_2^3$ in absolute value.

Using the above bounds for (1) and (2) we get that the left hand side of (A.5.12) is less than

$$\frac{\frac{O(m)}{n}(\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^3)}{(\frac{4m}{9n} z_{1-\alpha})^2 - \frac{8m}{9n} z_{1-\alpha} \|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4}.$$

Under the regime $\frac{m}{n} \longrightarrow \gamma$, we see that the expression in the above display can be made less than $\frac{1-\beta}{5}$ when $\|\Theta_\tau\|_2$ and n are large enough. $\square$

Proof of the claim when $c = 2$. Again, using independence, we find that

$$9f_{\mathbf{i},\mathbf{j}} = \underbrace{\sum 4\left(\mathbb{E}\left[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})\right]\right)^2}_{(1)} +$$
$$\underbrace{\sum \theta^2\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_j^{(p'q')})]}_{(2)} +$$
$$\underbrace{\sum 2\theta\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_k^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_j^{(p'q')})h_\tau(\mathbf{X}_j^{(pq)}, \mathbf{X}_l^{(pq)})]}_{(3)} +$$
$$\underbrace{\sum 2\theta\mathbb{E}[h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_j^{(p'q')}, \mathbf{X}_l^{(p'q')})]}_{(4)}, \quad \text{(A.5.15)}$$

where each summation is over all pairs $1 \leq p < q \leq m$ and $1 \leq p' < q' \leq m$, and i, j, k, l are any 4 distinct indices in $\{1, \dots, n\}$. We now derive bounds for the absolute values of the terms $(1), (2), (3), (4)$.

Term (1)*:* We claim that $|(1)| \leq O(m^2)(1 + \|\Theta_\tau\|_2^2)$. To show this, observe that (1) equals

$$\sum_{1\leq p<q\leq m} 4(\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_k^{(pq)})])^2 +$$
$$\sum_{\substack{|\{p,q\}\cap\{p',q'\}|=0 \\ 1\leq p<q\leq m \\ 1\leq p'<q'\leq m}} 4(\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})])^2. \quad \text{(A.5.16)}$$

Since $|h_\tau| \leq 1$, the first sum in (A.5.16) is bounded by a term of order $O(m^2)$. Considering the second sum, an expectation

$$\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_j^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_k^{(p'q')})] \quad \text{(A.5.17)}$$

with $\{p, q\} \neq \{p', q'\}$ equals 0 when $\theta = 0$ by independence. Moreover, $\mathbf{X}_i^{(pq)}$, $\mathbf{X}_j^{(pq)}$, $\mathbf{X}_i^{(p'q')}$, and $\mathbf{X}_k^{(p'q')}$ jointly follow an 8-variate normal distribution with block diagonal covariance matrix as in Lemma A.5.1. By that lemma and the fact that $\rho = \sin[\pi\theta/2]$, we obtain that (A.5.17) is bounded in absolute value by θ times a constant, hence the second sum in (A.5.16) is bounded in absolute value by a term equal to $O(m^2)\|\Theta_\tau\|_2^2$. Gathering the bounds for the two sums in (A.5.16) gives the claimed bound for the absolute value of term (1).

Term (2)*:* We claim that $|(2)| \leq O(m^2)\|\Theta_\tau\|_2^2$. Indeed, since $|h_\tau| \leq 1$, it is easy show that (2) is bounded in absolute value by $\binom{m}{2}^2\theta^2 = \binom{m}{2}\|\Theta_\tau\|_2^2 = O(m^2)\|\Theta_\tau\|_2^2$.

Terms (3) *and* (4)*:* We claim that $|(3)|, |(4)| \leq O(m^2)(\|\Theta_\tau\|_2 + \|\Theta_\tau\|_2^2)$. We give details for the proof of bound for $|(3)|$. The bound for (4) is analogous. We write (3) as

$$\sum_{\substack{|\{p,q\}\cap\{p',q'\}|\geq 1\\ 1\leq p<q\leq m\\ 1\leq p'<q'\leq m}} 2\theta\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_k^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_j^{(p'q')})h_\tau(\mathbf{X}_j^{(pq)}, \mathbf{X}_l^{(pq)})]+$$

$$\sum_{\substack{|\{p,q\}\cap\{p',q'\}|=0\\ 1\leq p<q\leq m\\ 1\leq p'<q'\leq m}} 2\theta\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_k^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_j^{(p'q')})h_\tau(\mathbf{X}_j^{(pq)}, \mathbf{X}_l^{(pq)})], \quad \text{(A.5.18)}$$

where the first sum is bounded by $2\theta\binom{m}{2}[\binom{m}{2} - \binom{m-2}{2}] = O(m^2)\|\Theta_\tau\|_2$ because $|h_\tau| \leq 1$. The expectation

$$\mathbb{E}[h_\tau(\mathbf{X}_i^{(pq)}, \mathbf{X}_k^{(pq)})h_\tau(\mathbf{X}_i^{(p'q')}, \mathbf{X}_j^{(p'q')})h_\tau(\mathbf{X}_j^{(pq)}, \mathbf{X}_l^{(pq)})]$$

equals 0 when $\{p,q\} \cap \{p',q'\}| = 0$, and Lemma A.5.1 can be invoked to show the second sum in (A.5.18) is bounded in absolute value by $O(m^2)\|\Theta_\tau\|_2^2$.

Having established the bounds for the terms $(1) - (4)$ in (A.5.15), we find that when $c = 2$ the left hand side of (A.5.12) is less than

$$\frac{O(m^2)n^{-2}(1 + \|\Theta_\tau\|_2 + \|\Theta_\tau\|_2^2)}{(\frac{4m}{9n}z_{1-\alpha})^2 - \frac{8m}{9n}z_{1-\alpha}\|\Theta_\tau\|_2^2 + \|\Theta_\tau\|_2^4},$$

which, under $\frac{m}{n} \longrightarrow \gamma$, can be made to be less than $\frac{1-\beta}{5}$ when $\|\Theta_\tau\|_2$ and n are large enough. □

Proof of the claim when $c \geq 3$. For $c = 3$ or $c = 4$, we may proceed similarly, using again the boundedness of h_τ and Lemma A.5.1. We note that if $c = 3$, then $|f_{\mathbf{i},\mathbf{j}}| \leq O(m^3)(1 + \|\Theta_\tau\|_2)$ and omit further details. □

Appendix B
PROOFS FOR CHAPTER 2

B.1 Proofs for Section 2.1

Proof of Lemma 2.1.1. We may assume $d \geq n$, otherwise J_f is never of full column rank. The implication $(i) \Rightarrow (ii)$ is obvious.

To show $(ii) \Rightarrow (iii)$, suppose for contradiction that J_f is not generically of full rank. Since f is polynomial, we then know that $\mathrm{Rank}(J_f) = r < n$ generically, that is, outside a proper algebraic subset $S' \subset \mathbb{R}^n$ the rank is constant r. By the rank theorem (Rudin, 1976, p. 229), for every point $s \in S \setminus (S' \cup \tilde{S})$, we can choose an open ball $\mathcal{B}(s)$ that contains s, is a subset of $S \setminus (S' \cup \tilde{S})$ and for which the restricted map $f|_{\mathcal{B}_s}$ has fibers of dimension $n - r > 0$, contradicting (ii).

It remains to show $(iii) \Rightarrow (i)$. We observe that since f is a polynomial we can assume $S = \mathbb{R}^n$. We then show that the set of points with an infinite fiber, denoted

$$\mathbb{F}_f := \{s \in \mathbb{R}^n : |\mathcal{F}_f(s)| = \infty\},$$

is contained in a proper algebraic subset of $\mathbb{R}^n$. We note that it suffices to assume $n = d$, for without loss of generality, we can permute the d component functions of f and assume that $\pi \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ has a generically full rank Jacobian matrix, where π is the projection onto the first n coordinates. Then $\mathbb{F}_f \subset \mathbb{F}_{\pi \circ f}$.

Now, assume $d = n$, and let $C = \{s \in \mathbb{R}^n : \det J_f(s) = 0\}$ be the set of critical points of f, where J_f is the Jacobian matrix of f. Note that by assumption C is a proper algebraic subset of $\mathbb{R}^n$.

Claim. If $y \in \mathbb{R}^n$ is a point such that $|\mathcal{F}_f(y)| = \infty$, then $\mathcal{F}_f(y) \cap C \neq \emptyset$.

Proof of the Claim. If an algebraic set like $\mathcal{F}_f(y)$ is infinite, then it has dimension $k > 0$. By semialgebraic stratification (Basu et al., 2006), one can see that there exists an open set $U \subset \mathbb{R}^k$ and a differentiable map $g : U \longrightarrow \mathcal{F}_f(y)$ such that the Jacobian of g has full rank on U. If $\mathcal{F}_f(y) \cap C = \emptyset$, then the chain rule yields that the composition $f \circ g : U \to \{y\}$ has Jacobian of positive rank. This, however, is a contradiction because $f \circ g$ is a constant function. Hence, $\mathcal{F}_f(y) \cap C \neq \emptyset$. $\square$

The claim implies that $\mathbb{F}_f \subset f^{-1}(f(C)) \subset f^{-1}(\overline{f(C)})$, where $\overline{f(C)}$ is the Zariski closure of the semialgebraic set $f(C)$. Since $\overline{f(C)}$ is algebraic, so is $f^{-1}(\overline{f(C)})$ given that f is a polynomial. To finish the proof we only need to show that $f^{-1}(\overline{f(C)})$ has dimension less than n, which is equivalent to $f^{-1}(\overline{f(C)}) \neq \mathbb{R}^n$. By Sard's theorem (Basu et al., 2006, p. 192), $f(C)$, and thus also $\overline{f(C)}$, has dimension less than n. If $f^{-1}(\overline{f(C)}) = \mathbb{R}^n$, then the inverse function theorem, which says that the restricted map $f|_{\mathbb{R}^n \setminus C}$ is a local diffeomorphism, is contradicted. $\square$

Proof of Theorem 2.1.5. Let $m = |V|$. For $i = 1, 2$, let $\overline{G}_i = (\overline{V}, \overline{E}_i)$ be the extended DAG of G_i, i.e., $\overline{V} = \{0, 1, \ldots, m\}$, and $\overline{E}_i = E_i \cup \{0 \to v : v \in \{1, \ldots, m\}\}$. By the well-known characterization that two DAGs are Markov equivalent if and only if they have the same skeleton and v-structures (Pearl, 2009), it is easy to see that $\overline{G}_1$ and $\overline{G}_2$ are also Markov equivalent.

For $i \in \{1, 2\}$, let $\Theta_i := \mathbb{R}_{E_i} \times \mathrm{diag}_m^+ \times \mathbb{R}^m$. Define

$$\Phi_{\overline{G}_i}\big((\Lambda, \Omega, \delta)\big) = (I_{m+1} - \overline{\Lambda}^T)^{-1} \overline{\Omega} (I_{m+1} - \overline{\Lambda})^{-1},$$

where $\Theta_i := \mathbb{R}_{E_i} \times \mathrm{diag}_m^+ \times \mathbb{R}^m$, $\overline{\Lambda}$ is a $(m+1) \times (m+1)$ matrix such that

$$\overline{\Lambda}_{vw} = \begin{cases} \delta_w & \text{if } v = 0, w = 1, \ldots, m, \\ \Lambda_{vw} & \text{if } v, w = 1, \ldots, m, \\ 0 & \text{otherwise,} \end{cases}$$

and $\overline{\Omega}$ is a diagonal matrix with $\overline{\Omega}_{00} = 1$ and $\overline{\Omega}_{vv} = \Omega_{vv}$ for $v = 1, \ldots m$. Then the image $\Phi_{\overline{G}_i}(\Theta_i)$ is the set of all covariance matrices of $(m+1)$-variate Gaussian distributions that obey the global Markov property of $\overline{G}_i$ and have the variance of node 0, which represents the latent variable L, equal to 1. Consider the projection

$$\pi(\Sigma) = \Sigma_{\{1,\dots,m\},\{1,\dots,m\}},$$

where Σ has its rows and columns indexed by $\{0,\dots,m\}$. Then the parametrization map for the latent variable model $\mathcal{N}_*(G_i)$ equals

$$\phi_{G_i} = \pi \circ \Phi_{\overline{G}_i}. \tag{B.1.1}$$

Since $\overline{G}_1$ and $\overline{G}_2$ are Markov equivalent, $\Phi_{\overline{G}_1}(\Theta_1) = \Phi_{\overline{G}_2}(\Theta_2)$. By Lemma 2.2.1, each map $\Phi_{\overline{G}_i}$ is injective on Θ_i with rational inverse defined on the common image $\Phi_{\overline{G}_1}(\Theta_1) = \Phi_{\overline{G}_2}(\Theta_2)$. From (B.1.1), we obtain that

$$\phi_{G_1} = \pi \circ \Phi_{\overline{G}_1} = \pi \circ \Phi_{\overline{G}_2} \circ \Phi^{-1}_{\overline{G}_2} \circ \Phi_{\overline{G}_1} = \phi_{G_2} \circ \left(\Phi^{-1}_{\overline{G}_2} \circ \Phi_{\overline{G}_1}\right).$$

Since $\Phi^{-1}_{\overline{G}_2} \circ \Phi_{\overline{G}_1} : \Theta_1 \longrightarrow \Theta_2$ is a diffeomorphism, the chain rule implies that the Jacobian of ϕ_{G_1} can be of full column rank if and only if the same is true for ϕ_{G_2}. Since ϕ_{G_i} are polynomial, the two Jacobians either both have generically full rank or are both everywhere rank deficient. By Lemma 2.1.1, ϕ_{G_1} is generically finite-to-one if and only if ϕ_{G_2} is so. $\square$

B.2 Proofs for Section 2.3

Proof of Lemma 2.3.3. We first give the structure of $J(\tilde{\varphi}_G)$ block by block.

(a) "$[J(\tilde{\varphi}_G)]_{D,\{\Psi,\Lambda,\gamma\}}$": For a given pair $(v,v) \in D$,

$$[\tilde{\varphi}_G(\Lambda,\Psi,\gamma)]_{vv} = \psi_v + \left(\sum_{w:v\to w\in E} \psi_w \lambda_{vw}^2\right) - \gamma_v^2.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{(v,v),\psi_w} = \begin{cases} 1 & \text{if } v = w, \\ \lambda_{vw}^2 & \text{if } v \to w \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{B.2.1}$$

$$[J(\tilde{\varphi}_G)]_{(v,v),\lambda_{wu}} = \begin{cases} 2\lambda_{wu}\psi_u & \text{if } v = w, \\ 0 & \text{otherwise}, \end{cases} \tag{B.2.2}$$

and

$$[J(\tilde{\varphi}_G)]_{(v,v),\gamma_u} = \begin{cases} -2\gamma_u & \text{if } v = u, \\ 0 & \text{otherwise}. \end{cases} \tag{B.2.3}$$

(b) "$[J(\tilde{\varphi}_G)]_{E,\{\Psi,\Lambda,\gamma\}}$": For any $v \to w \in E$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = -\lambda_{vw}\psi_w + \left(\sum_{u: \substack{v \to u \in E \\ w \to u \in E}} \lambda_{vu}\lambda_{wu}\psi_u \right) - \gamma_v\gamma_w.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v\to w,\psi_u} = \begin{cases} -\lambda_{vw} & \text{if } u = w, \\ \lambda_{vu}\lambda_{wu} & \text{if } v \to u \in E \text{ and } w \to u \in E, \\ 0 & \text{otherwise}, \end{cases} \tag{B.2.4}$$

$$[J(\tilde{\varphi}_G)]_{v\to w,\lambda_{ux}} = \begin{cases} -\psi_w & \text{if } v = u, w = x, \\ \lambda_{wx}\psi_x & \text{if } u = v,\ u \to x \in E \text{ and } w \to x \in E, \\ \lambda_{vx}\psi_x & \text{if } u = w,\ u \to x \in E \text{ and } v \to x \in E, \\ 0 & \text{otherwise}, \end{cases} \tag{B.2.5}$$

and

$$[J(\tilde{\varphi}_G)]_{v\to w,\gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u, \\ -\gamma_v & \text{if } w = u, \\ 0 & \text{otherwise}. \end{cases} \tag{B.2.6}$$

(c) "$[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$": For any $v \not\to w \in N$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = \left(\sum_{u: \substack{v \to u \in E \\ w \to u \in E}} \lambda_{vu}\lambda_{wu}\psi_u \right) - \gamma_v\gamma_w. \tag{B.2.7}$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v\not\leftrightarrow w,\psi_u} = \begin{cases} \lambda_{vu}\lambda_{wu} & \text{if } v \to u \in E \text{ and } w \to u \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{B.2.8}$$

$$[J(\tilde{\varphi}_G)]_{v\not\leftrightarrow w,\lambda_{ux}} = \begin{cases} \lambda_{wx}\psi_x & \text{if } u = v,\ u \to x \in E \text{ and } w \to x \in E, \\ \lambda_{vx}\psi_x & \text{if } u = w,\ u \to x \in E \text{ and } v \to x \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{B.2.9}$$

and

$$[J(\tilde{\varphi}_G)]_{v\not\leftrightarrow w,\gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u, \\ -\gamma_v & \text{if } w = u, \\ 0 & \text{otherwise.} \end{cases} \tag{B.2.10}$$

With slight abuse of notation, let $|\Psi|$, $|\gamma|$, $|\Lambda|$ denote the number of free variables in Ψ, γ and Λ respectively. Considering that $|D| = |\Psi|$ and $|E| = |\Lambda|$, we must have that $|N| \geq |\gamma|$ since $J(\tilde{\varphi}_G)$ is a tall matrix. Hence, if $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank, then there exists a subset $N' \subset N$ such that $|N'| = |\gamma|$ and the determinant of $J(\tilde{\varphi}_G)_{N',\gamma}$ is a nonzero polynomial in the variables of γ, in consideration of (B.2.10). Now it suffices to show that the $(2m+|E|) \times (2m+|E|)$ square submatrix $[J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}$ is generically of full rank.

Since the matrix concerned has polynomial entries, we need to show that the determinant of $[J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}$ is a nonzero polynomial. To this end, it is sufficient to show that the determinant is a nonzero polynomial in the entries of (Λ,γ) when we specialize $\psi_1 = \cdots = \psi_m = 1$. Noting that $|\Psi| + |\Lambda| + |\gamma| = |D| + |E| + |N'|$, let P denote the set of all permutation functions mapping from the set $D \cup E \cup N'$ to the set of free variables in Λ, Ψ and γ. Choose any ordering of the elements of domain and codomain so as to have a well-defined sign for the permutations. Then by Leibniz's formula, we have

$$\det\left([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\right) = \sum_{\sigma\in P} \operatorname{sgn}(\sigma) \prod_{s\in D\cup E\cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)}.$$

Let $\tilde{P}$ be the subset of all permutations $\sigma \in P$ with $\sigma((v,v)) = \psi_v$ for all $(v,v) \in D$ and $\sigma((v,w)) = \lambda_{vw}$ for all $(v,w) \in E$. Then we obtain that

$$\begin{aligned}
&\det\big([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\big) \\
&= \sum_{\sigma\in\tilde{P}} \operatorname{sgn}(\sigma) \prod_{s\in D\cup E\cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)} + \sum_{\sigma\in P\setminus\tilde{P}} \operatorname{sgn}(\sigma) \prod_{s\in D\cup E\cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)} \\
&= \pm\det\big(J(\tilde{\varphi}_G)_{N',\gamma}\big) + \sum_{\sigma\in P\setminus\tilde{P}} \operatorname{sgn}(\sigma) \prod_{s\in D\cup E\cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)}, \qquad \text{(B.2.11)}
\end{aligned}$$

where the equality in (B.2.11) follows from (B.2.1), (B.2.5) and the fact that $\psi_1 = \cdots = \psi_m = 1$. We also deduce from (B.2.1)-(B.2.10) that every summand in the second term of (B.2.11) is either zero or a polynomial term involving free variables of Λ. In contrast, $\det\big(J(\tilde{\varphi}_G)_{N',\gamma}\big)$ is a nonzero polynomial only in free variables of γ and can thus not be canceled by the second term in (B.2.11). □

B.3 Proofs for Section 2.5

Proof for Lemma 2.5.4. We first prove (i). Since $\mathcal{N}_*(G)$ is generically finitely identifiable by Theorem 2.1.3, there exists an algebraic subset Ξ' such that for all $\theta \in \Theta \setminus \Xi'$, $|\mathcal{F}_{\phi_G}(\theta)| < \infty$. Define Ξ to be the union of Ξ' and the set of triples $(\Lambda, \Omega, \delta) \in \mathbb{R}^{2m+|E|}$ with at least one coordinate $\delta_i = 0$. Let $\Sigma_0 = \phi_G(\Lambda_0, \Omega_0, \delta_0)$ and

$$S = (s_{ij}) := (I_m - \Lambda^T)\Sigma_0(I_m - \Lambda). \qquad \text{(B.3.1)}$$

Then for $1 \le i < j \le m$,

$$\begin{aligned}
s_{ij} &= \sum_{1\le k,k'\le m} \lambda_{ki}[\Sigma_0]_{kk'}\lambda_{k'j} - \sum_{1\le k\le m} [\Sigma_0]_{ik}\lambda_{kj} - \sum_{1\le k\le m} \lambda_{ki}[\Sigma_0]_{kj} + [\Sigma_0]_{ij} \\
&= \begin{cases} -\displaystyle\sum_{(k,m)\in E} [\Sigma_0]_{ik}\lambda_{km} + [\Sigma_0]_{im} & \text{if } j = m, \\ [\Sigma_0]_{ij} & \text{if } j < m, \end{cases}
\end{aligned}$$

where the last equality follows from the fact that λ_{ij} are nonzero only when $(i,j) \in E$. Hence, for any four indices $1 \le i < j < k < l \le m$, the tetrads

$$s_{ij}s_{kl} - s_{ik}s_{jl}, \qquad\qquad s_{il}s_{jk} - s_{ik}s_{jl}$$

are constant polynomials when $l < m$ and have degree 1 in the variables $\{\lambda_{vm} : (v,m) \in E\}$ when $l = m$. The equation system

$$\text{TETRADS}(S) = 0$$

is a thus a consistent linear system that can be represented as

$$C\lambda_{pa(m),m} = c, \tag{B.3.2}$$

where $\lambda_{pa(m),m} = (\lambda_{vm})^T_{v\in pa(m)}$ is the vector of all free Λ variables, C is a $2\binom{m-1}{3} \times |pa(m)|$ matrix and c is a $2\binom{m-1}{3}$-vector. Both C and c depend only on Σ_0.

To finish the proof, we now need to show that (B.3.2) is uniquely solvable in $\lambda_{pa(m),m}$. We will aim to contradict $|\mathcal{F}_{\phi_G}(\theta_0)| < \infty$ if (B.3.2) does not have a unique solution. Note that the solution set is an affine subspace $\mathcal{L} \subset \mathbb{R}^{|E|}$. For a contradiction, suppose that $\mathcal{L}$ is of positive dimension. Upon substituting $\Lambda = \Lambda_0$ into (B.3.1), we obtain

$$S_0 = (s^0_{ij}) \;=\; (I_m - \Lambda_0^T)\Sigma_0(I_m - \Lambda_0),$$

and in consideration of (2.5.3) in Theorem 2.5.2, it must be true that

$$s^0_{ii}s^0_{jk} - s^0_{ik}s^0_{ji} > 0, \;\text{ for all } i \ne j \ne k.$$

We may then pick an open ball $\mathcal{B}(\Lambda_0)$ such that for all solutions $\Lambda \in \mathcal{L} \cap \mathcal{B}(\Lambda_0)$, the matrix $S = (s_{ij})$ defined by (B.3.1) satisfies

$$s_{ii}s_{jk} - s_{ik}s_{ji} > 0 \text{ , for all } i \ne j \ne k.$$

It follows that $\mathcal{L} \cap \mathcal{B}(\Lambda_0)$ is an infinite set whose elements Λ all make the matrix $(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda)$ a Spearman matrix. Hence, the system

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta'$$

has infinitely many solutions, contradicting $|\mathcal{F}_{\phi_G}(\theta_0)| < \infty$.

The proof of (ii) is analogous. We first let $\Upsilon_0 = \varphi_G(\Lambda_0, \Psi_0, \gamma_0)$ and define

$$\tilde{S} = (\tilde{s}_{ij}) \; = \; (I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1}. \tag{B.3.3}$$

Noting that in this case $(I_m - \Lambda)^{-1} = I_m + \Lambda$, it can be easily seen that

$$\text{TETRADS}(\tilde{S}) = \text{TETRADS}\big((I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1}\big) = 0$$

is a linear system in the variables $\{\lambda_{1v} : v \in ch(1)\}$. Similar to the above arguments, we may use Theorem 2.1.3 and Theorem 2.5.3 to prove by contradiction that the system can only have a unique solution in $\{\lambda_{1v} : v \in ch(1)\}$. □

B.4 Proofs for Section 2.6

Proof of Theorems 2.6.1 and 2.6.2. For Theorem 2.6.1, one can partition the Jacobian matrix $J(\tilde{\varphi}_G)$ of $\tilde{\varphi}_G$ as in (2.3.7), only with γ replaced by $\gamma_{V\setminus V'} = \{\gamma_v : v \in V \setminus V'\}$. In analogy with Lemma 2.3.3, it can be shown that $J_{\tilde{\varphi}_G}$ is of full column rank if $[J(\tilde{\varphi}_G)]_{N,\gamma_{V\setminus V'}}$ is. The reasoning is then analogous to that in the proof of Theorem 2.1.3, the main step being the application of Lemma 2.3.2 where the graph defining the considered map becomes $(G^c)_{V\setminus V'}$.

The proof of Theorem 2.6.2 is analogous to the proof of Theorem 2.1.4. The only change is to replace G^c_{con}, $G^c_{|L,cov}$, γ and δ by $\widetilde{G}^c_{con}$, $\widetilde{G}^c_{|L,cov}$, $\gamma_{V\setminus V'}$ and $\delta_{V\setminus V'}$, respectively. □

VITA

Dennis Leung grew up in Hong Kong. After some detours he found his love for mathematics, and have been hooked up ever since. He spent part of his 20's on the beautiful shore of La Jolla. He obtained his PhD in statistics in 2016 at the University of Washington. In his spare time, he likes to run and work out to stay healthy.