Re: #2013:12:1126:0:1:REVIEW "Modified generalized method of moments for a robust estimation of polytomous logistic model".

Dear Editor,

Sincerest thanks for your response and reviewers comments on our manuscript. We sincerely apologize for the time it has taken us to respond to these comments, since I exerted a tremendous effort to find a correct way to adapt the kernel-weighted GMM for categorical data. It turns out that the kernel weighted GMM is not directly applicable to categorical data, just like that the general linear model does not work for binary outcome data. It worth a new paper about this topic.

We have modified the paper in response to the extensive and insightful reviewer comments. Furthermore we have rewritten sections of the manuscript and we hope that this comply with the referee's remarks. I hope that a revised version of the manuscript will still be considered by PeerJ. We will respond to the comments point counter point.

Thank you very much.

Xiaoshan Wang
Forsyth Institute
245 First Street, Cambridge MA 02142

Reviewer 1

Basic reporting

1. There are other weighted estimators, such as Kernel-weighted GMM estimator. You may want to refer to these estimators and compare with them in simulations.

Reply:

Thank you very much for an excellent point. Following the commenter's suggestion, we simulated data, and applied the kernel-weighted GMM to obtain parameter estimates. The results are listed in the following table. The simulation results indicates that the kernel weighting won't be able to reducing the bias in our case (Table r1, in which I reported two kernel weighting method. The results from other kernel weighting are similar.) This is probably due to the specific structures of polytomous responses, which is different from the scenario for which the kernel-weighted GMM developed for linear time-series responses (Kuersteiner 2012). In the case of time series data, we assume that the response at time t may be predicted by the response at time $t-1$. This is not the case in polytomous data, in which the response can take only one category. Such a specific correlation structure may require to have very specific kernel in order to achieve the goal of bias reduction.

Since the kernel-weighted GMM with current available kernels does not work well, we did not include the table into the manuscript, but we will do so (or to submit it as supplementary material) if the reviewer and editor would like us to. I added one sentence in the introduction section: "The method is different from the kernel-weighted GMM developed for linear time-series data by Kuersteiner (2012) in that this is a data-driven method for defining weights."

Table r1. Comparison between MLE and kernel-weighted GMM

n	Parameter	MLE			Tukey-Hanning kernel Weighted GMM			Bias reducing kernel Weighted GMM	
		Bias	MSE		Bias	MSE		Bias	MSE
100	β_{20}	0.1027	0.0963		0.1027	0.0106		0.1027	0.0106
	β_{30}	-0.0009	0.1627		-0.0009	0.0027		-0.0009	0.0027
	β_{21}	0.2539	0.1483		0.2539	0.0645		0.2539	0.0645
	β_{31}	-0.0909	0.1317		-0.0909	0.0083		-0.0909	0.0083
	β_{22}	0.4073	0.2316		0.4073	0.1659		0.4073	0.1659
	β_{32}	-0.1007	0.0956		-0.1007	0.0101		-0.1007	0.0101
1000	β_{20}	0.047	0.01		0.047	0.0022		0.047	0.0022
	β_{30}	0.0107	0.0143		0.0107	0.0038		0.0107	0.0038
	β_{21}	0.2905	0.0918		0.2905	0.0844		0.2905	0.0844
	β_{31}	-0.1402	0.0301		-0.1402	0.0197		-0.1402	0.0197
	β_{22}	0.4446	0.2035		0.4446	0.1977		0.4446	0.1977
	β_{32}	-0.0542	0.0104		-0.0542	0.0029		-0.0542	0.0029

2. In some applications, outlying continuous covariates are transformed into their log10 scales to remove outlying. How do you think of this alternative solution?

Reply:

Log-transformation may work to "remove outlying" when the transformation achieves variance stabilization to the variable. In this case, the estimation can be improved. But if the data generation process gives heterogeneous data such that the log-transformation cannot achieve variance stabilization, such a transformation may fail to control the impact of outliers (Luetkepohl and Xu 2009).

3. I think you did not finish section "Application", please check. It would be very interesting to present and compare coefficient estimates from your robust method and MLE in this section.

Reply:

Thanks for the reviewer to point out. I was intend to delete a sentence in the latex file. But I was mistakenly deleted a whole lot by putting a paragraph after the % sign in the latex file.

The following paragraph is put back now:

"As the results indicate, age, gender, and BMI all had significant impact on hypertension status. For example, one unit increase in BMI resulted in an increase of 1.26 (95% confidence interval: 1.16 - 1.35) times in likelihood to have Stage 2 hypertension when compared with the normal status. And with one year age increase, a subject was 1.07 (95% CI: 1.06 - 1.10) times more likely to have Stage 2 hypertension than to stay at the normal healthy status. Contrary to the MLE results for sodium intakes, which were difficult to make a conclusion due to inconsistent estimate, we now find that sodium intakes were statistically significant. When a daily intake of sodium increased one gram, a subject were 1.26 (95% CI: 1.15 - 1.37) times more likely to have Stage 1 hypertension, and 1.25 (95% CI: 1.17 - 1.35) times more likely to have Stage 2 hypertension. These results are consistent with the findings from previous studies".

Experimental design

1. The motivating real dataset is a 2006 study on hypertension, in which the outcomes are ordinal: normal, pre-hypertension, stage I and stage II. Could you please give out more justifications for using polytomous logistic regression model instead of other models that specifically designed for ordinal outcomes, such as cumulative logits model? Sine polytomous logistic regression model is designed for both ordinal and nominal outcomes, the estimation efficiency from this model is not optimal.

Reply:

As I stated in the last sentence of the last paragraph in the first page of the manuscript, the proportional odds assumption is violated. Based on the score test for the proportional odds assumption, we have $\chi^2 = 182.27$ with a degree of freedom of 8, which gives $p < 0.0001$. To avoid the problems caused by the violation of proportional odds assumption, I applied the polytomous logit model.

2. I interpret covariate effects from polytomous logistic regression model as conditional, according to Equation (1). I think you need more explanations on the real data analysis part since you claim \textit{"... research objectives is to examine the association between hypertension and risk factors \textbf{in the population}"}. I recommend you giving more introductions and discussions on the polytomous logistic regression model.

Reply:

Follow the review's suggestion, I added the following paragraph in the introduction section:

"we apply the generalized logit model, using the normal category as the reference level. In the case of J category, the generalized logit model have $J - 1$ comparisons. Each comparison have a set of parameters for all covariates in the model. Therefore, generalized logit model is not parsimonious. But the simultaneous estimation of all parameters is more efficient than separate models for each comparison. It is another option for ordinal response data, especially when a proportional odds model does not fit the data well."

The following sentences are added into the discussion section:

"A reasonable choice to fit ordinal response data is the proportional odds model if the proportional odds assumption is not violated. Proportional odds models can take the ordinal information into modeling. And it reduces the number of parameters which is needed by the generalized logit model. Unfortunately, our data does not met the fundamental assumption of proportional odds models, which makes us choose to treat the outcome as a nominal response."

3. I understand intuitively that the weight in Equation (11) will give observations with outliers smaller weights. Could you give intuitive explanations why these weights can also correct for outcome mis-classification?

Reply:

Since the moments contain the error term between observed response vector y and estimated probability vector $\hat{\pi}$, and the two distances (c_d and c_x) used for differentiating outliers are both summed over the response vector at subject level. Intuitively, a misclassification creates a relatively error elements in the error vector, which can be reflected by the distance, and therefore downweighted subsequently.

4. In the part "the generalized method of weighted moments" between Equations (11) and (12), to justify the choice of tuning parameter c_d, you claimed Rank(u_i^w(\bm{\beta}))=1, I am not following this conclusion. You may want to define Rank and give some proof of this conclusion.

Reply: u_i is a $(J-1)(p+1) \times 1$ column vector, since X_i is a $(J-1)(p+1) \times (J-1)$ matrix, and $y_i^* - \pi_i$ is a $(J-1) \times 1$ vector. So the results from Equation (10) is a scalar. To avoid misunderstanding, I changed the word in the text as "$u_i^W(\boldsymbol{\beta})$ is a column vector, and $d_i(\boldsymbol{\beta})$ is a scalar quadratic distance".

Validity of the findings

1. In the section "Discussion", you claim "the proposed method has good asymptotic behavior". I have seen the estimates are consistent from your simulations but how about their asymptotic distributions? You can include 95\% confidence interval coverage rates in Tables 3 and 4.

Reply: In the paragraph above the RESULTS section on Page 8, I stated "In the supporting document, we demonstrated…". The supporting document is actually theoretical proofs of GMWM's asymptotic property. The reviewer probably missed it since the supporting document may not be directly attached to the manuscript. I changed the supporting document into an appendix, and attach it directly to the manuscript.

Following the suggestion of the reviewer, 95% confidence interval coverage rates are added into Tables 3 and 4.

2. As for asymptotic behavior, could you please provide some theoretical references for the consistency and asymptotic distributions or give proof outlines?

Reply: See the reply to the above comment (1). The theoretical proofs are presented in the supporting document. I changed it into an appendix and make it directly attaching to the main manuscript.

3. In Table 2, I suggest you providing p.value instead of z-score.

Reply:

As suggested, p values replaced z-scores.

Reviewer 2 (Yi-Hui Zhou)

Basic reporting

No Comments

Experimental design

No Comments

Validity of the findings

No Comments

Comments for the author

This is a clearly written manuscript that describes a method to obtain robust estimation of polytomous logistic model. The approach is based on generalized method of moments. The author proved that the method works consistently. The only concern of mine is the starting values of parameters for the method. The author used the regular MLE estimates as starting values. When the data is contaminated, the regular MLE is biased. Such biased starting values might lead to convergence problem, especially when the starting values are seriously biased. It may need to find a better starting values.

Reply:

This is an excellent point. For some likelihood based methods, starting points of parameters is very important. This is because algorithms such as Newton-Raphson is very demanding for a close-to-truth starting point. For Newton-Raphson method, the neighborhood of parameters which can ensure its convergence is usually small. Ideally, the starting values will be close to the true value of parameters, which can also improve efficiency through saving the time by decreasing iterations. $\sqrt{n}$-consistent starting values, if we can find, are most welcomed in this case (Lehmann and Casella 1998; Dominitz and Sherman 2005). But finding such starting values sometimes is troublesome. But for GMM, the requirement for starting values of parameters is relaxed (Lung-fei Lee 2005). Our simulations suggest that ordinary MLE estimation as starting values work reasonably good. All simulations reached convergence, even though different elements of parameters may have different convergence rate (Lung-fei Lee 2005).

Reference:

Helmut Luetkepohl & Fang Xu, 2009. "The Role of the Log Transformation in Forecasting Economic Variables," CESifo Working Paper Series 2591, CESifo Group Munich.

Kuersteiner, Guido M., 2012. "Kernel-weighted GMM estimators for linear time series models," Journal of Econometrics, Elsevier, vol. 170(2), pages 399-421.

Lehmann, E. and Casella, G. (1998), Theory of Point Estimation (Springer Texts in Statistics), Springer.

Dominitz, J. and Sherman, R. P. (2005), \Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation," Econometric Theory, 21, 838-863.

Lung-fei Lee. (2005) Classical Inference with ML and GMM Estimates with Various Rates of Convergence. Accessed through http://www.econ.ohio-state.edu/lee/wp/ml-gmm-tests-05-june-3.pdf