


Article

Multiple Ordinal Correlation Based on Kendall's Tau Measure: A Proposal

Juan M. Muñoz-Pichardo ¹, Emilio D. Lozano-Aguilera ^{2,*}, Antonio Pascual-Acosta ¹ and Ana M. Muñoz-Reyes ¹

¹ Department of Estadística e I.O., Universidad de Sevilla, 410112 Seville, Spain; juanm@us.es (J.M.M.-P.); antoniop@us.es (A.P.-A.); anamun@us.es (A.M.M.-R.)

² Department of Statistics and O.R., University of Jaén, 23071 Jaén, Spain

* Correspondence: elozano@ujaen.es; Tel.: +34-953-211913

Abstract: The joint analysis of various ordinal variables is necessary in many experimental studies within research fields such as sociology and psychology. Therefore, the necessary measures of multiple ordinal dependence must be easy to interpret and facilitate the interpretation of multivariate models that fit ordinal data. The main objective of this article is to propose a multiple ordinal correlation measure based on a bivariate correlation measure: Kendall's tau. A sample version of the measure is proposed for its estimation. Furthermore, a confidence interval and a multiple ordinal independence test are proposed. The measure is applied to various simulations, covering a wide range of multiple ordinal dependency scenarios, in order to illustrate the adequacy of the measure and the proposed inferential techniques. Finally, the measure is applied to a real-world study based on a social survey of the levels of life satisfaction and the happiness index of a population.

Keywords: ordinal data; multiple ordinal dependence; Kendall's tau measure; ordinal logistic regression; happiness index; life satisfaction



Citation: Muñoz-Pichardo, J.M.; Lozano-Aguilera, E.D.; Pascual-Acosta, A.; Muñoz-Reyes, A.M. Multiple Ordinal Correlation Based on Kendall's Tau Measure: A Proposal. *Mathematics* **2021**, *9*, 1616. <https://doi.org/10.3390/math9141616>

Academic Editors: María Del Mar Rueda and Andrés Cabrera-León

Received: 18 May 2021

Accepted: 29 June 2021

Published: 8 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The presence of ordinal variables is common in studies and experiences based on structured questionnaires and evaluations by judges and expert observers. Thus, ordinal data frequently arise in many fields, especially in the social and biomedical sciences. However, it is also common for the statistical analysis of these variables to be carried out by applying techniques oriented to the analysis of nominal variables (if the number of ordered categories is small) or techniques oriented to the analysis of numerical or quantitative variables (if the number of ordered categories is large). One can suppose that this is due to the difficulty of interpreting ordinal statistical models, such as ordinal regression models, and especially the difficulty of transferring statistical and mathematical conclusions to conclusions in their specific field of study (economic, social, biomedical, etc.). Good works of presentation, development, and interpretation of techniques for modeling ordinal variables are presented by Agresti [1] and Fahremeir et al. [2].

This difficulty of interpretation does not arise when bivariate ordinal correlation measures are applied. The two most popular measures of association for ordinal random variables are Kendall's tau [3] and Spearman's rho [4]. However, to the best of our knowledge, it is necessary to propose new multiple ordinal correlation measures that are easily interpretable for researchers from different scientific areas. That is, measures describing the ordinal correlation of a target ordinal variable against a set of explanatory variables (whether they are ordinals or not). Agresti [5] proposes a generalization of Kendall's tau defined in terms of the proportional reduction in prediction errors obtained by predicting the ordering of pairs of observations on the objective variable based on orderings of the pairs on the explanatory variables. However, it does not have as simple an interpretation as bivariate Kendall's tau. Nagelkerke [6] proposes a general definition of a likelihood-based coefficient of determination, from which correlation or association measures can be

constructed. Specifically, Muñoz-Pichardo et al. [7] apply this idea to regression models for multivariate count data. In the case of continuous numerical variables, the sample multiple linear correlation coefficient (R^2) is available, which can be interpreted as a measure of linear statistical dependence and a measure of fit of a model.

In the present work, a multiple ordinal correlation measure is proposed that uses the concept of the Kendall's tau coefficient as a basis and is constructed following a similar path to the multiple linear correlation coefficient, considering the adjusted values through a model of ordinal regression. That is, it is not based on the model, nor on the prediction of the ordering through the joint ordering of the explanatory variables. The proposed measure is based on the prediction of the ordering through the ordering of the values fitted by the model. This model-based strategy of construction of statistical dependency measures can be applied to many models and types of variables. Once the measure is defined in what follows, then a methodology is proposed to develop its statistical inference, based on resampling methods to determine bias, confidence interval, and hypothesis testing. To illustrate the application of this methodology, simulations of various scenarios are carried out, from a scenario with statistical ordinal independence to a scenario of almost perfect ordinal dependence.

To illustrate the adequacy of the proposed measure and its easy interpretation, it is applied in the field of social studies. In recent years, there is an increasing need to evaluate the society with a focus on the population and with such essential objectives as happiness and satisfaction with life. Within this framework of social and scientific interest, a survey was carried out in the geographical area of Andalusia (southern Spain). Its main objective was to obtain a personal assessment of the level of happiness, the level of life satisfaction in general, and certain specific aspects of life (family, education, work, etc.). Hence, the study determined a happiness index and a level of satisfaction with life in general, in addition to a life satisfaction profile defined using ordinal measures on various topics or issues of personal life. The modeling of life satisfaction and the degree of happiness in relation to other specific aspects of life has been studied by other authors (for example, refer to the work of Chow [8] on life satisfaction among university students). Multiple ordinal dependence of the happiness index against the satisfaction profile and the degree of general satisfaction with the satisfaction profile will allow us to illustrate the adequacy of the ordinal correlation measure proposed in the present work.

2. Statistical Model: Ordinal Multiple Regression

A large number of parametric ordinal models can be found in the literature. A possible classification of these models determines three categories [9]: cumulative models, sequential models, and adjacent-category models. Bürkner and Vuorre [9] indicate that the adjacent-category models are literally chosen for their mathematical convenience rather than any quality of interpretation, because it is difficult to envisage a natural process leading to these models. Fahrmeir et al. [2] argue that, due to its construction, the sequential model is useful when the categories of the response variable are obtained sequentially, that is, a category can only be attained provided that the previous borderline category has been attained. Therefore, we have chosen to apply a cumulative model for the analysis of the experience presented in the introduction and developed in Section 5. This model is one that provides the best results in a data set associated with this experience. Furthermore, the methodological proposal to construct the multiple association measure can be adapted to the other ordinal regression models.

The cumulative model of ordinal regression [1] consists of relating the logit of the cumulative probability of the target variable with the linear predictor of the explanatory

variables. Thus, given the objective variable Y that takes the values $1, 2, \dots, K$, the advantages (odds) and its logarithmic transformation (logits) are considered:

$$\begin{aligned} \text{odds}(j; \mathbf{x}) &= \frac{\Pr[Y \leq j | \mathbf{x}]}{\Pr[Y > j | \mathbf{x}]}, \\ \text{logit}(j; \mathbf{x}) &= \log[\text{odds}(j; \mathbf{x})] = \beta_{0,j} - \boldsymbol{\eta}' \mathbf{x} = \beta_{0,j} - \eta_1 x_1 - \dots - \eta_p x_p, \end{aligned} \quad (1)$$

$j = 1, \dots, K - 1.$

Hence, Agresti [1] calls this a cumulative link model. Thus, the model represents each logit as a function of the linear predictor of the explanatory or predictive variables. Each cumulative logit has its own intercept: the $\{\beta_{0,j}\}$ are increasing in j , because $\Pr[Y \leq j | \mathbf{x}]$ increases in j for \mathbf{x} fixed. Furthermore, the model assumes identical effects of the covariates for each logit. If we denote by $\gamma_j = \exp(\beta_{0,j})$, the conditional probability distribution function that describes the model is given by:

$$\Pr[Y \leq j | \mathbf{x}] = \frac{\gamma_j \exp(-\boldsymbol{\eta}' \mathbf{x})}{1 + \gamma_j \exp(-\boldsymbol{\eta}' \mathbf{x})}, \quad j = 1, \dots, K - 1. \quad \Pr[Y \leq K | \mathbf{x}] = 1. \quad (2)$$

The model described by the conditional probability distribution is denoted as $\text{OrdMod}(Y | X_1, \dots, X_p)$. We can easily obtain the probability function:

$$\Pr[Y = 1 | \mathbf{x}] = \Pr[Y \leq 1 | \mathbf{x}], \quad (3)$$

$$\Pr[Y = j | \mathbf{x}] = \Pr[Y \leq j | \mathbf{x}] - \Pr[Y \leq j - 1 | \mathbf{x}], \quad j = 2, \dots, K. \quad (4)$$

In the case where the explanatory variables are not significant in the model, that is, the objective variable is not associated with the explanatory variables, the resulting model, which can be considered as the base model, has the following cumulative probabilities:

$$\Pr[Y \leq j] = \frac{\gamma_j}{1 + \gamma_j}, \quad j = 1, \dots, K - 1. \quad \Pr[Y \leq K] = 1. \quad (5)$$

Consequently, it must be verified that the constants γ_j , generators of the probabilities of the base model, must be ordered in the form: $0 \leq \gamma_1 \leq \dots \leq \gamma_{K-1}$. In the ordinal regression model, the generating constants of the conditional ordinal distribution $Y | \mathbf{x}$ are transformed to a function of the covariates:

$$\gamma_j(\mathbf{x}) = \gamma_j \exp(-\boldsymbol{\eta}' \mathbf{x}) = \gamma_j \exp(-\eta_1 x_1 - \dots - \eta_p x_p). \quad (6)$$

Thus, for a covariant X_1 binary factor (values 0,1), if $\eta_1 > 0$ then the presence of that factor leads to a contraction ($e^{-\eta_1}$) that affects all the generating constants. Similarly, if $\eta_1 < 0$ then the presence leads to an expansion of these constants. To facilitate an interpretation, let us consider the odds ratios (OR) associated with this covariate:

$$\log \text{OR}(j; X_1) = \text{logit}[j; (1, x_2, \dots, x_p)] - \text{logit}[j; (0, x_2, \dots, x_p)] = -\eta_1, \quad (7)$$

otherwise,

$$\text{OR}(j; X_1) = \frac{\text{odds}(j; X_1 = 1)}{\text{odds}(j; X_1 = 0)} = \exp(-\eta_1). \quad (8)$$

Consequently, if $\eta_1 > 0$ (equivalently, associated odds ratio less than 1), then

$$\frac{\Pr[Y \leq j | (1, x_2, \dots, x_p)]}{\Pr[Y > j | (1, x_2, \dots, x_p)]} < \frac{\Pr[Y \leq j | (0, x_2, \dots, x_p)]}{\Pr[Y > j | (0, x_2, \dots, x_p)]}, \quad j = 1, \dots, K - 1, \quad (9)$$

that is, the binary factor X_1 is a covariant whose “presence” leads to improve or increase the objective variable Y . In the opposite direction, the case $\eta_1 < 0$ can be interpreted.

Conversely, if X_1 is a quantitative or ordinal variable, with $\eta_1 > 0$, the increase of one unit in it leads to the increase of the objective variable and vice versa.

The estimators obtained through maximum likelihood, with approximate numerical methods, allow obtaining the fitted model. Specifically, the collection of parameters that must be estimated is $\{\beta_{0,1} \dots \beta_{0,K-1}, \eta_1 \dots \eta_p\}$. For the study of the significance of the parameters, the asymptotic likelihood ratio test (LRT) can be applied, as well as the global significance test of the model. Furthermore, as measures of goodness of fit, measures based on the likelihood of the model can be considered: the residual deviance statistic and the Akaike Information Criterion (AIC) [10]. The detailed development of the inference on this model is collected in the works of Agresti [1], McCullagh [11], and Fahrmeir et al. [2].

With such estimators, the model allows “predicting” the values of the target ordinal variable through the covariates:

$$\hat{y} = \arg \max_{j=1 \dots K} \hat{\Pr}[Y = j | \mathbf{x}]. \tag{10}$$

This makes it possible to analyze the residuals to evaluate the adjustment and the predictive capacity of the model $OrdMod(Y|X_1, \dots, X_p)$. Following the approach of Agresti (2010), we consider as a classification or prediction rule the response category that has the highest estimated probability or mode of the conditional distribution $Y|X = \mathbf{x}$. This approach may suffer from some problems such as:

Cases in which one outcome is much more likely than the others, this can result in always or nearly always predicting that category. This problem leads to the absence of multiple ordinal association.

Cases in which the conditional distribution of the ordinal response variable is bimodal. For the purpose of defining the measure, it has been chosen to randomize between modal categories. A detailed study of the problem may be the subject of future work.

Other prediction rules have been proposed for this model (see Agresti [12] and Marshall [13]), but we believe that prediction in the modal category is the most appropriate rule for defining the measure of association.

3. A Measure of Ordinal Multiple Correlation and Its Inference

3.1. Definition of the Model-Based Multiple Kendall's Tau

To complete the study of the ordinal regression model, it is necessary to obtain a measure of the multiple dependency or multiple association between the target ordinal variable and the collection of explanatory variables included in the model, that is, a measure similar to the multiple correlation coefficient in the multiple regression model under normal conditions. The multiple correlation coefficient between a quantitative random variable U and a quantitative random p -vector V is defined as the maximum correlation between U and any linear combination of V , $\beta'V$ [14]:

$$\rho^2(U|V) = \max_{\beta \in \mathbb{R}^p} \rho^2(U, \beta'V). \tag{11}$$

Another equivalent way of defining this coefficient is, under normal conditions, the linear correlation coefficient between U and the random variable $U^* = E[U|V]$, with support identical to U and a function of V :

$$\text{If } V = v \text{ then } U^* = E[U|V = v]. \tag{12}$$

That is,

$$\rho^2(U|V) = \rho^2(U, U^*) \tag{13}$$

given that,

$$U^* = \beta'_{OPT} V : \beta_{OPT} = \arg \max_{\beta \in \mathbb{R}^p} \rho^2(U, \beta'V). \tag{14}$$

Consequently, two relevant aspects can be highlighted: first, the multiple correlation coefficient is a measure of the linear dependence between one random variable and a certain collection of random variables; second, the linear dependence is measured with respect to the best linear combination that approximates U according to the multiple regression model [9,10]. As a result, the multiple correlation coefficient is the linear correlation coefficient between U and its prediction $\beta'V$ based on the model. It is generally considered squared since the sign of the correlation is not understandable in the multiple case.

From a sampling perspective, the estimator of the sample multiple correlation coefficient is obtained as the sample linear correlation coefficient between the observed values of U and their values adjusted for the model $\hat{U} = \hat{\beta}'V$ with $\hat{\beta}$ as the maximum likelihood estimator of the regression coefficients or parameters of the regression model.

In parallel with the analysis of multiple ordinal dependence, we propose a new measure based on the ordinal regression model, $OrdMod(Y|X_1, \dots, X_p)$, and Kendall's bivariate ordinal dependence measure or Kendall's τ coefficient.

The non-parametric correlation coefficient (or measure of association) known as Kendall's tau [3] is defined as follows: let (U_1, V_1) and (U_2, V_2) be independent two-dimensional random vectors with the same distribution as (U, V) , then

$$\tau(U, V) = \Pr[(U_1 - U_2)(V_1 - V_2) > 0] - \Pr[(U_1 - U_2)(V_1 - V_2) < 0]. \tag{15}$$

That is, the Kendall's τ coefficient is the probability of agreement minus the probability of disagreement between the variables. This coefficient can also be defined (see [15]) as the expected value of the random variable

$$A = \text{sgn}(U_1 - U_2) \text{sgn}(V_1 - V_2) \quad \text{where } \text{sgn}(w) = \begin{cases} -1 & \text{if } w < 0 \\ 0 & \text{if } w = 0 \\ 1 & \text{if } w > 0 \end{cases}. \tag{16}$$

Its estimation, through a sample $\{(u_i, v_i), i = 1 \dots n\}$ is obtained as $\hat{\tau} = \frac{2S}{n(n-1)}$, named Kendall's sample tau, where S is the number of concordant pairs minus the number of discordant pairs.

Thus, a measure of multiple ordinal dependence between an ordinal variable Y and a random p -vector $\mathbf{X} = (X_1, \dots, X_p)'$ can be defined as Kendall's τ coefficient between Y and the prediction with the best linear combination of \mathbf{X} with respect to the $OrdMod(Y|X_1, \dots, X_p)$ model. In other words, it can be called the model-based multiple Kendall's tau between Y and \mathbf{X} and is denoted by $\tau_M(Y, \mathbf{X})$. That is, $\tau_M(Y, \mathbf{X})$ is the Kendall's τ coefficient between the ordinal variable Y , with support $\{1, \dots, K\}$, and the ordinal variable Y^* with support identical to Y defined according to $OrdMod(Y|X_1, \dots, X_p)$, that is,

$$\text{if } \mathbf{X} = \mathbf{x} \text{ then } Y^* = \text{mode}[Y|\mathbf{X} = \mathbf{x}], \tag{17}$$

with $\text{mode}[Y|\mathbf{X} = \mathbf{x}]$ the modal value of conditioned distribution, or equivalently

$$\text{if } \mathbf{X} = \mathbf{x} \text{ then } Y^* = \arg \max_{j=1 \dots K} \Pr[Y = j | \mathbf{x}]. \tag{18}$$

Thus, $Y^* = \text{mode}[Y|\mathbf{X}]$ is a function of the random vector \mathbf{X} , as in the regression model under normal conditions, $U^* = E[U|V]$ is a function of the random vector V . Consequently, $\tau_M(Y, \mathbf{X}) = \tau(Y, Y^*)$.

Evidently, by the definition of $\tau_M(Y, \mathbf{X})$, $0 \leq \tau_M(Y, \mathbf{X}) \leq 1$. In the case of the ordinal independence of Y with respect to the random vector \mathbf{X} , the variable Y^* will be constant, $Y^* = \text{mode}(Y)$, consequently, by the definition of Kendall's tau coefficient, $\tau_M(Y, \mathbf{X}) = \tau(Y, Y^*) = 0$. In general, the converse is not true, but this disadvantage is shared by the multiple linear correlation coefficient. Appendix A illustrates this measure through a simple example.

If $\tau_M(Y, X) = 1$, the relationship is direct and perfect in the sense that,

$$Y_i < Y_j \text{ whenever } Y_i^* = \text{mode}[Y|X = x_i] < Y_j^* = \text{mode}[Y|X = x_j] \tag{19}$$

or

$$Y_i > Y_j \text{ whenever } Y_i^* > Y_j^*, \tag{20}$$

that is, perfect concordance or agreement between the target ordinal variable and the ordinal variable adjusted to the model $OrdMod(Y|X_1, \dots, X_p)$.

In practice, there will be scenarios where perfect concordance and independence are absent. In these scenarios, increasing degrees of concordance between Y and Y^* are reflected by increasing positive values of $\tau_M(Y, X)$.

Since the ordering of an ordinal variable is arbitrary, we can wonder if the definition is coherent with respect to a change of variable from Y to $K + 1 - Y$. The answer is positive since the τ_M 's Kendall measures the ordinal correlation between the target variable and the model prediction variable. We can also speculate about the problem of ties that occur often with ordinal data. Since the τ_M measure is defined as a bivariate measure, this problem can be approached identically to the bivariate case (see Agresti [12]).

3.2. Inference: Estimation, Confidence Interval, and Test

Proceeding analogously to the multiple regression model, an estimate of the aforementioned measure via a sample $\{(y_i, x_i), i = 1 \dots n\}$ can be the sample Kendall's tau of $\{(y_i, \hat{y}_i), i = 1 \dots n\}$, where \hat{y}_i are the fitted values of the model,

$$\hat{y}_i = \arg \max_{j=1 \dots K} \hat{P}r[Y = j|x_i]. \tag{21}$$

With this sample measure of multiple ordinal dependence, $\hat{\tau}_M(Y, X)$, it is possible to construct a test on the ordinal non-association of Y with respect to the vector X , that is, a test on the null hypothesis

$$H_0 : \tau_M(Y, X) = 0. \tag{22}$$

As noted by Bergsma and Dassios [16], a test of independence based on i.i.d. data can be obtained by application of the permutation test (or randomization test) to an estimator of τ_M . For moderately large sample sizes, it is not appropriate to use an exact permutation test. As with bootstrapping, a permutation test constructs the sampling distribution of statistics of the test, under the null hypothesis, by resampling the observed data through permutations. Specifically, it can be applied to the Monte Carlo approximation or resampling test, that is, a test of randomly selected permutations. In this case, starting from the initial sample $\{(y_i, x_i), i = 1 \dots n\}$, a number B of permutations of $\{1, \dots, n\}$ is determined,

$$\{p_k(1), \dots, p_k(n)\}, \quad k = 1, 2, \dots, B. \tag{23}$$

Through this, the B permuted samples are obtained of the form:

$$\left\{ \left(y_{p_k(i)}, x_i \right); i = 1 \dots n \right\}, \quad k = 1, 2, \dots, B. \tag{24}$$

The sample multiple Kendall's tau is calculated for each of the permuted samples, thus obtaining a collection of values $\{\hat{\tau}_{M(k)}; k = 1, \dots, B\}$ which allows obtaining a sampling distribution of the measure under the null hypothesis of ordinal independence between Y and X . This sampling distribution can be used to perform the test, since it is an approximate distribution to the distribution of the set of all $n!$ permutations. Specifically, an approximation to the permutation-achieved significance level (ASL) is given by

$$\frac{1}{B} \# \left\{ \hat{\tau}_{M(k)} \geq \hat{\tau}_M(Y, X) \right\}. \tag{25}$$

Finally, the confidence interval (CI) must be obtained to complete the inference on the parameter $\tau_M(Y, X)$. Given the difficulty of obtaining an explicit expression and in accordance with the work of Ruscio [17], the conclusion is that “coverage of bootstrap CIs was usually as good or better than coverage of analytic CIs”. Furthermore, with the application of the bootstrap methodology, estimates of the bias and the estimation error of the estimator $\hat{\tau}_M(Y, X)$ can be obtained. According to Canty [18], under suitable regularity conditions, given a theoretical distribution F , an associated functional or parameter $\theta(F)$ and an estimator $\hat{\theta} = \theta(\hat{F})$ defined on the empirical distribution function \hat{F} of a random sample, we can approximate the bias distribution of the estimator ($\hat{\theta} - \theta$) through the empirical distribution of $\hat{\theta}^* - \hat{\theta}$, where $\hat{\theta}^* = \theta(\hat{F}^*)$ is the estimator defined on the bootstrap distribution. That is, considering R bootstrap resamples of the original sample, the estimated bias is

$$(\hat{\theta}) = \overline{\hat{\theta}^*} - \hat{\theta}, \tag{26}$$

with $\overline{\hat{\theta}^*} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r^*$ and $\{\hat{\theta}_r^* : r = 1 \dots R\}$ being the collection of estimates obtained in the R bootstrap resamples. Similarly, an estimate of the variance of the estimator is

$$v(\hat{\theta}) = \frac{1}{R-1} \sum_{r=1}^R (\hat{\theta}_r^* - \overline{\hat{\theta}^*})^2 \tag{27}$$

and, consequently, its standard error of estimation is $v^{1/2}(\hat{\theta})$.

The boot package [18,19] has implemented various methods of obtaining bootstrap CIs: basic bootstrap CI (also known as reverse percentile bootstrap CI), percentile bootstrap CI, studentized bootstrap CI, and bias-corrected and accelerated bootstrap CI (BCa). We believe that this last method may be the most suitable for our objective. As noted by Jung et al. [20], to overcome the problems of overcoverage in the bootstrap confidence intervals based on percentiles [21,22], the BCa method corrects both the bias and the skewness of the bootstrap estimates of the parameters by incorporating a bias correction factor and an acceleration factor [18,19]. In the present study, $R = 2000$ bootstrap resamples were both generated and analyzed because this value is recommended in the bootstrap literature [23].

4. Simulation Study

In order to illustrate the adequacy of the proposed multiple ordinal correlation measure, the situations described below are randomly generated, beginning with the simple case ($p = 1$), with a strong association, either positive or negative. Subsequently, the multiple case is analyzed ($p = 3$), from a strong multiple association to ordinal independence. In all simulations, the sample size generated is $n = 500$.

All computations have been performed in R [24].

- **Simulation 1:** positive association with a single variable. An explanatory variable X_1 distributed according to $N(0, 1)$ and an ordinal variable Y , with support $\{1, 2, \dots, 10\}$ are generated following the pattern of the conditional distribution described in the $OrdMod(Y|X_1)$ with coefficient $\eta_1 = 1 > 0$. The base distribution of the ordinal variable Y is considered a symmetric distribution with probabilities.

$$\{Pr[Y = k]\}_{k=1}^{10} = \{0.05, 0.075, 0.10, 0.125, 0.15, 0.15, 0.125, 0.10, 0.075, 0.05\}$$

To avoid the overfitting of the model and the impossibility of carrying out the calculations to obtain the estimators, an error is randomly included in 2.5% of cases, in which the value of Y is generated randomly according to a uniform discrete distribution with support $\{1, 2, \dots, 10\}$. That is, the errors are generated independently of the explanatory variables (see Appendix B). This generation process is also applied in the following simulations.

- **Simulation 2:** negative association with a single variable. Identical to Simulation 1, with coefficient $\eta_1 = -1 < 0$.

- **Simulation 3:** association with a three-dimensional vector (2.5% error). Three explanatory variables X_1, X_2, X_3 that are independently distributed according to $N(0, 1)$ are generated and also an ordinal variable Y is generated with support $\{1, 2, \dots, 10\}$ following the pattern of the conditional distribution described in the model $OrdMod(Y|X_1, X_2, X_3)$ with negative and positive coefficients $\eta = (0.8, -1.0, 0.2)'$. This is intended to include positively (X_1, X_3) and negatively (X_2) associated variables, with both strong (X_1, X_2) and weak (X_3) intensity. The base distribution of Y is considered to be identical to the base distribution included in the previous simulations, with the specified error.
- **Simulation 4:** association with a three-dimensional vector (5% error). Simulation procedure is identical to that collected in Simulation 3, with error in 5% of the cases.
- **Simulation 5:** association with a three-dimensional vector (10% error). Simulation procedure identical to that in Simulation 3, with error in 10% of cases.
- **Simulation 6:** association with a three-dimensional vector (error 25%). Simulation procedure identical to that in Simulation 3, with error in 25% of cases.
- **Simulation 7:** association with a three-dimensional vector (50% error). Simulation procedure identical to that in Simulation 3, with error in 50% of cases.
- **Simulation 8:** association with a three-dimensional vector (75% error). Simulation procedure identical to that in Simulation 3, with error in 75% of cases.
- **Simulation 9:** association with a three-dimensional vector (90% error). Simulation procedure identical to that in Simulation 3, with error in 90% of cases.
- **Simulation 10:** null association or multiple ordinal independence. Three explanatory variables X_1, X_2, X_3 independently distributed in accordance with $N(0, 1)$ are generated and an ordinal variable Y is also generated according to a discrete uniform distribution with support $\{1, 2, \dots, 10\}$. That is, it can be considered to be generated following the ordinal model, including a 100% error.

The results of the simulations are shown in Table 1. For each simulation, the parameters used are indicated (error and coefficients of the linear predictor), the bivariate Kendall's tau coefficients are obtained with their corresponding level of significance (calculated through the function "cor.test ()" of R-Program [24]) and, finally, the multiple ordinal correlation coefficients and their significance obtained according to the permutations test with number of replications $N_{rep} = 2000$ are collected. For fitting of the models, the function *polr ()* of the MASS package [25] is applied.

The proposed measure adequately reflects the multiple ordinal correlation. In the scenario defined for Simulation 3, the data have been generated according to the multiple ordinal model, except for a disturbance in 2.5% of the data generated, obtaining an estimate of the multiple Kendall's tau measure $\hat{\tau}_M(Y, X) = 0.9501$, which reflects the almost functional relationship of the target variable versus the explanatory variables.

From Simulation 4 to the final simulation, there is an increase in the percentage of data generated that does not respect the model, with a gradual decrease in the measurement (seen in Simulation 4 $\hat{\tau}_M(Y, X) = 0.9397$ and Simulation 9 $\hat{\tau}_M(Y, X) = 0.0429$). Finally, in the last scenarios (Simulations 9 and 10), the data of the objective variable have been generated quasy-independent or independently of the predictor variables, obtaining estimates of the measure of multiple ordinal correlation that are practically null ($\hat{\tau}_M(Y, X) = 0.0429, 0.0297$), thus being able to accept the nullity of $\tau_M(Y, X)$ according to the permutations test (p -value $\approx 0.2715, 0.3095$).

Table 1. Results obtained in the simulations.

Simul.	Error	Variable	Coefficient η	Bivariate Kendall's τ		Multiple Kendall's τ	
				Estimate	p -Value ¹	Estimate	p -Value ²
1	2.5%	X_1	1.0	0.8607	0.0000	0.9085	0.0000
2	2.5%	X_1	-1.0	-0.8963	0.0000	0.9631	0.0000
3	2.5%	X_1	0.8	0.4141	0.0000	0.9501	0.0000
		X_2	-1.0	-0.5608	0.0000		
		X_3	0.2	0.0974	0.0021		
4	5%	X_1	0.8	0.4620	0.0000	0.9397	0.0000
		X_2	-1.0	-0.5331	0.0000		
		X_3	0.2	0.0982	0.0018		
5	10%	X_1	0.8	0.3595	0.0000	0.8291	0.0000
		X_2	-1.0	-0.4912	0.0000		
		X_3	0.2	0.0742	0.0189		
6	25%	X_1	0.8	0.3392	0.0000	0.6617	0.0000
		X_2	-1.0	-0.4301	0.0000		
		X_3	0.2	0.0440	0.1629		
7	50%	X_1	0.8	0.2019	0.0000	0.3578	0.0000
		X_2	-1.0	-0.2557	0.0000		
		X_3	0.2	0.0397	0.0207		
8	75%	X_1	0.8	0.0515	0.1006	0.1678	0.0000
		X_2	-1.0	-0.1242	0.0001		
		X_3	0.2	0.0034	0.9129		
9	90%	X_1	0.8	0.0298	0.3417	0.0429	0.2715
		X_2	-1.0	-0.0417	0.1840		
		X_3	0.2	-0.0039	0.8994		
10	100%	X_1	0.0	0.0266	0.3964	0.0297	0.3095
		X_2	0.0	0.0086	0.7826		
		X_3	0.0	0.0388	0.2157		

¹ Obtained using the function `cor.test()` from R-Program. ² Obtained through the permutations test.

In Scenarios 1 and 2, the data have been generated according to the $OrdMod(Y|X_1)$ model with positive and negative coefficients (η_1), respectively. The bivariate Kendall's tau coefficient does not coincide with the multiple Kendall's tau coefficient. This fact is logical since the former is a measure of the ability to predict the order of Y through the order of X_1 , $\tau(Y, X)$, while the latter is a measure of the ability to predict the order via the order of the variable fitted by the model, $\tau(Y, Y^*)$. Furthermore, given that the data have been generated with respect to this model, in both scenarios, the multiple measure is greater than the bivariate measure in absolute terms.

Table 2 shows the BCa bootstrap confidence intervals for each of the simulations. They have been obtained from the bootstrap library [15,16], generating $R = 2000$ bootstrap resamples. In the estimates of bias and standard error of estimation, the accuracy of the estimation procedure can be observed: the absolute value of the bias is always less than 0.01, except in Simulation 9, with $bias = 0.0208$; the values of the *std.error* are all less than 0.05. To carry out these calculations, problems have arisen in Simulations 1, 2 and 3, generated with only 2.5% error. That is, only 2.5% of the sample observations differ from the exact value generated by the model. Consequently, in a large number of bootstrap resamples, all the observations respect the model exactly. This situation means that the fit of the ordinal regression model for these resamples cannot be estimated due to overfitting in the model, or that the results in these simulations are not very stable. However, with exercising caution in these cases, the general conclusion to be drawn is that the estimation procedure, except for singular cases, is unbiased and very accurate.

Table 2. Inference on multiple Kendall's τ for data sets of simulations.

Simul.	Error	Multiple Kendall's τ				
		Estimate	Bias ¹	Std. Error ¹	95% BCa Bootstrap CI ¹	p -Value ²
1	2.5%	0.9085	−0.0001	0.0242	(0.8428, 0.9457)	0.0000
2	2.5%	0.9631				0.0000
3	2.5%	0.9501				0.0000
4	5%	0.9397	−0.0010	0.0143	(0.9046, 0.9642)	0.0000
5	10%	0.8291	−0.0019	0.0277	(0.7728, 0.8795)	0.0000
6	25%	0.6617	0.0022	0.0337	(0.5792, 0.7193)	0.0000
7	50%	0.3578	0.0072	0.0479	(0.2352, 0.4329)	0.0000
8	75%	0.1678	−0.0070	0.0413	(0.0908, 0.2506)	0.0000
9	90%	0.0429	0.0208	0.0452	(−0.0990, 0.1056)	0.2715
10	100%	0.0297	−0.0014	0.0478	(−0.0510, 0.1415)	0.3095

¹ Using the function "boot.ci ()" of the "boot" package. ² Obtained through the permutations test.

5. Application to a Social Study: Happiness Index and Life Satisfaction Level

As presented in the introduction, the data for the study consist of an anonymized data set obtained through a survey that was carried out in Andalusia (southern Spain) [26]. One of the objectives of the survey was the analysis of the level of happiness and the degree of life satisfaction of the population, as well as obtaining an approximation to the main problems that affect and concern that population. In this survey, the concepts of happiness and satisfaction with life, that were subjectively considered by the study participants who were interviewed, were assumed. As the concepts are subjective, consequently, they are not completely identical for all individuals, in the same way that there is no broad consensus in the scientific and philosophical fields. Although the concepts have different nuances for each individual, in a practical sense they coincide fundamentally for everyone, especially when restricted to a population under the same sociocultural and temporal patterns, which in our case is the population of Andalusia.

However, since the measures of happiness and life satisfaction are obtained through a survey, both are measures of the *hic et nunc* (here and now), filled with a wealth of sensations that provide spontaneity and possible volatility caused by poor reflection on the part of the individuals. Specifically, the people interviewed were asked about their overall satisfaction with life and about other dimensions related to living conditions, work, leisure, and social relationships. The respondents had to respond on a Likert-type scale of 0 to 10, where 0 is defined as completely dissatisfied and 10 is completely satisfied. Thus, it is possible to work with a life satisfaction profile through ordinal measurements of nine areas or aspects of personal life (see Table 3). Similarly, for the level of happiness the Likert-type scale from 0 to 10 is used, in which 0 is defined as completely unhappy and 10 is completely happy. Table 3 presents the variables considered.

The number of interviewees who answered all questions associated with the previous indices is $n = 445$. A statistical summary is included in Table 3.

The happiness index and the degrees of satisfaction can be considered as ordinal variables. Therefore, for the analysis of the bivariate correlation, Kendall's tau measure is considered. Given that the two target variables are Sgen and Ihap, the bivariate ordinal correlations are obtained using the remaining variables and applying the corrplot library [27]. The results are presented in Table 4. The degrees of satisfaction regarding the various aspects considered are positively correlated with the two objective variables. Furthermore, as one might suspect, the two target variables are correlated with each other, with sample Kendall's tau equal to 0.568.

Table 3. Statistical summary of the life satisfaction profile and happiness index.

Degree of Satisfaction with . . .	Mean	Q1	Median	Q3	Min	Max
Life satisfaction profile						
Sfam: family life	8.54	8	9	10	0	10
Sjob: job situation and work activity (including housework)	6.98	6	8	8	0	10
Sleis: leisure or recreational activities	7.38	6	8	9	0	10
Sfrie: relationships with close friends	8.13	8	8	9	0	10
Shea: health condition	7.34	6	8	9	0	10
Sphys: physical appearance	7.44	7	8	8	0	10
Seduc: education and professional training	6.86	5	8	9	0	10
Sbeha: their way of being and social behavior	8.08	7	8	9	3	10
Srela: couple and/or romantic relationships	7.05	6	8	9	0	10
Satisfaction with life in general						
Sgen: life in general	7.94	7	8	9	0	10
Happiness Index						
Ihap: happiness Index	7.70	7	8	9	0	10

Table 4. Sample bivariate Kendall's τ measures ¹.

	Sfam	Sjob	Sleis	Sfrie	Shea	Sphys	Seduc	Sbeha	Srela	Sgen	Ihap
Sgen	0.483	0.388	0.384	0.365	0.289	0.335	0.242	0.378	0.395	1	0.568
Ihap	0.429	0.310	0.274	0.264	0.236	0.261	0.175	0.290	0.346	0.568	1

¹ All measurements are significant with $p < 0.001$.

Two ordinal regression models are carried out to determine the dimensions related to the living conditions that have the most influence over the level of happiness and the degree of satisfaction with life in general.

First, Ihap is considered as the objective variable and the degrees of satisfaction with specific aspects of life are considered as explanatory variables: Sfam, Sjob, Sleis, Sfrie, Shea, Sphys, Seduc, Sbeha, and Srela. To estimate the model, the *polr* () function from the MASS package [25] of the R-Program is applied, thus obtaining the results shown in Table 5. Thus, the significant covariates in the model are the degrees of satisfaction with family, work, leisure, the way of being, and romantic relationships in addition to the intercepts. All the coefficients associated with these predictors are positive, that is, they are positively associated with the happiness index, particularly the degrees of satisfaction Sfam and Sbeha. A deeper and more extensive interpretation can be made regarding the field of sociology, but it is not the object of interest in the present study.

The proposed multiple ordinal correlation measure, the multiple Kendall's tau, attains the value of 0.501, with significance $p < 0.001$, this latter being obtained through the permutation test with 2000 permutations generated. That is, the ordinal correlation was improved through the adjustment of the model, with respect to the bivariate correlations (see Tables 4 and 7). Finally, the BCa method was applied to obtain a confidence interval. Table 7 shows the results obtained. The estimation of the bias and the standard error of estimation provides new evidence for the accuracy and unbiasedness of the estimation method.

The statistical interpretation of the adjusted model, the estimates of its parameters, and the multiple ordinal correlation measure can lead to an adequate sociological interpretation of the results toward determining the aspects of people's lives that most favorably influence their happiness and evaluating the capacity of the vital aspects considered in explaining the level of happiness.

Table 5. Model: Ihap on the life satisfaction profile ¹.

Variables	Coef.	Std. Error	t Value	p-Value
Sfam	0.569	0.076	7.450	0.000 ***
Sjob	0.133	0.043	3.066	0.002 **
Sleis	0.139	0.048	2.919	0.004 **
Sfrie	−0.001	0.065	−0.022	0.982
Shea	0.054	0.047	1.133	0.257
Sphys	0.075	0.064	1.171	0.242
Seduc	0.010	0.042	0.233	0.816
Sbeha	0.326	0.085	3.857	0.000 ***
Srela	0.138	0.030	4.565	0.000 ***
Intercepts				
0 2	5.174	0.917	5.644	0.000 ***
2 3	5.628	0.870	6.468	0.000 ***
3 4	6.216	0.834	7.451	0.000 ***
4 5	7.068	0.820	8.623	0.000 ***
5 6	8.442	0.834	10.120	0.000 ***
6 7	9.190	0.846	10.860	0.000 ***
7 8	10.697	0.885	12.085	0.000 ***
8 9	12.645	0.936	13.508	0.000 ***
9 10	14.152	0.970	14.594	0.000 ***

¹ Residual deviance: 1348.207; AIC: 1384.207; Multiple Kendall's tau = 0.501 (p -value < 0.001); *** p -value < 0.001; ** p -value < 0.01.

Similarly, the present study was carried out with consideration of the degree of satisfaction with life in general, Sgen, as an objective variable. According to the results of the model (see Table 6), the most relevant aspects are the degrees of satisfaction with family relationships, work, leisure and free time, physical appearance, the way of being and acting in society, and sentimental relationships. However, personal health and the level of professional training did not influence the degree of satisfaction with life.

Table 6. Model: Sgen on the life satisfaction profile ¹.

Variables	Coef.	Std. Error	t Value	p-Value
Sfam	0.621	0.080	7.773	0.000 ***
Sjob	0.131	0.047	2.799	0.005 **
Sleis	0.222	0.049	4.500	0.000 ***
Sfrie	0.114	0.067	1.693	0.090
Shea	0.020	0.049	0.400	0.689
Sphys	0.146	0.070	2.100	0.036 *
Seduc	0.039	0.042	0.921	0.357
Sbeha	0.524	0.091	5.766	0.000 ***
Srela	0.159	0.032	5.027	0.000 ***
Intercepts				
0 2	7.751	1.119	6.929	0.000 ***
2 3	8.864	0.984	9.009	0.000 ***
3 4	9.091	0.970	9.375	0.000 ***
4 5	9.710	0.942	10.312	0.000 ***
5 6	10.915	0.932	11.717	0.000 ***
6 7	12.011	0.948	12.666	0.000 ***
7 8	14.063	1.008	13.947	0.000 ***
8 9	16.779	1.100	15.256	0.000 ***
9 10	19.539	1.180	16.553	0.000 ***

¹ Residual deviance: 1057.027; AIC: 1093.027; Multiple Kendall's tau = 0.599 (p -value < 0.001); *** p -value < 0.001; ** p -value < 0.01; * p -value < 0.05; p -value < 0.10.

The multiple ordinal correlation measure reaches the value 0.599, exceeding the bivariate ordinal correlations (see Tables 4 and 7). Conclusions that are similar to the previous case can be obtained with the bias and the standard error of estimation [1].

Further, as with the previous model, the statistical interpretation of all these results allows obtaining a relevant sociological interpretation.

Table 7. Multiple ordinal correlation for life satisfaction profile.

Objective		Multiple Kendall's τ			
Variable	Estimate	Bias ¹	Std. Error ¹	95% BCa Bootstrap CI ¹	p-Value ²
IHap	0.5008	0.0080	0.0382	(0.4030, 0.5614)	0.000
SGen	0.5985	0.0246	0.0299	(0.5263, 0.6342)	0.000

¹ Using the function "boot.ci ()" of "boot" package. ² Obtained through the permutations test.

6. Discussion and Conclusions

In many economic, social, and biomedical studies, data sets containing ordinal variables are considered as a relevant part of the studies. However, statistical techniques not suitable for the ordinal scale are applied. This fact may seem strange given the set of ordinal regression models proposed in the body of literature. One of the reasons that their limited use may be justified is the difficulty of interpreting the models and their parameters. For this reason, it may be necessary to propose new parameters and functions that facilitate their interpretation and the transference of statistical conclusions to conclusions within the scientific scope of the study in a particular field.

The main contribution of this article is a multiple ordinal correlation measure and the inferential techniques necessary for its estimation. Important features of this contribution are considered to be the following. First, the measurement is based on a very intuitive idea, to measure the bivariate ordinal correlation between the target variable and the values fitted by an ordinal regression model. Kendall's tau coefficient is considered as a bivariate ordinal correlation measure and the cumulative model of ordinal regression is considered as a model. This intuitive idea, which is used in the definition of the multiple linear correlation coefficient, leads to easy interpretability of the correlation measure. Second, its easy interpretation can facilitate its use and the application of ordinal regression models in many scientific fields. It can also be considered as a measure of the adequacy of the model. Third, the idea of constructing a measure based on the model broadens the path, already opened by the multiple linear correlation coefficient, to construct other measures of multiple correlation and measures of multiple association at different scales (binary, nominal, ordinal, quantitative discrete, quantitative continuous). Fourth, the proposed inference methods are also easy to understand and apply. In particular, their point estimate is a sample version of the population measure. This estimate is almost unbiased according to the bias estimated through the bootstrap methodology, although its unbiasedness has not been mathematically demonstrated, to the best of our knowledge. Furthermore, this bootstrap methodology aids in obtaining confidence intervals with a computational procedure that is not complex and tedious. Finally, the significance test of the measure, which can be considered as a multiple ordinal independence test, is obtained through a permutation test that does not require excessive computation time.

Some aspects that are not in the present study need to be deepened and they should be the object of study in a future line of research. In the field of measurement inference, the aim is to obtain confidence intervals and significance tests, exact or asymptotic, which are not based on resampling techniques. However, the measurement construction methodology can be applied to other ordinal regression models (cumulative models, sequential models, and adjacent-category models), as well as with other ordinal correlation measures such as the Spearman rank correlation coefficient [4]. The methodology can also be extended to models based on binary and nominal variables through measures of bivariate association, as proposed in the body of literature. This methodology to construct the multiple association measure is easily applicable to the other models (for binary, categorical, ordinal, and continuous data). It is enough to combine two components: statistical model and bivariate

association measure. Such components should be appropriately in accordance with the scale of the data and the objective of the study.

Author Contributions: Investigation, J.M.M.-P., E.D.L.-A., A.P.-A. and A.M.M.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Simple example to illustrate the definition of the multiple Kendall’s tau measure.

Let X_1 and X_2 be two independent and identically distributed variables with support $\{1,2,3\}$ and probabilities $\{1/3, 1/3, 1/3\}$, respectively. Let Y be an ordinal random variable with support $\{1,2,3,4,5\}$. The following parameters of the model $OrdMod(Y|X_1, X_2)$ are considered:

$$\beta_{01} = -1.7, \beta_{02} = -0.5, \beta_{03} = 0.5, \beta_{04} = 1.7, \eta_1 = -0.5, \eta_2 = 0.5. \tag{A1}$$

Table A1 shows the conditioned probabilities $\{\Pr[Y = j | x], j = 1, \dots, 5\}$ according to the model. In addition, the modal value of each conditional distribution is included, that is, the value of the random variable $Y^* = mode[Y|X = x]$.

Table A1. A simple example: probability distributions associated with the model ⁽¹⁾.

X_1	X_2	$\Pr[Y = j x]$						
		$\Pr[X = x]$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	Y^*
1	1	0.11	0.1545	0.2231	0.2449	0.2231	0.1545	3
2	1	0.11	0.2315	0.2685	0.2311	0.1692	0.0998	2
3	1	0.11	0.3318	0.2906	0.1951	0.1195	0.0630	1
1	2	0.11	0.0998	0.1692	0.2311	0.2685	0.2315	4
2	2	0.11	0.1545	0.2231	0.2449	0.2231	0.1545	3
3	2	0.11	0.2315	0.2685	0.2311	0.1692	0.0998	2
1	3	0.11	0.0630	0.1195	0.1951	0.2906	0.3318	5
2	3	0.11	0.0998	0.1692	0.2311	0.2685	0.2315	4
3	3	0.11	0.1545	0.2231	0.2449	0.2231	0.1545	3

⁽¹⁾ $OrdMod(Y|X_1, X_2)$ with $\beta_{01} = -1.7, \beta_{02} = -0.5, \beta_{03} = 0.5, \beta_{04} = 1.7, \eta_1 = -0.5, \eta_2 = 0.5$.

Table A2 shows the joint probability distribution of (Y, Y^*) and the marginal distribution of Y^* , that is,

$$\Pr[Y^* = j] = \sum_{x: mode[Y|X=x]=j} \Pr[X = x] \tag{A2}$$

and,

$$\Pr[Y = k, Y^* = j] = \sum_{x: mode[Y|X=x]=j} \Pr[Y = k | x] \Pr[X = x]. \tag{A3}$$

Table A2. Joint probability distribution of (Y, Y^*) .

$\Pr[Y = k, Y^* = j]$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	Marginal of Y^*
$Y^* = 1$	0.0369	0.0323	0.0217	0.0133	0.0070	0.1111
$Y^* = 2$	0.0514	0.0597	0.0513	0.0376	0.0222	0.2222
$Y^* = 3$	0.0515	0.0744	0.0816	0.0744	0.0515	0.3333
$Y^* = 4$	0.0222	0.0376	0.0513	0.0597	0.0514	0.2222
$Y^* = 5$	0.0070	0.0133	0.0227	0.0323	0.0369	0.1111
Marginal of Y^*	0.1690	0.2172	0.2287	0.2172	0.1690	

The population multiple Kendall’s $\tau_M(Y, X)$ can be obtained as the bivariate Kendall’s τ of the joint distribution of (Y, Y^*) , that is, the population version of tau-b measure (see Agresti [12]):

$$\tau_b = \left(\Pi_c - \Pi_d \right) / \sqrt{(1 - \sum_k \Pr[Y=k]^2)(1 - \sum_j \Pr[Y^*=j]^2)} \tag{A4}$$

with Π_c and Π_d concordance and discordance probabilities, respectively. For the joint probability distribution included in Table A2, we obtain $\tau_M(Y, X) = 0.3205$.

Finally, in order to illustrate the estimation of this measure through the sample version, we randomly generated a sample of size 60 according to the probability distribution associated with the population model $OrdMod(Y|X_1, X_2)$ with the parameters listed above. The data set is included in Table A3. Fitting the model with the *polr* function of the MASS library package [25] provides the estimates:

$$\hat{\eta}_1 = -0.8414, \hat{\eta}_2 = 0.6558, \hat{\beta}_{01} = -2.2900, \hat{\beta}_{02} = -0.8678, \hat{\beta}_{03} = 0.3022, \hat{\beta}_{04} = 1.4829. \tag{A5}$$

These estimates allow us to estimate the probabilities of the conditional distributions and their modal values (fitted values of the model)

$$\hat{Y} = \arg \max_{j=1 \dots K} \hat{\Pr}[Y = j | x]. \tag{A6}$$

Table A3. Data set ($n = 60$).

		Y = 1	Y = 2	Y = 3	Y = 4	Y = 5	\hat{Y}
$X_1 = 1$	$X_2 = 1$	1	2	2	2	1	3
$X_1 = 1$	$X_2 = 2$	0	1	2	2	2	2
$X_1 = 1$	$X_2 = 3$	0	1	1	2	2	1
$X_1 = 2$	$X_2 = 1$	2	2	1	1	1	4
$X_1 = 2$	$X_2 = 2$	1	2	2	2	1	3
$X_1 = 2$	$X_2 = 3$	0	1	2	1	2	2
$X_1 = 3$	$X_2 = 1$	2	2	1	0	0	5
$X_1 = 3$	$X_2 = 2$	2	1	2	1	1	3
$X_1 = 4$	$X_2 = 3$	1	2	2	1	0	3

Finally, the crosstab between observed values and adjusted values (Table A4) allows us to obtain the sample version of tau-b measure (see Agresti [12]) as an estimate of the multiple ordinal correlation measure: $\hat{\tau}_M(Y, X) = 0.3243$.

Table A4. Joint frequency distribution of (Y, \hat{Y}) .

	Y = 1	Y = 2	Y = 3	Y = 4	Y = 5	Total
$\hat{Y} = 1$	2	2	1	0	0	5
$\hat{Y} = 2$	4	3	3	2	2	14
$\hat{Y} = 3$	3	7	8	6	4	28
$\hat{Y} = 4$	0	1	2	2	2	7
$\hat{Y} = 5$	0	1	1	2	2	6
Total	9	14	15	12	10	60

Appendix B

Simulation generation procedure (see Section 4).

Considering a vector of explanatory variables $\underline{X} = (X_1, \dots, X_p)'$ and the model $OrdMod(Y|X_1, \dots, X_p)$.

For error = $e\%$ and $\underline{X} = \underline{x}_i, i = 1, \dots, n (n = 500)$:

- For $j = 1, \dots, K$, calculate $p_j = \Pr[Y \leq j | x]$ according to the model $OrdMod(Y|X_1, \dots, X_p)$.

- Randomly generate $u \in [0, 1]$
 - If $u \geq e/100$, calculate $y = \arg \max_{j=1, \dots, K} p_j$. If bimodal distribution, randomize between modal categories.
 - If $u < e/100$, randomly generate $y \in \{1, 2, \dots, K\}$.

The error values considered are {2.5%, 5%, 10%, 25%, 50%, 75%, 90%, 100%}. Case $e = 0$ is not considered since the data will exactly fit the model and the `polr()` command of the MASS package [25] returns an overfitting error.

References

1. Agresti, A. *Categorical Data Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013.
2. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression: Models, Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2013.
3. Kendall, M.G. A New Measure of Rank Correlation. *Biometrika* **1938**, *30*, 81–93. [[CrossRef](#)]
4. Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
5. Agresti, A. A coefficient of multiple association based on ranks. *Commun. Stat. Theory Methods* **1977**, *6*, 1341–1359. [[CrossRef](#)]
6. Nagelkerke, N.J.D. A note on a general definition of the coefficient of determination. *Biometrika* **1991**, *78*, 691–692. [[CrossRef](#)]
7. Muñoz-Pichardo, J.M.; Pino-Mejías, R.; García-Heras, J.; Ruiz-Muñoz, F.; González-Regalado, M.L. A multivariate Poisson regression model for count data. *J. Appl. Stat.* **2021**, 1–17. [[CrossRef](#)]
8. Chow, H.P.H. Life Satisfaction among University Students in a Canadian Prairie City: A Multivariate Analysis. *Soc. Indic. Res.* **2005**, *70*, 139–150. [[CrossRef](#)]
9. Bürkner, P.-C.; Vuorre, M. Ordinal Regression Models in Psychology: A Tutorial. *Adv. Methods Pr. Psychol. Sci.* **2019**, *2*, 77–101. [[CrossRef](#)]
10. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [[CrossRef](#)]
11. McCullagh, P. Regression Models for Ordinal Data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1980**, *42*, 109–127. [[CrossRef](#)]
12. Agresti, A. *Analysis of Ordinal Categorical Data*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010.
13. Marshall, R.J. Classification to ordinal categories using a search partition methodology with an application in diabetes screening. *Stat. Med.* **1999**, *18*, 2723–2735. [[CrossRef](#)]
14. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2003.
15. Gibbons, J.D.; Chakraborti, S. *Nonparametric Statistical Inference*, 6th ed.; CRC Press: Boca Raton, FL, USA, 2020.
16. Bergsma, W.; Dassios, A. A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **2014**, *20*, 1006–1028. [[CrossRef](#)]
17. Ruscio, J.J. The College of New Jersey Constructing Confidence Intervals for Spearman's Rank Correlation with Ordinal Data: A Simulation Study Comparing Analytic and Bootstrap Methods. *J. Mod. Appl. Stat. Methods* **2008**, *7*, 416–434. [[CrossRef](#)]
18. Canty, A.J. Resampling Methods in R: The boot Package. *NewsL. R Proj.* **2002**, *2*, 2–7.
19. Canty, A.; Ripley, B. Boot: Bootstrap R (S-Plus) Functions. R Package Version 1.3-22. 2019. Available online: <https://CRAN.R-project.org/package=boot> (accessed on 15 January 2021).
20. Jung, K.; Lee, J.; Gupta, V.; Cho, G. Comparison of Bootstrap Confidence Interval Methods for GSCA Using a Monte Carlo Simulation. *Front. Psychol.* **2019**, *10*, 2215. [[CrossRef](#)] [[PubMed](#)]
21. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1998.
22. Efron, B. Better Bootstrap Confidence Intervals. *J. Am. Stat. Assoc.* **1987**, *82*, 171–185. [[CrossRef](#)]
23. DiCiccio, T.J.; Efron, B. Bootstrap Confidence Intervals. *Stat. Sci.* **1996**, *11*, 189–212. [[CrossRef](#)]
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: <http://www.R-project.org/> (accessed on 15 January 2021).
25. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S Plus*; Springer: New York, NY, USA, 2002.
26. Centro Andaluz de Prospectiva. Encuesta Sobre Preocupaciones y Satisfacción de Vida de los Andaluces EPSVA, 2019. Available online: <https://personal.us.es/juanm/EPsVAmetodoresult.htm> (accessed on 12 January 2021).
27. Wei, T.; Simko, V. R Package Corrplot: Visualization of a Correlation Matrix. (Version 0.84). 2017. Available online: <https://github.com/taiyun/corrplot> (accessed on 23 June 2020).