Assessing partial association between ordinal variables: quantification, visualization, and hypothesis testing

Dungang Liu[1], Shaobo Li, Yan Yu and Irini Moustaki

May 6, 2020

[1]Dungang Liu is the corresponding author and Assistant Professor (Email: dungang.liu@uc.edu), Department of Operations, Business Analytics and Information Systems, University of Cincinnati Lindner College of Business, Cincinnati, OH 45221. Shaobo Li is Assistant Professor, University of Kansas School of Business, Lawrence, KS 66045. Yan Yu is Joseph S. Stern Professor, Department of Operations, Business Analytics and Information Systems, University of Cincinnati Lindner College of Business, Cincinnati, OH 45221. Irini Moustaki is Professor, Department of Statistics, London School of Economics, United Kingdom. The authors thank Editor Hongyu Zhao, an anonymous Associate Editor and two referees for constructive comments.

Assessing partial association between ordinal variables: quantification, visualization, and hypothesis testing

Abstract

Partial association refers to the relationship between variables $Y_1, Y_2, \ldots, Y_K$ while adjusting for a set of covariates $\boldsymbol{X} = \{X_1, \ldots, X_p\}$. To assess such an association when Y_k's are recorded on ordinal scales, a classical approach is to use partial correlation between the latent continuous variables. This so-called polychoric correlation is inadequate, as it requires multivariate normality and it only reflects a linear association. We propose a new framework for studying ordinal-ordinal partial association by using surrogate residuals (Liu and Zhang, *JASA*, 2018). We justify that conditional on $\boldsymbol{X}$, Y_k and Y_l are independent if and only if their corresponding surrogate residual variables are independent. Based on this result, we develop a general measure ϕ to quantify association strength. As opposed to polychoric correlation, ϕ does not rely on normality or models with the probit link, but instead it broadly applies to models with any link functions. It can capture a non-linear or even non-monotonic association. Moreover, the measure ϕ gives rise to a general procedure for testing the hypothesis of partial independence. Our framework also permits visualization tools, such as partial regression plots and 3-D P-P plots, to examine the association structure, which is otherwise unfeasible for ordinal data. We stress that the whole set of tools (measures, p-values, and graphics) is developed within a single unified framework, which allows a coherent inference. The analyses of the National Election Study ($K = 5$) and Big Five Personality Traits ($K = 50$) demonstrate that our framework leads to a much fuller assessment of partial association and yields deeper insights for domain researchers.

Key words: covariate adjustment, multivariate analysis, partial regression plot, polychoric correlation, rating data, surrogate residual.

1 Introduction

In many research fields, data are not recorded on a quantitative scale, but a rating scale, such as (1,2,3,...) or (A,B,C,...). The observed symbol merely represents the rating, rather than the numeric value, of the effect size. For example, in oncology research, patients' general well-being and everyday activities are quantified by various scoring systems, such as the Zubrod score $(0, 1, 2, \ldots, 5)$ and the Karnofsky score $(0, 10, 20, \ldots, 100)$. Economic studies often use three rating levels (low, middle, high) to indicate the socioeconomic status of an individual. Political researchers may use a rating scale of "conservative, independent, and liberal" or a finer scale to describe a voter or party's political ideology. In social sciences, the Likert-type scale, with 5 to 10 levels, has become the standard in questionnaires. Collectively, data resulted from a rating scale are termed as ordinal data.

The nature of ordinal data presents special challenges in statistical inference, and inferential tools developed for quantitative variables may not be suitable for their analysis (Liu and Agresti, 2005; Agresti, 2010; Zhang, 2011; Tutz, 2012). There remain fundamental problems that have not been fully resolved. One such problem is: *how to fully assess the association between ordinal variables in the presence of a set of confounding covariates having to be controlled for?* A fuller assessment should at least include numerical quantification, graphical representation, and statistical testing, ideally all in a single framework. The goal of this paper is to establish such a unified framework and justify its utility for inference.

Generally speaking, association analysis investigates the relationship between two or more variables. It plays an indispensable role in almost all research fields. Often, the very first scientific question to address is whether or not the variables of interest are related, and if yes, how? A comprehensive assessment can help researchers direct further analysis, such as variable selection and model selection. For ordinal variables, there is a rich literature for assessing association (see Lapp et al., 1998; Kateri, 2014 and references therein). Well-known

are rank-based methods (Kendall, 1938; Goodman and Kruskal, 1954), polychoric correlation (Tallis, 1962), association models (Goodman, 1979, 1985), local/global odds ratios (Dale, 1984), and multivariate logistic models (Glonek and McCullagh, 1995), to mention a few.

In need of adjusting for a set of covariates, the assessment of *partial association* is more difficult. It has to remove the effect of the covariates (often termed as confounders), which may correlate with the variables of interest. Liu and Agresti (2005) surveyed non-model-based methods, where the adjustment is fulfilled by stratifying data using the covariates. For instance, the Mantel-Haenszel-type methods (Mantel and Haenszel, 1959; Liu and Agresti, 1996) weight the odds ratios of each stratum. The odds-ratio-type measure, unlike Pearson's rho, may not directly reflect the *strength* of association. In addition, data stratification is often subjective in the presence of quantitative covariates. When the number of quantitative covariates grows, the number of strata can grow exponentially. The consistency result requires that the number of strata increase proportional to the overall sample size (Liu and Agresti, 1996), which is nevertheless fixed in practice.

A widely used strategy for adjusting for covariates is to use regression models (Fisher, 1924; Baba et al., 2004; Li and Shepherd, 2010). Specifically, we assume that the variable of interest Y_k $(k = 1, 2, \ldots, K)$ has J_k categories with the order $1 < 2 < \cdots < J_k$, and a set of covariates $\boldsymbol{X} = \{X_1, \ldots, X_p\}$ need to be adjusted for. The adjustment can be realized by using cumulative link models (McCullagh, 1980)

$$G_k^{-1}\left(\Pr\{Y_k \leq j\}\right) = \alpha_{k,j} - f_k(\boldsymbol{X}, \boldsymbol{\beta}_k), \quad j = 1, \ldots, J_k. \tag{1}$$

Here, $G_k^{-1}(\cdot)$ is a link function, which is the inverse of a continuous cumulative distribution function G_k. The function $f_k(\boldsymbol{X}, \boldsymbol{\beta}_k)$ expresses the effect of $\boldsymbol{X}$ on the transformed probabilities $\Pr\{Y_k \leq j\}$. A special but commonly used form is $f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) = \boldsymbol{X}\boldsymbol{\beta}_k$. Model (1) contains a broad class of useful models including ordered probit models with the probit link $G_k^{-1}(\gamma) = \Phi^{-1}(\gamma)$ (McKelvey and Zavoina, 1975), proportional odds models (ordinal logit model) with the logit link $G_k^{-1}(\gamma) = \log(\gamma/(1 - \gamma))$ (McCullagh, 1980), proportional hazard

models with the complementary log-log link $G_k^{-1}(\gamma) = \log(-\log(1-\gamma))$ or negative log-log link $G_k^{-1}(\gamma) = -\log(-\log(\gamma))$ (Prentice and Gloeckler, 1978), relative risk models with the logarithmic link $G_k^{-1}(\gamma) = \log(\gamma)$ (Blizzard et al., 2013), and models with the t link (Albert and Chib, 1993) or Cauchit link $G_k^{-1}(\gamma) = \tan(\pi(\gamma - 0.5))$ (Agresti, 2010). Model (1) can be induced by assuming there exists a latent continuous trait Z_k that underlies the discrete outcome Y_k. Suppose $Z_k = f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + \epsilon_k$, where $\epsilon_k \sim G_k$ and it is independent of $\boldsymbol{X}$. Given a series of cutpoints on the continuous scale $-\infty = \alpha_{k,0} < \alpha_{k,1} < \cdots < \alpha_{k,J_k} = +\infty$, the observed outcome $Y_k = j$ if the latent trait Z_k falls in the j-th interval $(\alpha_{k,j-1}, \alpha_{k,j}]$.

The latent structure of Model (1) incurs a classical model-based measure of partial association, namely, polychoric correlation (Pearson and Pearson, 1922; Tallis, 1962). However, this measure relies on a strong assumption of multivariate normality, namely,

$$
\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{pmatrix} \sim \mathbf{MN} \left[\begin{pmatrix} f_1(\boldsymbol{X}, \boldsymbol{\beta}_1) \\ f_2(\boldsymbol{X}, \boldsymbol{\beta}_2) \\ \vdots \\ f_K(\boldsymbol{X}, \boldsymbol{\beta}_K) \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & 1 & \cdots & \rho_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{K1} & \rho_{K1} & \cdots & 1 \end{pmatrix} \right]. \tag{2}
$$

Such a latent multivariate structure requires the marginal distribution of Y_k follow Model (1) with the probit link, i.e., $G_k(\cdot) = \Phi(\cdot)$ for all k's. Polychoric correlation uses ρ_{kl} in (2) as a substitute to measure the partial association between Y_k and Y_l. It has been used widely in social and behavioral sciences (Olsson, 1979; Rigdon and Ferguson Jr, 1991; Jöreskog, 1994; Yiu and Poon, 2008). Nevertheless, it has at least three major limitations.

($\mathcal{L}$-1) It gauges correlation between the latent variables. It may misconvey the *true strength* of association between *ordinal variables*, as it remains constant regardless of the number and location of cutpoints (i.e., the values of $\alpha_{k,j}$).

($\mathcal{L}$-2) It requires the multivariate normality assumption and applies to Model (1) *with a probit link only*. It does not generalize to models with other link functions.

($\mathcal{L}$-3) It merely measures the strength of linear association in the latent structure, and it may not reflect a *non-linear or non-monotonic structure* of association.

4

Motivated by these major limitations, we propose a new framework for studying ordinal-ordinal partial association. We adjust for covariates by regressing each Y_k on the covariate $\boldsymbol{X}$, and use *surrogate residual* variables R_k's proposed by Liu and Zhang (2018). Our development relies on a key feature of surrogate residuals; that is, although Y_k takes ordinal values, its surrogate residual R_k is a *continuous variable*, which *preserves the residual information in the form of a continuous distribution*. The partial association between Y_k's is then studied by examining the relationship between the residual variables R_k's.

Our framework is built upon a fundamental result; that is, conditional on $\boldsymbol{X}$, Y_k and Y_l are independent if and only if their corresponding surrogate residual variables are independent. This result does not hold for other residuals known so far. The result allows us to develop a much fuller assessment of ordinal-ordinal partial association. Such an assessment encompasses numerical measures, graphical representation, and hypothesis testing.

Specifically, we develop a class of measures, labeled generally by ϕ, to quantify the strength of partial association. In contrary to polychoric correlation, ϕ has several advantages: (A1) its size reflects the strength of association between the ordinal variables, rather than the latent variables; (A2) it does not rely on the multivariate normality assumption or the probit link, but instead it broadly applies to models with any link functions; and (A3) it can capture non-linear or non-monotonic association, and it has potential to detect dependence of any complex structures. These properties are evidenced by examples with varying degrees of complexity of partial association. The measure ϕ also gives rise to a general procedure for testing the hypothesis of partial independence, which can strengthen our conclusion or facilitate our decision making. Moreover, our numerical assessment is complemented by visualization tools, such as partial regression plots and 3-D P-P plots, which were not effective for ordinal data. The graphical examination can reveal the structure of association, which is otherwise hidden as illustrated in our real data analysis. We stress that the measures, p-values, and graphics are all developed within a single framework. This

unified framework allows us to establish a coherent inference scheme.

2 Assessing ordinal-ordinal partial association

2.1 A general framework

We propose to study the partial association between Y_k's through examining the relationship between R_k's. Here, R_k ($k = 1, 2, \ldots, K$) is a *surrogate residual* variable (Liu and Zhang, 2018) obtained by regressing Y_k on the covariate $\boldsymbol{X}$ using Model (1).

To numerically assess partial association between Y_k and Y_l, we define a general measure

$$\phi = \phi(R_k, R_l). \tag{3}$$

The function $\phi(\cdot, \cdot)$ is a general association measure for *continuous* variables. Choices of such a function include

$$\phi_\rho = \rho(R_k, R_l) = \text{Cov}(R_k, R_l) \big/ \sqrt{\text{Var}(R_k)\text{Var}(R_l)}, \tag{4}$$

$$\phi_\tau = \tau(R_k, R_l) = \Pr\{(R_k - R_k^*)(R_l - R_l^*) > 0\} - \Pr\{(R_k - R_k^*)(R_l - R_l^*) < 0\}, \tag{5}$$

$$\phi_\sigma = 12 \iint_{[0,1]^2} |C(u, v) - uv| du dv, \text{ where } C(u, v) = \Pr\{G_k(R_k) \le u, G_l(R_l) \le v\}. \tag{6}$$

Here, ϕ_ρ is Pearson's correlation coefficient, ϕ_τ is Kendall's tau (Kendall, 1938) where R^* is an independent copy of R, and ϕ_σ is Schweizer-Wolff's sigma (Schweizer and Wolff, 1981) where $C(u, v) : [0, 1]^2 \to [0, 1]$ is a bivariate distribution function of $(G_k(R_k), G_l(R_l))$. Since $G_k(R_k) \sim U(0, 1)$, $C(u, v)$ is a bivariate copula function (see Section 3.4 for a detailed discussion). These three notions are known to have different utilities in quantifying association strength for quantitative variables. They have an increasing level of complexity, reflecting linear, monotonic, and the most general association. In Section 3, we will justify their utilities, both theoretically and numerically, for ordinal data. In Section 4, we will make use of these measures to develop procedures for testing the hypothesis of partial independence.

Our framework also enables us to use visualization tools that are developed for quantitative data. Throughout the sections, we will demonstrate their effectiveness in the analysis of ordinal data.

2.2 The tool: surrogate residual

Our general measure ϕ relies on Liu-Zhang's surrogate residuals. As they play a key role in our development, we briefly review the surrogate residuals, present their properties, and discuss their utility in the current context. The idea is to simulate a continuous variable S and use it as a "surrogate" of the ordinal variable Y (the subscript is omitted in this subsection for notational simplicity). Residuals are then defined based on S. For the cumulative link model in (1), a surrogate variable S can be defined as follows

$$
S \sim \begin{cases}
Z \mid -\infty < Z \leq \alpha_1 & \text{if } Y = 1, \\
Z \mid \alpha_1 < Z \leq \alpha_2 & \text{if } Y = 2, \\
\cdots \\
Z \mid \alpha_{J-1} < Z \leq +\infty & \text{if } Y = J,
\end{cases}
$$

where Z is the classical latent variable that underlies Y, as described in the introduction. Conditional on Y, the variable S is sampled from a truncated distribution of Z. An illustration is given in Figure 1, where we depict how to sample S for the model $G^{-1}\left(\Pr\{Y \leq j\}\right) = \alpha_j - X\beta, \ j = 1, 2, 3$. The entire curve represents the distribution of the latent variable Z, with $X\beta$ specifying its mean conditional on X and G specifying its shape. Given an observed discrete outcome, say $Y = 1$, S follows the truncated distribution represented by the thickened curve to the left of the vertical line at α_1. Similarly, when $Y = 2$ or 3, S follows the distribution truncated by the interval $(\alpha_1, \alpha_2]$ or $(\alpha_2, +\infty)$. Conforming to this rule, we can simulate a sample of S provided a sample of Y.

The definition of the surrogate variable S is not subject to the existence of the latent variable Z. Needed is merely a joint probability distribution of (Y, Z) implied by the latent structure of Model (1), rather than the realizations of Z. In fact, the construction of S can

be written without using Z, but three components of Model (1), namely, $f(\boldsymbol{X}, \boldsymbol{\beta})$, α_j's and G. Specifically,

$$S \mid (Y = j, \boldsymbol{X} = \boldsymbol{x}) \sim G(c - f(\boldsymbol{x}, \boldsymbol{\beta})) \text{ truncated by the interval } [\alpha_{j-1}, \alpha_j). \tag{7}$$

We use the variable S as a surrogate of Y in our analysis as it has three properties. ($\mathcal{P}$-1) S follows the same distribution as the latent variable Z, i.e., $\Pr\{S \leq c \mid \boldsymbol{X}\} \equiv \Pr\{Z \leq c \mid \boldsymbol{X}\}$ for any c and $\boldsymbol{X}$. ($\mathcal{P}$-2) Both S and Z give out the same outcome Y on the ordinal scale, i.e., $\alpha_{j-1} < Z \leq \alpha_j \Leftrightarrow Y = j \Leftrightarrow \alpha_{j-1} < S \leq \alpha_j$ for any j. ($\mathcal{P}$-3) S is observable whereas Z is not. On the continuous scale of S, Liu and Zhang (2018) defined a residual variable

$$R = S - E\{S \mid \boldsymbol{X}\}. \tag{8}$$

It is simply the difference between the surrogate variable S and its expectation, following the same principle for defining residuals for ordinary linear models. This principle, nevertheless, does not apply directly to the original ordinal data measured on a rating (not numeric) scale.

For cumulative link models in (1), the surrogate residual $R = S - E\{Z \mid \boldsymbol{X}\} = S - f(\boldsymbol{X}, \boldsymbol{\beta}) - \int_{-\infty}^{\infty} u \, dG(u)$. Liu and Zhang (2018) showed that R is independent of $\boldsymbol{X}$, and it follows the same distribution of the latent error ϵ. That is, $\Pr\{R \leq c \mid \boldsymbol{X}\} = \Pr\{R \leq c\} = G(c + \int u \, dG(u))$. Moreover, as illustrated in Figure 1, R summarizes *the direction and magnitude of the residual randomness of S after removing the effect of $\boldsymbol{X}$.*

2.3 Key results

Using surrogate residuals, we establish a necessary and sufficient condition for determining whether ordinal variables are conditionally independent or not. The results below lay the foundation of our entire framework for quantification, visualization, and hypothesis testing.

Theorem 1. *Assume that the ordinal variable Y_k $(k = 1, 2, \ldots, K)$ follows Model (1). Conditional on the covariate $\boldsymbol{X}$, Y_k and Y_l are independent if and only if the corresponding*

surrogate residual variables R_k and R_l, as defined in (8), are independent, i.e.,

$$(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Leftrightarrow (R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}.$$

Corollary 1. *Under the same condition of Theorem 1, if $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$ for any $\boldsymbol{X}$, the following results hold: (a) the conditional distribution of $(R_k, R_l) \mid \boldsymbol{X}$ is homogeneous across all values of $\boldsymbol{X}$, i.e., $\mathbb{P}(R_k \le r_k, R_l \le r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \le r_k)\mathbb{P}(R_l \le r_l)$; and (b) $R_k \perp\!\!\!\perp R_l$.*

Theorem 1 and Corollary 1 enable us to convert the task of checking conditional independence between Y_k's to that between the corresponding residual variables. This conversion is natural if Y_k's were continuous variables, but it is not trivial for discrete variables. When Y_k's are ordinal, alternative residuals R^{ALT} such as the sign-based residuals (Li and Shepherd, 2012) and the generalized residuals (Franses and Paap, 2001) cannot be used to establish similar results. In fact, the illustration below shows that $(Y_1 \perp\!\!\!\perp Y_2) \mid \boldsymbol{X} \not\Rightarrow R_1^{ALT} \perp\!\!\!\perp R_2^{ALT}$.

Figure 2 displays scatter plots of several types of residuals when Y_1 and Y_2 are simulated independently conditional on X. We observe that R_1^{ALT} and R_2^{ALT} are nevertheless not independent for the sign-based residuals, generalized residuals, and deviance residuals. For these residuals, the conditional distribution of R_2^{ALT} varies depending on the value of R_1^{ALT}. This is not a coincidence but reflects a fundamental defect of these alternative residuals, i.e., their distribution varies depending on the value of $\boldsymbol{X}$ (see Section 3.4 of Liu and Zhang, 2018). In contrast, the surrogate residual variable R_k is independent of $\boldsymbol{X}$. Without this property, we would not have been able to establish $R_k \perp\!\!\!\perp R_l$ in Corollary 1. Such an independence between the surrogate residuals is observed in at the bottom right of Figure 2.

Remark 1. *As a further explanation of the dependence of R_k^{ALT}'s as seen in Figure 2, we note that R_k^{ALT}'s cumulative distribution function (CDF) $F_{k,\boldsymbol{X}}(c) = \mathrm{Pr}_{\boldsymbol{X}}\{R_k^{ALT} \le c\}$ is a function of $\boldsymbol{X}$. Given that the CDFs $F_{k,\boldsymbol{X}}(c)$ and $F_{l,\boldsymbol{X}}(c)$ share a common $\boldsymbol{X}$, it is not unusual to observe the dependency between R_k^{ALT}'s even when Y_k's are partially independent.*

Remark 2. *Compared to the "if-and-only-if" result established for the surrogate residual in*

9

Theorem 1, the alternative sign-based residual R^{ALT} has a much weaker result, i.e., $(Y_k \perp\!\!\!\perp$
$Y_l) \mid \boldsymbol{X} \Rightarrow \mathbb{E}(R_k^{ALT} R_l^{ALT} \mid \boldsymbol{X}) = 0$ (Li and Shepherd, 2010), but not vice versa.

In the section below, we will establish properties of the association measures ϕ_ρ, ϕ_τ, and ϕ_σ in (4)-(6) with accompanying graphics. Their inference is deferred to Section 4.

3   Properties

First of all, an ideal assessment of partial association should be invariant to the labeling of the ordered categories of Y_k's. Given that Y_k has four categories, for instance, our inference should remain the same regardless of the labels used to indicate its categories, (0,1,2,3) or (1,2,5,10). The property below applies to the general ϕ in (3), including ϕ_ρ, ϕ_τ and ϕ_σ.

Property 1. *The association measure ϕ is invariant under strictly monotone transformations of Y_k and Y_l. Specifically, $\phi(Y_k, Y_l) \equiv \phi(g_k(Y_k), g_l(Y_l))$, where g_k and g_l are strictly increasing (decreasing) functions.*

In what follows, we consider association structures in an increasing order of complexity. For the ease of understanding, we use the distribution and dependence of latent variables to indicate the structural complexity. Studied are cases when the latent variables follow (3.1) normal distributions and a linear association; (3.2) non-normal distributions; (3.3) a nonlinear but monotonic association; and (3.4) a general non-normal non-monotonic association. To allow an in-depth discussion, we focus on the bivariate analysis of (Y_1, Y_2) in the simulated examples. The properties apply to any pair of (Y_k, Y_l) in a multivariate analysis.

3.1   Linear association

We begin with a setting where the latent variables follow a bivariate normal distribution and the dependence is linear, i.e., (Z_1, Z_2) follow Model (2). This is the setting where the

polychoric correlation ρ is applicable, but has limitations ($\mathcal{L}$-1)-($\mathcal{L}$-3) (see the introduction) in measuring the association between Y_k's. To illustrate ($\mathcal{L}$-1) and the advantage of our measure σ_ρ, we consider an example before presenting our theoretical properties.

Example 1. Consider three pairs of ordinal variables $(Y_1, Y_2^{(i)})$, $i = 1, 2, 3$, all following preset models. Of interest is the association in each pair after adjusting a covariate $X \sim N(0, 1)$. A computer simulation generates data, and we tabulate the first 20 observations in Table 1.

If we visually explore the relationship between Y_1 and $Y_2^{(i)}$, our common sense may lead to a rough conclusion that the association between $(Y_1, Y_2^{(2)})$ may be the weakest among the three pairs. The intuition is from the fact that we barely observe the variability of $Y_2^{(2)}$, nor its co-movement as Y_1 varies. However, if we use the polychoric correlation to measure such an association, it is $\rho(Y_1, Y_2^{(2)} \mid X) = 0.8$, which indicates a very strong correlation. The question is which is problematic, our intuition or the measure ρ?

Here is how we simulate the data. The ordinal variables $(Y_1, Y_2^{(i)})$ are defined as discretized outcomes of two continuous variables (Z_1, Z_2) that follow the bivariate normal model (2) with $\rho = 0.8$. The marginal distribution of Z_i and the cutpoints can be found in Table 2. The data of (Z_1, Z_2) are discarded and we only keep the data of $(Y_1, Y_2^{(i)})$ and X for inference. Since all the three pairs $(Y_1, Y_2^{(i)})$ are based on the same latent variables (Z_1, Z_2), their polychoric correlations are all the same, i.e., $\rho(Y_1, Y_2^{(i)} \mid X) = 0.8$ for any i. This constant measure does not reflect the varying number and positions of cutpoints among $Y_2^{(i)}$'s, as depicted in Figure 3. In particular, the two cutpoints for $Y_2^{(2)}$ are on the tails, far from the center, of the distribution of Z_2. As a result, the data of $Y_2^{(2)}$, as seen in Table 1, barely show any variability, nor the co-movement with Y_1. If the two cutpoints are further moved away from the center and toward (minus) infinity, $Y_2^{(2)}$ will eventually become a constant. In this situation, any sensible association measure is expected to converge to zero, as the degrees of co-movement of $(Y_1, Y_2^{(2)})$ will diminish. Nevertheless, the value of polychoric correlation remains the same, and it always indicates a strong association $\rho(Y_1, Y_2^{(2)} \mid X) = 0.8$. This is

11

an intrinsic problem of using polychoric correlation to measure partial association between ordinal variables, as *it ignores the impact of cutpoints on the variability of each variable and the heterogeneity of the data generating process.* Consequently, it may misconvey the strength of the ordinal-ordinal association. In fact, the ordinal outcome Y_k can be viewed as a function of the latent variable Z_k, i.e., $Y_k = T_k(Z_k)$ where T_k is determined by the cutpoints α_{kj}'s. With this in mind, we can conclude that it is fundamentally flawed to always use $\rho(Z_1, Z_2 \mid X)$ as a measure of the partial association between $Y_1 = T_1(Z_1)$ and $Y_2 = T_2(Z_2)$, regardless of the form (e.g., the number and locations of α_{kj}'s) of T_k's.

Unlike polychoric correlation, our association measure ϕ_ρ takes into account the influence of cutpoints. As observed in Table 2, if we use ϕ_ρ to gauge partial association, the pair $(Y_1, Y_2^{(2)})$ turns out to have the weakest association among the three pairs as $\phi_\rho(Y_1, Y_2^{(2)}) = 0.08$, in comparison with $\phi_\rho(Y_1, Y_2^{(1)}) = 0.48$ and $\phi_\rho(Y_1, Y_2^{(3)}) = 0.61$. The latter value also indicates that as the cutpoints become "dense", our ϕ_ρ tends towards the polychoric correlation ρ. All these observations are consistent with our intuition. We present below the theoretic properties of ϕ_ρ as well as its connection to polychoric correlation.

Property 2. *Under Model (2), the association measure ϕ_ρ has the following properties:*

(a) $\phi_\rho \in [-1, 1]$.

(b) $\phi_\rho = 0$ if $\rho = 0$.

(c) $\rho_1 = \rho_2 \nRightarrow \phi_{\rho_1} = \phi_{\rho_2}$.

(d) $\phi_\rho \to \rho$ as $J_k \to \infty$ and $\max |\alpha_{k,j+1} - \alpha_{k,j}| \to 0$, provided that $\int u^2 \, dG_k(u) < \infty$ (k=1,2).

If we rephrase Property 2, it says under Model (2), (a) our measure ϕ_ρ has the same range as the polychoric correlation ρ; (b) it is zero if the two latent variables Z_k's are independent; (c) its value is not uniquely determined by ρ, but instead, it also reflects the influence of the cutpoints α_{kj}'s; and (d) its value converges to ρ if we have infinitely many cutpoints and they are sufficiently dense on the real line. These proporties apply to the general measure ϕ as long as it has the form (3). With respect to ϕ_τ in (5), for instance, Table 2 shows that its

value varies, depending on the number and locations of the cutpoints. The smallest value of 0.05 occurs when the cutpoints are $\{-3, 3\}$, two points on the tail of the latent distribution as illustrated in the center panel of Figure 3. The value of ϕ_τ increases to 0.44 as we add 11 more equal-spaced cutpoints as illustrated in the right panel of Figure 3. These properties also hold in general for non-probit models, as similar patterns are observed in Table 2 where the latent errors (ϵ_1, ϵ_2) follow a bivariate Gumbel distribution. The following subsection will address the challenges that arise when the link function is not probit.

3.2 Link functions other than the probit

The second limitation ($\mathcal{L}$-2) of polychoric correlation is that it relies on a vulnerable assumption; that is, the latent variables Z_k's must follow a normal distribution. It will lose its theoretical ground if the link function in Model (1) is not probit. Moreover, it blindly enforces a *multivariate* structure, which requires any linear combination of $(Z_1, \ldots, Z_K)$ follows a normal distribution. Such a prerequisite can hardly be justified as we do not observe Z_k's. Bear in mind that the marginal distribution of Y_k specified in Model (1) does not imply any multivariate association structure. Our measure ϕ_ρ has a desirable property as below.

Property 3. *The association measure ϕ_ρ does not require any upfront specification of the bivariate structure of the joint distribution of (Y_k, Y_l) (e.g., the bivariate normality of (Z_k, Z_l)). It only requires a correct specification of the marginal distribution of each Y_k.*

Property 3 separates the task of assessing partial association between Y_k's from that of specifying marginal distributions for each Y_k. As a result, we do not have to impose any assumption on the multivariate association structure up front. To study ordinal-ordinal partial association, we only need to ensure that the marginal model for each Y_k is specified correctly. The association structure can be determined separately by examining the relationship between residual variables using our measures and graphics. This *division of labor* (i.e., the separation of a work process into a number of tasks, with each task performed separately)

is one of the most appealing features of our framework. Similar ideas have been used in the development of generalized estimating equation (GEE) and copula methods, where efforts of specifying marginal distributions and an association structure are divided completely and conducted separately. This division of labor frees us from the normality assumption. As for studying partial association, we are not confined to the probit link in Model (2) or the log-family link in multivariate logistic models (Glonek and McCullagh, 1995). In fact, we can use any link function in Model (1). This advantage broadens the applicability of our framework. It enables us to study partial association when the latent variable does not follow a symmetric and light-tailed distribution, as illustrated in the example below.

Example 2 (Complementary log-log link)**.** We consider the situation where the link function $G_k^{-1}(p) = \log(-\log(1-p))$ in the marginal models (1) for each Y_k. Similar to Example 1, the association between Y_1 and Y_2 is simulated through the corresponding latent variables Z_1 and Z_2. For complementary log-log links, the latent error variable ϵ_k follows the standard Gumbel distribution with a CDF $F(x) = 1 - \exp(-\exp(x))$. The two latent errors ϵ_1 and ϵ_2 are associated with each other through a Gaussian copula with a correlation of 0.8. For such models with a non-probit link, the polychoric correlation is not applicable, whereas our measure ϕ_ρ remains valid. As observed in Table 2, our measure ϕ_ρ again reflects the impact of the number and locations of the cutpoints. Comparing the values of ϕ_ρ with those when the latent variables follow a bivariate normal distribution, we see that our measure ϕ_ρ also captures the shape of the underlying distribution (or the link function). For example, when the cutpoints are $\{-3, 3\}$, the value of ϕ_ρ is 0.15 for the complementary log-log link which has a much longer lower tail, as compared to 0.08 for the thin-tailed probit link.

The property of "division of labor" applies to any association measure as long as it is defined in the general form of (3). The values of ϕ_τ, for instance, can be found in the bottom row of Table 2, where the pattern appears similar to that in Example 1.

14

3.3 Nonlinear but monotonic association

The division-of-labor feature allows us to examine partial association as if Y_k's were continuous variables. Recall that for continuous variables, we may use linear models to adjust for covariates and the ordinary residual (i.e., the difference between the fitted and observed values) to assess partial association. The association strength can be gauged by Pearson's rho, and the association structure can be inspected by a partial regression plot, where residuals of the two variables are plotted against each other. Borrowing these ideas, we demonstrate the utility of our framework in unveiling nonlinear structures of ordinal-ordinal association, which were scarcely discussed in the literature.

Example 3 (Exponential association). The two ordinal variables Y_1 and Y_2 are simulated by discretizing, respectively, the latent variables $Z_1 = 2X_1 + 2X_2 + \epsilon_1$ and $Z_2 = X_1 - 2X_2 + \epsilon_2$, where $X_1 \sim N(0,1)$ and $X_2 \sim U(0,1)$. We let $\epsilon_1 \sim \text{logistic}(0,1)$ and consider two scenarios of ϵ_2: (1) $\epsilon_2^{(1)} = \epsilon_1 + e$; and (2) $\epsilon_2^{(2)} = \exp(\epsilon_1) + e$. Here, $e \sim \text{logistic}(0,1)$ independently of ϵ_1. The scatter plots in the first column of Figure 4 illustrate the relationship between ϵ_1 and ϵ_2, which represents the partial association between Z_1 and Z_2 after adjusting $\boldsymbol{X}$. Different from the top plot, the bottom plot in this column exhibits a clear nonlinear association structure. This nonlinearity, however, is hidden behind the discrete data, and it would not be known unless we could directly observe the latent variables Z_k's. Using our approach, we provide in the middle column of Figure 4 the scatter plots of the surrogate residuals R_1 versus R_2 (when the number of categories is 5). We stress that these residuals are obtained from the discrete Y_k's, without requiring any observations of the latent Z_k's. It is evident that the scatter plots capture the association patterns of the hidden variables. In particular, the middle bottom plot reveals a nonlinear pattern, which otherwise may be concealed due to the ordinal nature of the data. When the number of categories increases to 10, the patterns remain similar (the right column). But the strength of the association measures ϕ_ρ and ϕ_τ is closer to that seen in the first column. This observation once again confirms that

when the number of categories increases, the association measures between ordinal variables approach their counterpart between latent variables (Property 2(d)). In fact, the proof of Property 2(d) does not require the normality of and linearity between latent variables.

Another appealing feature of our framework is that we are not restricted to the use of moment-based measures, but instead, we are allowed to use a wide scope of metrics to study more complex association structures. In the presence of nonlinearity, the measure ϕ_ρ studied previously may not be ideal in gauging association. Taking the right column of Figure 4 for example, $\phi_\rho = 0.06$ in the bottom plot does not necessarily indicate that the association is much weaker than that in the top plot where $\phi_\rho = 0.63$. It is simply because the moment-based ϕ_ρ does not fully reflect the strength of a nonlinear association. Applying Kendall's tau, a rank-based method, to the two residuals, we obtain $\phi_\tau = 0.47$ for the nonlinear case (the bottom plot), which is comparable to $\phi_\tau = 0.46$ for the linear case (the top plot).

Property 4. *Without the adjustment for the covariates, our measure ϕ_τ is identical to Kendall's τ, i.e., $\phi_\tau(R_k, R_l) \equiv \tau(Y_k, Y_l)$.*

In view of Property 4, our measure ϕ_τ generalizes Kendall' tau to the cases where an adjustment for covariates is applied. It makes it sensible to compare the value of $\tau(Y_k, Y_l)$ before the adjustment with the value of $\phi_\tau(R_k, R_l)$ after the adjustment. The difference reflects the amount of correlation eliminated by removing confounding effects. This point will be illustrated in our analysis of the National Election Study.

3.4 General association

When the association may not be linear or monotonic and the link function may not be probit, we show how to assess partial association through surrogate residuals. First, we have the following result due to Sklar's theorem (Sklar, 1959).

Theorem 2. *Let H be the joint distribution of the surrogate residual variables R_k and R_l*

16

in (8). There exists a unique copula $C(u, v)$ such that for all $(r_k, r_l) \in \mathbb{R}^2$

$$H(r_k, r_l) = C(G_k(r_k), G_l(r_l)), \tag{9}$$

where $G_k(\cdot)$ is the inverse of the link function in Model (1).

The establishment of Theorem 2 relies on a key fact that the surrogate residuals R_1 and R_2 are continuous variables and each follows an explicit distribution, i.e., $R_k \sim G_k(\cdot)$. For the alternative residual R^{ALT} mentioned in Section 2.3, it is difficult, if not impossible, to establish a simple and explicit copula expression, similar to (9), for the joint distribution H. The difficulty rises from the fact that the distribution of R^{ALT} may not be continuous, nor does its distribution have an explicit form (Section 3.4 of Liu and Zhang, 2018).

The implication of Theorem 2 is that given surrogate residuals, we can estimate the copula function $C(\cdot, \cdot)$ empirically. Based on $\hat{C}(\cdot, \cdot)$, measures and graphics can be developed to examine partial association in light of the corollary below.

Corollary 2. *If the ordinal variables Y_k and Y_l are partially independent, the copula C in (9) is a product copula, i.e., $Y_k \perp\!\!\!\perp Y_l \mid \boldsymbol{X}$ for any $\boldsymbol{X} \Rightarrow C(u, v) = C^{\perp}(u, v) = uv$.*

Corollary 2 supports the use of ϕ_σ in (6) as a partial association measure for the general setting. It calculates Schweizer-Wolff's sigma between two surrogate residual variables R_k and R_l. This ϕ_σ has potential to capture a non-monotonic association, which could be otherwise missed by the existing measures. Moreover, since the product copula $C^{\perp}(u, v) = uv$ corresponds to the case of independence, we can visualize the deviation $(C(u, v) - uv)$ to inspect the degree of independence. Specifically, we can plot $(\hat{C}(u, v) - uv)$ against (u, v) in a 3-D display, where $\hat{C}(u, v)$ is the empirical copula trained from the data. The example below illustrates that this 3-D plot extends the idea of the probability-probability (P-P) plot.

Example 4 (Half-circle association). We simulate two ordinal variables Y_1 and Y_2 by discretizing the latent variables $Z_1 = X + \epsilon_1$ and $Z_2 = 1.5X + \epsilon_2$, where $X \sim N(0, 1)$. As

depicted in Figure 5(a), we let the latent errors ϵ_1 and ϵ_2 have a half-circle association by setting $\epsilon_1 \sim \mathrm{U}(-3, 3)$ and $\epsilon_2 = -\sqrt{9 - \epsilon_1^2} + e$, where $e \sim N(0, 0.1^2)$, independent of ϵ_1. As the latent errors simulated as such are very different from the normal errors, the probit link is not appropriate for fitting Model (1) to the ordinal data of Y_k's. Assuming the true models as in (1), including the link functions, are known, we can obtain surrogate residuals R_1 and R_2. Figure 5(b) shows the partial regression plot of R_1 versus R_2. It reveals the hidden half-circle pattern. This non-monotonic association may not be unmasked by common Pearson's correlation $\phi_\rho = -0.01$ or Kendall's tau $\phi_\tau = -0.01$. In contrast, the copula-based measure $\phi_\sigma = 0.33$ indicates a nonignorable association. Displayed in Figure 6 is a 3-D plot of $(\hat{C}(u, v) - uv)$ against (u, v). If Y_k's are partially independent, the surface in Figure 6 is expected to fall onto the horizontal plane. This is similar to the P-P plot where the empirical line is expected to align with the theoretical line when the hypothesis holds. Evident in Figure 6 is the departure of the surface from the horizontal plan, which is another indicator of the dependency.

4   Inference

We have established the properties of $\phi = \phi(R_1, R_2) = \phi(R_{1,\boldsymbol{\theta}_1}, R_{2,\boldsymbol{\theta}_2})$, where the surrogate residual variable $R_{k,\boldsymbol{\theta}_k}$ relies on the unknown parameter $\boldsymbol{\theta}_k = (\alpha_{k,1}, \ldots, \alpha_{k,J_k-1}, \boldsymbol{\beta})$ in Model (1). To make inference of ϕ, we obtain a consistent estimate of $\boldsymbol{\theta}_k$ using the maximum likelihood method. Let $\tilde{\boldsymbol{\theta}}_k$ denote such an estimate and $\boldsymbol{r}_{k,\tilde{\boldsymbol{\theta}}_k}$ a sample of the variable $R_{k,\tilde{\boldsymbol{\theta}}_k}$. We discuss in this section how to use $\boldsymbol{r}_{k,\tilde{\boldsymbol{\theta}}_k}$ to carry out a variety of inferences of ϕ.

4.1   Estimation

We can obtain an empirical estimate of ϕ by using the empirical method of estimating Pearson's rho, Kendall's tau, and Schweizer-Wolff's sigma, namely, $\hat{\phi} = \hat{\phi}_{\mathrm{empirical}}(\boldsymbol{r}_{1,\tilde{\boldsymbol{\theta}}_1}, \boldsymbol{r}_{2,\tilde{\boldsymbol{\theta}}_2})$.

For example, the empirical estimate of Pearson's rho is

$$\hat{\phi}_{\text{empirical}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \sum_{i=1}^{n}(r_{1i} - \bar{r}_1)(r_{2i} - \bar{r}_2)/\sqrt{\sum_{i=1}^{n}(r_{1i} - \bar{r}_1)^2 \sum_{i=1}^{n}(r_{2i} - \bar{r}_2)^2}.$$

In Appendix A, we establish the consistency of the estimate $\hat{\phi}$ under certain mild conditions.

The estimate $\hat{\phi}$ uses surrogate residuals from a single draw from the sampling scheme (7). To reduce its variability and increase inference efficiency, we recommend to conduct M (> 1) draws and use the average as the final estimate (see e.g., Hong and He, 2010)

$$\hat{\phi}^{(M)} = \frac{1}{M}\sum_{m=1}^{M}\hat{\phi}_{\text{empirical}}(\boldsymbol{r}_{1,\tilde{\boldsymbol{\theta}}_1}^{(m)}, \boldsymbol{r}_{2,\tilde{\boldsymbol{\theta}}_2}^{(m)}). \tag{10}$$

Our numerical study adopts $M = 30$ as any larger number merely leads to a minimal improvement on the variance reduction.

4.2 Standard error, confidence interval, and p-value

To evaluate the uncertainty of $\hat{\phi}^{(M)}$ in (10), we propose a bootstrap algorithm as follows: [Step 1] obtain a size-n bootstrap sample $(\boldsymbol{x}^*, \boldsymbol{y}_k^*)$ by resampling from the given sample $(\boldsymbol{x}, \boldsymbol{y}_k)$; [Step 2] fit Model (1) to the bootstrap sample and obtain an estimate $\tilde{\boldsymbol{\theta}}_k^*$; [Step 3] obtain M samples of surrogate residuals $(\boldsymbol{r}_{k,\tilde{\boldsymbol{\theta}}_k^*}^{*(1)}, ..., \boldsymbol{r}_{k,\tilde{\boldsymbol{\theta}}_k^*}^{*(M)})$; and [Step 4] calculate the value of $\hat{\phi}^{(M)}$ in (10). Repeating Steps (1-4) B times, we have a bootstrap set $\{\hat{\phi}_1^*, \hat{\phi}_2^*, \ldots, \hat{\phi}_B^*\}$ (superscript (M) omitted). Its empirical distribution is denoted by $\hat{F}_B^*(\phi)$.

We can use the bootstrap distribution $\hat{F}_B^*(\phi)$ to draw a variety of inferences (Efron and Tibshirani, 1994). For example, its standard deviation is an estimate of the standard error of $\hat{\phi}^{(M)}$. The interval $(\hat{F}_B^{*(-1)}(\alpha/2), \hat{F}_B^{*(-1)}(1 - \alpha/2))$ yields a $100(1 - \alpha)\%$ confidence interval. For testing the hypothesis of independence, namely, $H_0 : \phi = 0$, the p-value is $2\min(\hat{F}_B^*(0), 1 - \hat{F}_B^*(0))$. More generally, we can use $2\min(\hat{F}_B^*(\delta), 1 - \hat{F}_B^*(-\delta))$ as a p-value to test a composite hypothesis such as $H_0 : |\phi| \le \delta$, where δ is a positive threshold representing

19

the association strength that is meaningful to domain researchers. The integrative use of these outcomes leads to a more informed conclusion, as they provide answers to the following questions: (i) how large is the association size? (ii) what are its confidence limits? (iii) is it statistically significant? and (iv) is it practically significant as well? The feature of allowing a comprehensive inference distinguishes our framework from the literature that only focuses on testing of the hypothesis of *independence* (e.g., Li and Shepherd, 2010; Zhu et al., 2012). A dichotomous decision (rejecting a simple null hypothesis or not) may be far from being adequate for understanding a scientific issue.

5   Simulation studies

Simulation I. As our inference relies on the bootstrap distribution $\hat{F}_B^*(\phi)$, we first and foremost numerically evaluate its approximation accuracy. We simulate data using the settings of Examples 1 and 2. For each simulated sample, we follow the algorithm in Section 4 and obtain $B = 2000$ bootstrap replicates $\{\hat{\phi}_1^*, \hat{\phi}_2^*, \ldots, \hat{\phi}_B^*\}$ of the estimate $\hat{\phi}^{(M)}$ in (10). We set $M = 30$ consistently in the simulation and omit this superscript. To compare the bootstrap distribution $\hat{F}_B^*(\phi)$ with the true distribution $F_0(\phi)$, we calculate their means, standard deviations, 2.5% quantiles, and 97.5% quantiles, respectively. The simulation is repeated 1000 times and the averages of those statistics are reported in Table 3. We see that in general, the mean and standard deviation (SD) of $\hat{F}_B^*(\phi)$ are very close to the true values. This approximation remains very well in cases of a small sample size ($n = 100$) and a non-normal (Gumbel) latent distribution, which supports the use of $\text{SD}(\hat{F}_B^*(\phi))$ to estimate the standard error of $\hat{\phi}$ in practice. Furthermore, Table 3 suggests the approximation is remarkably accurate even on the the distribution tails, as reflected by the close proximity of the estimated 2.5% and 97.5% quantiles to their true values. This observation confirms the validity of using $\hat{F}_B^*(\phi)$ to construct confidence intervals and p-values for hypothesis testing. In fact, Table 3 shows that the 95% confidence intervals have coverage probabilities close

to their nominal levels. In the rest of this section, we show how p-values should be used together with measures and graphics to achieve a full assessment of partial association.

Simulation II. We use exactly the same simulation setting of Li and Shepherd (2010). Specifically, Y_1 is generated from an ordered logit model $\{\text{logit}(P\{Y_1 \leq j | \boldsymbol{X}\}) = \alpha_{1,j} - \beta_1 \boldsymbol{X}\}$, and conditional on Y_1, Y_2 is generated from $\{\text{logit}(P\{Y_2 \leq j | \boldsymbol{X}\}) = \alpha_{2,j} - (\beta_2 \boldsymbol{X} + \eta_1 I_{(Y_1=1)} + ... + \eta_5 I_{(Y_1=5)})\}, j = 1, ..., J_k$. The parameter vector $\boldsymbol{\eta} = (\eta_1, ..., \eta_5)$ controls the degrees of the partial association between Y_1 and Y_2. Li and Shepherd (2010) considered four scenarios: (1) $\boldsymbol{\eta} = (0, 0, 0, 0, 0)$ (the null); (2) $\boldsymbol{\eta} = (-0.4, -0.2, 0, 0.2, 0.4)$ (linear effect); (3) $\boldsymbol{\eta} = (-0.30, 0.18, 0.20, 0.22, 0.24)$ (nonlinear-monotonic effect); and (4) $\boldsymbol{\eta} = (-0.2, 0, 0.2, 0, -0.2)$ (non-monotonic effect). We compare the type I error and power of our methods with those of the likelihood ratio test (LRT) and Li-Shepherd (LS) method. The LRT may be regarded as the gold standard in this setting as the partial association is determined solely by $\boldsymbol{\eta}$. Table 4 shows that compared to this gold standard, our method yields slightly lower type I error but even higher power in scenarios (2) and (3). We also observe that the power of our ϕ_ρ-based, ϕ_τ-based methods, and the LS method is close to each other, with the LS method being slightly more powerful under the logit model. This leads to a speculation that the *actual* associations might be approximately linear for all scenarios, despite that Scenarios (3)-(4) are termed "nonlinear-monotonic" and "non-monotonic" by Li and Shepherd (2010). The speculation is confirmed by our graphical inspection in Figure 7, where no clear signs of non-linearity are spotted in the three partial regression plots. For the "non-monotonic" case, the plot on the right of Figure 7 explains why our and LS methods all have low power as the lowess curve is almost flat. This insight is otherwise not available without our graphical inspection. In this case, the LRT method exhibits much higher power as seen in Table 4. As a further analysis, we replace the logit link with the complementary log-log link in the simulation models and repeat all the analysis. Similar results are observed in Table 4.

Simulation III. As our testing procedure is developed based on strength measures, we

are allowed to test whether or not the association is *negligible*. For instance, an association may be deemed as negligible if its strength is no more than a threshold δ. In this case, it is sensible to test $H_0 : |\phi| \leq \delta$ versus $H_1 : |\phi| > \delta$. Table 5 shows the type I error and power when $\delta = 0.1$. Considered are four scenarios including Scenario (2) (negligible linear effect) and Scenario (3) (negligible monotonic effect) used in the previous simulation, and (5) $\boldsymbol{\eta} = (-1, -0.5, 0, 0.5, 1)$ (non-negligible linear effect) and (6) $\boldsymbol{\eta} = (-1.2, 0, 0.3, 0.5, 0.6)$ (non-negligible monotonic effect). As the null is composite, the type I error of our methods is much below the nominal level of 0.05. On the other hand, at the alternatives, our methods retain reasonable power to claim that the association is not negligible. Our result points to the fact that it is not straightforward to use the LRT and LS methods to claim an effect being non-negligible without a measure of association strength.

Simulation IV. We carry out further analysis to examine testing power when the association is not linear. We follow the setting of Example 3 (Section 3.3) and simulate the latent errors as $\epsilon_1 \sim \text{logistic}(0, 1)$ and $\epsilon_2 = c \cdot e^{\epsilon_1} + e$, where $e \sim \text{logistic}(0, 1)$. The coefficient c controls the degree of non-linearity, with a large value indicating more notable non-linearity. For $c = \{1, 0.5, 0.1, 0.05, 0.01\}$, the power for testing $H_0: \phi = 0$ is shown in Table 6. When c is small (e.g., $c = 0.05, 0.1$), our ϕ_ρ-based method exhibits the highest power. This is not surprising as the association is approximately linear. Our ϕ_τ-based method is notably less powerful, so is the LS method. When c increases to 0.5 or 1, the power of our ϕ_τ-based method catches up and even surpasses that of the ϕ_ρ-based method. This is expected as the measure ϕ_τ is more capable of capturing a non-linear association. This point can be clearly seen in Table 7, where the estimates of ϕ_τ and ϕ_ρ are compared. When the degree of non-linearity is low ($c \leq 0.1$), the size of $\hat{\phi}_\tau$ remains slightly smaller than that of $\hat{\phi}_\rho$. But when $c = 1$, the effect size $\hat{\phi}_\tau$ is substantially larger (e.g, 0.302 versus 0.139 when $n = 200$). This simulation underlines the importance of having strength measures beyond testing results.

22

6 Real data examples

We analyze two survey data sets from the National Election Study and "BIG5" (personality traits) project. The analyses demonstrate the advantages of our entire framework over traditional methods, which often simply test independence and merely yield a dichotomous conclusion. We illustrate the integrative use of our measures, graphics, and p-values will result in substantial evidence, based on which a more informed inference can be made.

6.1 Analysis of the National Election Study

A traditional research question in political science is how voters' partisanship (e.g. party identification) relates to their voting behavior (Bartels, 2000). To answer this question, we analyze a data set of the 1996 American National Election (Rosenstone et al., 2016), which is collected by the American National Election Study project (`https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00003`). The sample consists of 944 individuals who responded to a survey. The variable "VOTE" indicates a respondent's voting preference for the two presidential candidates, Bill Clinton (Democratic) and Bob Dole (Republican). The variable "PID" is a respondent's party identification with 7 ordinal levels (from *strong Democratic* to *strong Republican*). The data are tabulated in Table 8.

The association between PID and VOTE in Table 8 is manifest as Republicans are inclined to support Dole. The estimate of Kendall's is 0.675 (0.012) (in the parentheses is an estimate of the standard error), which supports the common perception that VOTE is strongly correlated with PID. However, such a seemingly strong correlation may simply be a result of confounding factors. For instance, demographic factors, such as age, income and education, may correlate with both variables. Without adjusting for confounders and assess partial association, it remains unknown to what extent VOTE is *truly influenced by* PID.

6.1.1 Detailed bivariate analysis of VOTE and PID

We adjust for three potential confounders (age, income, and education) by applying Model (1) to VOTE and PID, respectively (see Table 9 for inference results). The probit link is chosen as commonly used in political science. This assumption of the latent variables being normal can be validated in model diagnostics (Liu and Zhang, 2018) by drawing a Q-Q plot of surrogate residuals versus theoretical normal quantiles (see Figure 8). We stress that although the probit link is used in each *marginal* model, we do not require that the two latent variables *jointly* follow a *bivariate* normal distribution. Such an assumption is, in fact, not supported by the data, as will be seen later.

To numerically assess the partial association between VOTE and PID, we obtain an estimate of the tau-based measure $\hat{\phi}_\tau(r_{VOTE}, r_{PID}) = 0.415\ (0.010)$ using the surrogate residuals r_{VOTE} and r_{PID}. The value of 0.415 indicates the association strength after adjusting for the covariates. Recall that before such an adjustment, $\hat{\phi}_\tau = \hat{\tau} = 0.675\ (0.012)$ (Property 4). Our analysis reveals a $(0.675\text{-}0.415)/0.675 = 38.5\%$ reduction in association strength with the confounding effects removed. In other words, these confounders account for 38.5% of the "total correlation" between VOTE and PID. Without eliminating such a notable amount of confounding effects, the influence of PID on VOTE could be overestimated.

To visualize the partial association between VOTE and PID, the partial regression plot in Figure 9 scatters $(r_{VOTE},\ r_{PID})$. The plot confirms a positive partial association, as indicated by the fitted cubic smoothing splines (red dashed curve). Another eye-catching pattern is that the majority of the points fall into Quadrant 1 ($r_{VOTE} > 0$ and $r_{PID} > 0$) and Quadrant 3 ($r_{VOTE} < 0$ and $r_{PID} < 0$), with much fewer points found in Quadrants 2 and 4. The pattern leads to a conjecture that the points in Quadrant 1 (or Quadrant 3) represent those who mostly voted for Dole (or Clinton). This conjecture is true once labels are placed on the points ($\times$ for Dole and $\circ$ for Clinton). A more careful observation reveals that those individuals with $r_{PID} < -1$ seldom voted for Dole, whereas quite a few supporters

of Clinton can be found among those with $r_{PID} > 1$. This may indicate that the influence of PID on VOTE is not symmetric after adjusting for age, income and education.

The partial regression plot in Figure 9 also exemplifies a case where the bivariate normality, as assumed by polychoric correlation, does not hold in general. Although the projection of the points in Figure 9 to each margin (r_{PID} or r_{VOTE}) approximates reasonably well to the normal distribution (see, Figure 8), the joint distribution of r_{PID} and r_{VOTE} does not appear to follow a *bivariate* normal distribution. To see this, we project the points in Figure 9 to a 60-degree straight line and provide a histogram of the projected sample. The distribution seen in Figure 10 appears to have two modes and apparently does not follow a normal distribution. Observations are similar if the sample is projected to 30- and 40-degree lines. We therefore conclude that it is not appropriate to blindly impose the bivariate normality assumption without a careful examination of the data using our framework.

The partial association between VOTE and PID can also be inspected through the lens of a 3-D P-P plot. Plotted in Figure 11 is $12(\hat{C}(u,v) - uv)$ against (u,v), where $\hat{C}(\cdot,\cdot)$ is an empirical copula trained from the data of (r_{PID}, r_{VOTE}). The whole surface stays well above the horizontal plane, which is another indication of a sizable positive association.

6.1.2 Expanded multivariate analysis of 5 discrete variables

Following the bivariate analysis of VOTE and PID, we expand the examination to a set of $K = 5$ discrete variables by incorporating *selfLR*, *ClinLR*, and *DoleLR* (i.e., Left-Right placement of the respondent, Clinton, and Dole). These three variables are recorded on the same seven-rating scales (from *extremely liberal* to *extremely conservative*). The same set of covariates (age, income, and education) is adjusted using Model (1) with the probit link (see Table 9 for the inference result). A comprehensive assessment of partial association is presented in Table 10, which shows (a) the estimates of ϕ_τ for each pair of variables; (b) the estimates of the standard errors of $\hat{\phi}_\tau$; and (c) the statistical significance for testing

$H_0 : \phi_\tau = 0$. All of these inference outcomes are obtained using our framework.

We observe in Table 10 that the partial association is significant for all pairs of the 5 variables except the pair of (PID, DoleLR). By comparing the size of $\hat{\phi}_\tau$ before and after the adjustment, we see that the strength of all the significant associations is weakened after adjusting for the covariates. The numerical assessments in Table 10 are accompanied by the partial regression plots in Figure 12. The last column of this plot matrix reveals that the partial association between DoleLR and other variables may not be linear or even monotonic. For instance, the red lowess curve in the scatter plot of (r_{DoleLR}, r_{VOTE}) exhibits an inverted-U shape, while that of (r_{DoleLR}, r_{ClinLR}) resembles the shape of a beach chair. Both are indications that the association may change its direction at a certain point, which warrants our attention and may lead to new insights in the domain research.

6.2 Analysis of Big Five personality traits

To further illustrate the usefulness of our framework, we analyze a Big-Five data set with a much higher dimension of variables ($K = 50$). The Big Five personality traits include extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience. They were originally proposed in Tupes and Christal (1961) to represent major aspects of an individual's personality. To study these traits, Goldberg (1992) developed 50 items for use in personality tests, 10 for each trait. For example, in the data set "BIG5" (`https://openpsychometrics.org/_rawdata/`), the 50 items are measured on a five-rating Likert scale (from 1=Disagree to 5=Agree), yielding a total of 50 columns of ordinal data.

A common research outcome often reported in social studies is the partial association between the Big Five traits while adjusting for other traits of an individual (see, e.g., Erdheim et al., 2006; Erdle et al., 2009). It reflects the bona fide connection between the Big Five traits with potential confounding effects removed. In the data set "BIG5", available covariates include *age*, *gender*, *engnat* (is English native language), *hand* (what hand does participant

use to write with), and *source* (how the participant came to the test). In what follows, we adjust for these covariates and perform partial association analysis of the Big Five traits. Our analysis focuses on the sample of 5287 Caucasian in the United States.

Seen in Figure 13 is a color-scaled partial association plot in the form of a 50×50 matrix. It visualizes the strength of the partial association between the 50 survey items. The five 10×10 diagonal matrices are much darker than those off-diagonal matrix blocks. This observation confirms that the survey items are more correlated to each other within each trait. As the color of the off-diagonal matrix blocks are light, we may conclude that the between-trait associations are weak, if they do exist. To decide if these weak associations are significant *statistically* and *practically*, we carry out hypothesis testing as below.

We test the hypothesis that the partial association is zero ($H_0 : \phi_\tau = 0$) between a survey item of extraversion and an item of conscientiousness. The testing results are presented in the left panel of Figure 14, where the significance (at $\alpha = 0.05$) is indicated using an orange dot. As the majority of the cells are filled, a quick conclusion may be that the two traits are partially correlated. Nevertheless, it is known that a practically negligible association may be statistically significant, simply as a result of a large sample size. We therefore carry out further analysis by testing $H_0 : |\phi_\tau| \leq 0.1$. The results in the right panel of Figure 14, with only one cell filled, may lead to a different conclusion; that is, the partial association between the two traits may be *insignificant in the practical sense*. This example once again demonstrates that our framework allows a fuller assessment of partial association.

7   Extension to general models

So far, we have focused on cumulative link models. In this section, we show that a straightforward extension to more general models naturally follows and the main result still holds.

Suppose the covariate adjustment is conducted through a parametric model. In its most

general form, the model can be written as

$$Y_k \sim F_k(y_k; \boldsymbol{X}, \boldsymbol{\beta}), \tag{11}$$

where $F_k(\cdot)$ is a discrete cumulative distribution function. This model is general enough to encompass commonly used models such as the adjacent-category logit model and the stereotype model. For the general model (11), Liu and Zhang (2018) defined a surrogate variable S_k using the jittering method on the probability scale. Specifically, $S_k \mid (Y_k = j) \sim U[F_k(j-1), F_k(j)]$. A residual variable is defined using the same formula (8), i.e., $R_k = S_k - E(S_k \mid \boldsymbol{X})$. This residual variable retains the key properties of R_k established for cumulative link models. For instance, under Model (11), it has (a) zero expectation ($E\{R_k \mid \boldsymbol{X}\} = 0$); (b) an explicit distribution ($R_k \mid \boldsymbol{X} \sim U(-1/2, 1/2)$); and (c) independence of $\boldsymbol{X}$ ($R_k \perp\!\!\!\perp \boldsymbol{X}$) (Liu and Zhang, 2018).

Below we establish theoretical results that are parallel to the key results in Section 2.3.

Theorem 3. *Assume that the ordinal variable Y_k ($k = 1, 2, \ldots, K$) follows Model (11). Conditional on the covariate $\boldsymbol{X}$, Y_k and Y_l are independent if and only if the corresponding surrogate residual variables R_k and R_l, as defined in (8), are independent, i.e.,*

$$(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Leftrightarrow (R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}.$$

Corollary 3. *Under the same condition of Theorem 3, if $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$ for any $\boldsymbol{X}$, the following results hold: (a) the conditional distribution of $(R_k, R_l) \mid \boldsymbol{X}$ is homogeneous across all values of $\boldsymbol{X}$, i.e., $\mathbb{P}(R_k \leq r_k, R_l \leq r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \leq r_k)\mathbb{P}(R_l \leq r_l)$; and (b) $R_k \perp\!\!\!\perp R_l$.*

Theorem 3 and Corollary 3 justify that the core idea of this article still applies to Model (11). In fact, the results here have reinforced our argument; that is, *for any parametric model*, we can convert the task of checking conditional independence between Y_k's to that between the corresponding residual variables. What follows are two examples where the adjacent-category logit model is used to carry out analysis.

28

Example 5 (partial independence). Given the value of $\boldsymbol{X}$, we simulate Y_1 and Y_2 independently from the adjacent-category logit model $\log(\pi_{kj}/\pi_{k,j+1}) = \alpha_{kj} + \boldsymbol{\beta}_k^T \boldsymbol{X}$, where $\pi_{kj} = \Pr(Y_k = j)$. In this case, we expect to see their corresponding residuals to be independent. Figure 15 displays partial regression plots of four types of residuals. We observe that the sign-based residuals, generalized residuals, and deviance residuals all exhibit certain patterns of dependence. The only exception is the plot of the surrogate residuals. This example demonstrates that Corollary 3 does not hold for other residuals known so far. This finding reinforces our conclusion established in Section 2.3; that is, the surrogate residual is so far the only one that satisfies $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Rightarrow R_k \perp\!\!\!\perp R_l$.

Example 6 (partial dependence). We re-analyze the national election data where VOTE and PID have been found to be dependent even after having adjusted for three confounders. We repeat the analysis except using the adjacent-category logit model to fit the data. A scatter plot of the surrogate residuals (r_{VOTE}, r_{PID}) is shown in Figure 16. Observed are (a) a positive association; (b) the majority of data points falling into Quadrants 1 and 3; and (c) more blue circles in Quadrant 4 than red crosses in Quadrant 2. These observations are similar to those found in Figure 9 where the cumulative link model was used to fit the data. The estimate of the measure ϕ_τ is 0.384 (0.011), which is also close to the previous estimate 0.415 (0.010) based on the cumulative link model. We therefore conclude that the inferences drawn out of the two models are similar. An expanded analysis of five discrete variables, similar to Section 6.1.2, is conducted in Appendix E.

8   Discussion

We have established a new framework for studying partial association between ordinal variables. Our theories, simulated examples, and real applications have justified that the framework has the following strengths: (1) **Generality.** The idea can be implemented using general ordinal models including all commonly used parametric models. (2) **Uniformity.**

It unifies quantification, visualization, and hypothesis testing in *a single framework*, which allows *a coherent inference*. As a result, the measures, *p*-values, and graphics can complement and strengthen each other. (3) **Capacity to load a vast tool set.** It allows us to study partial association between ordinal variables *in the same way* as we do for continuous variables. As a result, it permits us to use *the whole tool box* developed for continuous data (see, e.g., Balakrishnan and Lai, 2009). We conclude the paper with some final comments.

Choice of inference models

Our framework does not require the marginal model be the same for each Y_k. If the models use different link functions, the result is simply that the residuals R_k's will have different continuous distributions. It does not jeopardize the validity of applying our measures. The interpretation follows the convention of interpreting correlation measures for continuous variables with distinct distributions. What matters is the correct specification of the link function and other components of each model. In practice, we should use the model that best fits the data and well conforms to the domain conventions. More specifically, the model should be examined using diagnostic tools (Liu and Zhang, 2018). We should not adopt a model that exhibit clear evidence of being inconsistent with the data at hand. In cases where two choices (e.g., the logit link versus probit link) are indistinguishable for a given data set, we expect similar conclusions on the assessment of partial association. The final choice of which to present may be made based on domain knowledge and expert opinions.

Computational implementation

To implement our framework for multiple variables, we need a fitted model for each variable. The models can be fitted separately, which is equivalent to using a generalized estimating equation (GEE) with an identity correlation structure. As a result, the inference of $\phi = \phi(R_k, R_l)$ in a multivariate setup is the same as the inference given these two variables only. For a wide scope of GLMs, the R package "sure" is readily available to compute surrogate residuals (Greenwell et al., 2018). Li et al. (2020) have developed a new R package "PAsso"

to produce the measures, p-values, and graphics proposed in this paper.

References

Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, John Wiley & Sons: Hoboken, New Jersey, 2nd ed.

Albert, J. H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American statistical Association*, 88, 669–679.

Baba, K., Shibata, R., and Sibuya, M. (2004), "Partial correlation and conditional correlation as measures of conditional independence," *Australian & New Zealand Journal of Statistics*, 46, 657–664.

Balakrishnan, N. and Lai, C. D. (2009), *Continuous Bivariate Distributions*, Springer Science & Business Media: New York, NY.

Bartels, L. M. (2000), "Partisanship and voting behavior, 1952-1996," *American Journal of Political Science*, 44, 35–50.

Blizzard, C. L., Quinn, S. J., Canary, J. D., and Hosmer, D. W. (2013), "Log-link regression models for ordinal responses," *Open Journal of Statistics*, 3, 16–25.

Dale, J. R. (1984), "Local versus global association for bivariate ordered responses," *Biometrika*, 71, 507–514.

Efron, B. and Tibshirani, R. (1994), *An Introduction to the Bootstrap*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Erdheim, J., Wang, M., and Zickar, M. J. (2006), "Linking the Big Five personality constructs to organizational commitment," *Personality and Individual Differences*, 41, 959–970.

Erdle, S., Gosling, S. D., and Potter, J. (2009), "Does self-esteem account for the higher-order factors of the Big Five?" *Journal of Research in Personality*, 43, 921–922.

Fisher, R. A. (1924), "The distribution of the partial correlation coefficient," *Metron*, 3, 329–332.

Franses, P. H. and Paap, R. (2001), *Quantitative Models in Marketing Research*, Cambridge University Press: Cambridge, UK.

Glonek, G. F. and McCullagh, P. (1995), "Multivariate logistic models," *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 533–546.

Goldberg, L. R. (1992), "The development of markers for the Big-Five factor structure." *Psychological Assessment*, 4, 26.

Goodman, L. A. (1979), "Simple models for the analysis of association in cross-classifications having ordered categories," *Journal of the American Statistical Association*, 74, 537–552.

— (1985), "The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries," *The Annals of Statistics*, 13, 10–69.

Goodman, L. A. and Kruskal, W. H. (1954), "Measures of association for cross classifications," *Journal of the American Statistical Association*, 49, 732–764.

Greenwell, B. M., McCarthy, A. J., Boehmke, B. C., and Liu, D. (2018), "Residuals and diagnostics for binary and ordinal regression models: an introduction to the sure package," *The R Journal*, 10, 381–394.

Hong, H. G. and He, X. (2010), "Prediction of functional status for the elderly based on a new ordinal regression model," *Journal of the American Statistical Association*, 105, 930–941.

Jöreskog, K. G. (1994), "On the estimation of polychoric correlations and their asymptotic covariance matrix," *Psychometrika*, 59, 381–389.

Kateri, M. (2014), *Contingency Table Analysis: Methods and Implementation Using R*, Springer: New York, NY.

Kendall, M. G. (1938), "A new measure of rank correlation," *Biometrika*, 30, 81–93.

Lapp, K., Molenberghs, G., and Lesaffre, E. (1998), "Models for the association between ordinal variables," *Computational Statistics & Data Analysis*, 28, 387–411.

Li, C. and Shepherd, B. (2010), "Test of association between two ordinal variables while adjusting for covariates," *Journal of the American Statistical Association*, 105, 612–620.

— (2012), "A new residual for ordinal outcomes," *Biometrika*, 99, 473–480.

Li, S., Zhu, X., Chen, Y., and Liu, D. (2020), *PAsso: an R Package for Assessing Partial Association between Ordinal Variables*, R package version 4.0.2.

Liu, D. and Zhang, H. (2018), "Residuals and diagnostics for ordinal regression models: a surrogate approach," *Journal of the American Statistical Association*, 113, 845–854.

Liu, I. and Agresti, A. (2005), "The analysis of ordered categorical data: An overview and a survey of recent developments (with discussion)," *Test*, 14, 1–73.

Liu, I.-M. and Agresti, A. (1996), "Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response," *Biometrics*, 52, 1223–1234.

Mantel, N. and Haenszel, W. (1959), "Statistical aspects of the analysis of data from retrospective studies of disease," *Journal of the National Cancer Institute*, 22, 719–748.

McCullagh, P. (1980), "Regression models for ordinal data (with discussion)," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 109–142.

McKelvey, R. D. and Zavoina, W. (1975), "A statistical model for the analysis of ordinal level dependent variables," *Journal of Mathematical Sociology*, 4, 103–120.

Olsson, U. (1979), "Maximum likelihood estimation of the polychoric correlation coefficient," *Psychometrika*, 44, 443–460.

Pearson, K. and Pearson, E. S. (1922), "On polychoric coefficients of correlation," *Biometrika*, 14, 127–156.

Prentice, R. L. and Gloeckler, L. A. (1978), "Regression analysis of grouped survival data with application to breast cancer data," *Biometrics*, 34, 57–67.

Rigdon, E. E. and Ferguson Jr, C. E. (1991), "The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data," *Journal of Marketing Research*, 28, 491–497.

Rosenstone, S. J., Kinder, D. R., Miller, W. E., and National Election Studies. University of Michigan. Center for Political Studies (2016), "American National Election Study, 1996: Pre- and Post-Election Survey," `https://doi.org/10.3886/ICPSR06896.v6`.

Schweizer, B. and Wolff, E. F. (1981), "On nonparametric measures of dependence for random variables," *The Annals of Statistics*, 9, 879–885.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Tallis, G. (1962), "The maximum likelihood estimation of correlation from contingency tables," *Biometrics*, 18, 342–353.

Tupes, E. and Christal, R. (1961), "Recurrent personality factors based on trait ratings (Tech. Rep.)," *Lackland Air Force Base, TX: USAF*.

Tutz, G. (2012), *Regression for Categorical Data*, Cambridge University Press: Cambridge,UK.

Yiu, C.-F. and Poon, W.-Y. (2008), "Estimating the polychoric correlation from misclassified data," *British Journal of Mathematical and Statistical Psychology*, 61, 49–74.

Zhang, H. (2011), "Statistical analysis in genetic studies of mental illnesses," *Statistical Science*, 26, 116–129.

Zhu, W., Jiang, Y., and Zhang, H. (2012), "Nonparametric covariate-adjusted association tests based on the generalized Kendall's tau," *Journal of the American Statistical Association*, 107, 1–11.

Appendix A. Technical details for Section 4 of Inference

In this appendix, we show a general procedure by which one can verify the consistency of the estimate $\hat{\phi} = \hat{\phi}_{\text{empirical}}(\boldsymbol{r}_{1,\tilde{\boldsymbol{\theta}}_1}, \boldsymbol{r}_{2,\tilde{\boldsymbol{\theta}}_2})$ defined in Section 4.

For a specific measure ϕ of interest. we define an intermediate quantity

$$\tilde{\phi} = \phi(R_{1,\tilde{\boldsymbol{\theta}}_1}, R_{2,\tilde{\boldsymbol{\theta}}_2}). \tag{12}$$

The following result shows that $\tilde{\phi}_\rho$ converges to ϕ_ρ in probability under mild conditions. Similar results can be established for ϕ_τ and ϕ_σ.

Theorem 4. *Assuming that $\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k = o_p(1)$ and the conditions below hold*

(C1) $f_k(\boldsymbol{X}, \boldsymbol{\beta}_k)$ and $\partial f_k(\boldsymbol{X}, \boldsymbol{\beta}_k)/\partial \boldsymbol{\beta}_k$ are bounded in probability;

(C2) The density function $g_k(u) = G'_k(u)$ is bounded from above.

Then, the random variable $\tilde{\phi}_\rho = \phi_\rho + o_p(1)$.

Given $\tilde{\phi} \to \phi$ as seen in Theorem 4, the general result below states that the estimate $\hat{\phi} = \hat{\phi}_{\text{empirical}}(\boldsymbol{r}_{1,\tilde{\boldsymbol{\theta}}_1}, \boldsymbol{r}_{2,\tilde{\boldsymbol{\theta}}_2})$ is consistent provided that the condition (C3) holds.

(C3) $\hat{\phi} \to \tilde{\phi}$ in probability uniformly in a neighborhood $\mathcal{B}(d_1, d_2) = \{(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2) : \parallel \tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1 \parallel \leq d_1, \parallel \tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2 \parallel \leq d_2\}$ of the true value $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

Here, the uniform convergence means that for any $\varepsilon > 0$ and $\delta > 0$, there exists an N such that for all $n > N$ and $(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2) \in \mathcal{B}(d_1, d_2)$, $\Pr\{|\hat{\phi} - \tilde{\phi}| > \varepsilon\} < \delta$.

Theorem 5. *The estimate $\hat{\phi} = \hat{\phi}_{\text{empirical}}(\boldsymbol{r}_{1,\tilde{\boldsymbol{\theta}}_1}, \boldsymbol{r}_{2,\tilde{\boldsymbol{\theta}}_2})$ converges to ϕ in probability, provided that the result in Theorem 4 and the condition (C3) hold.*

Appendix B: Proofs

Proofs for Section 2.3: A key result

Proof of Theorem 1. First, we prove $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Rightarrow (R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}$. By the definition, the surrogate variable $S_k \sim (Z_k \mid \alpha_{k,Y_k=y_k-1} < Z_k \leq \alpha_{k,Y_k=y_k})$ where $Z_k = f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + \epsilon_k$. Thus, if $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$, we have $(S_k \perp\!\!\!\perp S_l) \mid \boldsymbol{X}$. It follows immediately $(R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}$ as $R_k = S_k - f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) - \int_{-\infty}^{\infty} u \, \mathrm{d}G_k(u)$.

On the other side, we show that $(R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X} \Rightarrow (S_k \perp\!\!\!\perp S_l) \mid \boldsymbol{X} \Rightarrow (Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$. The first "$\Rightarrow$" is a result of $S_k = f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + \int_{-\infty}^{\infty} u \, \mathrm{d}G_k(u) + R_k$. The second "$\Rightarrow$" is due to the fact that $Y_k = j$ if and only if $\alpha_{k,j-1} < S_k \leq \alpha_{k,j}$. This completes the proof. $\square$

Proof of Corollary 1. By Theorem 1, $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Rightarrow (R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}$. Thus,

$$\mathbb{P}(R_k \leq r_k, R_l \leq r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \leq r_k \mid \boldsymbol{X})\mathbb{P}(R_l \leq r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \leq r_k)\mathbb{P}(R_l \leq r_l).$$

The last equation is due to the fact that the distribution of R_k does not depend on $\boldsymbol{X}$, i.e., $R_k \perp\!\!\!\perp \boldsymbol{X}$ (see Theorem 2(b) in Liu and Zhang (2018)). If $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$ for any $\boldsymbol{X}$, the equation above holds for any $\boldsymbol{X}$, which establishes $R_k \perp\!\!\!\perp R_l$. $\square$

Proofs for Section 3.1: Linear association

Proof of Property 2. The result in Part(a) holds by definition. To see Part(b), note that under Model 2, $\rho = 0 \Rightarrow Y_1 \perp\!\!\!\perp Y_2 \mid \boldsymbol{X} \Rightarrow R_1 \perp\!\!\!\perp R_2$, by Theorem 1. Thus, $\phi_\rho = \rho(R_1, R_2) = 0$. The conclusion in Part(c) is justified by the numerical result in Table 2.

In what follows, we prove the result in Part(d). It is known that the surrogate residual R_k has the same distribution as the latent error ϵ_k, $k = 1, 2$ (Liu and Zhang, 2018, Theorem 3). Hence, $\mathbb{E}(R_k) = \mathbb{E}(\epsilon_k) = 0$ and $\mathrm{Var}(R_k) = \mathrm{Var}(\epsilon_k)$.

Given $\boldsymbol{X}$, we write the surrogate residual R_k as

$$R_1 \mid \boldsymbol{X} = \tilde{\epsilon}_{11}\mathbf{1}(\alpha^*_{1,0} < \epsilon_1 \le \alpha^*_{1,1}) + \tilde{\epsilon}_{12}\mathbf{1}(\alpha^*_{1,1} < \epsilon_1 \le \alpha^*_{1,2}) + \cdots + \tilde{\epsilon}_{1J_1}\mathbf{1}(\alpha^*_{1,J_1-1} < \epsilon_1 \le \alpha^*_{1,J_1}),$$

$$R_2 \mid \boldsymbol{X} = \tilde{\epsilon}_{21}\mathbf{1}(\alpha^*_{2,0} < \epsilon_2 \le \alpha^*_{2,1}) + \tilde{\epsilon}_{22}\mathbf{1}(\alpha^*_{2,1} < \epsilon_2 \le \alpha^*_{2,2}) + \cdots + \tilde{\epsilon}_{2J_2}\mathbf{1}(\alpha^*_{2,J_2-1} < \epsilon_2 \le \alpha^*_{2,J_2}),$$

where $\alpha^*_{k,j} = \alpha_{k,j} - f_k(\mathbf{X}, \boldsymbol{\beta}_k)$, and $\tilde{\epsilon}_{kj}$ represents a truncated random variable, i.e., $\tilde{\epsilon}_{kj} \overset{\text{ind.}}{\sim} \epsilon_k | \alpha^*_{k,j-1} < \epsilon_k \le \alpha^*_{k,j}$, independent of all other variables. Then,

$$
\begin{aligned}
\mathbb{E}(R_1 R_2 \mid \boldsymbol{X}) &= \mathbb{E}\left[\sum_{i=1}^{J_1} \sum_{j=1}^{J_2} \tilde{\epsilon}_{1i}\mathbf{1}(\alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i})\tilde{\epsilon}_{2j}\mathbf{1}(\alpha^*_{2,j-1} < \epsilon_2 \le \alpha^*_{2,j}) \right] \\
&= \sum_{i=1}^{J_1} \sum_{j=1}^{J_2} \mathbb{E}\tilde{\epsilon}_{1i}\mathbb{E}\tilde{\epsilon}_{2j}\mathbb{P}\left(\alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i}, \alpha^*_{2,j-1} < \epsilon_2 \le \alpha^*_{2,j} \right) \\
&= \sum_{i=1}^{J_1} \sum_{j=1}^{J_2} \mathbb{E}\left[\epsilon_1 | \alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i} \right] \mathbb{E}\left[\epsilon_2 | \alpha^*_{1,j-1} < \epsilon_2 \le \alpha^*_{2,j} \right] \\
&\quad \times \mathbb{P}\left(\alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i}, \alpha^*_{2,j-1} < \epsilon_2 \le \alpha^*_{2,j} \right).
\end{aligned}
$$

Let $M_k(\boldsymbol{X})$ be a discrete random variable, taking the value of $\mathbb{E}\left[\epsilon_k | \alpha^*_{k,j-1} < \epsilon_k \le \alpha^*_{k,j} \right]$ with a probability of $\mathbb{P}\left(\alpha^*_{k,j-1} < \epsilon_k \le \alpha^*_{k,j} \right)$. Then, $\mathbb{E}(R_1 R_2 \mid \boldsymbol{X}) = \mathbb{E}(M_1(\boldsymbol{X})M_2(\boldsymbol{X}) \mid \boldsymbol{X})$. In what follows, we will show that $\mathbb{E}(M_1(\boldsymbol{X})M_2(\boldsymbol{X}) \mid \boldsymbol{X}) \to \mathbb{E}(\epsilon_1\epsilon_2)$ at the same convergence rate for any value of $\boldsymbol{X}$. To see this, we write

$$
\begin{aligned}
\mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X})M_2(\boldsymbol{X}) - \epsilon_1\epsilon_2| &\le \mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X})(M_2(\boldsymbol{X}) - \epsilon_2)| + \mathbb{E}_{\cdot|\boldsymbol{X}}|\epsilon_2(M_1(\boldsymbol{X}) - \epsilon_1)| \\
&\le \sqrt{\mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X})|^2\mathbb{E}_{\cdot|\boldsymbol{X}}|M_2(\boldsymbol{X}) - \epsilon_2|^2} + \sqrt{\mathbb{E}|\epsilon_2|^2\mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X}) - \epsilon_1|^2},
\end{aligned}
$$

where $\mathbb{E}_{\cdot|\boldsymbol{X}}$ denotes the conditional expectation given $\boldsymbol{X}$. The last inequality is from Cauchy-Schwarz inequality. We can also establish the following two inequalities

$$
\begin{aligned}
\mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X})|^2 &= \sum_{i=1}^{J_1} \left[\mathbb{E}\left(\epsilon_1 | \alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i} \right) \right]^2 \mathbb{P}\left(\alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i} \right) \\
&\le \sum_{i=1}^{J_1} \mathbb{E}\left(\epsilon_1^2 | \alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i} \right) \mathbb{P}\left(\alpha^*_{1,i-1} < \epsilon_1 \le \alpha^*_{1,i} \right) = \mathbb{E}|\epsilon_1|^2;
\end{aligned}
$$

and

$$\mathbb{E}_{\cdot|\boldsymbol{X}}|M_k(\boldsymbol{X}) - \epsilon_k|^2 = \int_\Omega |M_k(\boldsymbol{X})(\omega) - \epsilon_k(\omega)|^2 d\mathbf{P}(\omega)$$

$$\leq \int_\Omega \max_j |\alpha_{k,j+1} - \alpha_{k,j}|^2 d\mathbf{P}(\omega) = \max_j |\alpha_{k,j+1} - \alpha_{k,j}|^2.$$

Therefore

$$\mathbb{E}_{\cdot|\boldsymbol{X}}|M_1(\boldsymbol{X})M_2(\boldsymbol{X}) - \epsilon_1\epsilon_2| \leq \left(\sqrt{\mathbb{E}(\epsilon_1^2)} + \sqrt{\mathbb{E}(\epsilon_2^2)} \right) \max_j |\alpha_{k,j+1} - \alpha_{k,j}|$$

Since the bound on the right side of the above inequality does not depend on $\boldsymbol{X}$, it also holds for the unconditional expectation $\mathbb{E}|M_1(\boldsymbol{X})M_2(\boldsymbol{X}) - \epsilon_1\epsilon_2|$. As a result, $\mathbb{E}(R_1 R_2) = \mathbb{E}(M_1(\boldsymbol{X})M_2(\boldsymbol{X})) \to \mathbb{E}(\epsilon_1\epsilon_2)$, as long as both ϵ_1 and ϵ_2 have finite second moments. This completes the proof. □

Proofs for Section 3.3: Nonlinear but monotonic association

Proof of Property 4. Define the following events:

$$C_Y = \{\omega : (Y_1 - Y_1^*)(Y_2 - Y_2^*) > 0\}, \qquad C_R = \{\omega : (R_1 - R_1^*)(R_2 - R_2^*) > 0\}$$

$$D_Y = \{\omega : (Y_1 - Y_1^*)(Y_2 - Y_2^*) < 0\}, \qquad D_R = \{\omega : (R_1 - R_1^*)(R_2 - R_2^*) < 0\}$$

$$E_Y = \{\omega : (Y_1 - Y_1^*)(Y_2 - Y_2^*) = 0\}, \qquad E_R = \{\omega : (R_1 - R_1^*)(R_2 - R_2^*) = 0\}$$

In the absence of $\boldsymbol{X}$, we have $y_k < y_k^* \Rightarrow r_k < r_k^*$ and $y_k > y_k^* \Rightarrow r_k > r_k^*$. Thus,

$$\mathbb{P}(C_R) = \mathbb{P}(C_R \cap C_Y) + \mathbb{P}(C_R \cap D_Y) + \mathbb{P}(C_R \cap E_Y) = \mathbb{P}(C_Y) + \mathbb{P}(C_R \cap E_Y).$$

Similarly, we have $\mathbb{P}(D_R) = \mathbb{P}(D_Y) + \mathbb{P}(D_R \cap E_Y)$. Given $y_k = y_k^*$, the conditional probabilities $\mathbb{P}(r_k < r_k^* \mid y_k = y_k^*) = \mathbb{P}(r_k > r_k^* \mid y_k = y_k^*)$, since r_k is a random sample given y_k. As a result, $\mathbb{P}(C_R \cap E_Y) = \mathbb{P}(D_R \cap E_Y)$. Therefore,

$$\phi_\tau(R_1, R_2) = \tau(R_1, R_2) = \mathbb{P}(C_R) - \mathbb{P}(D_R) = \mathbb{P}(C_Y) - \mathbb{P}(D_Y) = \tau(Y_1, Y_2).$$

This completes the proof. □

Proofs for Section 3.4: General association

Proof of Theorem 2. This is a direct result of Sklar's theorem, since both R_1 and R_2 are continuous variables and the marginal distribution of R_k is G_k. $\qquad\square$

Proof of Corollary 2. In view of Theorem 1, Y_1 and Y_2 are partially independent if and only if the surrogate residual variables R_1 and R_2 are independent. The later condition is equivalent to requiring the copula C in Theorem 2 is a product copula, i.e., $C(u, v) = uv$. $\qquad\square$

Proofs for Section 7: Extension to general models

Proof of Theorem 3. As $R_k = S_k - E(S_k \mid \boldsymbol{X}) = S_k - 1/2$, we only need to show $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Leftrightarrow (S_k \perp\!\!\!\perp S_l) \mid \boldsymbol{X}$. By the definition of $S_k \mid Y_k = j \sim U[F_k(j-1), F_k(j))$, $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Rightarrow (S_k \perp\!\!\!\perp S_l) \mid \boldsymbol{X}$. One the other hand, because $F_k(j-1) \leq S_k < F(j) \Rightarrow Y_k = j$, we immediately have $(S_k \perp\!\!\!\perp S_l) \mid \boldsymbol{X} \Rightarrow (Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$. $\qquad\square$

Proof of Corollary 3. By Theorem 3, $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X} \Rightarrow (R_k \perp\!\!\!\perp R_l) \mid \boldsymbol{X}$. Thus,

$$\mathbb{P}(R_k \leq r_k, R_l \leq r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \leq r_k \mid \boldsymbol{X})\mathbb{P}(R_l \leq r_l \mid \boldsymbol{X}) = \mathbb{P}(R_k \leq r_l)\mathbb{P}(R_k \leq r_l).$$

The last equation is due to the fact that the distribution of R_k does not depend on $\boldsymbol{X}$, i.e., $R_k \perp\!\!\!\perp \boldsymbol{X}$ (see Theorem 4(b) in Liu and Zhang (2018)). If $(Y_k \perp\!\!\!\perp Y_l) \mid \boldsymbol{X}$ for any $\boldsymbol{X}$, the equation above holds for any $\boldsymbol{X}$, which establishes $R_k \perp\!\!\!\perp R_l$. $\qquad\square$

Proofs for the theorems in Appendix A

Proof of Theorem 4. First, we show that $E(R_{k,\tilde{\boldsymbol{\theta}}_k}) = E(R_{k,\boldsymbol{\theta}_k} \mid \boldsymbol{\theta}_k = \tilde{\boldsymbol{\theta}}_k)$, as a function of $\tilde{\boldsymbol{\theta}}_k$, satisfies that

$$E(R_{k,\tilde{\boldsymbol{\theta}}_k}) = E(R_{k,\boldsymbol{\theta}_k}) + o_p(1). \tag{13}$$

40

Without loss of generality, we assume $\int_{-\infty}^{\infty} u \, dG_k(u) = 0$. Then,

$$E(R_{k,\tilde{\boldsymbol{\theta}}_k} \mid \boldsymbol{X}) = \sum_{j=1}^{J_k} E_{\tilde{\boldsymbol{\theta}}_k}(R_k \mid Y_k = j, \boldsymbol{X}) \Pr(Y_k = j \mid \boldsymbol{X})$$

$$= \sum_{j=1}^{J_k} \left\{ E_{\tilde{\boldsymbol{\theta}}_k}(S_k \mid Y_k = j, \boldsymbol{X}) - f_k(\boldsymbol{X}, \tilde{\boldsymbol{\beta}}_k) \right\} \Pr(Y_k = j \mid \boldsymbol{X}).$$

Note that $f_k(\boldsymbol{X}, \tilde{\boldsymbol{\beta}}_k) \doteq f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + \partial f_k(\boldsymbol{X}, \boldsymbol{\beta}_k)/\partial \boldsymbol{\beta}_k'(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) = f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + o_p(1)$, due to the condition (C1). Let $\gamma_{k,j} = \alpha_{k,j} - f_k(\boldsymbol{X}, \boldsymbol{\beta}_k)$ and $\tilde{\gamma}_{k,j} = \tilde{\alpha}_{k,j} - f_k(\boldsymbol{X}, \tilde{\boldsymbol{\beta}}_k)$, then $\tilde{\gamma}_{k,j} = \gamma_{k,j} + o_p(1)$. Therefore, we have

$$\left\{ E_{\tilde{\boldsymbol{\theta}}_k}(S_k \mid Y_k = j, \boldsymbol{X}) - f_k(\boldsymbol{X}, \tilde{\boldsymbol{\beta}}_k) \right\} - \left\{ E_{\boldsymbol{\theta}_k}(S_k \mid Y_k = j, \boldsymbol{X}) - f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) \right\}$$

$$= \int_{\tilde{\gamma}_{k,j-1}}^{\tilde{\gamma}_{k,j}} u \, dG_k(u) - \int_{\gamma_{k,j-1}}^{\gamma_{k,j}} u \, dG_k(u)$$

$$= \int_{\gamma_{k,j}}^{\tilde{\gamma}_{k,j}} u \, dG_k(u) - \int_{\gamma_{k,j-1}}^{\tilde{\gamma}_{k,j-1}} u \, dG_k(u)$$

$$= v_{k,j} g_k(v_{k,j})(\tilde{\gamma}_{k,j} - \gamma_{k,j}) - v_{k,j-1} g_k(v_{k,j-1})(\tilde{\gamma}_{k,j-1} - \gamma_{k,j-1})$$

by the mean value theorem, where $v_{k,j}$ is a value between $\tilde{\gamma}_{k,j}$ and $\gamma_{k,j}$. Since the condition (C2) holds, the last quantity above is $o_p(1)$. We immediately have

$$E(R_{k,\tilde{\boldsymbol{\theta}}_k} \mid \boldsymbol{X}) = \sum_{j=1}^{J_k} \left\{ E_{\boldsymbol{\theta}_k}(S_k \mid Y_k = j, \boldsymbol{X}) - f_k(\boldsymbol{X}, \boldsymbol{\beta}_k) + o_p(1) \right\} \Pr(Y_k = j \mid \boldsymbol{X})$$

$$= E(R_{k,\boldsymbol{\theta}_k} \mid \boldsymbol{X}) + o_p(1).$$

Since the $o_p(1)$ term in the last quantity above does not depend on the value of $\boldsymbol{X}$, we conclude that the equality in (13) holds.

Similarly, we can establish

$$E(R_{k,\tilde{\boldsymbol{\theta}}_k}^2) = E(R_{k,\boldsymbol{\theta}_k}^2) + o_p(1); \tag{14}$$

$$E(R_{1,\tilde{\boldsymbol{\theta}}_1} R_{2,\tilde{\boldsymbol{\theta}}_2}) = E(R_{1,\boldsymbol{\theta}_1} R_{2,\boldsymbol{\theta}_2}) + o_p(1). \tag{15}$$

Combining the results in (13), (14) and (15), we can see

$$
\begin{aligned}
\tilde{\phi}_\rho = \rho(R_{1,\tilde{\boldsymbol{\theta}}_1}, R_{2,\tilde{\boldsymbol{\theta}}_2}) &= \mathrm{Cov}(R_{1,\tilde{\boldsymbol{\theta}}_1}, R_{2,\tilde{\boldsymbol{\theta}}_2}) \Big/ \sqrt{\mathrm{Var}(R_{1,\tilde{\boldsymbol{\theta}}_1})\mathrm{Var}(R_{2,\tilde{\boldsymbol{\theta}}_2})} \\
&= \{\mathrm{Cov}(R_{1,\boldsymbol{\theta}_1}, R_{2,\boldsymbol{\theta}_2}) + o_p(1)\} \Big/ \left\{ \sqrt{\mathrm{Var}(R_{1,\boldsymbol{\theta}_1})\mathrm{Var}(R_{2,\boldsymbol{\theta}_2})} + o_p(1) \right\} \\
&= \rho(R_{1,\boldsymbol{\theta}_1}, R_{2,\boldsymbol{\theta}_2}) + o_p(1) \\
&= \phi_\rho + o_p(1).
\end{aligned}
$$

This completes the proof. $\qquad \square$

Proof of Theorem 5. For any given $\varepsilon > 0$, we write

$$
\Pr\{|\hat{\phi} - \phi| > \varepsilon\} \leq \Pr\{|\hat{\phi} - \tilde{\phi}| + |\tilde{\phi} - \phi| > \varepsilon\} \leq \Pr\{|\hat{\phi} - \tilde{\phi}| > \varepsilon/2\} + \Pr\{|\tilde{\phi} - \phi| > \varepsilon/2\}.
$$

Since $\Pr\{|\tilde{\phi} - \phi| > \varepsilon/2\} \to 0$, we only need to show $\Pr\{|\hat{\phi} - \tilde{\phi}| > \varepsilon/2\} \to 0$. Notice that

$$
\Pr\{|\hat{\phi} - \tilde{\phi}| > \varepsilon/2\} \leq \Pr\{|\hat{\phi} - \tilde{\phi}| > \varepsilon/2 \,\cap\, (\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2) \in \mathcal{B}(d_1, d_2)\} + \Pr\{(\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2) \notin \mathcal{B}(d_1, d_2)\}.
$$

Both terms on the right side go to zero. This completes the proof. $\qquad \square$

Appendix C. Figures

Section 2.2. The tool: surrogate residual
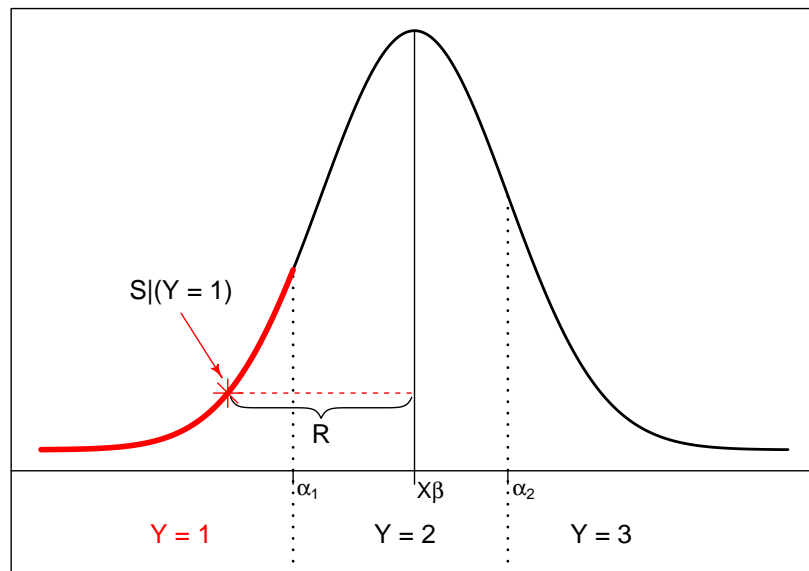
An illustration of the surrogate idea

Figure 1: An illustration of the surrogate variable S and surrogate residual R. Considered is the model $G^{-1}\left(\Pr\{Y \leq j\}\right) = \alpha_j - X\beta, \ j = 1, 2, 3$. The entire curve represents the density function corresponding to G, the inverse of the link function.

Section 2.3 Key results

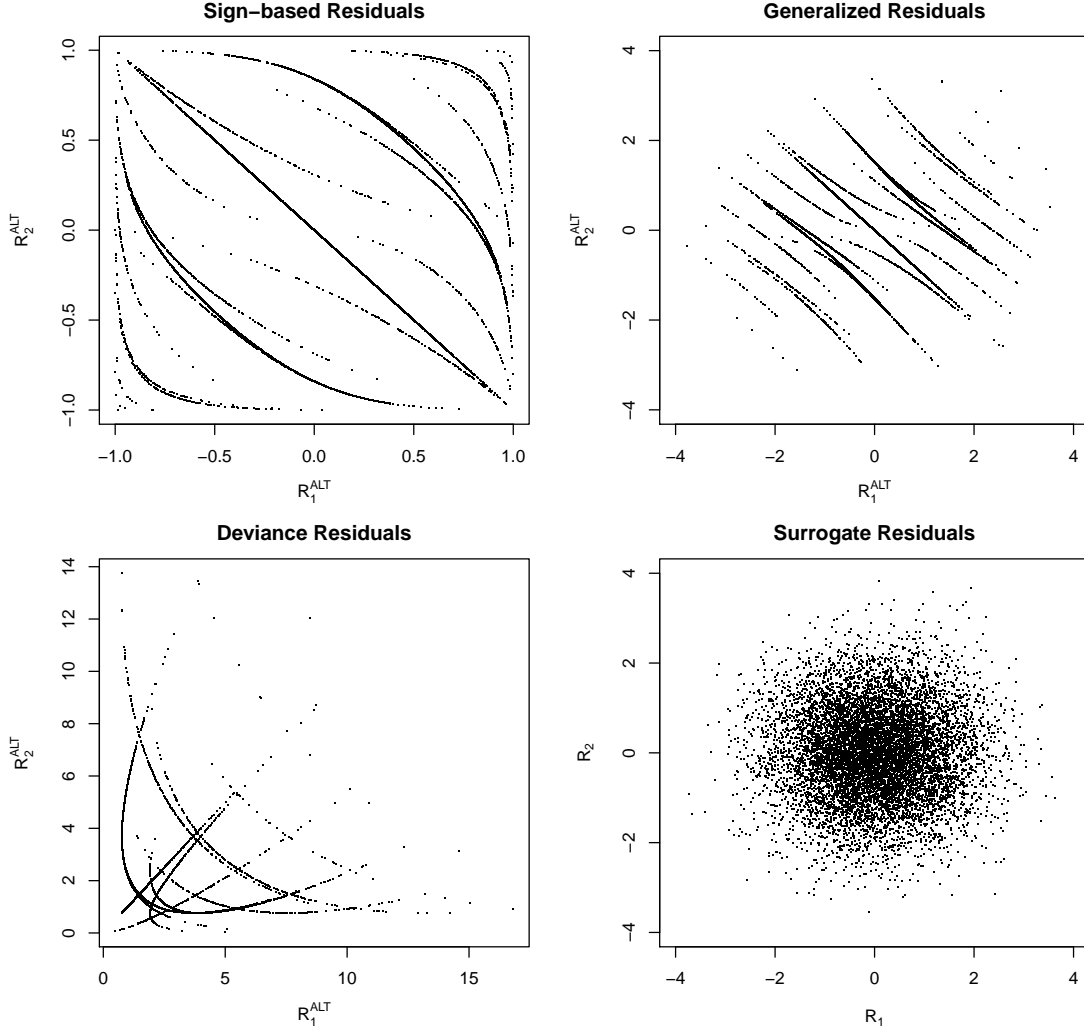An illustration of residuals when two ordinal variables are partially independent

Figure 2: An illustration of the association between the residual variables when the two ordinal variables Y_1 and Y_2 are partially independent. We simulate two partially independent latent variables $Z_1 = X + \epsilon_1$ and $Z_2 = -X + \epsilon_2$, where $X \sim N(0, 1)$ and $(\epsilon_1, \epsilon_2) \sim BN(\mathbf{0}, \mathrm{I})$. The ordinal outcomes Y_1 and Y_2 are obtained by applying the cutpoints $\{-3, -2, 0, 2, 3\}$ and $\{-2, 0, 2\}$ to Z_1 and Z_2, respectively. Displayed is a random draw of 10000 observations.

Section 3.1 Linear association

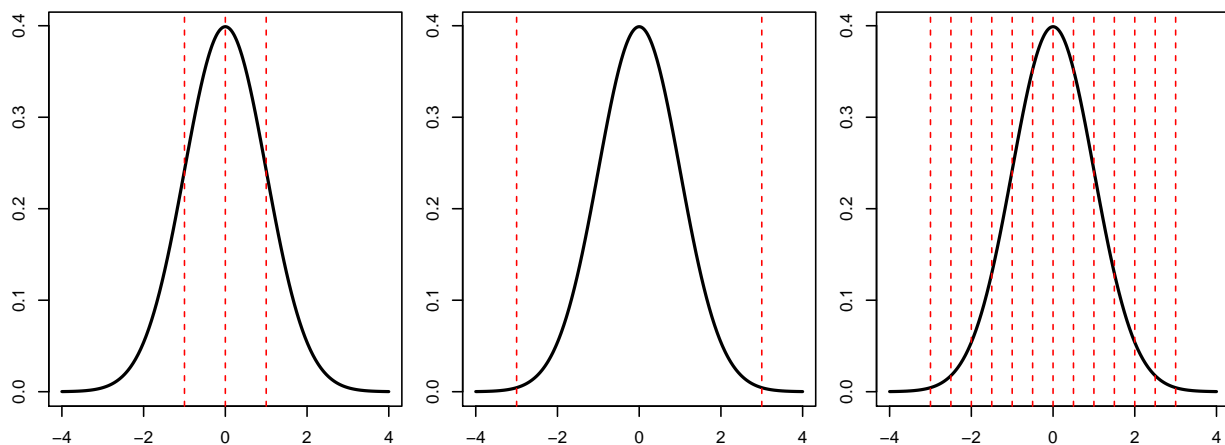An illustration of the influence of cutpoints

Figure 3: An illustration of the influence of cutpoints on the variability of ordinal variables and the strength of association as discussed in Example 1.

Section 3.3 Nonlinear but monotonic association

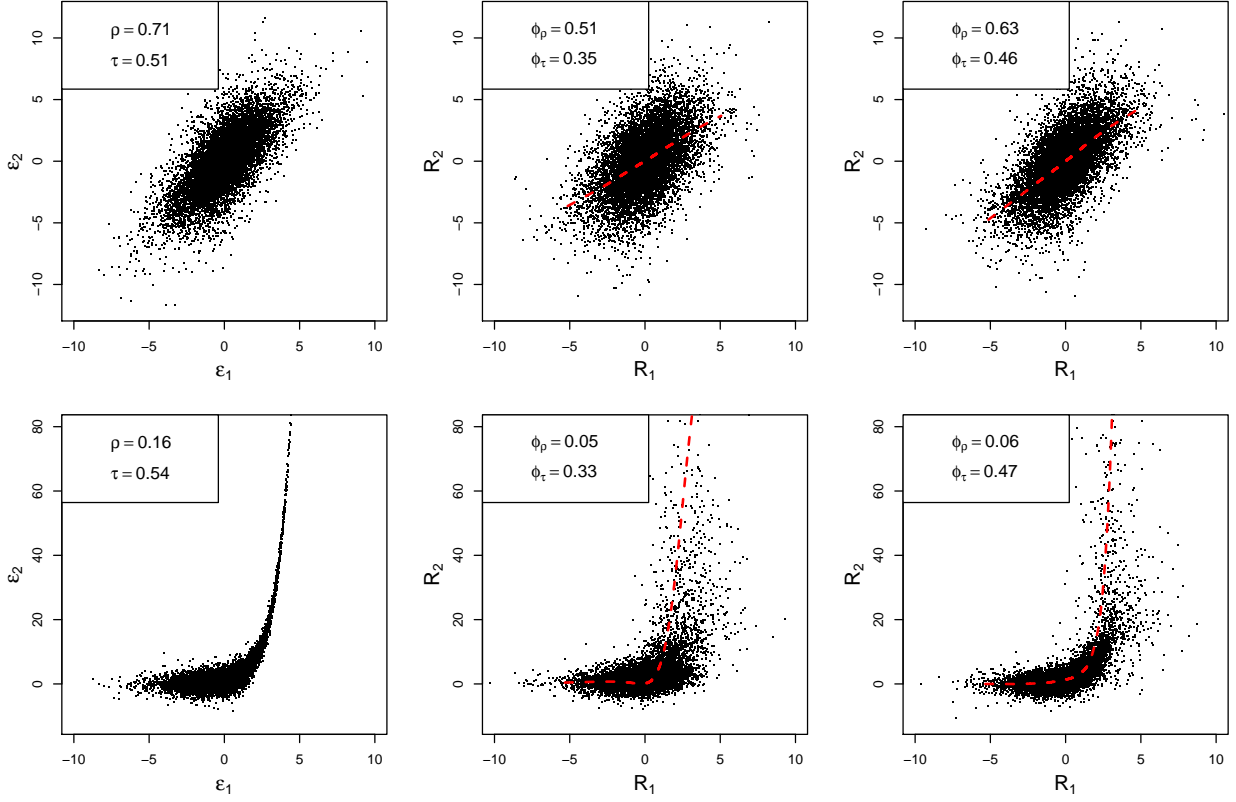An illustration of linear and nonlinear partial associations

Figure 4: An illustration of partial association between two latent variables (left column) and that between two ordinal variables (middle and right columns) as discussed in Example 3. Considered are two scenarios : (1) $\epsilon_2^{(1)} = \epsilon_1 + e$ (top row); and (2) $\epsilon_2^{(2)} = \exp(\epsilon_1) + e$ (bottom row). Displayed are scatter plots of latent errors (left column) and surrogate residuals when the ordinal variables have 5 categories (middle column) or 10 (right column). The dashed red curves are obtained from cubic smoothing splines.

Section 3.4 General association

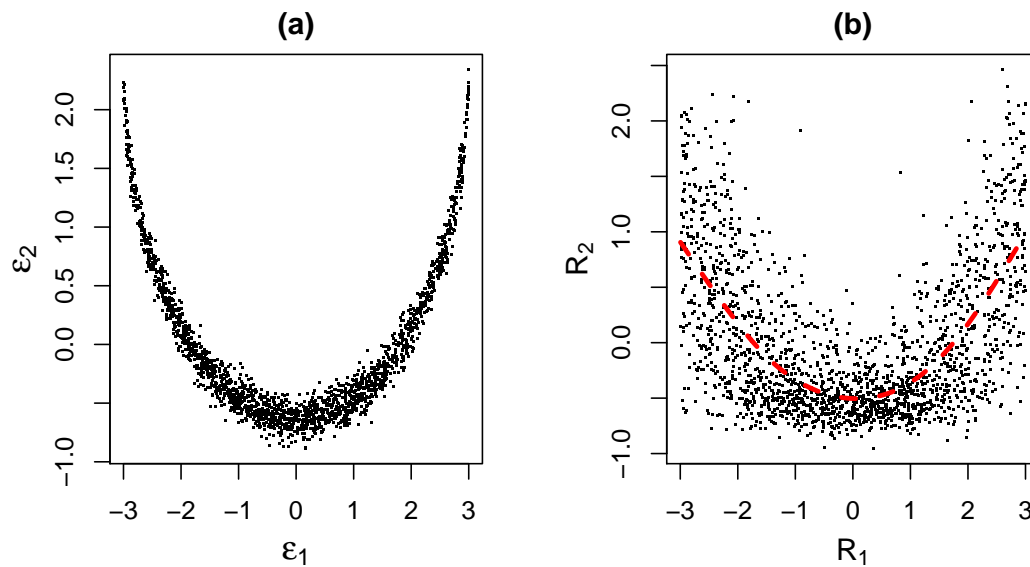An illustration of a non-monotonic partial association

Figure 5: An illustration of a half-circle association, which is not linear nor monotone, as discussed in Example 4. Depicted are scatter plots of 2000 observations of (a) the latent errors ϵ_1 and ϵ_2, and (b) the surrogate residuals R_1 and R_2 when the ordinal variables have 8 categories. The dashed red curve in (b) is obtained from cubic smoothing splines.

Section 3.4 General association

<u>A 3-D P-P plot in the presence of non-monotonic partial dependence</u>
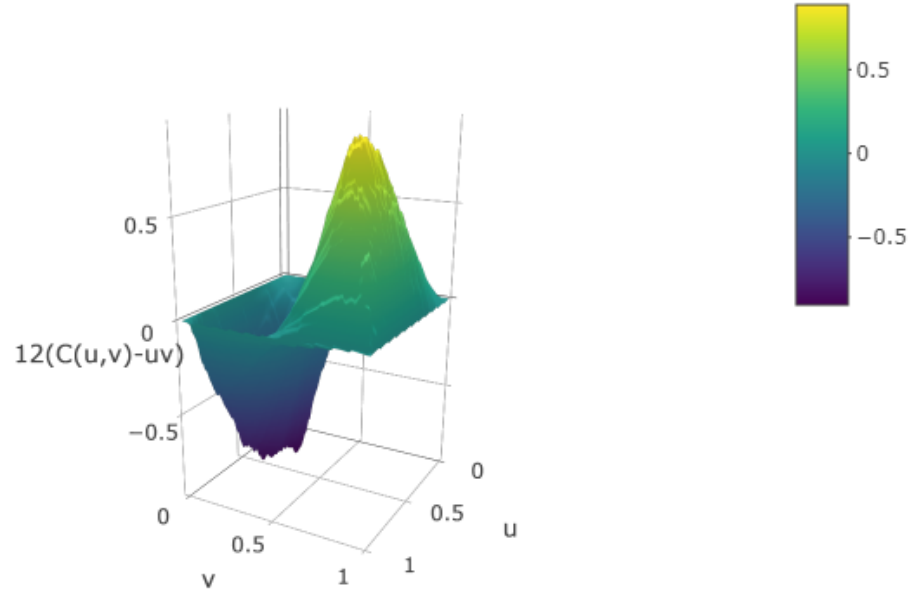

Figure 6: A 3-D plot of $12(C(u,v) - uv)$ where the empirical copula $C(u,v)$ is obtained from surrogate residuals R_1 and R_2 as discussed in Example 4.

Section 5 Simulation studies – Simulation II

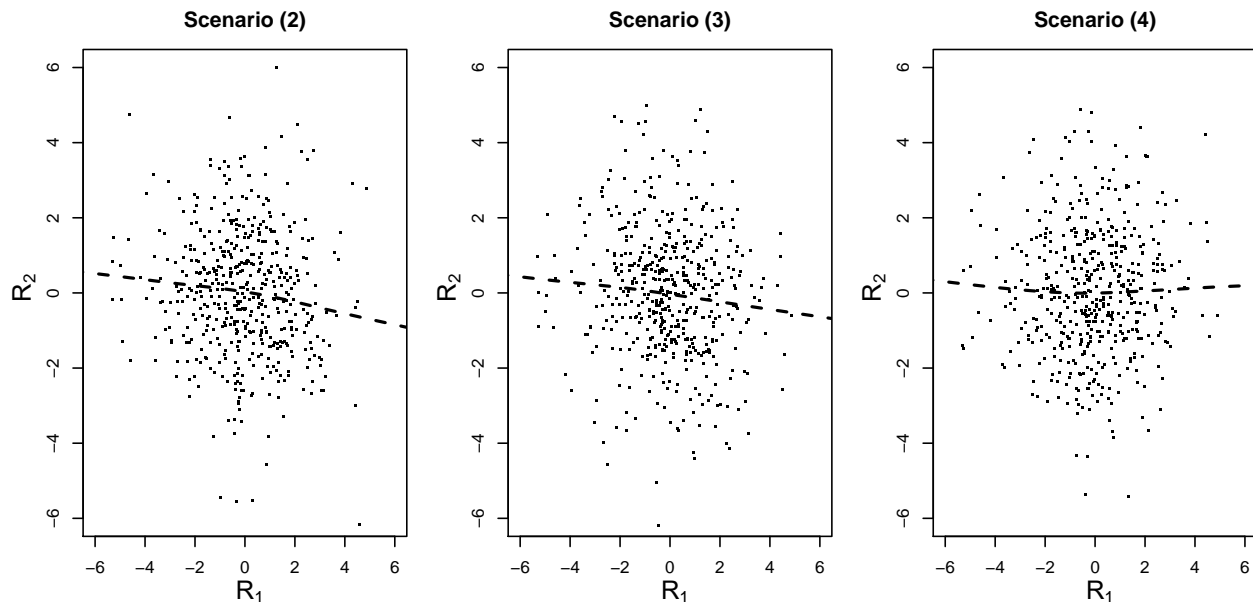A graphical inspection of the scenarios considered by Li and Shepherd (2010)

Figure 7: Partial regression plots for inspecting partial association for the simulation scenarios used by Li and Shepherd (2010). Displayed are scatter plots of surrogate residuals for Scenario (2) $\boldsymbol{\eta} = (-0.4, -0.2, 0, 0.2, 0.4)$ (linear effect); (3) $\boldsymbol{\eta} = (-0.30, 0.18, 0.20, 0.22, 0.24)$ (nonlinear-monotonic effect); and (4) $\boldsymbol{\eta} = (-0.2, 0, 0.2, 0, -0.2)$ (non-monotonic effect).

Section 6.1 Analysis of the National Election Study

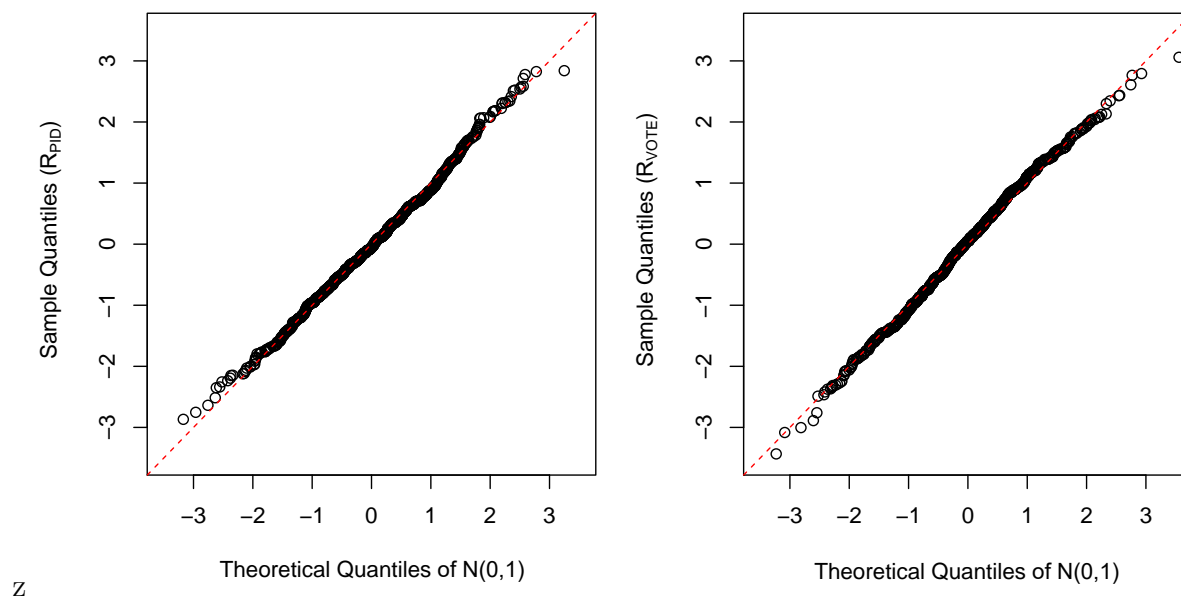Diagnostics of the probit models used for VOTE and PID

Figure 8: Q-Q plots of surrogate residuals versus theoretical normal quantiles. Examined are the probit models for PID (left) and VOTE (right).

Section 6.1 Analysis of the National Election Study

Visualizing the partial association between VOTE and PID
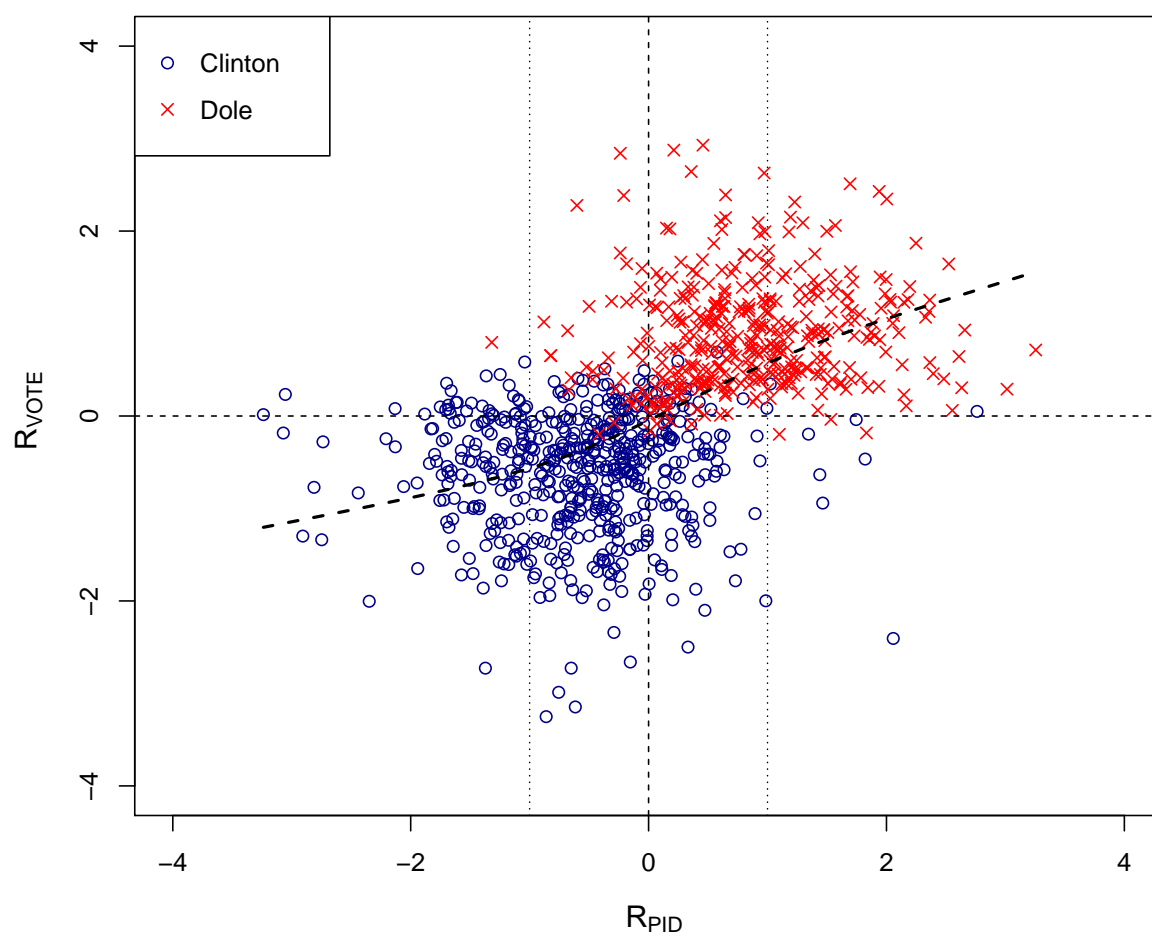
Figure 9: Partial regression plot of VOTE versus PID after adjusting for age, income, and education. Shown is a scatter plot of surrogate residuals for VOTE and PID. Blue circles (or red crosses) indicate those who voted for Clinton (or Dole). The black dashed curve is obtained from cubic smoothing splines.

Section 6.1 Analysis of the National Election Study

Evidence that r_{PID} and r_{VOTE} do not jointly follow a bivariate normal distribution
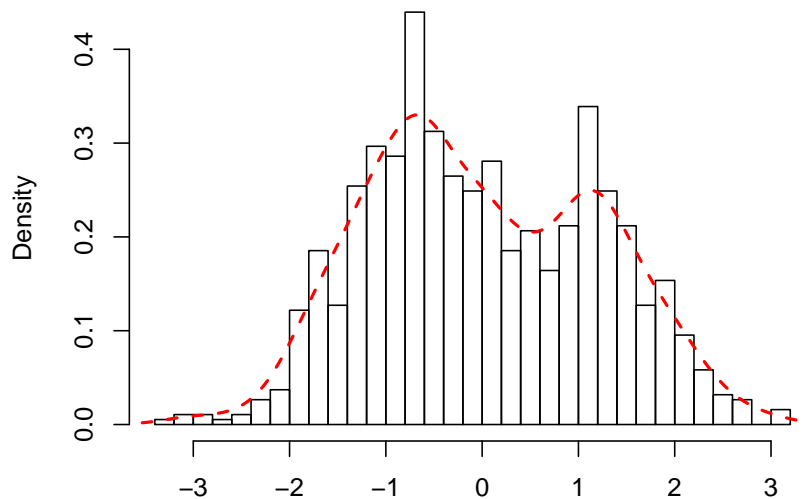
Figure 10: Histogram of a sample projected from (r_{PID}, r_{VOTE}) in Figure 9 to a 60-degree line. The red dashed curve represents the density function fitted using the kernel method.

Section 6.1 Analysis of the National Election Study

Visualizing partial association through the lens of a 3-D P-P plot
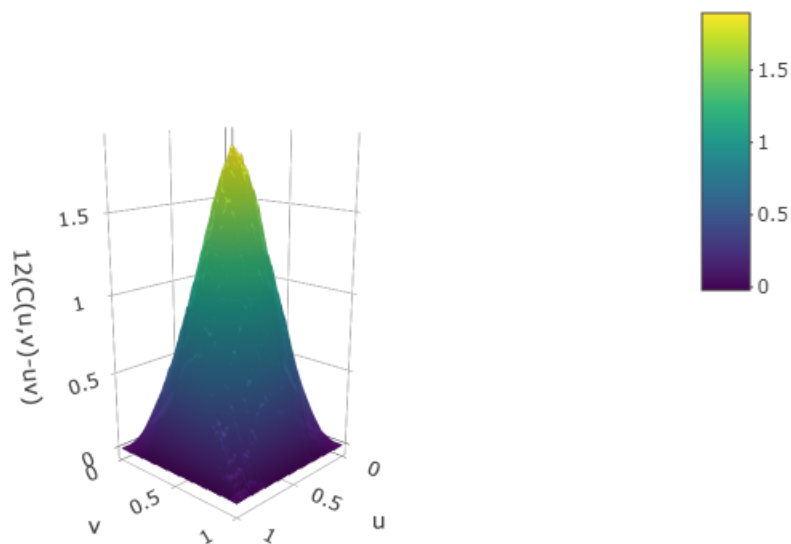
Figure 11: A 3-D P-P plot for inspecting the partial association between PID and VOTE in the National Election Study. Plotted is $12(\hat{C}(u,v) - uv)$ against (u,v), where $\hat{C}(\cdot,\cdot)$ is an empirical copula trained from the data of (r_{PID}, r_{VOTE}).

Section 6.1 Analysis of the National Election Study

Visualizing pairwise partial association between five discrete variables
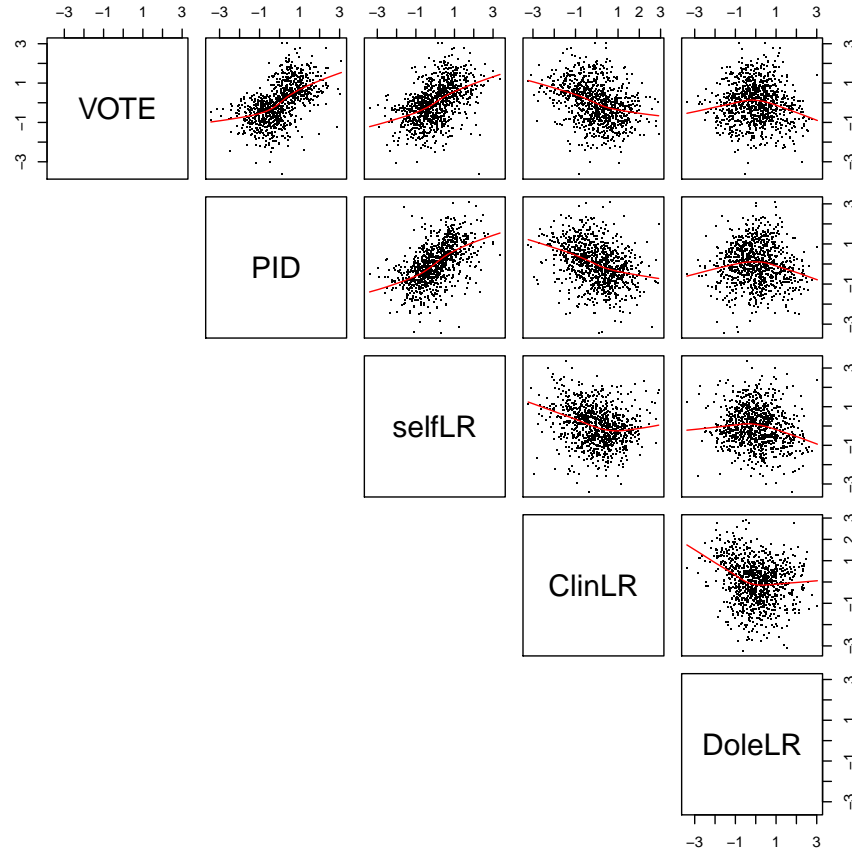
Figure 12: Partial regression plot matrix for 5 discrete variables in the National Election Study after adjusting for age, income, and education. Shown are pairwise scatter plot of surrogate residuals. Added red curves are obtained from cubic smoothing splines.

Section 6.2 Analysis of Big Five personality traits

Partial correlation matrix for 50 survey variables measuring personality traits
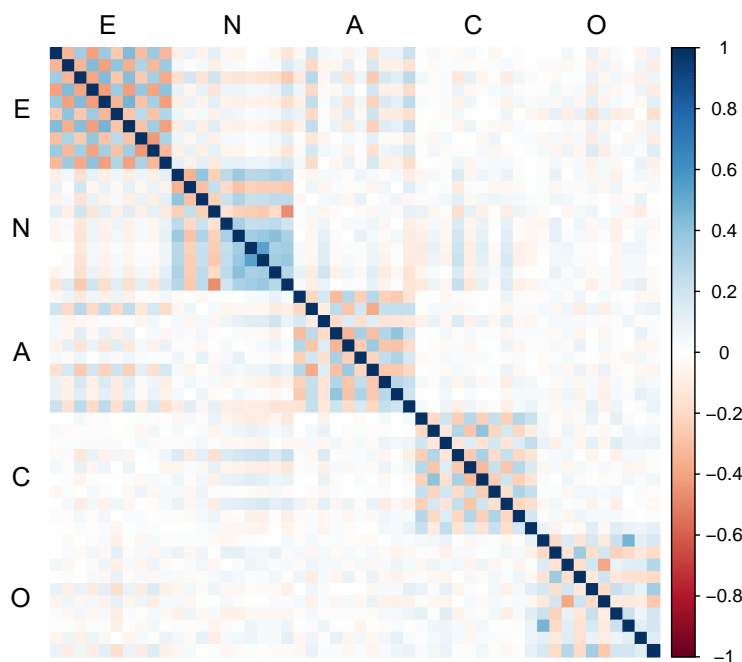
Figure 13: A color-scaled 50×50 matrix of the partial association between 50 survey questions used to characterize Big Five personality traits. Displayed traits are Extraversion (E), Neuroticism (N), Agreeableness (A), Conscientiousness (C), and Openness to experience (O).

Section 6.2 Analysis of Big Five personality traits

Statistical significance for testing the nulls of no or a negligible partial association
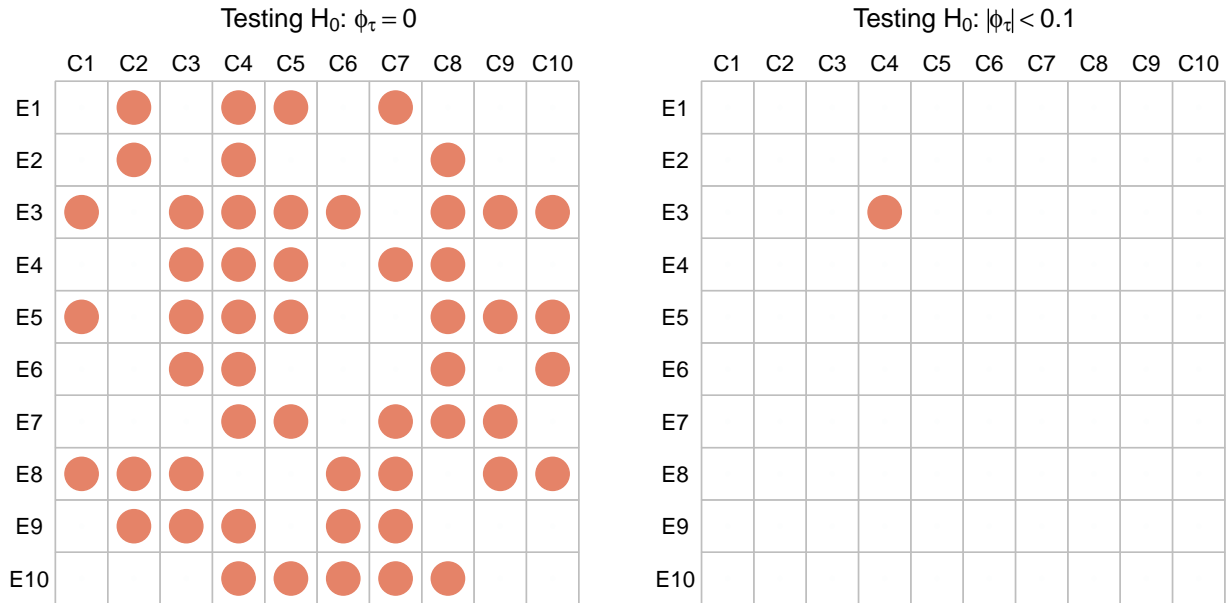
Figure 14: Hypothesis testing of the partial independence between two personality traits Extraversion (E) and Conscientiousness (C). Each trait is measured by 10 survey questions (C1-C10 or E1-E10). Every pair of the survey outcomes is tested with an orange dot indicating a p-value less than 0.05. Displayed are the results for testing H_0: $\phi_\tau = 0$ (left panel) and H_0: $|\phi_\tau| < 0.1$ (right panel).

Section 7 Extension to general models

An illustration of residuals for adjacent-category logit models when two ordinal variables are partially independent
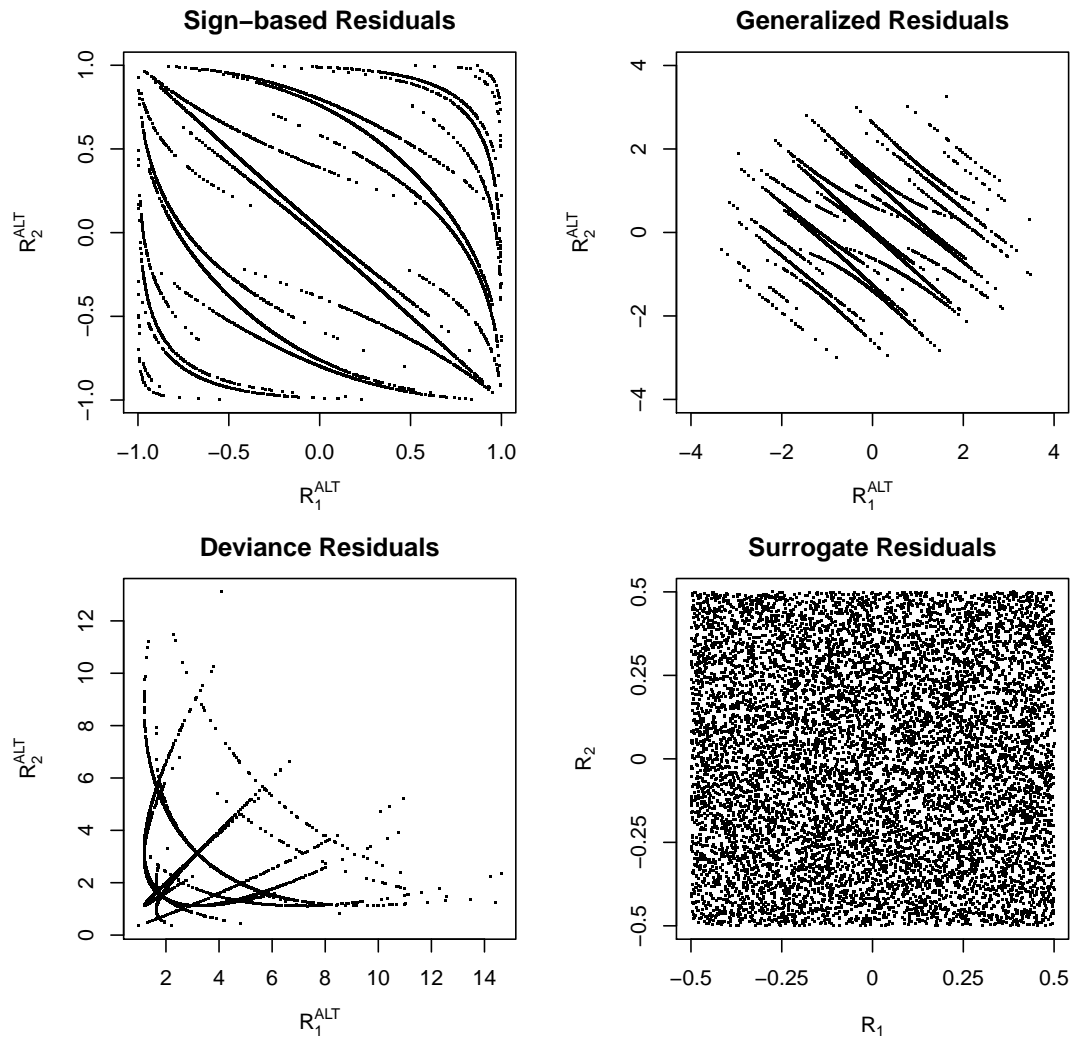
Figure 15: An illustration of the association between the residual variables when the two ordinal variables Y_1 and Y_2 are partially independent. Conditional on $\boldsymbol{X}$, Y_k's are independently simulated from the adjacent-category logit model $\log(\pi_{kj}/\pi_{k,j+1}) = \alpha_{kj} + \boldsymbol{\beta}_k^T \boldsymbol{X}, j = 1, ..., J_k$, where $X \sim N(0,1)$, $\alpha_{1j} = \{-3, -2, 0, 2, 3\}$, $\beta_1 = 1$, $\alpha_{2j} = \{-2, 0, 2\}$, and $\beta_2 = -1$. Displayed is a random draw of 10000 observations.

Section 7 Extension to general models

Visualizing the partial association between VOTE and PID when adjacent-category logit models are used
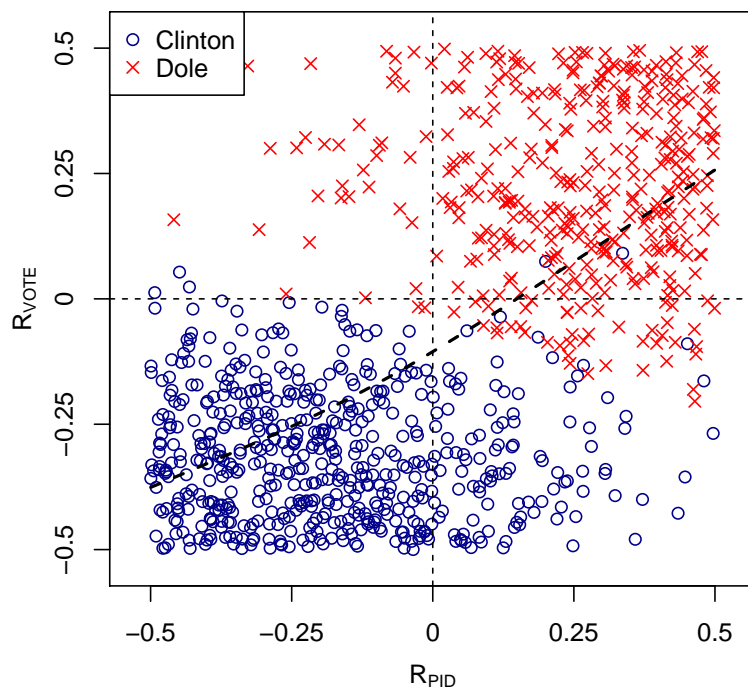
Figure 16: Partial regression plot of VOTE versus PID after adjusting for age, income, and education using the adjacent-category logit model. Shown is a scatter plot of surrogate residuals defined for general models in Section 7. Blue circles (or red crosses) indicate those who voted for Clinton (or Dole). The black dashed curve is obtained from cubic smoothing splines.

Appendix D. Tables

Section 3.1 Linear association

The data set to illustrate the influence of cutpoints (Example 1)

Table 1: Part of the data of ordinal variables $(Y_1, Y_2^{(i)}), i = 1, 2, 3$

Y_1	3	4	3	4	2	2	2	1	2	4	3	2	2	2	1	2	2	4	1	4
$Y_2^{(1)}$	3	3	2	4	2	2	1	1	2	4	3	1	2	3	1	1	2	4	1	3
$Y_2^{(2)}$	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	3	2	2
$Y_2^{(3)}$	9	8	6	10	7	7	4	5	7	13	9	5	6	8	1	4	7	14	4	8

Section 3.1 Linear association

A comparison of partial association measures when the number and positions of cutpoints very (Example 1)

Table 2: Partial association between ordinal variables $(Y_1, Y_2^{(i)}), i = 1, 2, 3$. The three pairs are discretized from the same set of latent variables (Z_1, Z_2) but with different cutpoints. The latent variables (Z_1, Z_2) follow a bivariate normal distribution (Example 1) or a bivariate Gumbel distribution (Example 2).

Ordinal variable	Y_1	$Y_2^{(1)}$	$Y_2^{(2)}$	$Y_2^{(3)}$
	{1, 2, 3, 4}	{1, 2, 3, 4}	{1, 2, 3}	{1, 2,..., 14}
Latent Variable	$Z_1 = X + \epsilon_1$	$Z_2 = 1.5X + \epsilon_2$		
cutpoints α_{kj}'s	{-1, 0, 1}	{-1, 0, 1}	{-3, 3}	{-3, 3, by 0.5}
(ϵ_1, ϵ_2) follow a bivariate normal distribution				
Polychoric correlation ρ		0.80	0.80	0.80
Proposed ϕ_ρ		0.48	0.08	0.61
Proposed ϕ_τ		0.34	0.05	0.44
(ϵ_1, ϵ_2) follow a bivariate Gumbel distribution				
Polychoric correlation ρ		NA	NA	NA
Proposed ϕ_ρ		0.44	0.15	0.54
Proposed ϕ_τ		0.34	0.08	0.43

Section 5 Simulation studies – Simulation I

Approximation accuracy of the bootstrap distribution used for our unified inference

Table 3: Approximation accuracy of the bootstrap distribution $\hat{F}_B^*(\phi)$ to the true distribution of $\hat{\phi}$. Compared are functionals of the two distributions including the mean, standard deviation (SD), 2.5% and 97.5% quantiles, and coverage probability (CP). The study follows the same simulation settings used in Examples 1 and 2.

	$\hat{\phi}_\rho$					$\hat{\phi}_\tau$				
	Mean	SD	2.5%Q	97.5%Q	CP	Mean	SD	2.5%Q	97.5%Q	CP
(ϵ_1, ϵ_2) follow a bivariate normal distribution										
$n = 100$										
True	0.479	0.056	0.363	0.582		0.335	0.045	0.240	0.421	
Bootstrap	0.476	0.054	0.364	0.574	0.941	0.332	0.044	0.243	0.416	0.942
$n = 200$										
True	0.482	0.039	0.401	0.550		0.338	0.031	0.278	0.395	
Bootstrap	0.477	0.038	0.399	0.549	0.938	0.334	0.031	0.271	0.394	0.932
$n = 500$										
True	0.483	0.025	0.436	0.529		0.339	0.020	0.302	0.378	
Bootstrap	0.483	0.024	0.435	0.529	0.951	0.339	0.020	0.300	0.377	0.953
(ϵ_1, ϵ_2) follow a bivariate Gumbel distribution										
$n = 100$										
True	0.441	0.054	0.331	0.536		0.339	0.043	0.250	0.418	
Bootstrap	0.446	0.053	0.338	0.545	0.957	0.342	0.043	0.254	0.424	0.957
$n = 200$										
True	0.437	0.037	0.364	0.510		0.339	0.030	0.280	0.400	
Bootstrap	0.439	0.037	0.364	0.510	0.952	0.341	0.031	0.279	0.400	0.958
$n = 500$										
True	0.436	0.024	0.389	0.483		0.341	0.020	0.300	0.379	
Bootstrap	0.437	0.024	0.389	0.482	0.960	0.340	0.019	0.302	0.378	0.963

Section 5 Simulation studies – Simulation II

Power comparison in the same setting of Li and Shepherd (*JASA*, 2010)

Table 4: Power analysis of Simulation II. The ordinal variable Y_1 is generated from $G^{-1}(P\{Y_1 \leq j_1 | \boldsymbol{X}\}) = \alpha_{1,j_1} - \beta_1 \boldsymbol{X}$, and conditional on Y_1, Y_2 is generated from $G^{-1}(P\{Y_2 \leq j_2 | \boldsymbol{X}\}) = \alpha_{2,j_2} - (\beta_2 \boldsymbol{X} + \eta_1 I_{(Y_1=1)} + ... + \eta_5 I_{(Y_1=5)})$. The parameters $\alpha_1 = (-1, 0, 1, 2)$, $\beta_1 = 1$, $\alpha_2 = (-1, 0, 1)$ and $\beta_2 = -0.5$. Considered are four scenarios: (1) $\boldsymbol{\eta} = (0, 0, 0, 0, 0)$ (the null); (2) $\boldsymbol{\eta} = (-0.4, -0.2, 0, 0.2, 0.4)$ (linear effect); (3) $\boldsymbol{\eta} = (-0.30, 0.18, 0.20, 0.22, 0.24)$ (nonlinear-monotonic effect); and (4) $\boldsymbol{\eta} = (-0.2, 0, 0.2, 0, -0.2)$ (non-monotonic effect). The results are calculated based on 1000 simulation replicates with a sample size of 500.

	ϕ_ρ-based	ϕ_τ-based	LS	LRT
	Logit link			
Null	0.039	0.041	0.042	0.055
Linear	0.845	0.849	0.853	0.706
Nonlinear-monotonic	0.595	0.605	0.609	0.548
Nonmonotonic	0.070	0.067	0.070	0.301
	Complementary log-log link			
Null	0.059	0.057	0.057	0.057
Linear	0.859	0.899	0.912	0.797
Nonlinear-monotonic	0.892	0.882	0.899	0.934
Nonmonotonic	0.354	0.321	0.384	0.666

Note: "LS" stands for Li and Shepherd (2010)'s method and "LRT" the likelihood ratio test.

Section 5 Simulation studies – Simulation III

Power comparison for testing the null of a negligible association

Table 5: Type I error and power for Simulation III where the null hypothesis is H_0: $|\phi| \leq 0.1$. Results are calculated based on 1000 replicates.

	$n = 200$		$n = 500$	
	ϕ_ρ-based	ϕ_τ-based	ϕ_ρ-based	ϕ_τ-based
Negligible linear effect (null)	0.024	0.003	0.006	0.000
Negligible monotonic effect (null)	0.005	0.001	0.006	0.000
Non-negligible linear effect	0.506	0.250	0.877	0.539
Non-negligible monotonic effect	0.627	0.378	0.946	0.743

Section 5 Simulation studies – Simulation IV

Power comparison with a varying degree of nonlinearity of partial association

Table 6: Power for testing partial independence in Simulation IV. The coefficient c represents the degree of non-linearity. The results are calculated using the significance level $\alpha = 0.01$ and 1000 simulation replicates.

c	$n = 60$			$n = 100$			$n = 200$		
	ϕ_ρ-based	ϕ_τ-based	LS	ϕ_ρ-based	ϕ_τ-based	LS	ϕ_ρ-based	ϕ_τ-based	LS
$c = 1$	0.942	0.971	0.962	0.996	1.000	1.000	1.000	1.000	1.000
$c = 0.5$	0.798	0.804	0.783	0.956	0.975	0.974	0.999	1.000	1.000
$c = 0.1$	0.163	0.129	0.126	0.298	0.229	0.217	0.658	0.581	0.572
$c = 0.05$	0.069	0.047	0.047	0.133	0.094	0.093	0.287	0.215	0.208
$c = 0.01$	0.025	0.021	0.021	0.022	0.015	0.010	0.022	0.018	0.012

Note: "LS" stands for Li and Shepherd (2010)'s method.

Our measures with a varying degree of nonlinearity of partial association

Table 7: Estimates of our measures for gauging partial dependence in Simulation IV. The coefficient c represents the degree of non-linearity. Estimated standard errors are reported in parenthesis. The results are calculated based on 1000 simulation replicates.

	$n = 60$		$n = 100$		$n = 200$	
	$\hat{\phi}_\rho$	$\hat{\phi}_\tau$	$\hat{\phi}_\rho$	$\hat{\phi}_\tau$	$\hat{\phi}_\rho$	$\hat{\phi}_\tau$
$c = 1$	0.207 (0.041)	0.297 (0.058)	0.178 (0.027)	0.302 (0.045)	0.139 (0.015)	0.302 (0.031)
$c = 0.5$	0.222 (0.055)	0.244 (0.067)	0.194 (0.037)	0.248 (0.050)	0.157 (0.022)	0.248 (0.036)
$c = 0.1$	0.148 (0.091)	0.105 (0.073)	0.144 (0.067)	0.106 (0.056)	0.134 (0.043)	0.111 (0.039)
$c = 0.05$	0.106 (0.100)	0.069 (0.073)	0.101 (0.079)	0.066 (0.060)	0.100 (0.050)	0.069 (0.040)
$c = 0.01$	0.037 (0.103)	0.022 (0.074)	0.033 (0.079)	0.019 (0.056)	0.034 (0.053)	0.020 (0.038)

Section 6.1 Analysis of the National Election Study

The data for VOTE and PID

Table 8: Two-way contingency table for VOTE and PID

	strDem	weakDem	indDem	indind	indRep	weakRep	strRep
Clinton (labeled as 0)	197	169	101	26	24	26	8
Dole (labeled as 1)	3	11	7	11	70	124	167

Adjusting covariates through probit regression models

Table 9: Model inference results for the binary variable VOTE and ordinal variables PID, selfLR, ClinLR and DoleLR in the analysis of 1996 National Election Study. Shown are the maximum likelihood estimates of the regression coefficients and their standard deviations (in the parentheses). The symbol * indicates statistical significance at the 0.05 level.

	Age	Income	Education	Intercepts					
	(β_1)	(β_2)	(β_3)	(α_1)	(α_2)	(α_3)	(α_4)	(α_5)	(α_6)
VOTE	0.006*	0.008*	0.012	-1.019*					
	(0.003)	(0.002)	(0.017)	(0.282)					
PID	0.000	0.008*	0.011	-0.287	0.287	0.587*	0.688*	0.955*	1.460*
	(0.002)	(0.001)	(0.014)	(0.233)	(0.232)	(0.233)	(0.233)	(0.234)	(0.236)
selfLR	0.004*	0.004*	-0.055*	-2.597*	-1.600*	-1.017*	-0.293	0.199	1.379*
	(0.002)	(0.001)	(0.014)	(0.250)	(0.232)	(0.228)	(0.228)	(0.229)	(0.235)
ClinLR	-0.005*	-0.005*	-0.051*	-2.432*	-1.347*	-0.677*	-0.046	0.418	0.930*
	(0.002)	(0.001)	(0.014)	(0.238)	(0.234)	(0.232)	(0.231)	(0.231)	(0.238)
DoleLR	-0.002	0.003*	0.037*	-1.660*	-1.121*	-0.757*	-0.313	0.323	1.769*
	(0.002)	(0.001)	(0.014)	(0.246)	(0.236)	(0.235)	(0.236)	(0.237)	(0.238)

Section 6.1 Analysis of the National Election Study

Correlation matrix before and after the covariate adjustment

Table 10: Analysis of association and partial association between 5 discrete variables in the National Election Study. Shown are the estimates of our measure ϕ_τ and their standard errors (in the parenthesis). The symbol * indicates statistical significance at the 0.05 level.

	Marginal association					Partial association			
	PID	selfLR	ClinLR	DoleLR		PID	selfLR	ClinLR	DoleLR
VOTE	0.675*	0.533*	-0.447*	-0.077*		0.415*	0.329*	-0.260*	-0.049*
	(0.012)	(0.020)	(0.023)	(0.029)		(0.010)	(0.013)	(0.014)	(0.016)
PID		0.511*	-0.331*	-0.025			0.421*	-0.248*	-0.032
		(0.020)	(0.023)	(0.026)			(0.017)	(0.019)	(0.020)
selfLR			-0.194*	-0.103*				-0.151*	-0.069*
			(0.026)	(0.024)				(0.021)	(0.019)
ClinLR				-0.172*					-0.130*
				(0.029)					(0.021)

Appendix E. Re-analysis of the National Election Study using the adjacent-category logit model (Section 7)

In this Appendix, we re-analyze the national election data examined in Section 6.1. Instead of using a cumulative link model, we use the adjacent-category logit model to adjust for covariates. The goal is to further demonstrate, beyond the bivariate analysis in Example 6, that the proposed surrogate framework applies to general models.

We examine the partial association between $K = 5$ discrete variables including *VOTE*, *PID*, *selfLR*, *ClinLR*, and *DoleLR*. Adjusted are three covariates: age, income, and education. We carry out the same analysis as seen in Section 6.1.2, except that the adjacent-category logit model is used to fit the data and obtain surrogate residuals. Our assessment of partial association is summarized in Table 11, which shows (a) the estimates of ϕ_τ for each pair of variables; (b) the estimates of the standard errors of $\hat{\phi}_\tau$; and (c) the statistical significance for testing $H_0 : \phi_\tau = 0$. The result from the cumulative probit model is included for the comparison purpose. It is evident that the new estimates $\hat{\phi}_\tau$'s, on the right panel of Table 11, are all close to their counterparts on the left panel. In addition, both models identify the pair (PID, DoleLR) as not being partially associated at the 0.05 significance level.

Table 11: Comparison of the partial association analyses of the National Election Study using two different models. Shown are the estimates of our measure ϕ_τ and their standard errors (in the parenthesis). The symbol * indicates statistical significance at the 0.05 level.

	Cumulative probit model				Adjacent-category logit model			
	PID	selfLR	ClinLR	DoleLR	PID	selfLR	ClinLR	DoleLR
VOTE	0.415*	0.329*	-0.260*	-0.049*	0.383*	0.310*	-0.242*	-0.056*
	(0.010)	(0.013)	(0.014)	(0.016)	(0.015)	(0.014)	(0.015)	(0.016)
PID		0.421*	-0.248*	-0.032		0.421*	-0.249*	-0.034
		(0.017)	(0.019)	(0.020)		(0.017)	(0.019)	(0.020)
selfLR			-0.151*	-0.069*			-0.157*	-0.070*
			(0.021)	(0.019)			(0.021)	(0.019)
ClinLR				-0.130*				-0.130*
				(0.021)				(0.021)

The pairwise partial regression plots are displayed in Figure 17. Comparing this figure with Figure 12 where the cumulative probit model is used, we see remarkable similarity in terms of the direction and shape of the lowess curve. In particular, we once again observe an inverted-U shape of the red lowess curve in the scatter plot of (r_{DoleLR}, r_{VOTE}), and the shape of a beach chair in (r_{DoleLR}, r_{ClinLR}). Our findings all together have strengthened the conclusion made in Example 6; that is, the inferences drawn out the two models are similar.
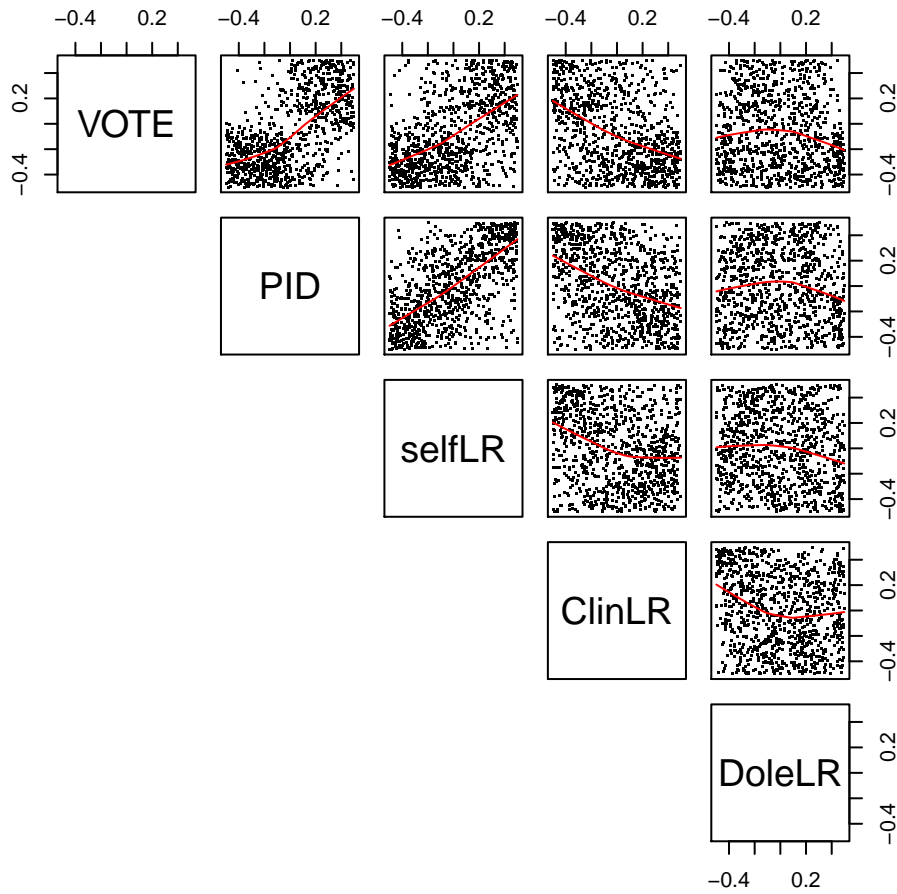

Figure 17: Partial regression plot matrix for 5 discrete variables in the National Election Study using the adjacent-category logit model. Shown are pairwise scatter plot of surrogate residuals. Added red curves are obtained from cubic smoothing splines.