Working Paper Series, N. 4, January 2014

Pairwise likelihood inference for multivariate categorical responses with application to customer satisfaction

Euloge Clovis Kenne Pagui
Department of Statistical Sciences
university of padua
Italy

Antonio Canale
Department of Economics and Statistics
University of Turin and Collegio Carlo Alberto
Italy

Abstract: A common practice in customer satisfaction analysis is to administer surveys where subjects are asked to express opinions on a number of statements, or satisfaction scales, by use of ordered categorical responses. Motivated by this application, we propose a pseudo-likelihood approach to estimate the dependence structure among multivariate categorical variables. As it is commonly done in this area, we assume that the responses are related to latent continuous variables which are truncated to induce categorical responses. A Gaussian likelihood is assumed for the latent variables leading to the so called ordered probit model. Since the calculation of the exact likelihood is computationally demanding, we adopt an approximate solution based on pairwise likelihood. To asses the performance of the approach, simulation studies are conducted comparing the proposed method with standard likelihood methods. A parametric bootstrap approach to evaluate the variance of the maximum pairwise likelihood estimator is proposed and discussed. An application to customer satisfaction survey is performed showing the effectiveness of the approach in the presence of covariates and under other generalizations of the model which can make a difference in real data situations.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Euloge Clovis Kenne Pagui
tel: +39 049 827 4124
kenne@stat.unipd.it
http://www.stat.unipd.it/~kenne

Pairwise likelihood inference for multivariate categorical responses with application to customer satisfaction

Euloge Clovis Kenne Pagui
Department of Statistical Sciences
university of padua
Italy

Antonio Canale
Department of Economics and Statistics
University of Turin and Collegio Carlo Alberto
Italy

Abstract: A common practice in customer satisfaction analysis is to administer surveys where subjects are asked to express opinions on a number of statements, or satisfaction scales, by use of ordered categorical responses. Motivated by this application, we propose a pseudo-likelihood approach to estimate the dependence structure among multivariate categorical variables. As it is commonly done in this area, we assume that the responses are related to latent continuous variables which are truncated to induce categorical responses. A Gaussian likelihood is assumed for the latent variables leading to the so called ordered probit model. Since the calculation of the exact likelihood is computationally demanding, we adopt an approximate solution based on pairwise likelihood. To asses the performance of the approach, simulation studies are conducted comparing the proposed method with standard likelihood methods. A parametric bootstrap approach to evaluate the variance of the maximum pairwise likelihood estimator is proposed and discussed. An application to customer satisfaction survey is performed showing the effectiveness of the approach in the presence of covariates and under other generalizations of the model which can make a difference in real data situations.

1 Introduction

Surveys and questionnaires are a fundamental research tool in many application areas such as psychometry or marketing research among others. Responses in customer satisfaction surveys, for example, are generally in the form of rating scale or Likert scale (Likert, 1932). When responding to a Likert questionnaire, customers specify their level of agreement or disagreement for each of the q questions or items. Usually each item consists in $K = 5$ ordered response levels, but sometimes also $K > 5$ can be adopted. Each item aims to measure one different aspect of the overall phenomenon under study (e.g. global customer satisfaction) and hence the responses are clearly correlated.

A single item, in Likert questionnaires, can be considered as an ordered-categorical outcome. An approach to model ordinal categorical responses which, albeit wrong, is commonly adopted, consists in treating them as continuous observation and to use standard methods such as linear regression to estimate covariate dependence, t-test or ANOVA models to test for differences among groups, and so forth. A more rigorous and elegant approach consists in considering that the observed categorical data are related to continuous latent variables. The relation between the latent continuous variables and the observed categorical variables is usually induced by use of thresholds partitioning the latent sample space into a series of regions corresponding to each ordinal category. If a Gaussian distribution is assumed for the latent variables, an ordered probit model is induced (Agresti, 2002).

Multivariate ordered probit models are the straightforward multivariate extension of ordered probit models where both the latent and the observed variables have dimension $q \geq 2$, as it is the case for customer satisfaction surveys. With regard to this class of models, we discuss the problem of the estimation of the dependence structure, following a frequentist likelihood-based approach. The multivariate ordered probit model has clear computational problems related to the calculation of a q dimensional integral for each single likelihood contribution. Such a procedure is computationally demanding or even unfeasible also for moderate q. For that reason we adopt an approximate solution based on a simple pseudolikelihood belonging to the class of composite likelihoods (Lindsay, 1988): the pairwise likelihood (PL) (Cox and Reid, 2004). The use of the PL when the full likelihood is computationally unmanageable is becoming a commonly adopted procedure (Varin et al., 2011). Applications of pairwise likelihood to ordered probit models are not new, and significant contributions can be found in De Leon (2005), where the dependence among the ordered variables is evaluated in terms of polychoric correlation, i.e. the linear correlation of the latent variables, and in Varin and Vidoni (2006) and Varin and Czado (2010) in the context of longitudinal data and focusing on the mean evolution in time. Our contribution to the PL approach in the case of ordered probit model is building up over De Leon (2005), but it differs for three main aspects. First, we estimate different measures of dependence and not only the polychoric correlation, i.e. the linear correlation coefficient and the L measure coefficient (Lu, 2011). The polychoric correlation is indeed based on the latent Gaussianity assumption which we consider more as a computational solution rather than an exact characteristic of the model. Second, the dimensions that we consider mimic real contexts and are greater than $q = 3$. This has clearly no theoretical difference, but it increases the computational burden. In addition, this causes the number of model parameters to leaven, leading to the practical impossibility of evaluate the Fisher's and Godambe's information matrices and forcing us to study other approaches to evaluate the variability of the maximum likelihood estimator. Third, we apply the methods to a concrete customer satisfaction application, introducing subject-specific covariates and other complications. We discuss a bootstrap-based approach which allows us to perform classical hypothesis testing on each pairwise variable dependence and on the significance of the regression coefficients. This bootstrap-based approach is, to our knowledge, the first attempt to asses the variance of the maximum likelihood estimator for this class of problems.

The rest of the paper is organized as follows. In the next section we review the multivariate ordered probit model and introduce our solution based on pairwise likelihood. In Section 3 an extensive simulation experiment is conducted comparing the point estimates obtained maximizing the full and pairwise likelihoods. A discussion on the computational feasibility of the two procedures is also reported. In Section 4 the model is generalized including the effect of subject specific covariates, motivated by applications to customer satisfaction surveys. Data from a survey conducted by a company operating in the sector of information technology services and consulting are then analyzed and discussed, introducing a bootstrap-based solution to asses the estimator standard errors. Section 5 summarizes the main results.

2 Multivariate ordered probit model

Let $Y_i = (Y_{i1}, \ldots, Y_{iq})^{\mathrm{T}}$, $Y_{ij} \in \{1, 2, \ldots, K\}$ for $j = 1, \ldots, q$ be a q-dimensional ordinal categorical random vector with joint density $g(Y_i)$ depending on some unknown parameter θ, with $i = 1, \ldots, n$ defining a collection of iid random vectors. The ordered probit model assumes that, for each $i = 1, \ldots, n$, there exists a latent random vector $Z_i = (Z_{i1}, \ldots, Z_{iq})^{\mathrm{T}}$, with $Z_i \overset{iid}{\sim} N(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \cdots & \rho_{1,q} \\ & 1 & \cdots & \cdots & \rho_{2,q} \\ & & 1 & \rho_{rs} & \vdots \\ & & & 1 & \rho_{q-1,q} \\ & & & & 1 \end{pmatrix}$$

is an unknown q-dimensional positive definite correlation matrix. In observing y_{ij}, the following relation is assumed

$$y_{ij} = k \text{ if and only if } \; z_{ij} \in (a_{k-1}, a_k], \tag{1}$$

where $-\infty = a_0 < a_1 < \cdots < a_K = \infty$ is a sequence of real numbers defining a disjoint partition of $\mathbb{R}$. The likelihood function for a single observation is then proportional to

$$\mathrm{pr}(Y_{i1} = y_{i1}, \ldots, Y_{iq} = y_{iq}) = \int_{a_{y_{i1}-1}}^{a_{y_{i1}}} \cdots \int_{a_{y_{iq}-1}}^{a_{y_{iq}}} \phi_\Sigma(z_{i1}, \ldots, z_{iq}) dz_{i1} \ldots dz_{iq},$$

where $\phi_\Sigma(\cdot)$ denotes the multivariate Gaussian distribution with mean zero and variance Σ. The full log-likelihood for the n observations is therefore given by

$$\ell(\theta) = \sum_{i=1}^{n} \log \left\{ \int_{a_{y_{i1}-1}}^{a_{y_{i1}}} \cdots \int_{a_{y_{iq}-1}}^{a_{y_{iq}}} \phi_\Sigma(z_{i1}, \ldots, z_{iq}) dz_{i1} \ldots dz_{iq} \right\}, \tag{2}$$

where the parameters of the model are the thresholds $(a_1, \ldots, a_{K-1})$ and the latent correlations $\{\rho_{r,s}\}$ for $r, s = 1, \ldots q$, and $r < s$, and $\theta = (a_1, \ldots, a_{K-1}, \rho_{1,2}, \ldots, \rho_{q-1,q})$ is the joint vector of all parameters with $\hat{\theta}$ denoting its maximum likelihood estimator. Equation (2) presents a q-dimensional integral for each observation. In fact this function is not easily tractable, since its direct evaluation requires the computation of n integrals of a q-dimensional Gaussian distribution.

2.1 Pairwise likelihood for multivariate probit model

The pairwise likelihood that we discuss in this paper, is a pseudolikelihood constructed from bivariate margins. Specifically, the pairwise log-likelihood for our ordered categorical data is

$$\begin{aligned}\ell^P(\theta) &= \sum_{i=1}^{n}\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} w_{rs}\log\left\{\mathrm{pr}(Y_{ir}=y_{ir},Y_{is}=y_{is})\right\}\\ &= \sum_{i=1}^{n}\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} w_{rs}\log\left\{\int_{a_{y_{ir}-1}}^{a_{y_{ir}}}\int_{a_{y_{is}-1}}^{a_{y_{is}}}\phi_{\Sigma(\rho_{r,s})}\left(z_{ir},z_{is}\right)dz_{ir}dz_{is}\right\},\end{aligned}$$

where w_{rs} are nonnegative weights to be chosen and $\Sigma(\rho)$ denotes the 2×2 correlation matrix with off-diagonal entries equal to ρ. The PL approach is particularly attractive because it substitutes a log-likelihood involving high dimensional integrations with a sum of bivariate integrals, which can be easily evaluated with standard available software. Similar PL has been considered by De Leon (2005) for likelihood inference within group continuous models, that is models for multivariate ordinal data. While the latent model is the same, we are not interested in estimating just the polychoric correlations (i.e. the linear correlation if the latent variables) but also directly a measure of relation among the ordered categorical responses. To this end we use both the standard correlation coefficient and the L measure defined as

$$\mathrm{Cor}(U,V) = \frac{\mathrm{Cor}(U,V)}{\sqrt{\mathrm{Var}(U)\mathrm{Var}(V)}} \tag{3}$$

$$L(U,V) = \left[1-\exp\left\{\frac{-2I(U,V)}{1-I(U,V)/\min\{H(U)H(V)\}}\right\}\right]^{1/2}, \tag{4}$$

where $I(U,V)=\sum_{x,y}\mathrm{pr}(u,v)\log\{\mathrm{pr}(u,v)/\mathrm{pr}(u)\mathrm{pr}(v)\}$ is the mutual information between U and V and $H(U)=-\sum_u \mathrm{pr}(u)\log(\mathrm{pr}(u))$ is the entropy of U. The L measure is a positive and symmetric index of dependence between any possible random variable. It equals 0 if and only if the random variables are independent and 1 if there is a strict dependence between the random variables. The L measure is defined using the mutual information index $I(\cdot,\cdot)$, it is invariant under marginal one-to-one transformations of the random variables, and it equals the correlation coefficient if the two random variables are normally distributed. Note also that definition (4) exploits the fact that, in our case, each random variable is discrete. For further details see Lu (2011).

Generally, the composite likelihood inferential procedures are similar to the ones based on the standard likelihood. For example, the pairwise score vector, $U^P(\theta)=\partial\ell^P(\theta)/\partial\theta$, is still unbiased being the sum of the score vector based on the likelihood contribution of each pair of observations. As for the standard likelihood, the maximum pairwise likelihood estimator $\hat\theta^P$ can be obtained either by maximizing $\ell^P(\theta)$ or by solving the pairwise score equation $U^P(\theta)=0$. Under the usual regularity conditions (Molenberghs and Verbeke, 2005), the maximum pairwise likelihood estimator is consistent and asymptotically normal as $n\to\infty$. Specifically

$$\hat\theta^P \dot\sim N(\theta, G(\theta)^{-1})$$

where $G(\theta) = K(\theta)J(\theta)^{-1}K(\theta)$ is known as the Godambe information or sandwich information, $K(\theta) = \mathrm{E}_\theta\{-\partial U^P(\theta)/\partial\theta\}$ is the sensitivity matrix and $J(\theta) = \mathrm{Var}_\theta\{U^P(\theta)\}$ is the variability matrix.

The use of a pseudo-likelihood usually leads to a loss of efficiency when the length of the random vector increases (see, for example, Cox and Reid, 2004, Section 1; Zhao and Joe, 2005, Section 6 and Xu and Reid, 2011, Section 3 for some illustrations). On the other hand, there are few classes of models in which the maximum pairwise likelihood estimator is fully efficient (Mardia et al., 2009). An appropriate choice of weights might affect on efficiency. For the pairwise likelihood defined above, it seems plausible to use equal weights.

With regard to the present model, both the standard and pairwise likelihood involve a numerical integration and thus the evaluation of the Fisher information, $K(\theta)$, and $J(\theta)$ may be a daunting task. Particularly in the applications we have in mind which involve high dimensions q. To this end, in Section 4 we adopt a bootstrap-based approach.

3 Computation and simulation

To asses the performance of the procedure based on the PL we conduct a simulation study. Before describing the simulation study and commenting its results we briefly discuss the computational issues related to the proposed model stressing when it is preferable to the full-likelihood approach.

3.1 Computational issues

Our code is implemented in the R software (R Core Team, 2013) with call to C functions for the most demanding operations. To evaluate the Gaussian integrals, the package uses the Fortran 77 subroutine `sadmvnt.f` by Genz (1992). As discussed in Varin and Vidoni (2006), to assure the ordering of the thresholds $a_0, < \cdots < a_K$, we reparametrize the model defining

$$\delta_k = \log(a_k - a_{k-1})$$

for $k = 2, \ldots, K-1$. Clearly, it is not possible to identify both the mean of the latent variables and the first threshold a_1. To deal with this, we fix the latent mean to be zero. For optimization, we consider as starting values for the thresholds $a_1 = 0$ and $\delta_k = 0$ for $k = 2, \ldots, K-1$. This is equivalent to assume that the thresholds are equally spaced with distance one. As for the covariance components, we consider as starting values the sample covariances of the observed categorical variables treated as continuous. Optimization of the PL and of the full-likelihood functions is performed via quasi-Newton box-constrained optimization algorithm, with a relative tolerance of 1×10^{-10}.

A comparison between the PL and the full-likelihood is reported in the next section. However such a comparison is feasible only for moderate q. Indeed our motivation to use the PL approach lies both in speeding up computations, and in avoiding the numerical instability intrinsically related to the calculation of q-dimensional integrals. Figure 1 gives an idea of the former aspect. We plot the

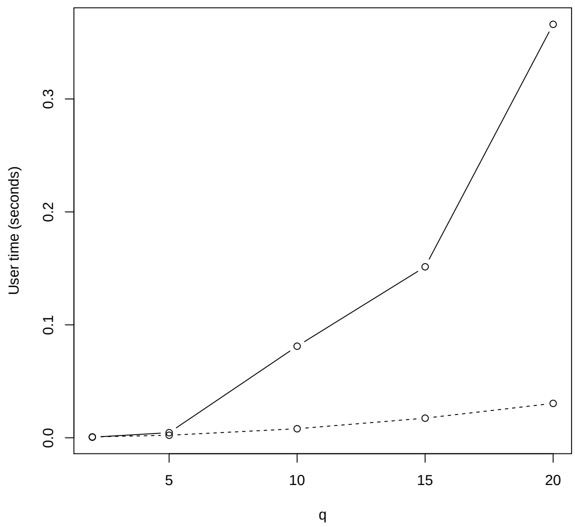

Figure 1: Execution time of a single likelihood evaluation for different dimensions q for the full likelihood (continuous line) and pairwise likelihood (dashed line) for a sample of size $n = 50$.

executions time of a single evaluations of the likelihood functions for samples of size $n = 50$ and different dimensions $q = 2, 5, 10, 15, 20$. The computational burden of the full likelihood with respect to the PL is clearly comparable for $q = 2$ while it is double for $q = 5$ and even ten times bigger for $q > 5$. Note that in maximizing the likelihood function numerically, several evaluation of the likelihood functions are needed leading to a leavening of the differences between the two approaches. In particular, by our experience, we have noted that for $q > 5$ the maximization of the full likelihood turns out to be extremely slow. The times are in seconds and calculated when running the algorithms in R version 3.0.1 on a 64bit Mac OS X 10.6 machine with a 2.4 Ghz Intel Core 2 processor with 4 Gb of RAM.

3.2 Simulation experiment

To assess the performance of the proposed approach we simulate data from three scenarios. All scenarios consider $q = 5$ and $K = 5$. As previously discussed the computations for the full likelihood when $q > 5$ are demanding and hence we decide to fix $q = 5$ for sake of comparison. However, the PL proposal is feasible also for higher q as we show in the application of Section 4. In the first scenario we first simulate latent Z_i from a multivariate normal with mean vector equal to $(2, \ldots, 2)$ and the unconstrained variance matrix Σ_1 reported in the Appendix, and then round Z_i with thresholds $-\infty, 1, 2, 3, 4, \infty$. In the second scenario we substitute Σ_1 with a constrained covariance matrix with all correlations equal to $\rho = 0.8$ while in the last one we use Σ_3, a sparse matrix which is reported in the Appendix. For each scenario we simulate 1000 samples of sizes $n = 50, 100, 200$ and then estimate

Table 1: Mean square error of the polychoric correlation coefficients under the three scenarios

	n	full likelihood				pairwise likelihood			
Scenario 1	50	0.0195	0.0192	0.0232	0.0184	0.0250	0.0245	0.0218	0.0184
		-	0.0009	0.0225	0.0142	-	0.0008	0.0255	0.0152
		-	-	0.0208	0.0089	-	-	0.0243	0.0097
		-	-	-	0.0101	-	-	-	0.0122
	100	0.0088	0.0025	0.0112	0.0062	0.0122	0.0028	0.0116	0.0045
		-	0.0047	0.0129	0.0091	-	0.0058	0.0127	0.0131
		-	-	0.0117	0.0026	-	-	0.0110	0.0023
		-	-	-	0.0115	-	-	-	0.0109
	200	0.0027	0.0025	0.0054	0.0032	0.0060	0.0052	0.0060	0.0061
		-	0.0001	0.0044	0.0046	-	0.0008	0.0023	0.0059
		-	-	0.0049	0.0049	-	-	0.0059	0.0054
		-	-	-	0.0015	-	-	-	0.0063
Scenario 2	50	0.0053	0.0044	0.0048	0.0044	0.0047	0.0050	0.0048	0.0048
		-	0.0050	0.0050	0.0049	-	0.0045	0.0050	0.0049
		-	-	0.0044	0.0038	-	-	0.0051	0.0054
		-	-	-	0.0047	-	-	-	0.0050
	100	0.0020	0.0020	0.0019	0.0020	0.0021	0.0022	0.0020	0.0021
		-	0.0021	0.0022	0.0022	-	0.0023	0.0023	0.0023
		-	-	0.0022	0.0021	-	-	0.0024	0.0021
		-	-	-	0.0020	-	-	-	0.0021
	200	0.0010	0.0011	0.0011	0.0010	0.0011	0.0011	0.0011	0.0011
		-	0.0011	0.0011	0.0010	-	0.0011	0.0011	0.0010
		-	-	0.0011	0.0010	-	-	0.0012	0.0010
		-	-	-	0.0011	-	-	-	0.0012
Scenario 3	50	0.0255	0.0095	0.0200	0.0253	0.0247	0.0090	0.0212	0.0255
		-	0.0261	0.0233	0.0089	-	0.0255	0.0249	0.0079
		-	-	0.0208	0.0270	-	-	0.0222	0.0268
		-	-	-	0.0172	-	-	-	0.0185
	100	0.0123	0.0084	0.0114	0.0107	0.0123	0.0083	0.0117	0.0109
		-	0.0122	0.0111	0.0079	-	0.0123	0.0126	0.0084
		-	-	0.0117	0.0114	-	-	0.0118	0.0125
		-	-	-	0.0022	-	-	-	0.0020
	200	0.0061	0.0048	0.0062	0.0058	0.0061	0.0048	0.0062	0.0058
		-	0.0062	0.0069	0.0058	-	0.0062	0.0069	0.0058
		-	-	0.0049	0.0064	-	-	0.0049	0.0063
		-	-	-	0.0060	-	-	-	0.0060

the vector of parameters θ. From the estimated values of parameters, we derive the polychoric correlation coefficients and calculate with a plug-in approach, the pairwise correlation and L measure coefficients as (3)–(4). The Monte Carlo mean squared errors for the polychoric correlations, linear correlations and L measures are reported in Table 1, 2 and 3.

As expected, in all scenarios and considering the three measures of dependence, the mean squared errors decrease towards zero as the sample size increases, both in the full and in the pairwise likelihood approach. Consider Table 1. In the second scenario, namely when we have all the polychoric correlations fixed to ρ the mean squared errors between the full likelihood and the PL approach are almost identical also for $n = 50$. Also in scenario 3 the resulting errors converge but only for greater sample size. Consider now Table 2 and 3. Here the PL always shows greater mean

Table 2: Mean square error of the correlation coefficients under the three scenarios

	n	full likelihood				pairwise likelihood			
Scenario 1	50	0.0153	0.0153	0.0183	0.0147	0.0195	0.0193	0.0173	0.0148
		-	0.0011	0.0179	0.0118	-	0.0010	0.0201	0.0128
		-	-	0.0168	0.0078	-	-	0.0194	0.0085
		-	-	-	0.0088	-	-	-	0.0107
	100	0.0069	0.0023	0.0088	0.0054	0.0095	0.0026	0.0091	0.0040
		-	0.0042	0.0102	0.0072	-	0.0051	0.0101	0.0103
		-	-	0.0093	0.0025	-	-	0.0088	0.0023
		-	-	-	0.0092	-	-	-	0.0088
	200	0.0023	0.0022	0.0042	0.0026	0.0048	0.0042	0.0047	0.0048
		-	0.0002	0.0036	0.0036	-	0.0008	0.0020	0.0047
		-	-	0.0040	0.0039	-	-	0.0047	0.0043
		-	-	-	0.0014	-	-	-	0.0049
Scenario 2	50	0.0053	0.0045	0.0049	0.0045	0.0048	0.0051	0.0048	0.0049
		-	0.0051	0.0050	0.0049	-	0.0045	0.0051	0.0050
		-	-	0.0045	0.0039	-	-	0.0052	0.0055
		-	-	-	0.0048	-	-	-	0.0051
	100	0.002	0.0021	0.0020	0.0021	0.0021	0.0023	0.0020	0.0022
		-	0.0021	0.0022	0.0022	-	0.0024	0.0023	0.0023
		-	-	0.0022	0.0021	-	-	0.0025	0.0022
		-	-	-	0.0021	-	-	-	0.0022
	200	0.001	0.0011	0.0011	0.0010	0.0011	0.0012	0.0012	0.0011
		-	0.0011	0.0011	0.0010	-	0.0011	0.0012	0.0010
		-	-	0.0012	0.0010	-	-	0.0012	0.0010
		-	-	-	0.0011	-	-	-	0.0012
Scenario 3	50	0.0200	0.0083	0.0159	0.0198	0.0193	0.0079	0.0168	0.0200
		-	0.0205	0.0183	0.0079	-	0.0200	0.0195	0.0070
		-	-	0.0165	0.0212	-	-	0.0176	0.0210
		-	-	-	0.0142	-	-	-	0.0152
	100	0.0096	0.0069	0.0089	0.0084	0.0096	0.0068	0.0092	0.0086
		-	0.0095	0.0087	0.0065	-	0.0097	0.0098	0.0068
		-	-	0.0093	0.0090	-	-	0.0093	0.0098
		-	-	-	0.0021	-	-	-	0.0020
	200	0.0048	0.0038	0.0048	0.0045	0.0048	0.0038	0.0049	0.0045
		-	0.0048	0.0054	0.0045	-	0.0049	0.0054	0.0045
		-	-	0.0039	0.0050	-	-	0.0039	0.0050
		-	-	-	0.0047	-	-	-	0.0047

squared errors, if compared to the full likelihood approach. For increasing n, the mean squared errors are however more comparable.

4 Application to customer satisfaction analysis

Customer satisfaction surveys are useful tools to measure customers' opinions on products and services. For example, a typical surveys structure to measure the quality of services, follows the so called SERVQUAL structure (Parasuraman et al., 1988). The main idea behind this approach is that, in order to guarantee a good quality of service, it is necessary to go beyond a customer's expectations. Thus, it is important to measure the gap between expected services and experienced services. To this end, questionnaires are typically divided into two separate blocks of

Table 3: Mean square error of the L measures under the three scenarios

	n	full likelihood				pairwise likelihood			
Scenario 1	50	0.0086	0.0117	0.0172	0.0158	0.0133	0.0121	0.0172	0.0165
		-	0.0010	0.0130	0.0138	-	0.0009	0.0125	0.0150
		-	-	0.0148	0.0100	-	-	0.0158	0.0108
		-	-	-	0.0112	-	-	-	0.0133
	100	0.0038	0.0032	0.0082	0.0071	0.0053	0.0037	0.0091	0.0054
		-	0.0052	0.0093	0.0055	-	0.0063	0.0097	0.0074
		-	-	0.0096	0.0036	-	-	0.0094	0.0033
		-	-	-	0.0100	-	-	-	0.0097
	200	0.0029	0.0028	0.0032	0.0031	0.0045	0.0044	0.0045	0.0041
		-	0.0001	0.0040	0.0026	-	0.0012	0.0027	0.0045
		-	-	0.0043	0.0031	-	-	0.0041	0.0044
		-	-	-	0.0019	-	-	-	0.0046
Scenario 2	50	0.0075	0.0063	0.0068	0.0063	0.0067	0.0072	0.0068	0.0069
		-	0.0071	0.0070	0.0068	-	0.0063	0.0071	0.0070
		-	-	0.0063	0.0055	-	-	0.0073	0.0077
		-	-	-	0.0067	-	-	-	0.0072
	100	0.0029	0.0030	0.0029	0.0029	0.003	0.0032	0.0029	0.0032
		-	0.0031	0.0032	0.0032	-	0.0034	0.0033	0.0033
		-	-	0.0032	0.0030	-	-	0.0035	0.0031
		-	-	-	0.0030	-	-	-	0.0031
	200	0.0014	0.0016	0.0016	0.0015	0.0015	0.0016	0.0017	0.0016
		-	0.0016	0.0015	0.0014	-	0.0016	0.0017	0.0015
		-	-	0.0017	0.0014	-	-	0.0017	0.0015
		-	-	-	0.0016	-	-	-	0.0017
Scenario 3	50	0.0206	0.0109	0.0138	0.0204	0.0199	0.0105	0.0146	0.0206
		-	0.0212	0.0189	0.0107	-	0.0207	0.0202	0.0097
		-	-	0.0155	0.0218	-	-	0.0158	0.0217
		-	-	-	0.0160	-	-	-	0.0167
	100	0.0099	0.0077	0.0053	0.0086	0.0098	0.0076	0.0054	0.0087
		-	0.0097	0.0089	0.0073	-	0.0099	0.0101	0.0077
		-	-	0.0093	0.0091	-	-	0.0094	0.0100
		-	-	-	0.0030	-	-	-	0.0028
	200	0.0049	0.0041	0.0040	0.0046	0.0048	0.0040	0.0040	0.0046
		-	0.0049	0.0055	0.0029	-	0.0050	0.0055	0.0029
		-	-	0.0041	0.0051	-	-	0.0041	0.0051
		-	-	-	0.0032	-	-	-	0.0032

questions. In the first, customers are asked to say how important are some characteristics of a given service in general. In the second, customers are asked to report their actual satisfaction on the same characteristics, for the experienced service.

We focus on data from $n = 324$ questionnaires collected in 2009 by an Italian company operating in the sector of information technology services. The survey consists of $q = 14$ questions, divided into two blocks of seven items each. In the first block, the question was “how important are the following aspects”, while in the second block the question was “how satisfied are you regarding the following aspects”. The seven aspects are: efficiency of the service, reliability of the service, flexibility of the service, velocity in giving solutions, ability to satisfy the customers' needs, ability to implement technological innovation, velocity in adapting to changing legislation. Each item is measured on a scale of $K = 10$ levels. In addition to the responses above, subject specific covariates are available, namely gender, age, and educational level (elementary school, high school, and university) of the customer. Let x_i the vector of the covariates for customer i.

From an application point of view, we are mainly interested in the mean responses, accounting for meaningful dependence structure between and within the two blocks. Specifically, we want to allow for different means for each different items in order to measure the discrepancy between experienced and expected quality of service, while controlling for any effect due to the covariates. The company hopes to record similar means in the two block and possibly higher scores in the experienced quality block and positive correlation between an item of the first block and the relative item in the second block.

To this end we modify the model of Section 2, assuming

$$Z_i = (Z_{i1}, \dots, Z_{iq})^{\mathrm{T}}, \;\; E[Z_{ij}] = \xi_j + x_i^T \beta, \;\; \mathrm{Var}[\mathbf{Z}_i] = \Sigma \tag{5}$$

for all $i = 1, \dots, n$ and $j = 1, \dots, q$, with Σ being a q-dimensional positive definite correlation matrix to be estimated. Let furthermore $\rho = (\rho_{1,2}, \dots, \rho_{q-1,q})^T$, the $q(q-1)/2$ vector containing all the upper diagonal elements of Σ and $\xi = (\xi_1, \dots, \xi_q)^T$, the vector of the item-specific means. We assume to observe $y_{ij} = k$ if $z_{ij} \in (a_{k-1}, a_k]$ and let $a = (a_2, \dots, a_{K-1})^T$, be the vector of $K-2$ thresholds. Note that it is not possible to identify both the mean of the latent variables and all the thresholds. Thus a_1 is fixed to zero. We fix the first thresholds rather than the means because the generalization (5) has a different mean for each item and subject. For the same reason, we cannot compute the actual correlations and L-measures, and hence we focus on the polychoric correlation estimation only.

The parameter vector of model (5) is thus made of $K-2$ thresholds, $q(q-1)/2$ correlations, q item-specific intercept levels, and p regression coefficients. The PL log-likelihood for the parameters is

$$\ell^P(\rho, \xi, \beta, a) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \log \left\{ \mathrm{pr}(Y_{ir} = y_{ir}, Y_{is} = y_{is}) \right\}$$

where

$$\mathrm{pr}(Y_{ir} = y_{ir}, Y_{is} = y_{is}) = \int_{a_{y_{ir}-1}}^{a_{y_{ir}}} \int_{a_{y_{is}-1}}^{a_{y_{is}}} \phi_{\Sigma(\rho_{rs})} \left(z_{ir} - \mu_{ir}, z_{is} - \mu_{is} \right) dz_{ir} dz_{is},$$

and $\mu_{ik} = \xi_k + x_i^T \beta$. Clearly, the high number of parameters (117, for our dataset) makes the likelihood optimization a demanding task. By our experience, the quasi-Newton box-constrained optimization algorithm, greatly benefits from the following initialization. As before, we consider as starting values for the polychoric correlations, the sample covariances of the observed categorical variables. Then, for the β coefficients we first compute the average scores over the items for each subject, then we standardize them and after that, we perform a linear regression between the standardized-average-scores and the covariates. The least squares estimates for the regression coefficients act as starting values for the β parameters. For ξ, we simply consider the sample mean for each item over the subjects. For what concerns the initialization of the thresholds we first compute the sample mean ($\bar{y}$), standard deviation (s_y), and relative frequencies of the categories ($\hat{f}_1, \ldots, \hat{f}_{10}$) of the whole sample. Then we calculate the quantiles of levels $\hat{f}_j$ for $j = 2, \ldots, 10$ of a Gaussian distribution with mean $\bar{y}$ and standard deviation s_y and consider those values as stating values for the thresholds.

In this applied problem, we are facing a very complex model both from the analytical and computational viewpoints. Despite the fact that we unburden the problem with the use of the pairwise likelihood, it still remains the question of evaluating the unmanageable pairwise likelihood quantities such as $K(\theta)$ and $J(\theta)$, necessary for inference on the parameters. The nonparametric bootstrap methods reveal to be of great help to provide an alternative solution. The steps of the methodology are briefly summarized in what follows. Several independent resample from the original dataset are carried out with replacement obtaining bootstrap parameter estimates, $\hat{\theta}^*_b$ for each $b = 1 \ldots, B$, where B refers to the number of bootstrap replicates. The empirical distribution of $\hat{\theta}^*$ is used as a good estimate for the distribution of the estimator $\hat{\theta}$ from which we can derive for example bias, standard error, and so forth. Thanks to the independence among the bootstrap replicates, the procedure strongly benefits from a parallel implementation on multicores machines or clusters of CPUs. For this problem, we took $B = 1400$.

To calculate confidence intervals only for correlation parameters, we use the bias corrected and accelerated confidence intervals (BCa) which are recommended for complex models (Efron and Tibshirani, 1994; DiCiccio and Efron, 1996). The BCa method has the advantage to be transformation respecting and second order accurate. A $100(1-\alpha)\%$ BCa percentile confidence interval is based upon the quantiles of the bootstrap distribution of the estimates and is given by $\left[\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right]$, where $\hat{\theta}^{*(\alpha)}$ is the empirical percentile of level α and

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha/2)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha/2)})}\right), \; \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha/2)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha/2)})}\right)$$

where $\hat{z}$ is a bias correction constant defined as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}^*_b < \hat{\theta})}{B}\right),$$

where and # means "number of". To calculate the accelerated constant $\hat{a}$ we follow (Efron and Tibshirani, 1994, p. 186):

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6\{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\}^{3/2}},$$

where $\hat{\theta}_{(i)}$ is the estimate of θ based on the observed data with the i-th observation deleted and $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(i)}/n$. The estimates for the polychoric correlations and their 95% confidence intervals are also reported in Table 4.

The results reported in Table 5 and 6 are obtained thanks to the normal approximation of estimators. From Table 5, we comment that age and sex are not linearly

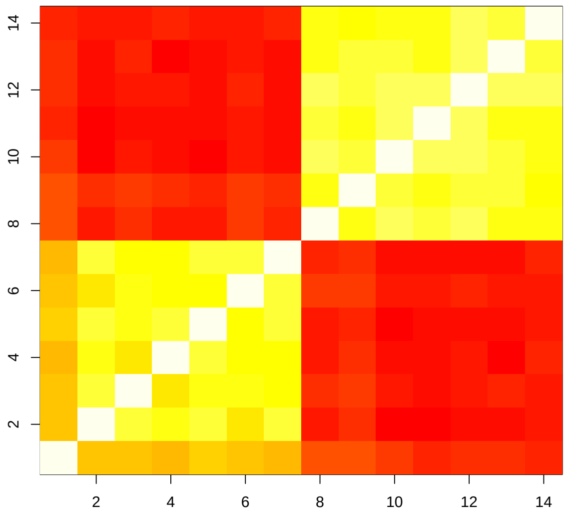

Figure 2: Maximum pairwise likelihood estimates for the polychoric correlations represented on a color scale ranging from red (polychoric correlation equal to 0) to white (polychoric correlation equal to 1).

related to the global item scores. On the other side higher scores are more prominent among subjects with higher educational level (p-value lower than 5% level). To visualize the relations among the items, consider the plot in Figure 2 in which we represent the estimated polychoric correlations with a color ranging from white (correlation equal to one) to red (correlation equal to zero). Note that we do not have any negative estimate. Our estimates indicate that the correlations of the latent Gaussian variables are higher within the two blocks of questions and lower between the two blocks. This is a very important evidence, since it means that high rating in the first item, i.e. high expected quality for the first item, is strongly related to the expected quality of other items, but not so related to its experienced quality. To confirm that, consider Table 6, reporting the differences $\xi_{j+7} - \xi_j$, namely the gap between experienced quality and expected quality for the j-th characteristic. For all $j = 1, \ldots, 7$, this gap is negative with high significance (p-value lower than 10^{-40}), meaning that the the experienced quality is clearly below customers' expectations.

5 Discussion

In this paper we have studied the pairwise likelihood approach for multivariate ordinal probit models, showing that its computational gain is dramatic if compared to the standard likelihood approach. The method has been applied to customer satisfaction data, introducing several complications to the basic model, such as covariates dependence and different means parameter for each latent variable. Despite the considerable number of parameters, the likelihood optimization is feasible in reasonable time. A bootstrap-based approach to evaluate the estimators' standard errors has been adopted. We hope that our contribution will encourage the use of pairwise likelihood for ordered probit models in several applied fields such as marketing, psychometry, and medical research.

Appendix

The matrices used in the simulation experiment of Section 3 are

$$\Sigma_1 = \begin{pmatrix} 1 & -0.06 & 0.16 & -0.30 & -0.36 \\ & 1 & 0.93 & 0.11 & 0.49 \\ & & 1 & 0.21 & 0.60 \\ & & & 1 & 0.59 \\ & & & & 1 \end{pmatrix}$$

$$\Sigma_3 = \begin{pmatrix} 1 & 0 & -0.65 & -0.26 & 0 \\ & 1 & 0 & 0 & -0.69 \\ & & 1 & -0.23 & 0 \\ & & & 1 & 0.41 \\ & & & & 1 \end{pmatrix}$$

References

Agresti, A., 2002. Categorical data analysis. Vol. 359. John Wiley & Sons.

Cox, D. R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. Biometrika 91, 729–737.

De Leon, A., 2005. Pairwise likelihood approach to grouped continuous model and its extension. Statistics & probability letters 75 (1), 49–57.

DiCiccio, T. J., Efron, B., 1996. Bootstrap confidence intervals. Statistical Science, 189–212.

Efron, B., Tibshirani, R. J., 1994. An introduction to the bootstrap. Vol. 57. CRC press.

Genz, A., 1992. Numerical computation of multivariate normal probabilities. Journal of computational and graphical statistics 1 (2), 141–149.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of psychology 22 (140).

Lindsay, B., 1988. Composite likelihood methods. Comtemporary Mathematics 80, 221–240.

Lu, S., 2011. Measuring dependence via mutial information. Master's thesis, Queen's University.

Mardia, K. V., Kent, J. T., Hughes, G., Taylor, C. C., 2009. Maximum likelihood estimation using composite likelihoods for closed exponential families. Biometrika 96 (4), 975–982.

Molenberghs, G., Verbeke, G., 2005. Models for discrete longitudinal data. Springer Verlag.

Parasuraman, A., Zeithaml, V., L.L., B., 1988. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64 (1), 12–37.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `http://www.R-project.org/`

Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. Biostatistics 11 (1), 127–138.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statistica Sinica 21, 5–42.

Varin, C., Vidoni, P., 2006. Pairwise likelihood inference for ordinal categorical time series. Computational Statistics & Data Analysis 51 (4), 2365–2373.

Xu, X., Reid, N., 2011. On the robustness of maximum composite likelihood estimate. Journal of Statistical Planning and Inference 141 (9), 3047–3054.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. Canadian Journal of Statistics 33, 335–356.

Table 4: Estimated polychoric correlations and 95 % accelerated bias corrected bootstrap confidence intervals for the quality of service dataset.

Index	Estimate	CI	Index	Estimate	CI
(1, 2)	0.516	(0.406,0.625)	(5, 6)	0.120	(0.565,0.737)
(1, 3)	0.446	(0.331,0.562)	(5, 7)	0.127	(0.642,0.806)
(1, 4)	0.730	(0.357,0.556)	(5, 8)	0.154	(0.109,0.334)
(1, 5)	0.462	(0.42,0.611)	(5, 9)	0.139	(0.079,0.313)
(1, 6)	0.683	(0.372,0.567)	(5, 10)	0.177	(0.04,0.275)
(1, 7)	0.629	(0.365,0.56)	(5, 11)	0.165	(0.018,0.243)
(1, 8)	0.516	(0.174,0.39)	(5, 12)	0.756	(0.017,0.248)
(1, 9)	0.738	(0.075,0.307)	(5, 13)	0.769	(0.002,0.245)
(1, 10)	0.684	(0.057,0.294)	(5, 14)	0.797	(0.039,0.274)
(1, 11)	0.720	(-0.019,0.206)	(6, 7)	0.149	(0.7,0.832)
(1, 12)	0.473	(0.025,0.251)	(6, 8)	0.173	(0.165,0.395)
(1, 13)	0.590	(-0.007,0.231)	(6, 9)	0.194	(0.105,0.332)
(1, 14)	0.673	(-0.022,0.213)	(6, 10)	0.198	(0.059,0.305)
(2, 3)	0.661	(0.66,0.799)	(6, 11)	0.133	(0.055,0.286)
(2, 4)	0.650	(0.556,0.77)	(6, 12)	0.189	(0.056,0.294)
(2, 5)	0.470	(0.666,0.807)	(6, 13)	0.171	(-0.012,0.24)
(2, 6)	0.711	(0.501,0.671)	(6, 14)	0.813	(-0.02,0.231)
(2, 7)	0.661	(0.629,0.787)	(7, 8)	0.799	(0.146,0.368)
(2, 8)	0.641	(0.157,0.361)	(7, 9)	0.835	(0.105,0.338)
(2, 9)	0.719	(0.167,0.383)	(7, 10)	0.841	(0.076,0.309)
(2, 10)	0.767	(0.035,0.257)	(7, 11)	0.127	(0.042,0.267)
(2, 11)	0.291	(0.009,0.216)	(7, 12)	0.162	(0.047,0.279)
(2, 12)	0.267	(0.057,0.275)	(7, 13)	0.182	(0.054,0.291)
(2, 13)	0.300	(0.039,0.277)	(7, 14)	0.126	(0.073,0.312)
(2, 14)	0.245	(0.055,0.28)	(8, 9)	0.124	(0.692,0.82)
(3, 4)	0.219	(0.512,0.718)	(8, 10)	0.126	(0.753,0.861)
(3, 5)	0.291	(0.599,0.761)	(8, 11)	0.176	(0.692,0.824)
(3, 6)	0.260	(0.588,0.751)	(8, 12)	0.702	(0.765,0.861)
(3, 7)	0.206	(0.575,0.739)	(8, 13)	0.780	(0.628,0.78)
(3, 8)	0.276	(0.179,0.416)	(8, 14)	0.820	(0.592,0.763)
(3, 9)	0.238	(0.113,0.351)	(9, 10)	0.720	(0.791,0.875)
(3, 10)	0.229	(0.075,0.319)	(9, 11)	0.829	(0.71,0.83)
(3, 11)	0.199	(0.01,0.243)	(9, 12)	0.099	(0.738,0.853)
(3, 12)	0.231	(0.081,0.303)	(9, 13)	0.175	(0.723,0.835)
(3, 13)	0.226	(0.062,0.302)	(9, 14)	0.160	(0.589,0.772)
(3, 14)	0.756	(0.035,0.275)	(10, 11)	0.197	(0.726,0.863)
(4, 5)	0.192	(0.642,0.797)	(10, 12)	0.165	(0.784,0.884)
(4, 6)	0.154	(0.584,0.74)	(10, 13)	0.114	(0.769,0.869)
(4, 7)	0.196	(0.53,0.736)	(10, 14)	0.204	(0.678,0.833)
(4, 8)	0.204	(0.135,0.364)	(11, 12)	0.678	(0.793,0.879)
(4, 9)	0.161	(0.124,0.341)	(11, 13)	0.677	(0.643,0.801)
(4, 10)	0.195	(0.096,0.312)	(11, 14)	0.750	(0.656,0.824)
(4, 11)	0.201	(0.033,0.269)	(12, 13)	0.746	(0.774,0.881)
(4, 12)	0.808	(0.086,0.311)	(12, 14)	0.830	(0.778,0.887)
(4, 13)	0.835	(-0.001,0.258)	(13, 14)	0.811	(0.748,0.902)
(4, 14)	0.102	(0.074,0.315)			

Table 5: Regression coefficients for the effect of the covariates

	Estimate	Std. Error	z value	Pr(>|z|)
Male	0.0130	0.0086	1.5148	0.1298
Age	-0.2002	0.1645	-1.2170	0.2236
High school	1.2857	0.6051	2.1249	0.0336
University	1.4260	0.6222	2.2919	0.0219

Table 6: Differences between the item specific intercepts in the expected and experienced blocks

	Estimate	Std. Error	z value	Pr($< z$)
Gap 1	-2.0277	0.1587	-12.7737	<1e-40
Gap 2	-2.7384	0.1862	-14.7066	<1e-40
Gap 3	-2.7536	0.1736	-15.8601	<1e-40
Gap 4	-3.3125	0.1985	-16.6898	<1e-40
Gap 5	-2.9145	0.1907	-15.2857	<1e-40
Gap 6	-2.6237	0.1909	-13.7470	<1e-40
Gap 7	-2.9021	0.1892	-15.3385	<1e-40

Gap 1: efficiency of the service; Gap 2: reliability of the service; Gap 3: flexibility of the service; Gap 4: velocity in giving solutions; Gap 5: ability of satisfy the customer's needs; Gap 6: ability to implement technological innovation; Gap 7: velocity in adapting to changing legislation;

Acknowledgements

The authors thank Bruno Scarpa for generously providing the data and Nicola Sartori for fruitful comments.

Department of Statistical Sciences
University of Padua
Italy