WIN RATIO FOR PARTIALLY ORDERED DATA

Lu Mao

Department of Biostatistics and Medical Informatics,

School of Medicine and Public Health,

University of Wisconsin-Madison

Abstract: The win ratio, initially developed for time-to-event data, is extended to any data type equipped with a partial order. We study this extension in both nonparametric inference and semiparametric regression. We begin by formulating the win ratio as an estimand that contrasts between two populations with partially ordered responses, which reduces to the familiar odds ratio in the case of binary data. For hypothesis testing, we prove that the empirical two-sample win ratio test is consistent against stochastically ordered distributions and efficient against proportional odds alternatives under a total order. In regression, we model the marginal win ratio multiplicatively against covariates, extending logistic regression from binary to partially ordered responses. This model is implied by a generalized continuation-ratio logit model but requires fewer assumptions on the relationship between response levels. To make inference, we construct a class of weighted U-statistic estimating equations and derive pseudo-efficient weights to improve efficiency. Simulation studies demonstrate that the proposed procedures perform well in both testing and regression under

finite samples. As illustrations, we analyze bivariate radiologic assessments in a recent liver disease study and subject smoking status in a youth tobacco use study, treating them both as partially ordered outcomes. The proposed methodology is implemented in the R package `poset`, publicly available on GitHub at `https://lmaowisc.github.io/poset` and on the Comprehensive R Archive Network (CRAN).

Key words and phrases: Continuation ratio; Logistic regression; Odds ratio; Ordinal data; Stochastic order; U-statistic.

1. Introduction

Partially ordered data, a common variant of ordinal data, frequently arise in medical and sociological studies. Unlike totally ordered tumor grades or Likert scales, partially ordered data are not necessarily pairwise comparable. For instance, in the tumor-node-metastasis cancer staging system (Edge et al., 2010), primary and metastatic tumors are scored on four- and three-level severity scales, respectively. Two patients can be compared in overall severity only if one scores higher on both scales (Lin et al., 2013). Similar partial orders are found in groups of radiologic ratings or survey responses.

Despite their prevalence, partially ordered data have received only limited methodological attention. In the nonparametric setting, Rosenbaum (1991) discussed desirable properties for statistics measuring the association

between two partially ordered responses (where one may be a binary treatment). He demonstrated that the Wilcoxon rank sum and Spearman's rank correlation statistics satisfy these properties in the special case of totally ordered data. Mondal & Hinrichs (2016) proposed a class of association tests by conceptualizing an underlying total order consistent with the partial order and then computing rank statistics based on that latent order. However, the existence of such a compatible total order is not guaranteed, and a direct measure of between-group difference is not provided.

For regression analysis, Zhang & Ip (2012) proposed a class of partitioned conditional models, constructed in three consecutive steps. First, a nominal (i.e., multinomial) regression model is used to estimate the probabilities of any disjoint networks across which no two elements are comparable (if more than one such network exist). Next, each network is partitioned into a sequence of weakly parallel antichains (Trotter, 1992), i.e., subsets containing mutually incomparable elements, whose conditional probabilities are modeled using ordinal regression, such as the proportional odds model (McCullagh, 1980; Agresti, 2010). Finally, the mutually incomparable elements within each antichain are modeled using nominal regression. Peyhardi et al. (2016) extended this approach to general categorical responses. Although the partitioned conditional models elegantly reduce complex par-

tial orders into layers of familiar ordinal or nominal structures, they involve extensive model assumptions that are difficult to verify. Moreover, the large number of regression coefficients, many of which are not of direct interest, complicates the interpretation of covariate effects.

A potential solution lies in the win ratio, which has recently gained popularity in the analysis of composite time-to-event outcomes (Pocock et al., 2012). Originally, the win ratio was defined as the ratio of wins to losses among all pairs between treatment and control, based on a specific comparison rule. This concept is well-suited for partially ordered data, with the partial order serving as the comparison rule and incomparable pairs treated as "ties". Unlike the elaborate partitioned conditional models, the win ratio focuses on dimensions with a clear order between outcomes, providing a succinct summary of treatment effect. Even for totally ordered outcomes, the win ratio may offer a distinct alternative to standard ordinal regression (Agresti, 2010), as it imposes no constraints on the relationships between response levels. While the two-sample win ratio has been informally applied to partially ordered data (Wittkowski et al., 2004; Bebu & Lachin, 2016), its statistical properties as estimators and tests remain unexplored. Additionally, regression modeling of partially ordered data using the win ratio has yet to be investigated in the literature.

In this work, we conduct a detailed study of the win ratio for partially ordered data in both nonparametric inference and regression analysis. Specifically, we define the win ratio formally as a model-free estimand and explore the operating characteristics of its empirical estimator in two-sample testing. Quantitative covariates are incorporated via a multiplicative model, in which the regression coefficients are interpreted as log-win ratios resulting from unit increases in the covariates. Estimating functions are constructed based on covariate-weighted pairwise residuals, with efficient weights informed by parallels with logistic regression and its well-known efficient score function.

2. Two-sample estimation and testing

2.1 Set-up

Let $(\mathcal{Y}, \preceq)$ denote a partially ordered set, or poset (Trotter, 1992), where the set $\mathcal{Y}$ is equipped with a partial order $\preceq$. Throughout, we assume that $\mathcal{Y}$ is discrete with at most countably many elements. The partial order $\preceq$ is a relation between pairs of elements in $\mathcal{Y}$ that satisfies the following properties: reflexive ($y \preceq y$ for all $y \in \mathcal{Y}$), anti-symmetric ($y \preceq y^*$ and $y^* \preceq y$ imply $y = y^*$), and transitive ($y \preceq y^*$ and $y^* \preceq y^{**}$ imply $y \preceq y^{**}$, for all y, y^*, and $y^{**} \in \mathcal{Y}$). Occasionally, the relation $y \preceq y^*$ will

be equivalently expressed as $y^* \succeq y$. If neither $y \preceq y^*$ nor $y^* \preceq y$, the pair is said to be incomparable. If every pair in $\mathcal{Y}$ is comparable, the poset is said to be totally ordered.

Example 1 (Ordinal data)**.** Ordinal data, such as tumor grades and Likert scales, are totally ordered. If $\mathcal{Y}$ consists of m totally ordered elements, we can, without loss of generality, let $(\mathcal{Y}, \preceq) = (\mathbb{N}_{m-1}, \leq)$, where $\mathbb{N}_{m-1} = \{0, 1, \ldots, m-1\}$. Binary data are a special case with $m = 2$.

Example 2 (Multivariate ordinal data)**.** A more complex data type arises when multiple ordinal attributes are combined, with two observations being comparable if and only if the same order holds across all components. In poset terminology, this is referred to as the product order (Garg, 2015). Alternatively, certain components may be prioritized so that other components are compared only when the prioritized ones are tied. This is known as a lexicographical order (similar to how words are arranged in a dictionary) (Garg, 2015). In general, the poset for outcomes with K ordinal components can be represented by $(\mathcal{Y}, \preceq) = (\prod_{k=1}^{K} \mathbb{N}_{m_k-1}, \preceq)$, where $\prod_{k=1}^{K} \mathbb{N}_{m_k-1} = \mathbb{N}_{m_1-1} \times \cdots \times \mathbb{N}_{m_K-1}$ (assuming the kth component has m_k levels). In the case of a product order, $\preceq$ can be replaced by $\leq$, which is understood to operate component-wise.

Every partial order $\preceq$ induces a corresponding strict partial order $\prec$.

Namely, if $y \preceq y^*$ and $y \neq y^*$, then we say $y \prec y^*$, or equivalently, $y^* \succ y$. By definition, the strict order is irreflexive, i.e., $y \not\prec y$ for all $y \in \mathcal{Y}$, and transitive. In Example 1, the strict order $\prec$ is simply the strict inequality $<$; in Example 2 with the product order, $y_i \prec y_j$ if $y_i \leq y_j$ component-wise with strict inequality for at least one component. Without loss of generality, we use the symbol $\succ$ as the win operator; that is, we say that y_i wins against y_j if $y_i \succ y_j$. In both Examples, this means a higher-ordered outcome is more favorable.

Given $y \in \mathcal{Y}$, the upper and lower closures of y are defined by $U[y] = \{y^* \in \mathcal{Y} : y^* \succeq y\}$ and $D[y] = \{y^* \in \mathcal{Y} : y^* \preceq y\}$, respectively. Similarly, the strict upper and lower closures are defined by $U(y) = \{y^* \in \mathcal{Y} : y^* \succ y\}$ and $D(y) = \{y^* \in \mathcal{Y} : y^* \prec y\}$, respectively. The greatest or least element of a subset $A \subset \mathcal{Y}$, if it exists, is defined as the unique element $s \in A$ such that $s^* \preceq s$ or $s^* \succeq s$, respectively, for all $s^* \in A$. A lattice is a special type of poset such that for any two elements y_i and y_j, a least element exists for $U[y_i] \cap U[y_j]$ and a greatest element exists for $D[y_i] \cap D[y_j]$ (Garg, 2015). It is easy to see that the data spaces described in Examples 1 and 2 are both lattices. Non-lattice posets are also common in applications. For instance, Zhang & Ip (2012) described a national longitudinal study on youth tobacco use with six levels of smoking behavior ordered in a non-lattice structure.

For a probability measure ν on $\mathcal{Y}$, we define some shorthand notation as follows. With a random element $Y \sim \nu$, denote $\nu(y) = \text{pr}(Y = y)$ for $y \in \mathcal{Y}$, $\nu(A) = \text{pr}(Y \in A)$ for $A \subset \mathcal{Y}$, and $\nu(f) = E\{f(Y)\}$ for an integrable function $f : \mathcal{Y} \to \mathbb{R}$. For every ν in question, it is always assumed that ν is supported on $\mathcal{Y}$, i.e., $\nu(y) > 0$ for all $y \in \mathcal{Y}$.

2.2 Nonparametric estimand and estimator

Let ν_1 and ν_0 denote the distributions of a partially ordered outcome in the treatment and control groups, respectively. An expedient measure of the difference between the two groups is to assign a numeric score to each outcome level and consider the difference in the average score. More specifically, let $f : \mathcal{Y} \to \mathbb{R}$ be a non-decreasing function in the sense that $y_i \preceq y_j$ implies $f(y_i) \leq f(y_j)$. The treatment effect can then be measured by $\nu_1(f) - \nu_0(f)$. Alternatively, one could model the relationship between ν_1 and ν_0 parametrically, for example using the proportional odds model in the case of ordinal outcomes, and then evaluate the model-based effect size, such as the odds ratio.

Using a nonparametric estimand does not rely on such arbitrary scoring or modeling. Define it as

$$\mathcal{R}(\nu_1, \nu_0) = \frac{\mathcal{W}(\nu_1, \nu_0)}{\mathcal{W}(\nu_0, \nu_1)} \equiv \frac{\int\int I(y_1 \succ y_0)\nu_1(\mathrm{d}y_1)\nu_0(\mathrm{d}y_0)}{\int\int I(y_0 \succ y_1)\nu_1(\mathrm{d}y_1)\nu_0(\mathrm{d}y_0)}, \tag{2.1}$$

where $I(\cdot)$ is the indicator function. If $Y_z \sim \nu_z$ $(z = 1, 0)$ and $Y_1 \perp\!\!\!\perp Y_0$,

we have $\mathcal{W}(\nu_z, \nu_{1-z}) = \text{pr}(Y_z \succ Y_{1-z})$, which represents the probability of group z winning against $1-z$. Hence, the ratio $\mathcal{R}(\nu_1, \nu_0)$ can be interpreted as the fold change in the likelihood of winning by the treatment compared to the control. This interpretation is similar to that of the odds ratio for a binary outcome. In fact, the metric reduces to the odds ratio in the binary case.

Proposition 1 (Equivalence of win and odds ratios).

$$\mathcal{R}(\nu_1, \nu_0) = \frac{\nu_1(1)\{1 - \nu_0(1)\}}{\nu_0(1)\{1 - \nu_1(1)\}}.$$

Proof. The result follows by $\mathcal{W}(\nu_z, \nu_{1-z}) = \nu_z(1)\{1 - \nu_{1-z}(0)\}$ $(z = 1, 0)$. □

Let $\hat{\nu}_z$ denote the empirical distribution in group z of size n_z $(z = 1, 0)$. Then, a natural estimator for $\mathcal{R}(\nu_1, \nu_0)$ is plug-in $\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)$, which is just a compact notation for the empirical win ratio statistic based on pairwise comparisons. It then follows that $\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)$ is an efficient estimator for $\mathcal{R}(\nu_1, \nu_0)$ due to the efficiency of the empirical distribution functions and the efficiency-preservation property of $\mathcal{R}(\cdot, \cdot)$ as a smooth functional (Bickel et al., 1993). Like the Wilcoxon rank sum statistic, $\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)$ also satisfies the decreasing property of Rosenbaum (1991) for association metrics between an ordered outcome and a binary treatment. In fact, if we think of the $\mathcal{W}(\hat{\nu}_z, \hat{\nu}_{1-z})$, i.e., the empirical win and loss proportions, as the partial-

order equivalent of ranks, we can view the win ratio as an extension of the Wilcoxon rank statistic (Mao, 2022). Finally, the asymptotic normality of $\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)$ and its variance are derived in the supplementary material, using similar U-statistic techniques to those used for time-to-event outcomes (Luo, 2015; Bebu & Lachin, 2016; Dong et al., 2016). The results can also be obtained as a special case of the multiplicative regression model described in Section 3, with the treatment indicator as the covariate.

2.3 Operating characteristics of the test

In addition to measuring treatment effect, $\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)$ can also be used to test the null hypothesis $H_0 : \nu_1 = \nu_0$. This test will be particularly sensitive if ν_0 and ν_1 are stochastically ordered.

Definition 1 (Stochastic order) *A subset $B \subset \mathcal{Y}$ is called an up-set if $y \in B$ implies $U[y] \subset B$. For two probability measures ν_0 and ν_1 on $\mathcal{Y}$, ν_1 is stochastically greater than ν_0, denoted by $\nu_0 \preceq \nu_1$, or $\nu_1 \succeq \nu_0$, if*

$$\nu_1(B) \geq \nu_0(B) \text{ for every up-set } B \subset \mathcal{Y}.$$

This definition extends the concept of stochastic order from Euclidean spaces to posets, capturing the intuitive notion that one random element "tends" to win against another. An important implication of $\nu_1 \succeq \nu_0$ is

that $\nu_1(f) \geq \nu_0(f)$ for every non-decreasing function f (Kamae & Krengel, 1978). An example of such a function on $(\prod_{k=1}^{K} \mathbb{N}_{m_k-1}, \leq)$ is the average score $f(y) = K^{-1} \sum_{k=1}^{K} s_{ki_k}$, where $y = (i_1, \ldots, i_K)^{\mathrm{T}}$ and $s_{k0} \leq s_{k1} \leq \cdots \leq s_{k,m_k-1}$ is an ordered sequence of scores assigned to the m_k levels of the kth component $(k = 1, \ldots, K)$.

Write $n = n_1 + n_0$ and assume that $n_1/n \to$ [illegible] $(0, 1)$ as $n \to \infty$. [illegible] derivation in the supplementary material, we find [illegible] log transf[illegible] and normalized test statistic $S_n = \hat{\sigma}_n^{-1} n^{1/2} \log\{\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)\} \to_d N(0, 1)$ [illegible] der H_0, where $\hat{\sigma}_n^2$ is a nonparametric variance estimator [illegible] the supplementary material. A two-sided, asymp[illegible]ic level-α test rejects H_0 if $|S_n| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$[illegible] quantile of the standard normal distribution.

We show that this test is [illegible]nt against the alternative hypothesis $H_A : \nu_1 \succ \nu_0$, i.[illegible] $\nu_1 \neq \nu_0$. For a nontrivial poset $\mathcal{Y}$, [illegible]symptotic va[illegible]ce [illegible] $n^{1/2} \log\{\mathcal{R}(\hat{\nu}_1, \hat{\nu}_0)\}$ is finite, so the noncentrality parameter fo[illegible] is [illegible]rder $O[n^{1/2} \log\{\mathcal{R}(\nu_1, \nu_0)\}]$. As a result, the Rejec[illegible] tends to 1 as $n \to \infty$ if $\mathcal{R}(\nu_1, \nu_0) > 1$ The following lemma establish[illegible] intermediate result for showing that $\mathcal{R}(\nu_1, \nu_0) > 1$ under $\nu_1 \succ \nu_0$. The proof can be found in the Appendix.

Lemma 1. *If $\nu_1 \succ \nu_0$, then*

$$\nu_1\{U(y)\} \geq \nu_0\{U(y)\} \text{ for all } y \in \mathcal{Y},$$

with strict inequality for at least one y.

In addition to consistency, we can also show that the win ratio test, when applied to ordinal data, is asymptotically efficient under proportional odds alternatives.

Theorem 1. *The win ratio test has the following properties.*

(a) When $\nu_1 \succ \nu_0$, for every $\alpha \in (0, 1)$,

$$\text{pr}(|S_n| > z_{1-\alpha/2}) \to 1 \text{ as } n \to \infty.$$

(b) With a totally ordered $\mathcal{Y}$ [illegible] conform to a proportional odds model with a non-unit odds ratio, the win ratio test achieves the highest asymptotic power possible. The precise statement is given in the supplementary material.

Proof. To prove (a), by earlier arguments it suffices to show that $\mathcal{W}(\nu_1, \nu_0) > \mathcal{W}(\nu_0, \nu_1)$. But

$$\mathcal{W}(\nu_1, \nu_0) = \int \nu_1\{U(y)\}\nu_0(\mathrm{d}y) > \int \nu_0\{U(y)\}\nu_0(\mathrm{d}y)$$
$$\geq \int \nu_0\{U(y)\}\nu_1(\mathrm{d}y) = \mathcal{W}(\nu_0, \nu_1),$$

where the first equality follows by Lemma 1 and the second by $\nu_1 \succeq \nu_0$ and $\nu_0\{U(y)\}$ being a non-increasing function of y. The proof of (b) is more involved and is relegated to the supplementary material, where the asymptotic power function is derived explicitly. □

3. Regression analysis

3.1 A multiplicative win ratio model

With a p-dimensional covariate $Z \in \mathbb{R}^p$, we aim to assess its effect on $Y \in \mathcal{Y}$. Like in the two-sample case, direct modeling of the conditional distribution, denoted by $\nu(\cdot \mid Z)$, would likely involve unnecessary assumptions and nuisance parameters non-essential to the covariate effects. In keeping with the two-group win ratio development, we focus on the covariate-specific win ratio $\mathcal{R}\{\nu(\cdot \mid Z_i), \nu(\cdot \mid Z_j)\}$, where Z_i and Z_j are independent copies of Z. For convenience, we denote this quantity as $\mathcal{R}(Z_i, Z_j)$. Fixing Z_i and Z_j, $\mathcal{R}(Z_i, Z_j)$ conceptually indistinguishable from the two-group win ratio studied in Section 2.2, only with the groups redefined as the subpopulations with covariates Z_i and Z_j. Specifically, if Y_i and Y_j are outcomes drawn independently from covariate groups Z_i and Z_j, respectively, then $\mathcal{R}(Z_i, Z_j) = \mathrm{pr}(Y_i \succ Y_j \mid Z_i, Z_j)/\mathrm{pr}(Y_j \succ Y_i \mid Z_i, Z_j)$, which represents the fold change in the likelihood of winning for subpopulation Z_i

as compared to Z_j.

With multi-dimensional Z, especially when it contains continuous components, a parsimonious model for $\mathcal{R}(Z_i, Z_j)$ is desired to avoid the curse of dimensionality. Since $\mathcal{R}(\cdot,\cdot) \in \mathbb{R}^+$ and must satisfy $\mathcal{R}(Z_i, Z_j)\mathcal{R}(Z_j, Z_i) \equiv 1$, it is natural to consider

$$\log\{\mathcal{R}(Z_i, Z_j)\} = \beta^{\mathrm{T}}(Z_i - Z_j), \qquad (3$$

where β is a p-dimensional vector of regression coefficients. The components of β are thus the log-win ratios resulting from unit increases in the corresponding components of Z. If Z contains dummy variables encoding a categorical covariate, then (3.2) becomes a saturated model with $\exp(\beta)$ containing the nonparametric win ratios as defined in Section 2.2, comparing each level with the reference.

In general, model (3.2) implies that the covariates have multiplicative effects on the win ratio and will hence be called the multiplicative win ratio regression model. It is semiparametric in the sense that, apart from the parameterized conditional win ratio, other aspects of the conditional distribution of Y given Z are left unspecified.

Remark. (3.2) is similar in structure and interpretation to the proportional win-fractions model of Mao & Wang (2021) for composite time-to-event outcomes. The latter, however, involves stricter assumptions on

the constancy of the win ratio as follow-up goes on (hence the proportionality). If instead one chooses to model the win ratio under a fixed time frame, then the (possibly component-prioritized) composite outcomes would likely follow a partial order, and model (3.2) would apply.

3.2 Model-generating mechanisms

While model (3.2) allows for a simple and intuitive interpretation of ate effects, its pairwise formulation hides the underlying response-cova relationship. For transparency, we explore the possible er which (3.2) holds.

First, we show that under $(\mathbb{N}_1, \leq)$, model (3. valent to the standard logistic regression with t sam ressi coefficients. In this case, β in (3.2) can be interpreted as r th g win ratio or log-odds ratio, extending the two-s ple i result in Proposition 1 to regression models. The proof is str tfo rd and can be found in the Appendix.

Proposition Equivalence with logistic regression for binary data)**.** *For $(\mathbb{N}_1, \leq)$, model is equivalent to the logistic regression model*

$$1 \mid Z) = \frac{\exp(\gamma + \beta^{\mathrm{T}} Z)}{1 + \exp(\gamma + \beta^{\mathrm{T}} Z)} \quad \textit{for some } \gamma \in \mathbb{R}. \tag{3.3}$$

Next, we show that in general, model (3.2) is implied by a logit model

generalized from the continuation-ratio model for ordinal responses (Armstrong & Sloan, 1989), as proved in the Appendix.

Proposition 3 (Sufficiency of generalized continuation-ratio model)**.** *Model* (3.2) *is implied by the following generalized continuation-ratio model:*

$$\mathrm{pr}(Y \succ y \mid Y \succeq y; Z) = \frac{\exp\{\gamma(y) + \beta^{\mathrm{T}} Z\}}{1 + \exp\{\gamma(y) + \beta^{\mathrm{T}} Z\}} \quad \text{for } \gamma : \mathcal{Y} \to \mathbb{R}. \quad (3.4)$$

Under model (3.4), the conditional continuation ratio at y is $\mathrm{pr}(Y \succ y \mid Y \succeq y; Z)/\mathrm{pr}(Y = y \mid Y \succeq y; Z) = \exp\{\gamma(y) + \beta^{\mathrm{T}} Z\}$. This is the odds that a strictly "better" outcome given that it is comparable to any level than y. The model specifies that the covariate effect is invariant to the level $y \in \mathcal{Y}$, which imposes a strong constraint across outcome levels. In contrast, the multiplicative win-ratio model specifies only the overall relationship between the response and covariates without level-specific constraints, making it more relaxed.

Proposition 4 (Non-necessity of generalized continuation-ratio model)**.** *There exist multiplicative win ratio models that do not satisfy* (3.4).

Proof. Consider a counter-example with poset outcomes in $(\mathbb{N}_{m-1}, \leq)$ $(m > 2)$ and a covariate $Z \in \{1, 0\}$. Model (3.2) then becomes a saturated model, which holds trivially with $\beta = \log\{\mathcal{R}(\nu_1, \nu_0)\}$, where $(Y \mid Z = z) \sim \nu_z$ $(z = 1, 0)$. On the other hand, model (3.4) imposes a real constraint. To

see this, write $\lambda_{zk} = \mathrm{pr}(Y > k \mid Y \geq k; Z = z)$. With $\mathrm{logit}(x) = \log\{x/(1-x)\}$, the model implies that $\mathrm{logit}(\lambda_{1k}) - \mathrm{logit}(\lambda_{0k})$, the log-continuation odds ratio at level k, is constant across $k = 0, 1, \ldots, m-2$, which need not be true as $(\lambda_{z0}, \ldots, \lambda_{z,m-2})^{\mathrm{T}}$ can vary freely in $(0,1)^{\otimes(m-1)}$. □

3.3 Estimating equations

Let (Y_i, Z_i) $(i = 1, \ldots, n)$ be a random n-sample of (Y, Z). To fit [illegible] (3.2), consider the pairwise error term

$$\mathcal{E}_{ij}(\beta) = I(Y_i \succ Y_j)\exp(\beta^{\mathrm{T}} Z_j) - I(Y_j \succ Y_i)\exp(\beta^{\mathrm{T}} Z_i), \quad i, j = 1, \ldots, n.$$

Under (3.2), it is easy to see that $E[\mathcal{E}_{ij}(\beta) \mid Z_i, Z_j] = 0$. This motivates the covariate-weighted U-estimation [illegible]

$$\mathcal{U}_n(\beta; \hat{W}) = \binom{n}{2}^{-1} \sum_{i<j} (Z_i - Z_j)\hat{W}(Z_i, Z_j; \beta)\mathcal{E}_{ij}(\beta), \tag{3.5}$$

where $\hat{W}(\cdot, \cdot; \beta)$ is [illegible] dependent symmetric function on $\mathcal{Z}^{\otimes 2}$ and $\mathcal{Z} \subset \mathbb{R}^p$ is the [illegible] space. The estimator $\hat{\beta}$ solves $\mathcal{U}_n(\hat{\beta}; \hat{W}) = 0$. Under suitable regularity conditions, finding $\hat{\beta}$ can be reformulated as a minimization problem of a strictly convex function (see supplementary material for details), allowing us to use the standard Newton–Raphson algorithm. With a binary Z, it is easily seen that $\hat{\beta}$ under every $\hat{W}$ reduces to the log of the

empirical two-sample win ratio of Section 2.2.

To estimate the variance of $\hat{\beta}$, the correlations between the terms in $\mathcal{U}_n(\beta; \hat{W})$ must be accounted for. The following theorem, proved in the supplementary material, establishes the asymptotic properties of $\hat{\beta}$ using uniform central limit theorems for U-processes (e.g., Arcones & Giné, 1993, Theorem 4.10) under the following regularity conditions. Let β_0 denote the true value of β.

(C1) The covariate space $\mathcal{Z}$ is bounded and the covariance matrix of $\dot{\mathcal{U}}$ is positive definite.

(C2) With $(Y_i, Z_i) \perp\!\!\!\perp (Y_j, Z_j)$, we have that $p(\Delta_{ij} > 0 \mid Z_i, Z_j) > 0$ with probability one, where $\Delta_{ij} = I(Y_i \succ Y_j) + I(Y_j \succ Y_i)$.

(C3) The weight function $\hat{W}$ satisfies

$$||\hat{W} - w||_\infty = \sup_{z^*) \in \mathcal{Z}^2, \beta \in \mathbb{R}^p} |\hat{W} - w|(z, z^*; \beta) \to_p 0, \tag{3.6}$$

for some limiting function $w(z, z^*; \beta)$ that is strictly positive, uniformly continuous in β at β_0.

Remark 3. The boundedness of covariates in (C1) is imposed only to simplify the proof and may not be necessary in practice. As shown in the simulations in Section 4, unbounded distributions like the Gaussian

distribution work well so long as their tail probabilities are not too heavy (see, e.g., Andersen & Gill, 1982). Condition (C2) can be viewed as non-degeneracy of the outcome distribution with respect to the partial order, and is satisfied if there are at least two comparable points in $\mathcal{Y}$ with strictly positive conditional probabilities. Condition (C3) basically ensures that the weight function is asymptotically stable with a well-behaved limit.

Theorem 2. *Under conditions (C1)–(C3), we have*

$$n^{1/2}(\hat{\beta} - \beta_0) = 2n^{-1/2}\Omega^{-1}\sum_{i=1}^{n}\psi(Y_i, Z_i) + o_p(1) \quad \text{as } n \to \infty,$$

where $\Omega = E\left[R_{ij}(Z_i - Z_j)^{\otimes 2}w(Z_i, Z_j; \beta_0)\{\exp(\beta_0^{\mathrm{T}} Z_i) + \exp(-\beta_0^{\mathrm{T}} Z_j)\}^{-1}\right]$ *and*

$$\psi(y, z) = E\left[(z - Z)w(z, Z; \beta_0)\{I(\quad \prec y) - I(Y \succ y)\exp(\beta_0^{\mathrm{T}} z)\}\right].$$

Estimators $(\hat{\Omega}, \hat{\psi})$ of (Ω, ψ) are constructed by replacing the expectations with their empirical analogs, w with $\hat{W}$, and β_0 with $\hat{\beta}$. Then the asymptotic variance of $\hat{\beta}$ can be consistently estimated by the second moment of the influence function (Bickel et al., 1993) on the right hand side of (3.7), i.e., $\hat{\Sigma} = 4\hat{\Omega}^{-1}\{n^{-1}\sum_{i=1}^{n}\hat{\psi}(Y_i, Z_i)^{\otimes 2}\}\hat{\Omega}^{-1}$. We can thus make inference on β_0 based on the consistency and asymptotic normality of $\hat{\beta}$ along with this variance estimator.

It is clear from Theorem 2 that the asymptotic distribution of $\hat{\beta}$ depends on $\hat{W}(\cdot,\cdot;\beta)$ only through its limit $w(\cdot,\cdot;\beta)$ at $\beta=\beta_0$. Consequently, if we substitute β in the weight function for some initial consistent estimator $\hat{\beta}_{\text{init}}$, the asymptotic property of the resulting estimator should remain the same. This substitution can save considerable computation when the target weight function $\hat{W}(\cdot,\cdot;\beta)$ itself requires an iterative numerical procedure to compute for each β, such as $\hat{W}_{\text{eff}}$ in (3.8) below. In such cases, solving the original equation $\mathcal{U}_n(\beta;\hat{W})=0$ would necessitate an inner loop to recompute $\hat{W}(\cdot,\cdot;\beta)$ at each iteration of the Newton–Raphson algorithm.

Corollary 1. *Let $\hat{W}(\cdot,\cdot;\beta)$ be a weight function satisfying condition (C3) and let $\hat{\beta}_{\text{init}}$ be an initial estimator such that $\hat{\beta}_{\text{init}}\to\beta_0$. Write $\hat{W}_{\text{init}}(\cdot,\cdot)=\hat{W}(\cdot,\cdot;\hat{\beta}_{\text{init}})$. Then the estimator solving $\mathcal{U}_n(\beta;\hat{W})=0$ and $\mathcal{U}_n(\beta;\hat{W}_{\text{init}})=0$ are asymptotically equivalent.*

3.4 Efficiency consideration

The simplest choice for the weight is just $\hat{W}\equiv 1$, but this choice may not produce an asymptotically efficient $\hat{\beta}$. To improve efficiency, we exploit the equivalence between the multiplicative win ratio and logistic regression models in the case of a binary outcome (Proposition 2).

The basic strategy is outlined as follows. We first rewrite the effi-

cient score of logistic regression in the pairwise form of (3.5), with weight $\hat{W}(Z_i, Z_j; \beta) = \hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$ for some $\hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$. By standard likelihood theory, $\hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$ must be the efficient weight in the binary case. In the general case, because model (3.2) does not completely specify the likelihood, we construct pseudo-efficient weights by mimicking the form of $\hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$. While the efficiency of pseudo-efficient weights is not theoretically guaranteed, we hope that they will at least improve upon weights like $\hat{W} \equiv 1$. As the first step, the following lemma exhibits the form of $\hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$, with proof provided in the supplemental.

Lemma 2 (Efficient weight for binary data). *Under $(\mathbb{N}_1, \leq)$, let*

$$\hat{W}_{\text{eff}}(Z_i, Z_j; \beta) = [1 + \exp\{\hat{\gamma}(\beta) + \beta^{\mathrm{T}} Z_i\}]^{-1} [1 + \exp\{\hat{\gamma}(\beta) + \beta^{\mathrm{T}} Z_j\}]^{-1}, \tag{3.8}$$

where $\hat{\gamma}(\beta)$ solves

$$\sum_{i=1}^{n} \left[Y_i - \frac{\exp\{\hat{\gamma}(\beta) + \beta^{\mathrm{T}} Z_i\}}{1 + \exp\{\hat{\gamma}(\beta) + \beta^{\mathrm{T}} Z_i\}} \right] = 0. \tag{3.9}$$

Then, the U-statistic function $\mathcal{U}_n(\beta; \hat{W}_{\text{eff}})$ reduces to a constant multiple of the efficient score function for β in the logistic regression model (3.3) and is hence optimal.

In the general setting, the form of $\hat{W}_{\text{eff}}(Z_i, Z_j; \beta)$ in (3.8) applies without change. However, the definition of $\hat{\gamma}(\beta)$ in (3.9) needs adaptation, as the

outcome Y is no longer binary or numeric. We resolve this by mapping the partially ordered Y monotonically onto a numeric scale.

Proposition 5 (Pseudo-efficient weights in general)**.** *Let $r : \mathcal{Y} \to [0,1]$ be a strictly increasing scoring function in the sense that it maps the least and greatest elements in $\mathcal{Y}$, if existent, to 0 and 1, respectively, and the $r(y_i) < r(y_j)$ for any $y_i \prec y_j$. Let $\hat{\beta}_{\text{init}}$ denote a consistent initial estimator for β_0, e.g., $\hat{\beta}$ obtained under the naive weight $\hat{W} \equiv 1$. Define the efficient weight by*

$$\hat{W}_{\text{pseff}}(Z_i, Z_j) = \left[1 + \exp\{\hat{\gamma}(\hat{\beta}_{\text{init}}) + \hat{\beta}_{\text{init}}^{\mathrm{T}} Z_i\}\right]^{-1} \left[1 + \exp\{\hat{\gamma}(\hat{\beta}_{\text{init}}) + \hat{\beta}_{\text{init}}^{\mathrm{T}} Z_j\}\right]^{-1}, \tag{3.10}$$

where $\hat{\gamma}(\cdot)$ is defined as in (3.9) except with Y_i replaced by $r(Y_i)$. Then, the estimating function $\mathcal{U}_n(\beta; \hat{W}_{\text{pseff}})$ leads to a consistent estimator for β under $(\mathbb{N}_1, \leq)$.

The proof of Proposition 5 is given in the Appendix. For $(\prod_{k=1}^{K} \mathbb{N}_{m_k-1}, \leq)$ in Example 1, any averaging score of the form $r(y) = K^{-1} \sum_{k=1}^{K} r_{ki_k}$ with $y = (i_1, \ldots, i_K)$ and $0 = r_{k0} < r_{k1} < \cdots < r_{k,m_k-1} = 1$ satisfies the requirements for a strictly increasing scoring function. Even in the general case, it is usually straightforward to construct a proper scoring function by following through a Hasse diagram (Trotter, 1992) for the partial order. An

example is offered in the supplementary material for the non-lattice poset from the smoking behavior study described in Section 2.1 (Zhang & Ip, 2012).

Both the naive and pseudo-efficient analyses of win ratio regression models are implemented in the R-package `poset`, which is publicly available on GitHub at `https://lmaowisc.github.io/poset` and on the Comprehensive R Archive Network (CRAN).

4. Simulation studies

We first assessed the empirical power of the win ratio test under the proportional odds model for ordinal data in comparison with standard tests. We considered two scenarios, one under N_2 with $\nu_0(0) = \nu_0(1) = 0.3$ and $\nu_0(2) = 0.4$, and the other under $(N_4, \leq)$ with $\nu_0(0) = \nu_0(3) = 0.1$, $\nu_0(1) = 0.2$, and $\nu_0(2) = 0.3$. We compared the win ratio with four other tests:

1. The parametric threshold test based on the maximum-likelihood log-odds ratio estimator under the proportional odds model.

2. The Pearson chi-square test on the nominal levels.

3. The binomial test on the responses dichotomized by $\{0\}$ vs $\{1, 2\}$ for

$(\mathbb{N}_2, \leq)$ and $\{0, 1\}$ vs $\{2, 3, 4\}$ for $(\mathbb{N}_4, \leq)$.

Under varying odds ratios, we computed and plotted the empirical powers of the four tests for $n = 200$ in Figure 1. In both scenarios, the asymptotic power function for the efficient test given in the supplementary material provides accurate approximation to the empirical powers of both the win ratio and parametric tests, which outperform those of the other two tests by considerable margins. These results confirm the efficiency of the win ratio under proportional odds alternatives, as established in Theorem 1 (b).

Next, we assessed the inference procedures described in Sections 3.3 and 3.4 for the multiplicative regression model. $(\mathcal{Y}, \preceq) = (\mathbb{N}_2 \times \mathbb{N}_2, \leq)$ and $Z = (Z_1, Z_2)^{\mathrm{T}}$, where $Z_1 \sim$ [illegible] $2 \times \text{Bernoulli}(0.5) - 1$. To generate Y given Z, we used the generalized continuation-ratio logit model in (3.4). It can be shown in the supplementary material, that model (3.4) in this case completely determines the conditional distribution of the outcome. Set $\gamma_{00} = 2.0$, $\gamma_{01} = \gamma_{10} = 1.0$, $\gamma_{02} = \gamma_{20} = \gamma_{11} = 0.2$, and $\gamma_{21} = \gamma_{12} =$ [illegible], where γ_y is a shorthand notation for $\gamma(y)$. Under this setting, the conditional probabilities for the nine levels of the outcome given $Z = (0, 0)^{\mathrm{T}}$ are roughly bounded between 0.05 and 0.20. For $n = 200, 500, 1000$ and $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (0, 0)^{\mathrm{T}}, (0.25, -0.25)^{\mathrm{T}}, (0.5, -0.5)^{\mathrm{T}}$, we

assessed the estimation of β_1 using both the naive estimator (derived from $\mathcal{U}_n(\beta; \hat{W})$ with $\hat{W} \equiv 1$) and the pseudo-efficient estimator. The latter was constructed using the naive estimator as $\hat{\beta}_{\text{init}}$ to form the pseudo-efficient weight in (3.10) with the average scoring function $r\{(i_1, i_2)^{\mathrm{T}}\} = 4^{-1}(i_1+i_2)$.

The results for both estimators are summarized in Table 1. Both estimators exhibit minimal bias, with largely accurate standard error estimates and confidence intervals. Under $\beta = (0, 0)^{\mathrm{T}}$, the relative efficiency between the two estimators is close to one. This is unsurprising, as the pseudo-efficient weight in (3.10) is asymptotically constant in this case, making it equivalent to the naive weight. As the magnitude of β increases, the pseudo-efficient estimator becomes more efficient, with efficiency gains of over 30% under $\beta = (0.5, -0.5)^{\mathrm{T}}$ compared to the naive estimator. The efficiency difference is expected to widen under stronger covariate-response associations. Interestingly, the pseudo-efficient estimator and the associated variance estimator also appear to be more accurate than the naive versions for smaller sample sizes.

Additional simulations in the supplementary material explore other scoring functions for the pseudo-efficient weight, including transformations that produce skewed scores (e.g., square or square-root transformations). The resulting pseudo-efficient estimators perform similarly and all outper-

Table 1: Simulation results for estimation of β_1 in the multiplicative win ratio model.

		Naive				Pseudo-efficient				
n	β_1	Bias	SE	SEE	CP	Bias	SE	SEE	CP	RE
200	0	0.000	0.117	0.115	0.945	0.000	0.114	0.115	0.951	1.05
	0.25	0.008	0.123	0.121	0.948	0.001	0.1	18	0.956	1.14
	0.50	0.016	0.147	0.137	0.932	−0.002	(	25	0.126	0.952	1.3
500	0	0.000	0.072	0.072	0.948	0.000	0.0			
	0.25	0.001	0.077	0.076	0.950	−0.002	0.073	0.073	0.952	1
	0.50	0.008	0.090	0.088	0.940	0.000	0.078	0.07		3
1000	0	0.000	0.050	0.050	0.953	0.00	0.050	0.050	0.952	1.01
	0.25	0.001	0.054	0.0 4	0.951	0.000	0	051	0.953	1.12
	0.50	0.004	0.064	0.063	0.94	0.000		0.055	0.946	1.35

SE, empirical standard error of the e ator pirical average of the standard error estimator; CP, empiric cov ge rate the 95% confidence interval. RE, relative efficiency, i.e., inverse em al variance, comparing the pseudo-efficient sus naive estim Eac try is based on 5,000 replicates.

form the naive pro A case with ordinal outcomes is also considered in the supplemen material.

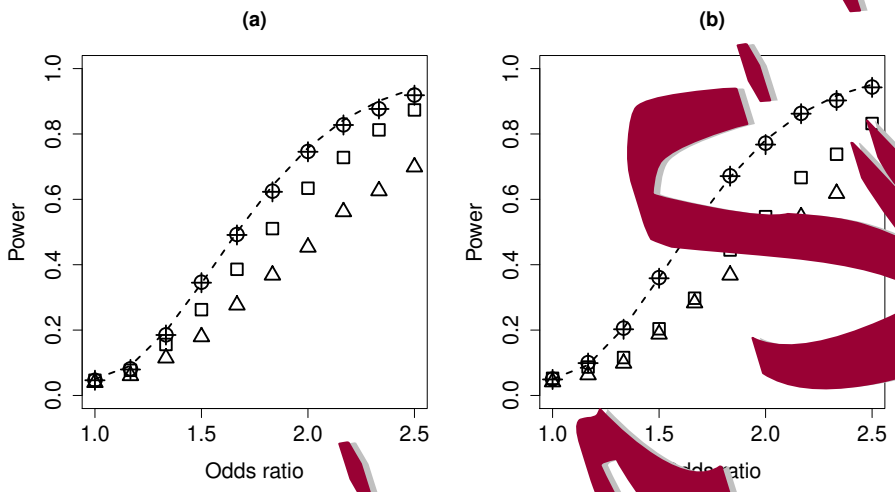

Figure 1: Empirical power as a function of the odds ratio under the proportional odds model for (a) three-level and (b) five-level ordinal outcomes with $n = 200$. Dashed line: theoretical power for the efficient test; plus sign: empirical power of the score test; circle: parametric Wald test; square: chi-square test at nominal levels; triangle: binomial test on dichotomized outcome. Each point of the empirical power is based on $2,000$ replicates.

5. Real examples

5.1 A radiologic study of liver disease

A total of 186 patients with non-alcoholic fatty liver disease (NAFLD) were recruited at the University of Wisconsin Hospitals in 2017. The patients underwent computed tomography scan of the liver for presence of non-alcoholic steato-hepatitis (NASH), the most severe form of NAFLD. The images were subsequently assessed by two radiologists using a scale of 1 to 5, with higher values indicating a greater likelihood of disease. Descriptive statistics on the study cohort are tabulated in the supplementary material.

For win ratio analysis of the reader assessments, we invert the scores so that higher values indicate a lower likelihood of disease and are thus more favorable. Because the two readers have similar levels of experience, we apply the product order to the bivariate scores. This results in an outcome space represented by $\mathcal{T} = \mathbb{N}_4 \times \mathbb{N}_4$. Under the product order, the win ratio measures the fold change in the probability of achieving a concordant consensus of the two readers.

We employ the multiplicative win ratio model of Section 3 to examine the relationship between the radiologists' scores and several covariates: patient sex, the presence of advanced fibrosis (AF), and quantitative biomark-

Table 2: Win ratio regression analysis of the non-alcoholic fatty liver disease study data.

	Naive			Pseudo-efficient		
	Estimate	Std error	p-value	Estimate	Std error	p-value
Sex (f v. m)	−0.144	0.270	0.593	−0.151	0.262	0.563
AF (y v. n)	−0.890	0.307	0.004	−	0.296	0.002
Steatosis (%)	−0.026	0.006	<0.001	−	0.005	<
Gray level	−0.010	0.006	0.070	−0.010		
LSN score	−0.062	0.134	0.646	−0.074	0.130	0.5

ers such as percent of steatosis (liver fat cont nt), liver mean gray level intensity, and liver surface nodularity (LSN) sc e. n the simulations, we first fit the model using t naiv p cedu followed by the pseudo-efficient one utilizing an average s ng n similar to that in Section 4. The results are mma ze bl 2. While the point estimates of the regression coefficien are mparable between the two methods, the pseudo-effic nt t ha nsistently exhibit smaller standard errors and ica enhanced efficiency over the naive estimators.

By the recom ded pseudo-efficient approach, advanced fibrosis status and of steatosis are strongly and significantly associated with the likelihood of NASH. In particular, patients with advanced fibrosis are

$\exp(-0.89) = 41.1\%$ times as likely to receive favorable assessments than those without. Additionally, one percent increase in steatosis results in $1 - \exp(-0.026) = 2.6\%$ reduction in the likelihood of favorable assessments.

For comparison, we use the continuation-ratio model to analyze the sum of the two radiologists' scores against the same set of covariates. As detailed in the supplementary material, the regression coefficients obtained are comparable to those in Table 2. Notably, the effects of advanced fibrosis and steatosis remain highly significant, while the association with mean gray level intensity becomes less pronounced. These findings suggest that, despite the fewer assumptions inherent in the win ratio regression, it maintains a level of efficiency comparable to that of parametric methods.

5.2 The youth tobacco use study

The youth tobacco use study mentioned in Section 2.1 was analyzed using the partial proportional odds model (PCM) (Zhang & Ip, 2012). The response is a subject's smoking status classified into six partially ordered levels ($y = 0, 1, \ldots, 5$), as illustrated in the Hasse diagram in Figure 2. Nonsmoker was considered the most desirable status, while heavy frequent smoker was the least desirable. The PCM consisted of three or-

dinal/nominal sub-models, generating three sets of regression coefficients. Many of these coefficients, such as the log-odds between light frequent vs. heavily infrequent smoking, may not be of direct interest if the primary goal is to assess risk factors for undesirable behavior.

As a comparison, we apply the win ratio approach to a mock datas t of the study (Zhang & Ip, 2012). The data consi 3370 male a l 54 female youths aged 12 to 16 years. Predictors clude sex, race, whether living with parents, having a nonsupportive mother, having strict parents, attending school, having a n ude towards discipline, and having smoking peers. scriptive statistics of these variables, as well as the outcome, are summariz Table 3.

We use the multiplicative w rati del t analyze the effects of these predictors on the smoking sta fi t the model using the naive weight. Then, with t m ring function $r(0) = 0, r(1) = r(2) = 1/3, r(3) = r(4) = 2/3$, and $r(5) = 1$, we compute the pseudo-efficient estimators. s lt summarized in Table 4. Not surprisingly, the pseud nates are similar to the naive ones in magnitude, but with generally r standard errors. Except for age, all predictors are highly significantly associated with smoking behavior. In particular, male youths are $1 - \exp(-1.087) = 66.3\%$ less likely to have a desirable smoking

Table 3: Descriptive statistics for the National Longitudinal Study of Youth on tobacco use.

		Female	Male	Overall
Age	(years)	14.0 (13.1, 14.9)	13.9 (13.[illegible]	14.0 (13.0,15.0)
Nonwhite		954 (28.3%)	1174 ([illegible].7%)	2128 ([illegible])
Live with parents		1119 (33.2%)	1601 ([illegible])	[illegible]
NS mother		2305 (68.4%)	4061 (75.1%)	6366 (72.5%)
Strict parents		1032 (30.6%)	1268 (23.4%)	[illegible]
Attend school		2704 (80.2%)	[illegible] (89.1%)	7527 (85.7%)
Neg. discipline		[illegible]398 (71.2%)	42[illegible]8	6616 (75.3%)
Smoking peers		[illegible]609 (77[illegible]%)	46[illegible]%)	7256 (82.6%)
Smoking status	0	141[illegible]%)	(19.5%)	799 (9.1%)
	1	[illegible].1%	147 (4.4%)	205 (2.3%)
	2	[illegible]	52 (1.5%)	88 (1%)
	[illegible]	95[illegible].6%)	923 (27.4%)	1878 (21.4%)
	4	334 (6.2%)	168 (5%)	502 (5.7%)
	5	3887 (71.8%)	1422 (42.2%)	5309 (60.5%)

Categorical variable[illegible]e summarized by $N(\%)$ and quantitative variables by median (inter[illegible]).

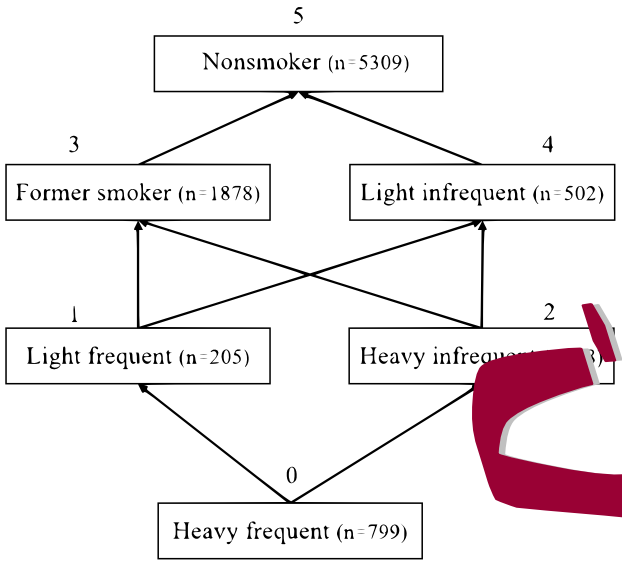

Figure 2: Hasse diagram from Zhang & Ip (2012) for the National Longitudinal Study of Youth on tobacco use.

behavior than female youths. As sensitivity analysis, an alternative scoring function is considered in the supplementary material, which shows similar results to the original pseudo-score analysis.

Concluding Remarks

The new can be immediately extended to a stratified analysis, which is desirable if there is considerable between-strata heterogeneity. Let X denote the categorical variable to be stratified on, e.g., sex, race, or age groups, where $\mathcal{X}$ is a discrete space. To restrict comparisons within

Table 4: Win ratio regression analysis of smoking behavior in the National Longitudinal Study of Youth on tobacco use.

	Naive			Pseudo-efficient		
	Estimate	Std error	p-value	Estimate	Std error	
Male	−1.087	0.046	<0.001	−1.156	0.043	<0.001
Nonwhite	0.623	0.053	<0.001	[illegible]	0.047	<0.001
Age (years)	0.014	0.021	[illegible]	[illegible]	0.018	0.155
Live w. parents	0.442	[illegible]	[illegible]	0.508	0.044	<0.001
NS mother	−0.537	[illegible]	[illegible]	−0.602	0.045	<0.001
Strict parents	[illegible]	[illegible]	[illegible]	0.623	0.045	<0.001
Attend school	[illegible]	[illegible]	<0.001	1.221	0.056	<0.001
Neg. discipline	[illegible]	0.053	<0.001	−0.722	0.046	<0.001

each stratum, simply treat $\mathcal{Y} \times \mathcal{X}$ as the new outcome space, equipped with the partial order $\preceq_{\mathrm{s}}$ such that $(y_i, x_i) \preceq_{\mathrm{s}} (y_j, x_j)$ if and only if $y_i \preceq y_j$ and $x_i = x_j$. Under this formulation, the results for nonparametric inference and semiparametric regression carry over without change.

Compared to standard parametric models, the win ratio provides a parsimonious approach to treatment or covariate effects, by focusing only on the global favorability of outcomes under the partial order. Nevertheless, there may be situations where parametric methods are preferred, particularly if the model suits the context of application. For example, the continuation-ratio model (3.4) allows us to evaluate the odds of moving to a higher category given the current status, providing intuitive interpretations for outcomes that proceed in stages, like educational attainment (Agresti, 2010). A smaller sample may also favor parametric methods (pending further investigation). On the other hand, an elaborate model increases the risk of misspecification, and residual analysis should be used to check model adequacy whenever possible.

The analysis of partially ordered data has received insufficient attention in the literature, and Theorem 1 and Proposition 5 provide only preliminary results on this topic. For the multiplicative win ratio model, in particular, the pseudo-efficient estimators show improvements over the

naive ones but are not guaranteed to be globally optimal. A complete semiparametric efficiency theory, traditionally relying on the concept of influence functions (Bickel et al., 1993), is complicated by the pairwise formulation of the model, which makes it hard to isolate the influence of each individual. Recently, Vermeulen et al. (2023) examined the efficiency problem for the probabilistic index model (Thas et al., 2012), a pairwise defined model for continuous and ordinal outcomes in strategies to help shed light on the structurally similar win ratio regression.

Appendix

Proof of Lemma 1

The inequalities follow by Definition 1 and $U(y)$ being an up-set. For strictness, suppose for a contradiction that $\nu_1\{U(y)\} = \nu_0\{U(y)\}$ for all y. Since the $U[y]$ are atomic posets, we must have that $\nu_1(U[y]) \geq \nu_0(U[y])$. This means that $\nu_1(\cdot) \succ \nu_0(\cdot)$ for all $y \in \mathcal{Y}$, which is impossible unless $\nu_1(y) = \nu_0(y)$ for all $y \in \mathcal{Y}$, contradicting the assumption that $\nu_1 \succ \nu_0$.

Proof of Proposition 2

For sufficiency of (3.3), use calculations similar to the proof of Proposition 1 to find that

$$\mathcal{R}(Z_i, Z_j) = \frac{\mathrm{pr}(Y_i = 1 \mid Z_i)\mathrm{pr}(Y_j = 0 \mid Z_j)}{\mathrm{pr}(Y_i = 0 \mid Z_i)\mathrm{pr}(Y_j = 1 \mid Z_j)} = \frac{\exp(\gamma + \beta^{\mathrm{T}} Z_i)}{\exp(\gamma + \beta^{\mathrm{T}} Z_j)}$$
$$= \exp\{\beta^{\mathrm{T}}(Z_i - Z_j)\}, \tag{6.11}$$

where the second equality follows by $\mathrm{pr}(Y = 1 \mid Z)/\mathrm{pr}(Y = 0 \mid Z) = \exp(\gamma + \beta^{\mathrm{T}} Z)$ under model (3.3). For necessity, take $Z_i = Z$ and $Z_j = z_0$ for some fixed z_0 in (3.2) and use the first equality in (6.11) to obtain (3.3) with $\gamma = \log\{\mathrm{pr}(Y = 1 \mid z_0)/\mathrm{pr}(Y = 0 \mid z_0)\} - \beta^{\mathrm{T}} z_0$.

Proof of Proposition 3

Let $S(y \mid Z) = \mathrm{pr}(Y \succeq y \mid Z)$. Under model (3.4), the numerator of $\mathcal{R}(Z_i, Z_j)$ is

$$\mathrm{pr}(Y_i \succ Y_j \mid Z_i, Z_j)$$
$$= \sum_{y \in \mathcal{Y}} \mathrm{pr}(Y_i \succ Y_j \mid Y_i \succeq y; Z_i, Z_j)\mathrm{pr}(Y_i \succeq y \mid Z_i)$$
$$= \sum_{y \in \mathcal{Y}} \mathrm{pr}(Y_i \succ Y_j \mid Y_i \succeq y; Z_i)\mathrm{pr}(Y_j = y \mid Y_j \succeq y; Z_j)S(y \mid Z_i)S(y \mid Z_j)$$
$$= \exp(\beta^{\mathrm{T}} Z_i) \sum_{y \in \mathcal{Y}} \frac{\exp\{\gamma(y)\}S(y \mid Z_i)S(y \mid Z_j)}{[1 + \exp\{\gamma(y) + \beta^{\mathrm{T}} Z_i\}][1 + \exp\{\gamma(y) + \beta^{\mathrm{T}} Z_j\}]}. \tag{6.12}$$

By symmetry, the denominator of $\mathcal{R}(Z_i, Z_j)$ is the far right hand side of (6.12) with the factor $\exp(\beta^{\mathrm{T}} Z_i)$ replaced with $\exp(\beta^{\mathrm{T}} Z_j)$. This yields the desired form of $\mathcal{R}(Z_i, Z_j)$ in (3.2).

Proof of Propostion 5

Any qualified scoring function r reduces to the [illegible] under $(\mathbb{N}_1, \leq)$. By Lemma (2), the weight defined by the [illegible] hand side of [illegible] with $\hat{\beta}_{\text{init}}$ replaced by β is efficient. But by Corollary 1, the replacement [illegible] by $\hat{\beta}_{\text{init}}$ does not alter the asymptotic distribution of the resu[illegible]r.

Supplementary Materials

Supplementary materials inclu[illegible] tech[illegible] resu[illegible] and additional numerical studies. An R-package `poset` th[illegible] ple[illegible]nts the proposed methodology is available on GitH[illegible] at h[illegible] `aowisc.github.io/poset` as well as th[illegible] Comprehensiv[illegible] Ar[illegible]e N[illegible]work (CRAN), both with a tutorial based on the liver [illegible] i[illegible] S[illegible] 5.1.

Acknowledgeme[illegible]

This research was supported by the U.S. National Institutes of Health grant R01HL149875 and National Science Foundation grant DMS-2015526.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. New York: John Wiley & Sons.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10**, 1100-1120.

Arcones, M. A. and Giné, E. (1993). Limit theorems for *U*-p... *Annals of Probabilit...* **21**, 1494–1542.

Armstrong, B. G. and Sloan, M. (1989). Ordinal regression models for epid... *American Journal of Epidemiology* **129**, 191–204.

Bebu, I. and Lachin, J. M. (2016). Large sample inference for a win ratio analysis of a composite outcome based on pri...itized components. ...o... **17**, 178–187.

Bickel, P. J., Klaassen, C. A., Ri...ov, Y. ...d We... J. A. (1993). *Efficient and Adaptive Estimation for Semiparame... Mod...* ...more: Johns Hopkins University Press.

Dong, G., Li, D., Ba...erst... S. ... VandeMeulebroecke, M. (2016). A generalized analytic s...tic... d...he... ...o to analyze a composite endpoint considering the clinical ... no... ...mponents. *Pharmaceutical Statistics* **15**, 430–437.

Edge, S., Byrd, D. ... ...ompton, C. C., Fritz, A. G. and Greene, F. L. (2010). *AJCC Cancer Staging Handbook: From the AJCC Cancer Staging Manual*. New York: Springer.

Fill, J. A. and Machida, M. (2001). Stochastic monotonicity and realizable monotonicity.

Annals of Probability **29**, 938–978.

Garg, V. K. (2015). *Introduction to Lattice Theory with Computer Science Applications*. Hoboken: Wiley.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**, 293–325.

Kamae, T. and Krengel, U. (1978). Stochastic partial ordering. *Annals of Probability* 1044–1049.

Lin, Y., Wang, S. & Chappell, R. J. (2018). Lasso tree for cancer staging with survival data. *Biostatistics* **14**, 327–339.

Luo, X., Tian, H., Mohanty, S. and Tsai, W. Y. (2015). An alternative approach to confidence interval estimation for the win ratio statistic. *Biometrics* **71**, 139–145.

Mao, L. (2018). On causal estimation using UA proxy. *Psychometrika* **105**, 215-220.

Mao, L. (2022). On the relative efficiency of the intent-to-treat Wilcoxon–Mann–Whitney test in the presence of non-compliance. *Biometrika*, **109**, 873–880.

Mao, L. and Wang, T. (2021). A class of proportional win-fractions regression models for composite outcomes. *Biometrics*, **77**, 1265–1275.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, **42**, 109–127.

Mondal, D. and Hinrichs, N. (2016). Rank tests from partially ordered data using importance

and MCMC sampling methods. *Statistical Science*, **31**, 325–347.

Peyhardi, J., Trottier, C. and Guédon, Y. (2016). Partitioned conditional generalized linear models for categorical responses. *Statistical Mododelling* **16**, 297–321.

Pocock, S. J., Ariti, C. A., Collier, T. J. and Wang, D. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical prioriti *European Heart Journal* **33**, 176–182.

Rosenbaum, P. R. (1991). Some poset statistics. *Annals of Sta*

Thas, O., Neve, J. D.,Clement, L. and Ottoy, J. P. (2012). Probabilistic index *Journal of the Royal Statistical Society, Series B*, **74**, 623–671.

Trotter, W. T. (1992). *Combinat ics and Partially Or Dimension Theory.* Baltimore: Johns Hopkins University Press.

van der Vaart, A. (1998). *Asymp* ambridge: Cambridge University Press.

van der Vaart, A. W. A A. (1996). *Weak Convergence and Empirical Processe* New Yo rlag.

De Neve, J., Thas, O. and Vansteelandt, S. (2023). Semiparametric es ation of probabilistic index models: efficiency and bias. *Statistica Sin* 24.

Wittkowski, K. M., Lee, E., Nussbaum, R., Chamian, F. N. and Krueger, J. G. (2004).

Combining several ordinal measures in clinical studies. *Statistics in Medicine* **23**, 1579–1592.

Zhang, Q. and Ip, E. H. (2012). Generalized linear model for partially ordered data. *Statistics in Medicine* **31**, 56–68.

Department of Biostatistics and Medical Informatics, School [illegible]e and Public Health, University of Wisconsin-Madison.

E-mail: lmao@biostat.wisc.edu