The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

Brief research report: A Monte Carlo simulation study of small sample bias in ordered logit model under multicollinearity

Eric Nimako Aidoo, Simon K. Appiah & Alexander Boateng

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2019.1708233

MEASUREMENT, STATISTICS, AND RESEARCH DESIGN

Brief research report: A Monte Carlo simulation study of small sample bias in ordered logit model under multicollinearity

Eric Nimako Aidoo[a], Simon K. Appiah[a], and Alexander Boateng[b]

[a]Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana;
[b]Department of Statistics, University of Cape Coast, Cape Coast, Ghana

ABSTRACT

This study investigated the small sample biasness of the ordered logit model parameters under multicollinearity using Monte Carlo simulation. The results showed that the level of biasness associated with the ordered logit model parameters consistently decreases for an increasing sample size while the distribution of the parameters becomes less variable with low extreme values. In the presence of multicollinearity, the level of biasness increases and this issue is particularly severe for small sample sizes. By comparing three different approaches for dealing with the multicollinearity problem in the model, the study demonstrated that the use of penalized maximum likelihood estimation technique provides better results which is interpretable compared to the other approaches considered.

KEYWORDS

Multicollinearity; ordered logit model; penalized mle; principal component analysis; simulation; small sample

CATEGORICAL DATA ANALYSIS is often conducted in many disciplines to describe the effect of a set of explanatory variables on a categorical dependent variable. Such a dependent variable can be of three forms: count, nominal or ordinal (Long & Freese, 2011). When the dependent variable is ordinal (such as rating systems as excellent, good, fair, or poor), there are many suitable models available to describe such a situation and the choice of a specific model over the others may be influenced by some underlying assumptions (Fullerton, 2009; Long & Freese, 2011). Among the various models for the ordinal dependent variable with more than two outcomes, the most commonly used one in many disciplines is the ordered logit model (Fagerland & Hosmer, 2016; Fullerton, 2009; Lipsitz et al., 2013).

The ordered logit model (also known as cumulative logit or proportional odds model) has been widely used because it allows for linear modeling of ordinal dependent variable with several predictors (Tamura & Liu, 2015) and its interpretation is simple. The parameters of the ordered logit model are usually obtained using the maximum likelihood estimation technique. The maximum likelihood estimators are known to be unbiased and consistent when large sample size is involved (Murphy, Rossini, & van der Vaart, 1997). However, in many disciplines such as behavioral and medical sciences, small samples are usually common because of many reasons including financial constraints (McNeish, 2017). As observed in other known models (Bergtold, Yeager, & Featherstone, 2018; Ye & Lord, 2014), the maximum likelihood estimation of ordered logit model parameters using small sample can produce bias estimates with high standard errors (Zahid & Ramzan, 2012).

CONTACT Eric Nimako Aidoo en.aidoo@yahoo.com Department of Mathematics, Kwame Nkrumah University of Science and Technology, Private Mail Bag, KNUST, Kumasi, Ghana.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/vjxe.

Apart from small sample size, multicollinearity or collinearity also causes biasness in the maximum likelihood estimator (Zahid & Ramzan, 2012). Multicollinearity describes a situation where there exists a high correlation between two or more independent variables in a model. The existence of high level of multicollinearity will cause inconsistency or bias in the model parameters and can also inflate the estimated standard errors associated with the model parameters (Lavery, Acharya, Sivo, & Xu, 2017; Sinan & Alkan, 2015; Zahid & Ramzan, 2012). Multicollinearity can also cause difficulty in the interpretation of the model parameters (Mackinnon & Puterman, 1989). Although the problem of small sample bias (Bergtold et al., 2018; Nemes, Jonasson, Genell, & Steineck, 2009; Ye & Lord, 2014) and multicollinearity bias (Lavery et al., 2017) have been studied for other modeling approaches, little is known about the ordered logit model, particularly the combined effect. Such information is important for researchers to understand the effect or the level of bias these two problems can have in their conclusions. Although, different studies with regards to multicollinearity and small sample biasness have been conducted (Lavery et al., 2017; Zainodin, Noraini, & Yap, 2011), it is difficult to know at what sample size a researcher should be worried about multicollinearity problem.

For the past decades, researchers have tried to provide remedies to multicollinearity problems. One of the simplest and most frequently used among the various approaches is the *drop-variable* approach (Chen, 2012). This approach involves the removal of one of the collinear predictors from the model. Another approach which is also in use is the principal component analysis (PCA). The PCA method extract from the collinear variables a new variable called "component" which is a linear combination of the collinear variables (Jolliffe, 2002). The collinear variables in the model are then replaced with the component. In addition to the aforementioned approaches, an estimator based method called penalized maximum likelihood (penalized MLE) estimation is also used. The penalized MLE approach also known as ordinal ridge regression resolve multicollinearity problem in a model by combing log-likelihood function with penalty term in order to obtain parameter estimates with good compromise between bias and variance (Zahid & Ramzan, 2012).

Although, all these methods exist to remedy the multicollinearity problem, there is limited literature that compares their performance with regards to their suitability for different sample sizes. This information is particular important as previous study argued that none of the method really solve the problem (Chen, 2012). This study seeks to address the following questions: 1) How does small sample affect the parameter distribution? 2) Does the presence of multicollinearity exacerbate the parameter bias due to small sample? 3) How different are the results from the three methods of resolving multicollinearity problem, particular with regards to small sample size?

The study aimed to examine the sample bias in the estimated parameters of ordinal logit model under multicollinearity using Monte Carlo simulation techniques. Though this introduction has focused on small sample and multicollinearity bias, we also extend the investigation to ascertain the extent to which small sample cause biasness under multicollinearity. In addition, the study aim to compare the three methods of handling multicollinearity problem in ordered logit model to ascertain the extent to which each approach can yield less biased model parameter estimates.

Ordered logit model

The ordered logit model (also known as cumulative logit or proportion logit model) is an extension of the logistic regression model and was proposed by McCullagh (1980) to handle ordinal dependent variable with more than two outcomes. Although, McCoullagh discussed a general class of possible models for ordinal data, the ordered logit model is the most widely used in practice because of its simplicity in terms of interpretation. Let Y represents an ordinal response variable with k ordered outcomes and $\boldsymbol{X}$ represents a vector of explanatory variables, then the ordered logit model describing the relationship between these variables can be expressed Agresti, 2010):

$$\text{logit}[P(y \leq j)] = \log\left[\frac{P(y \leq j)}{1 - P(y \leq j)}\right] = \alpha_j + \beta X, \quad \textit{for } j = 1, 2, \ldots, k-1 \tag{1}$$

where $P(y \leq j) = p_1 + p_2 + \cdots + p_j$ are the cumulative probabilities, α_j represents the intercept parameter (also known as cutoff point or threshold parameter) of the cumulative probability j and β is a column vector of parameters that describe the effects of the explanatory variable(s) on the dependent variable. From (1), since the relationship between all pairs of categories is the same, we obtain only one coefficient (beta) for the all the categories in the estimated model and different intercepts (alpha) for each category. Model (1) can then be expressed in terms of cumulative probabilities as:

$$P(y \leq j) = \frac{\exp(\alpha_j + \beta X)}{1 + \exp(\alpha_j + \beta X)}, \quad \textit{for } j = 1, 2, \ldots, k-1 \tag{2}$$

The parameters in the model (2) are estimated using the maximum likelihood estimation method (Agresti, 2010; McCullagh, 1980).

Simulation setting

To investigate small sample and multicollinearity bias in the ordered logit model, a Monte Carlo simulation studies was conducted. In the simulation 1,000 random samples (replications) of sizes $n = 30$; 50; 100; 150; 200; 300; 500; 1000; 1500; 2000; and 5000 were generated based on model (1). During data generation process three versions of ordered logit models were specified to examine the small sample bias and multicollinearity bias of ordered logit model. These three versions of data were simulated as follows:

In the first version, the following routing was done:

a. Generate three independent normal covariates $x_1, x_2 \text{and} x_3$ from a multivariate normal distribution with mean vector $\boldsymbol{\mu}' = (0, 2, 3)$ and covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

b. Use information in (a) to generate a latent variable y^* from logistic distribution with location parameter $\mu = -0.2x_1 + 0.6x_2 + 0.4x_3$ and scale parameter of 1.

c. Generate the observed response variable y which relates to the latent variable through the following structural model:

$$y = \begin{cases} 1 & \textit{if } \; y^* < \tau_1 \\ 2 & \textit{if } \; \tau_1 \leq y^* < \tau_2 \\ 3 & \textit{if } \; y^* \geq \tau_2 \end{cases} \tag{3}$$

where τ_1 and τ_2 represent 0.3 and 0.7 quantile of the distribution of y^*, respectively.

d. Fit ordered logit model $(y \sim x_1 + x_2 + x_3)$ to the data obtain in (a) and (c).

e. Repeat the process from (a) to (d) 1,000 times for different samples and calculate the appropriate parameters.

In the second version:

a. To allow for multicollinearity in the model, the three normal covariates x_1, x_2 and x_3 were generated from a multivariate normal distribution with mean vector $\boldsymbol{\mu}' = (0, 2, 3)$ and covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \rho = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 \text{ and } 0.99.$$

b. All other procedures remained the same as outlined in the first version.

In the third version:

a. To compared the three methods of handling multicollinearity described in the previous section, the three normal covariates x_1, x_2 and x_3 were generated from a multivariate normal distribution with mean vector $\boldsymbol{\mu}' = (0, 2, 3)$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.95 & 0 \\ 0.95 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

b. All other procedures remained the same as outlined in the first version.

In general, 1000 replications of each possible combination of 11 sample size and 12 collinearity conditions were considered. This led to a total of 132,000 simulation replications.

The results for different sample sizes and multicollinearity level were compared using root mean square error (RMSE) and mean absolute error (MAE). The RMSE is calculated as the square root of the average squared distance between the true value and estimated parameter from the model in each replication whiles the MAE is calculated as the average of the absolute distance between the true value and estimated parameter from the model in each replication. The average is performed over the 1000 replications. The computation was repeated for each sample size and collinearity condition across all replications of that condition. Estimates with smaller RMSE and MAE are preferred. All the analysis including data generation, simulation and estimation of model parameters were conducted using *R* (R Core Team, 2018).

Results and discussion

The small sample bias in the parameters of ordered logit model was investigated using the first version simulated data. The mean of the estimated parameters and their corresponding standard error as well the empirical bias for different sample sizes is presented in Table 1. The bias was measured using RMSE and MAE. The results suggest that the estimated parameters asymptotically approach the true value on average whilst their corresponding standard error also decreased for increasing sample size on average. The two measures of bias also indicated that the bias for the three parameters decrease for increasing sample size (Table 1). The results are consistent with

Table 1. Mean of the estimated parameter and their corresponding standard errors for ordered logit model with three independent covariates and different sample sizes.

	Coefficient			Standard error			MSE			MAE		
Sample size	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
N = 30	−0.216	0.676	0.445	0.398	0.422	0.406	0.456	0.483	0.464	0.353	0.376	0.353
N = 50	−0.223	0.661	0.425	0.291	0.308	0.297	0.311	0.347	0.310	0.245	0.269	0.244
N = 100	−0.198	0.620	0.411	0.196	0.208	0.201	0.203	0.218	0.199	0.161	0.170	0.155
N = 150	−0.213	0.615	0.411	0.160	0.168	0.163	0.165	0.173	0.167	0.131	0.137	0.133
N = 200	−0.200	0.607	0.404	0.137	0.144	0.140	0.144	0.144	0.137	0.115	0.112	0.108
N = 300	−0.200	0.608	0.408	0.111	0.117	0.114	0.113	0.118	0.119	0.090	0.095	0.094
N = 500	−0.200	0.602	0.404	0.086	0.090	0.087	0.086	0.093	0.085	0.069	0.074	0.068
N = 1,000	−0.205	0.604	0.401	0.061	0.064	0.062	0.061	0.066	0.060	0.049	0.054	0.048
N = 1,500	−0.200	0.604	0.401	0.049	0.052	0.050	0.049	0.053	0.050	0.039	0.043	0.040
N = 2,000	−0.200	0.599	0.401	0.043	0.045	0.044	0.044	0.046	0.045	0.035	0.037	0.036
N = 5,000	−0.200	0.599	0.400	0.027	0.028	0.028	0.027	0.029	0.028	0.021	0.023	0.022

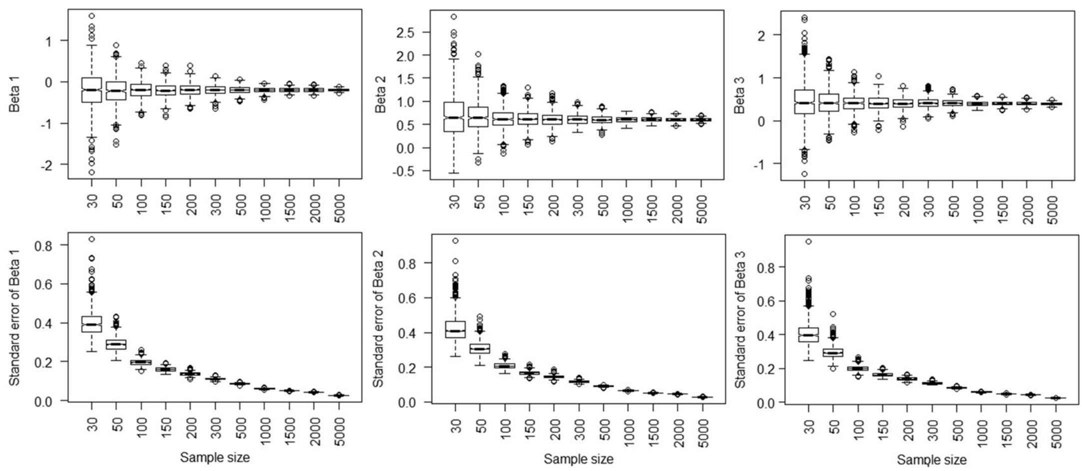

Figure 1. Boxplot of the simulated distribution of the parameters (top panel) and their standard error (bottom panel) for ordered logit model with three independent covariates and for different sample sizes.

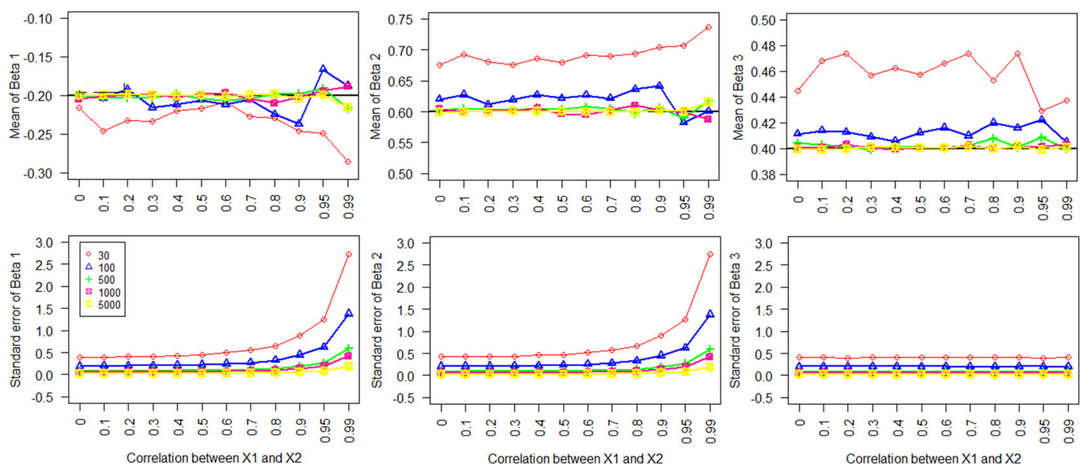

Figure 2. The estimated parameter (top panel) and their corresponding standard errors (bottom) for ordered logit model with three covariates of which two (x_1 and x_2) of them are collinear and for different sample sizes.

the existing findings on the behavior of maximum likelihood estimation for increasing sample size (Ye & Lord, 2014). It was also found that the small sample size did not only leads to more bias in the model parameters, but it also translate to more variability and more extreme values in the distribution of the model parameters and their corresponding standard errors (Figure 1).

The effect of multicollinearity was also investigated in addition to the sample size bias in the ordered logit model parameters using the second version simulated data. The results showed that the higher the level of multicollinearity the more variable the model coefficient (Top panel of Figure 2). This was particularly true for small sample size. However, when the sample size increases, the impact of the multicollinearity toward the model parameter becomes less severe. For a sample of 500 and above, the model parameters are less variable for high level of multicollinearity. The standard error of the model parameters also increases for an increasing level of multicollinearity (Bottom panel of Figure 2). Again, it was more pronounced for small sample size, particularly below 500. For large sample size, the effect of multicollinearity in the standard error becomes pronounced only in high values of multicollinearity (about 0.9 and above). These findings confirm the notion that an increase in multicollinearity makes model parameters less stable

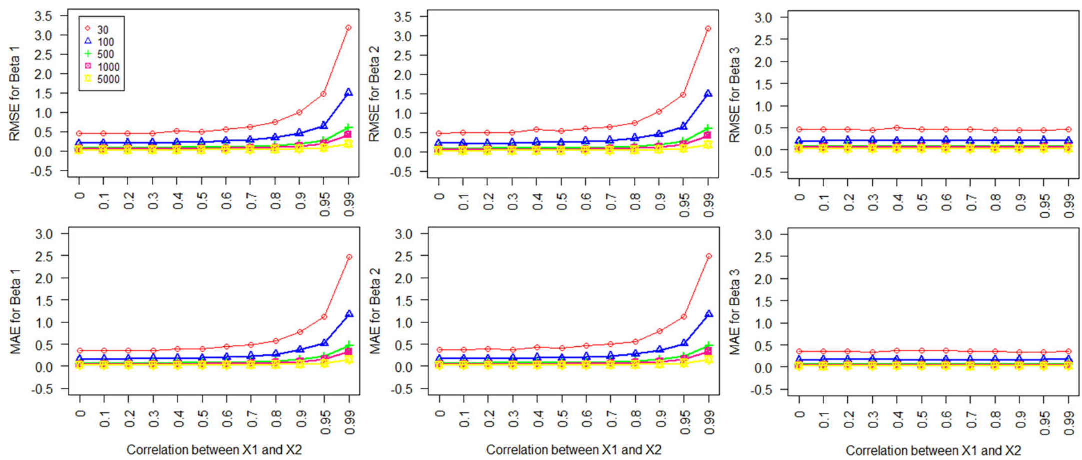

Figure 3. The RMSE (top panel) and MAE (bottom) of the estimated parameters of ordered logit model with three covariates of which two (x_1 and x_2) are collinear and for different sample sizes.

Table 2. Estimated parametrs for ordered logit model with multicollinearity. The paramter are presented for the three approaches described to handle multicollinearity problem.

	Drop variable x_1		Drop variable x_2		PCA		Penalized MLE		
Sample size	β_2	β_3	β_1	β_3	Component	β_3	β_1	β_2	β_3
N = 30	0.435	0.436	0.394	0.436	−0.009	0.436	−0.105	0.536	0.435
N = 50	0.427	0.440	0.379	0.436	0.001	0.438	−0.172	0.590	0.442
N = 100	0.416	0.410	0.375	0.407	0.012	0.409	−0.159	0.567	0.411
N = 150	0.418	0.409	0.371	0.407	0.013	0.408	−0.217	0.625	0.410
N = 200	0.411	0.404	0.369	0.401	−0.001	0.402	−0.181	0.583	0.405
N = 300	0.416	0.405	0.374	0.402	−0.002	0.404	−0.182	0.589	0.405
N = 500	0.413	0.399	0.371	0.396	0.008	0.398	−0.190	0.594	0.399
N = 1,000	0.411	0.402	0.367	0.400	0.000	0.401	−0.209	0.609	0.403
N = 1,500	0.410	0.401	0.367	0.399	0.009	0.400	−0.196	0.596	0.402
N = 2,000	0.410	0.399	0.368	0.397	0.002	0.398	−0.198	0.599	0.399
N = 5,000	0.409	0.399	0.366	0.396	0.022	0.398	−0.198	0.597	0.399

across samples thereby increasing the associated standard error (Bergtold et al., 2018; Kutner, Nachtsheim, Neter, & Li, 2005; Lavery et al., 2017). The results confirms already established findings that the presence of multicollinearity cause biasness in model parameters (Lavery et al., 2017). Also, the bias associated for collinear variables were found to be for higher levels of multicollinearity (Figure 3).

In an attempt to minimize the level of biasness in model parameters due to multicollinearity, three different approaches were explored. To begin with, the drop-variable approach which involves omitting of the collinear variables was first explored. The results show that, excluding one of the collinear variables from the model leads to a significant change in the coefficient of the remaining collinear variable in the model (Table 2). Similar results were obtain irrespective of the variable excluded from the model (i.e., either dropping variable x_1 or x_2). In addition, there is the tendency for the sign of the coefficient to change and this may leads to incorrect interpretation of the model parameter. For instance as observed in Table 2, the sign of the parameter β_1 changed from negative to positive after dropping the variable x_2 from the model. The sign remained positive irrespective of the changes in the sample sizes. This result may be influenced by the fact that the parameters of a model formed from collinear covariates are unreliable and interchangeable, thereby making it more difficult to determine the contribution of each predictor in the model (Lavery et al., 2017). The level of biasness associated with the remaining variables in

Table 3. Estimated RMSE for ordered logit model with multicollinearity. The paramter are presented for the three approaches described to handle multicollinearity problem

	Drop variable x_1		Drop variable x_2		PCA		Penalized MLE		
Sample size	β_2	β_3	β_1	β_3	Component	β_3	β_1	β_2	β_3
N = 30	0.763	0.463	0.733	0.467		0.436	1.081	1.068	0.441
N = 50	0.693	0.351	0.650	0.350		0.313	0.821	0.821	0.318
N = 100	0.645	0.279	0.607	0.280		0.204	0.601	0.592	0.206
N = 150	0.639	0.251	0.594	0.253		0.163	0.493	0.501	0.164
N = 200	0.626	0.242	0.585	0.244		0.141	0.431	0.436	0.141
N = 300	0.627	0.225	0.585	0.228		0.113	0.346	0.350	0.113
N = 500	0.619	0.221	0.577	0.223		0.091	0.264	0.270	0.091
N = 1,000	0.613	0.208	0.570	0.210		0.064	0.190	0.193	0.064
N = 1,500	0.612	0.205	0.569	0.207		0.050	0.154	0.155	0.050
N = 2,000	0.612	0.205	0.569	0.208		0.043	0.135	0.135	0.043
N = 5,000	0.609	0.203	0.567	0.206		0.028	0.084	0.085	0.028

Table 4. Estimated MAE for ordered logit model with multicollinearity. The paramter are presented for the three approaches described to handle multicollinearity problem.

	Drop variable x_1		Drop variable x_2		PCA		Penalized MLE		
Sample size	β_2	β_3	β_1	β_3	Component	β_3	β_1	β_2	β_3
N = 30	0.661	0.374	0.623	0.375		0.481	0.864	0.851	0.342
N = 50	0.628	0.287	0.584	0.286		0.440	0.643	0.649	0.250
N = 100	0.616	0.231	0.575	0.232		0.407	0.480	0.469	0.163
N = 150	0.618	0.211	0.571	0.213		0.398	0.391	0.398	0.131
N = 200	0.611	0.208	0.569	0.211		0.406	0.346	0.351	0.111
N = 300	0.616	0.198	0.574	0.201		0.405	0.278	0.281	0.092
N = 500	0.613	0.202	0.571	0.205		0.393	0.210	0.213	0.072
N = 1,000	0.611	0.198	0.567	0.200		0.400	0.152	0.157	0.051
N = 1,500	0.610	0.199	0.567	0.201		0.391	0.123	0.125	0.040
N = 2,000	0.610	0.201	0.568	0.203		0.398	0.108	0.108	0.034
N = 5,000	0.609	0.201	0.566	0.204		0.378	0.068	0.068	0.022

the models particularly the collinear variables was also high (Tables 3 and 4) compared to the estimated bias for the same parameters in Table 1.

The use of principal component analysis to form a new variable from the set of correlated variables was explored. The results showed that the estimated coefficients vary from the true parameter values and this makes it difficult for interpretation (Table 2). In addition, the sign of the coefficient kept on changing for different sample sizes, which would tend to affect its interpretation. Although it was not possible to calculate the level of biasedness for the estimated parameter of the component, the estimated bias for β_3 was also high (Tables 3 and 4) compared to the estimated bias for the same parameter in Table 1.

Lastly, the use of penalized maximum likelihood estimation (penalized MLE) technique also known as ordinal ridge regression (Zahid & Ramzan, 2012) was explored. From the results the penalized MLE approach produced more realistic parameters compared to the other two approaches discussed (Table 2). The estimated bias for all the three parameters from the penalized MLE was least compared to the other two approaches (Tables 3 and 4). The results also showed that the level of biasness reduces for an increasing sample size.

Conclusion

This study investigated small sample biasness of ordered logit model under multicollinearity using Monte Carlo simulation. Such study is important, particularly for practitioners to allow them to understand the consequences of using a specific sample size. In addition, it will also allow them to understand the effect of multicollinearity on model parameters and how to deal with them if present. The results from the simulation studies suggest that the empirical parameter of the ordered logit model asymptotically approaches the true value. The bias of the model parameter as determined by the two accuracy measures (RMSE and MAE) decreases for an increasing sample size. The use of small sample in modeling does not only leads to bias but it also translates to more variability and more extreme values in the empirical distribution of the model parameters. Multicollinearity was found to increase the level of biasness in the model parameter, particularly for small sample size. However, when sample size increase the effect of multicollinearity becomes less severe. A comparison of the three approaches for handling multicollinearity in ordered logit model showed that the use of penalized maximum likelihood approach for estimating the model parameter is better compared to either excluding one of the collinear variables or applying principal component analysis. It was also found that the use of penalized estimation method provided results which are easily interpretable. The results from such approach have similar characteristics to the model without multicollinearity problem. Hence, the use of such approach is more recommendable in handling multicollinearity in ordered logit model compared to other two approaches.

These findings of the study suggest that in applied research a sample size of 500 or more will be appropriate when multicollinearity is expected and the coefficient of correlation is 0.9 or less. In such a situation the multicollinearity will not have a detrimental effect on the model output and the interpretation. On the other hand, if the expected coefficient of correlation is above 0.9, it is recommended that a sample size above 500 should be used. In a situation where such a sample cannot be obtained, it is recommended that the penalized maximum likelihood approach should be used instead of the conventional mle approach.

Disclosure statement

The authors declare that they have no conflict of interest.

Funding

This research received no grant from any funding agency in the public, commercial or not-for-profit sectors.

ORCID

Eric Nimako Aidoo http://orcid.org/0000-0003-2663-6747

References

Agresti, A. (2010). *Analysis of ordinal categorical data*, 2nd Edition. Ontario: John Wiley & Sons, Inc.

Bergtold, J. S., Yeager, E. A., & Featherstone, A. M. (2018). Inferences from logistic regression models in the presence of small samples, rare events, nonlinearity, and multicollinearity with observational data. *Journal of Applied Statistics*, *45*(3), 528–546. doi:10.1080/02664763.2017.1282441

Chen, G. J. (2012). A simple way to deal with multicollinearity. *Journal of Applied Statistics*, *39*(9), 1893–1909. doi: 10.1080/02664763.2012.690857

Fagerland, M. W., & Hosmer, D. W. (2016). Tests for goodness of fit in ordinal logistic regression models. *Journal of Statistical Computation and Simulation*, *86*(17), 3398–3418. doi:10.1080/00949655.2016.1156682

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods & Research*, *38*(2), 306–347. doi:10.1177/0049124109346162

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical model, 5th Edition*. New York: McGraw-Hill/Irwin.

Lavery, M. R., Acharya, P., Sivo, S. A., & Xu, L. (2017). Number of predictors and multicollinearity: What are their effects on error and bias in regression? *Communications in Statistics - Simulation and Computation*, *48*(1), 27–38. doi:10.1080/03610918.2017.1371750

Lipsitz, S. R., Fitzmaurice, G. M., Regenbogen, S. E., Sinha, D., Ibrahim, J. G., & Gawande, A. A. (2013). Bias correction for the proportional odds logistic regression model with application to a study of surgical complications. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(2), 233–250. doi:10.1111/j.1467-9876.2012.01057.x

Long, J. S., & Freese, J. (2011). *Regression model for categorical dependent variables unisng stata*. College Station, TX: Stata Corperation, USA.

Mackinnon, M. J., & Puterman, M. L. (1989). Collinearity in generalized linear models. *Communications in Statistics - Theory and Methods*, *18*(9), 3463–3472. doi:10.1080/03610928908830102

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological))*, *42*(2), 109–142. doi:10.1111/j.2517-6161.1980.tb01109.x

McNeish, D. (2017). Brief Research Report: Growth Models With Small Samples and Missing Data. *The Journal of Experimental Education*, *86*(4), 690–701. doi:10.1080/00220973.2017.1369384

Murphy, S. A., Rossini, A. J., & van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, *92*(439), 968–976. doi:10.2307/2965560

Nemes, S., Jonasson, J. M., Genell, A., & Steineck, G. (2009). Bias in odds ratios by logistic regression modelling and sample size. *BMC Medical Research Methodology*, *9*(1), 56. doi:10.1186/1471-2288-9-56

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Sinan, A., & Alkan, B. B. (2015). A useful approach to identify the multicollinearity in the presence of outliers. *Journal of Applied Statistics*, *42*(5), 986–993. doi:10.1080/02664763.2014.993369

Tamura, R. N., & Liu, X. (2015). Power estimation using the exemplary dataset for ordinal data in the presence of nonproportional odds. *Statistics in Biopharmaceutical Research*, *7*(2), 106–114. doi:10.1080/19466315.2015.1037965

Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic Methods in Accident Research*, *1*, 72–85. doi:10.1016/j.amar.2013.03.001

Zahid, F. M., & Ramzan, S. (2012). Ordinal ridge regression with categorical predictors. *Journal of Applied Statistics*, *39*(1), 161–171. doi:10.1080/02664763.2011.578622

Zainodin, H. J., Noraini, A., & Yap, S. J. (2011). An alternative multicollinearity approach in solving multiple regression problem. *Trends in Applied Sciences Research*, *6*(11), 1241–1255. doi:10.3923/tasr.2011.1241.1255