A peer-reviewed version of this preprint was published in PeerJ on 1 July 2014.

View the peer-reviewed version (peerj.com/articles/467), which is the preferred citable publication unless you specifically need to cite this preprint.

Wang X. 2014. Modified generalized method of moments for a robust estimation of polytomous logistic model. PeerJ 2:e467
https://doi.org/10.7717/peerj.467

Modified generalized method of moments for a robust estimation of polytomous logistic model

Xiaoshan Wang[1,2*],

1. Department of clinical and translational research/Forsyth Institute, Cambridge, MA, USA

2. Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Cambridge, MA, USA

∗ **E-mail: xwang@forsyth.org**

Abstract

The maximum likelihood estimation (MLE) method, typically used for polytomous logistic regression, is prone to bias due to both misclassification in outcome and contamination in the design matrix. Hence, robust estimators are needed. In this study, we propose such a method for nominal response data with continuous covariates. A generalized method of weighted moments (GMWM) approach is developed for dealing with contaminated polytomous response data. In this approach, distances are calculated based on individual sample moments. And Huber weights are applied to those observations with large distances. Mellow-type weights are also used to downplay leverage points. We describe theoretical properties of the proposed approach. Simulations suggest that the GMWM performs very well in correcting contamination-caused biases. An empirical application of the GMWM estimator on data from a survet demonstrates its usefulness.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.299v1 | CC-BY 3.0 Open Access | received: 23 Mar 2014, published: 23 Mar 2014

Introduction

Polytomous logistic regression models for multinomial data are a powerful technique for relating dependent categorical responses to both categorical and continuous explanatory covariates [1, 2]. In practice, however, the model building process can be highly influenced by peculiarities in the data. The maximum likelihood estimation (MLE) method, typically used for the polytomous logistic regression model (PLRM), is prone to bias due to both misclassification in outcome and contamination in the design matrix [3, 4]. Hence, robust estimators are needed.

For categorical covariates, we may apply MGP estimator [5], ϕ-divergence estimator [6], and robust quadratic distance estimator [7]. The least quartile dfference estimator can deal with overdispersion problem [8]. But all these methods are difficult to be adapted for continuous covariates.

A generalized method of moments (GMM) estimation can be formed as a substitute of MLE. The GMM is particularly useful when the moment conditions are relatively easy to obtain. GMM has been extensively studied in econometrics [9–13]. Under some regularity conditions, the GMM estimator is consistent [9]. With an appropriately chosen weight matrix, GMM achieves the same efficiency as the MLE [14]. Furthermore, under certain circumstances, GMM provides more flexibility, such as dealing with endogeneity through instrumental variables [15].

Like MLE, GMM estimation can be easily corrupted by aberrant observations [16]. Such observations can bring up disastrous bias on standard parameter estimates if they are not properly accounted for, see [17], [18], and [19]. So we propose a modified estimation method based on an outlier robust variant of GMM. The method is different from the kernel-weighted GMM developed for linear time-series data by [20] in that this is a data-driven method for defining weights. The new approach is evaluated using asymptotic theory, simulations, and an empirical example.

The robust GMM estimator is motivated by the data from a 2006 study on hypertension in a sample of the Chinese population. 520 people completed the survey. Observed variables included demographics, social-economic status, weight, height, blood pressure, and food consumption. Sodium intakes were calculated based on overall food consumption. Among those covariates, age, body mass index (BMI), and sodium intakes are all continuous. Based on blood pressure measurements, subjects were classified into 4 categories: Normal, Pre-hypertension, Stage 1 and Stage 2 hypertension. Table 1 lists the summary statistics of the sample. One of the research objectives is to examine the association between hypertension

and risk factors in the population. Since the proportional odds assumption is violated (Score test for the proportional odds assumption gives $\chi^2 = 182.27$ with a degree of freedom of 8, $p < 0.0001$), we apply the polytomous logistic model, using the normal category as the reference level. In the case of J category, the polytomous logit model have $J - 1$ comparisons. Each comparison have a set of parameters for all covariates in the model. Therefore, the generalized logit model is not parsimonious when comparing with the proportional odds model. 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. Table 2 lists the output from the model estimated by MLE. It is obvious that, if MLE is used, the estimates is inconsistent for sodium intakes, particularly the negative coefficient of sodium intake for the odds between the Stage 2 hypertension and the Normal categories. The inconsistency is more obvious when we plot the odds with respect to the sodium intake, the downward trend of the odds in Figure 2.A. This result contradicts the previous finding that there is a strong relationship between sodium intake and hypertension, see for example [21], [22] and references therein. Besides, Figure 2.A also shows another strange situation: the higher starting points for the odds between the Pre-hypertension and the Normal categories. The scatter plot (Figure 1) between distances and leverages suggests some observations are possible outliers: Observations 21, 33, 85, 92, 194, 274, 336, 414, 459, 483, and 489 have large distances, which are blue-colored, and Observations 37, 83, 263, 459, 483, 485, and 490 have large leverages, which are red-colored.

Table 1. Summary statistics for surveyed subjects.

		Hypertension categories			
Covariate		Normal	Pre-hypertension	Stage 1	Stage 2
Gender	Male	138	104	29	8
	Female	87	114	31	9
Age	Mean	43.2	48.8	54.3	60.3
	Std. Dev.	13.7	13.8	12.2	13.4
BMI	Mean	43.2	48.8	54.3	60.3
	Std. Dev.	13.7	13.8	12.2	13.4
Sodium Intake	Mean	3.7	3.7	4.6	2.7
	Std. Dev.	3.0	2.4	5.0	2.1

The paper is set up as follows. In the next section we presents the basic notations, model, and

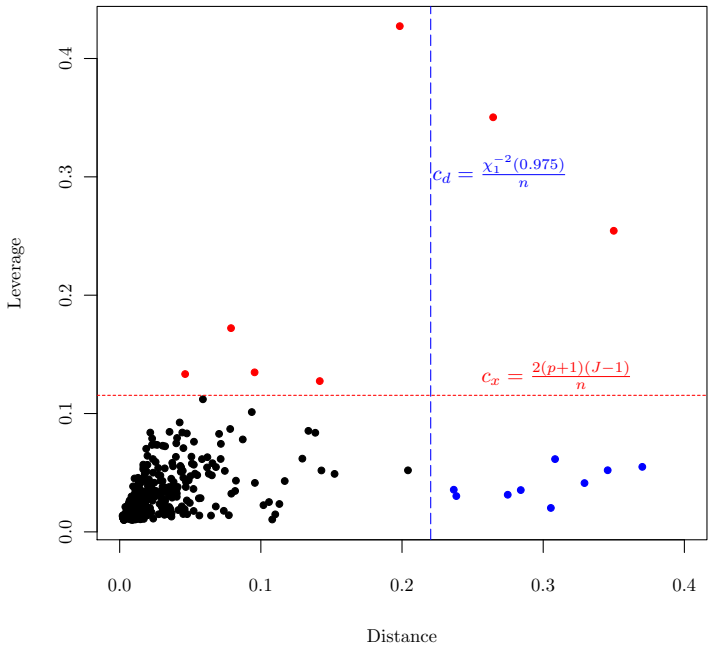

Figure 1. Scatter plot of distance vs. leverages, which are based on MLE. Criteria c_d for the distance and c_x for the leverage are demonstrated.

Table 2. Polytomous logistic regression of a hypertension data: coefficient estimates and standard errors from GMWM and MLE.

		MLE			GMWM		
Var	Coefs	Estimates	Std.Err	p value	Estimates	Std.Err	p value
Sex	β_{21}	0.7062	0.2022	0.0002	1.3339	0.2269	< 0.0001
	β_{31}	0.9789	0.3235	0.0012	1.0368	0.3013	0.0003
	β_{41}	1.4193	0.5746	0.0068	0.6753	0.2195	0.0010
Age	β_{22}	0.0350	0.0075	< 0.0001	0.0671	0.0086	< 0.0001
	β_{32}	0.0715	0.0121	< 0.0001	0.1139	0.0133	< 0.0001
	β_{42}	0.1096	0.0216	< 0.0001	0.0753	0.0103	< 0.0001
BMI	β_{23}	0.1147	0.0316	0.0001	0.1681	0.0360	< 0.0001
	β_{33}	0.2422	0.0474	< 0.0001	0.4382	0.0538	< 0.0001
	β_{43}	0.4351	0.0884	< 0.0001	0.2279	0.0388	< 0.0001
Sodium	β_{24}	0.0104	0.0349	0.3829	0.1831	0.0355	< 0.0001
	β_{34}	0.0919	0.0426	0.0155	0.2315	0.0486	< 0.0001
	β_{44}	-0.2699	0.1580	0.9562	0.2294	0.0353	< 0.0001

Std.Err = standard error.

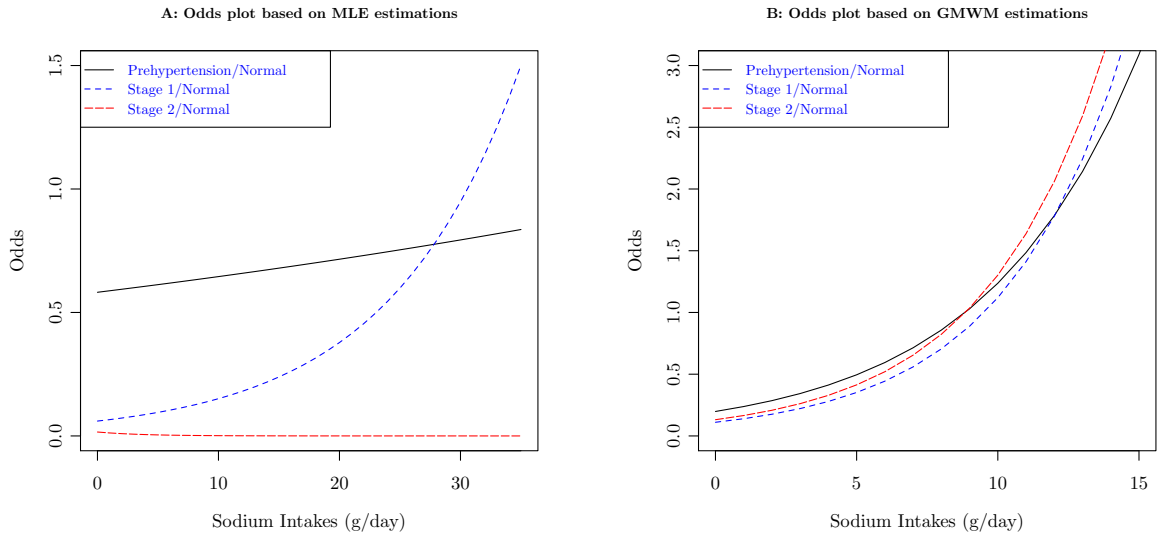

Figure 2. Compare odds plots of sodium intakes between MLE estimates and GMWM estimates on the population of Female, Age=40, and BMI=23.

standard GMM. Section "A robust GMM" introduces the outlier robust GMM estimator, and gives a detailed exposition of its implementation. In Section "Results", we compares the performance of the standard MLE with the new estimator using a Monte-Carlo experiment. And we apply both estimators to real epidemiological data, and illustrate the usefulness of the robust estimator for application oriented researchers. We conclude with a discussion of advantages and limitations of the approach. The supporting document gathers the proofs of the asymptotic property.

Materials and Methods

The baseline-category logit model

Assume a random sample of size n from a large population. Each element in the population may be classified into one of J categories, denoted by $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iJ})$ the multinomial trial for subject i, where $y_{ij} = 1$ when the response is in category j and $y_{ij} = 0$ otherwise, $i = 1, ..., n$, $j = 1, ..., J$. Thus, $\sum_j y_{ij} = 1$. Suppose p explanatory covariates, with at least one of them being continuous, are observed. Define $\mathbf{x}_i = (1, x_{i1}, ..., x_{ip})$, and $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$. We assume that $(\mathbf{y}_i, \mathbf{x}_i)$ are independently

and identically distributed ($i.i.d$). Let $\pi_{ij} = \pi_j(\mathbf{x}_i) = P(Y_i = j|\mathbf{x}_i)$,, denote the probability that the observation of Y belongs to category j, given covariates $\mathbf{x}_i$, we assume the relationship between the probability π_j and $\mathbf{x}$ can be modeled as:

$$log\left\{\frac{\pi_j(\mathbf{x}_i)}{\pi_J(\mathbf{x}_i)}\right\} = \mathbf{x}_i^T\beta_j, \quad j = 2, ..., J \tag{1}$$

where $\beta_j^T = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})$. Here we set the first category as reference class. This model is called a baseline-category logit model [23] or generalized logit model [24]. MLE is usually used for obtaining parameter estimation of this model. Here we present an alternative estimation method formed with the GMM.

Estimation using GMM

The baseline-category logit model can be viewed as a multivariate model. Define $\mathbf{y}_i^{*T} = (y_{i2}, ..., y_{iJ})$, since y_{i1} is redundant. Let $\mathbf{X}^T = (X_1^T, ..., X_n^T)$ is a $n(J-1) \times (p+1)(J-1)$ matrix, with X_i^T, a $(J-1) \times (p+1)(J-1)$ matrix, defined as:

$$X_i^T = \begin{pmatrix} \mathbf{x}_i^T & & & \\ & \mathbf{x}_i^T & & \\ & & \cdots & \\ & & & \mathbf{x}_i^T \end{pmatrix} \tag{2}$$

In the GMM framework, we define

$$u(\boldsymbol{\beta}) = X_i(\mathbf{y}_i^* - \boldsymbol{\pi}_i), \quad i = 1, ..., n. \tag{3}$$

where $\boldsymbol{\pi}_i^T = (\pi_{i2}, \pi_{i3}, ..., \pi_{iJ})$. And $\boldsymbol{\beta}^T = (\beta_2^T, \beta_3^T, ..., \beta_J^T)$ is the $(p+1)(J-1)$ vector of unknown parameters. The population moment condition is

$$E\{u(\boldsymbol{\beta})\} = 0,$$

with the corresponding sample moment condition

$$U_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} u(\boldsymbol{\beta}). \tag{4}$$

The GMM estimation of $\hat{\boldsymbol{\beta}}_M$ can be obtained by minimizing the following quadratic objective function

$$Q_n(\boldsymbol{\beta}) = U_n^T(\boldsymbol{\beta})\Sigma_n^{-1}(\boldsymbol{\beta})U_n(\boldsymbol{\beta}),$$

where $\Sigma_n(\boldsymbol{\beta})$ can be the empirical variance-covariance matrix given by

$$\Sigma_n(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n} u^T(\boldsymbol{\beta})u(\boldsymbol{\beta}) - \frac{1}{n}U_n(\boldsymbol{\beta})U_n^T(\boldsymbol{\beta}).$$

Or, for the best efficiency of the GMM estimation, we can take the information matrix of the polytomous logit model (PLRM), that is,

$$\Sigma_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} X_i(D_i - \boldsymbol{\pi}_i\boldsymbol{\pi}_i^T)X_i^T. \tag{5}$$

where $D_i = diagonal(\boldsymbol{\pi}_i)$.

In general, $\hat{\boldsymbol{\beta}}_M$ can be computed via an iterative procedure [25]. Under standard regularity conditions, the GMM estimator $\hat{\boldsymbol{\beta}}_M$ exists and converges in probability to the true parameter $\boldsymbol{\beta}_0$ [9]. A proof of asymptotic normality of GMM can be found in Page 2148 of [13].

A robust GMM

In this section we introduce the outlier robust GMM estimator. In the following subsection, we specify moment conditions used for robust estimation. And the details on the implementation of the estimator follows.

The generalized method of weighted moments

The main principle used in the robust GMM estimator is that we replace moment conditions by a set of observation weighted moment conditions. Instead of Equation (3), we define

$$u^w(\boldsymbol{\beta}) = w_i X_i(\mathbf{y}_i^* - \boldsymbol{\pi}_i) - c_i, \quad i = 1, ..., n. \tag{6}$$

where $c_i = E\{w_i X_i(\mathbf{y}_i^* - \boldsymbol{\pi}_i)\}$. Then the estimation can be based on the moment conditions

$$E\{u^w(\boldsymbol{\beta})\} = 0.$$

Consequently, the generalized method of weighted moments (GMWM) estimates can be defined by

$$\hat{\beta}^w = \arg\min_{\beta \in \mathcal{B}} Q_n^w(\boldsymbol{\beta}). \tag{7}$$

where

$$Q_n^w(\boldsymbol{\beta}) = [U_n^w(\boldsymbol{\beta})]^T \{\Sigma_n^w(\boldsymbol{\beta})\}^{-1} U_n^w(\boldsymbol{\beta}), \tag{8}$$

with

$$U_n^w(\boldsymbol{\beta}) = \sum_{i=1}^{n} u^w(\boldsymbol{\beta}). \tag{9}$$

Here we take the summation as the sample moment condition. The advantage of using the summation is that it can lead us to a direct estimation of covariance matrix.

It is clear to see that this definition is analogous to the standard GMM. If we choose $w_i = 1$ and $c_i = 0$ for all observations, the moment conditions in (6) are reduced to the standard moment conditions. Therefore, the standard GMM is a special case of the GMWM.

In order to specify the weights for the robust GMM estimator, we need the following definition of a distance, which is based on individual moment conditions:

$$d_i(\boldsymbol{\beta}) = [u_i^w(\boldsymbol{\beta})]^T \{\Sigma_n^w(\boldsymbol{\beta})\}^{-1} u_i^w(\boldsymbol{\beta}), \quad i = 1, ..., n. \tag{10}$$

The weight is assigned based on $d_i(\boldsymbol{\beta})$, that is, $w_d = w\left(d_i(\boldsymbol{\beta})\right)$. There are several alternative specifications of weight functions available in the literature [17, 18]. In this study, the Huber's weights are applied:

$$w\left(d_i(\boldsymbol{\beta})\right) = \min\left(1, \frac{c_d}{d_i(\boldsymbol{\beta})}\right). \tag{11}$$

The above specification of weight function requires a value of the tuning constant c_d. Both the outlier sensitivity and the efficiency of the estimator are determined by the constant. On the one hand, the estimator should be reasonably efficient if the sample contains no outlier. On the other hand, the estimator

should be insensitive to outliers. To determine c_d, understanding the distribution of $d_i(\boldsymbol{\beta})$ is critical. Clearly, $u_i^w(\boldsymbol{\beta})$ is a column vector, and $d_i(\boldsymbol{\beta})$ is a scalar quadratic distance, so we set $c_d = \chi_1^{-2}(0.975)/n$, where $\chi_p^{-2}(\cdot)$ is the quantile of the χ^2 distribution with p degrees of freedom.

If we take the information matrix (5) of the PLRM as $\Sigma_n^w(\boldsymbol{\beta})$, we can computer a leverage for each observation:

$$H_i = X_i\{\Sigma_n^w(\boldsymbol{\beta})\}^{-1}X_i^T\sigma_i^w, \quad i = 1, ..., n. \tag{12}$$

where σ_i^w is the ith component of $\Sigma_n^w(\boldsymbol{\beta})$. Then, a Mallows-type weight can be defined based on $trace(H_i)$; that is, $w_x = w(trace(H_i))$, to downplay the observations with high leverages. [26] suggest that the practical rule for isolating leverage points might set $c_x = 2(p+1)(J-1)/n$. In this study, we give observations with large leverages 0 weights,

$$w_x = w(trace(H_i)) = \begin{cases} 1 & \text{if } trace(H_i) \leq \frac{2(p+1)(J-1)}{n} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

An approach often used to combine the two weights is $w_i = w_d \cdot w_x$. [27].

The consistency correction vector c_i is defined as

$$c_i = \left(w\left(d_i^{(1)}(\boldsymbol{\beta})\right) - w\left(d_i^{(0)}(\boldsymbol{\beta})\right)\right)/diag\left(\Sigma_n^w(\boldsymbol{\beta})\right), \quad i = 1, ..., n.$$

where $w\left(d_i^{(h)}(\boldsymbol{\beta})\right) = w\left(X_i\{h - \pi_i(\boldsymbol{\beta})\}/diag\left[\Sigma_n^w(\boldsymbol{\beta})\right]^{-1}\right)$ with $h = \{0, 1\}$, is the weight for y_i^*.

Implementation of the estimator

The continuous updating estimation method is applied in this study for estimating the regression coefficients and corresponding variance. The procedure is detailed as follows:

1. Apply an initial value $\beta^{(0)}$ for computing $\Sigma_n(\boldsymbol{\beta})$.
2. Compute $d_i(\boldsymbol{\beta})$ using Equation (10) and H_i using Equation (12); assign weights correspondingly based on (11) and (13).
3. With the combined weights, calculate $\Sigma_n^w(\boldsymbol{\beta})$ and $U_n^w(\boldsymbol{\beta})$ in Equation (9).
4. Obtain the estimator $\hat{\boldsymbol{\beta}}_w^{(1)}$ by minimizing Q_n^w of Equation (8).

5. Go back to Step 1, replace $\boldsymbol{\beta}^{(0)}$ with the estimator $\hat{\boldsymbol{\beta}}_w^{(1)}$ in computing $\Sigma_n^w\left(\hat{\boldsymbol{\beta}}_w^{(1)}\right)$, and move to the next iteration.

6. Continue this procedure until convergence criteria are met.

For the starting value $\beta^{(0)}$, a reasonable choice is the MLE estimation based on the original data.

In the appendix, we proved that, under some regularity assumptions, we can have that $\hat{\boldsymbol{\beta}}_w$ is consistent for $\boldsymbol{\beta}_0$. And by studying the behavior of the weighted moment equations in a neighborhood of $\boldsymbol{\beta}_0$, we showed that the asymptotic linearity ensures the applicability of the central limit theorem for the asymptotic normality of GMWM.

Results

Monte Carlo simulations

In this section we investigate the properties of the GMWM estimator using a Monte-Carlo study. We generate data with three response categories and two covariates which are from multivariate normal distribution with 0 mean and identity covariance. The true coefficient matrix $\boldsymbol{\beta}_0$ is

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \beta_{10} & \beta_{20} & \beta_{30} \\ \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{pmatrix} = \begin{pmatrix} 0 & 1.0 & -0.3 \\ 0 & -0.8 & 0.7 \\ 0 & -1.0 & -0.5 \end{pmatrix}.$$

Based on the specified coefficient values and using the probability based on the model (1), we compute the category-specific probabilities for each subject. Then, using the computed probabilities, we determine the most likely category to which each subject belongs. This decision is made through random generation from the multinomial distribution with the probability vector as a parameter. For instance, multinomial categories in R-Language are generated using $rmultinorm(n_i, N_i, \pi(\mathbf{x}_i))$ function, where $\pi(\mathbf{x}_i) = (\pi_1(\mathbf{x}_i), ..., \pi_J(\mathbf{x}_i))$ is the probability vector, n_i is the number of random vectors to draw, and N_i is the total number of objects that are put into J-categories. In our case, $n_i = N_i = 1$ for all subjects and $J = 3$.

Two sample sizes, 100 and 1000, are examined. For each sample size, we run the simulation 1000 times. Average biases and MSEs are calculated and tabulated. Table 3 shows the results from randomly

generated data with no outliers added. When the sample size is small, GMWM will give greater biases on β_{30} and β_{31} compared to the MLE method. For the sample size 1000, biases on these two parameters increase too, but not so obviously. Variances will also be inflated due to the weights we applied.

Table 3. Bias of parameter estimates and MSE from randomly generated data without outliers.

			MLE			GMWM		
n	Parameter	True	Bias	MSE	Coverage	Bias	MSE	Coverage
100	β_{20}	1.0	0.0666	0.1030	0.945	0.0488	0.1986	0.949
	β_{30}	-0.3	-0.0059	0.1206	0.957	**-0.1440**	0.5578	0.952
	β_{21}	-0.8	-0.0654	0.1190	0.938	-0.0513	0.2550	0.961
	β_{31}	0.7	0.0566	0.1892	0.963	**0.2318**	0.5468	0.923
	β_{22}	-1.0	-0.0853	0.1764	0.969	-0.0691	0.2380	0.950
	β_{32}	-0.5	-0.0624	0.1453	0.945	0.0203	0.3195	0.964
1000	β_{20}	1.0	0.0050	0.0087	0.956	0.0043	0.0181	0.962
	β_{30}	-0.3	-0.0055	0.0105	0.984	**-0.0106**	0.0333	0.950
	β_{21}	-0.8	-0.0039	0.0099	0.943	-0.0013	0.0251	0.956
	β_{31}	0.7	0.0081	0.0160	0.968	**0.0162**	0.0401	0.954
	β_{22}	-1.0	-0.0071	0.0145	0.987	-0.0025	0.0258	0.948
	β_{32}	-0.5	-0.0047	0.0122	0.948	0.0041	0.0361	0.947

Outliers are generated from a multivariate normal distribution with the mean vector $= (2, 3)$ and identity covariance $\mathbf{I}_2$. For these outliers, their responses are intentionally misclassified, that is, they are placed within a different category from those predicted categories based on the true parameters.

Table 4 lists simulation results with outliers added. For estimations from datasets with 5% outliers, bias correction from the GMWM is excellent. However, when the datasets have 10% outliers, biases on estimations of some parameters (β_{21} and β_{22} in this simulation) are decreased, but not completely corrected.

Table 4. Comparison between GMWM and MLE estimation from randomly generated data with outliers added.

		5% contamination						10% contamination					
		GMWM			MLE			GMWM			MLE		
Size	Parameter	Bias	MSE	Coverage	Bias	MSE	Coverage	Bias	MSE	Coverage	Bias	MSE	Coverage
100	β_{20}	0.0568	0.1102	0.956	0.0860	0.0884	0.957	0.0489	0.0999	0.971	0.0868	0.0819	0.970
	β_{30}	-0.0038	0.1427	0.954	-0.0055	0.1528	0.949	-0.0057	0.1510	0.945	-0.0431	0.1461	0.814
	β_{21}	-0.0392	0.1464	0.949	0.2377	0.1360	0.785	0.0319	0.1227	0.946	0.3607	0.1933	0.579
	β_{31}	0.0175	0.2020	0.944	-0.1072	0.1270	0.921	-0.0235	0.1770	0.943	-0.1631	0.1283	0.949
	β_{22}	0.0374	0.1207	0.949	0.3848	0.2115	0.578	0.0207	0.0968	0.945	0.6088	0.4151	0.526
	β_{32}	-0.0548	0.1572	0.956	-0.0964	0.0904	0.964	-0.0817	0.1349	0.977	-0.1069	0.0803	0.967
1000	β_{20}	0.0172	0.0189	0.939	0.0490	0.0102	0.932	0.0451	0.0202	0.944	0.0657	0.0120	0.900
	β_{30}	0.0012	0.0340	0.945	0.0124	0.0075	0.952	-0.0071	0.0336	0.952	-0.0111	0.0063	0.822
	β_{21}	0.0260	0.0242	0.937	0.2874	0.0885	0.101	0.0164	0.0207	0.936	0.3876	0.1545	0.002
	β_{31}	-0.0058	0.0356	0.950	-0.1423	0.0345	0.697	-0.0497	0.0346	0.917	-0.2269	0.0658	0.521
	β_{22}	0.0366	0.0237	0.936	0.4390	0.2032	0.000	0.0238	0.0182	0.938	0.6500	0.4322	0.000
	β_{32}	-0.0106	0.0292	0.951	-0.0538	0.0103	0.940	-0.0434	0.0250	0.953	-0.0629	0.0106	0.902

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.299v1 | CC-BY 3.0 Open Access | received: 23 Mar 2014, published: 23 Mar 2014

Application

For the hypertension data, the criterion for identifying observations with large distances is $c_d = 0.22$, and the criterion for identifying leverage points is $c_x = 0.12$. Applying the GMWM estimator, those blue-colored points in Figure 1 are automatically downweighted, and red-colored points have 0 weight. The GMWM method indeed eliminates those inconsistencies: the coefficient of sodium intake for the odds model between the Stage 2 hypertension and the Normal categories is no longer negative, see the right side of Table 2.

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 [21, 22].

Discussion

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.

A datum with a nominal response and some continuous covariates is commonly seen in many scientific areas, such as sociology, economy, and biomedical studies. In order to be able to deal with outliers, we modified the GMM estimator to replace the standard moment conditions with weighted moment conditions, so that aberrant observations automatically receive less weight. We proved that the proposed method has good asymptotic behavior. When outliers are present, the GMWM estimator give much

199 smaller biases than the estimations derived from the traditional MLE method. This method can be
200 adapted to check whether results obtained with the traditional MLE approach are driven only by a few
201 outlying observations. The weights produced from the robust procedure can be used to diagnose the
202 cause of the differences and to indicate routes for model re-specification.

Appendix: Consistency and asymptotic normality

In this appendix, we introduce the assumptions for the asymptotic analysis of GMWM, and outline the derivations on the main asymptotic properties of GMWM.

We make the following sets of regularity assumptions regarding properties of the moment functions and identification assumptions.

Assumption I

I1. $\mathcal{B}$ is a compact parametric space.

I2. Σ is a positive definite matrix.

I3. It holds that $E[u^w(\boldsymbol{\beta})] = 0$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, and for any $\epsilon > 0$, that

$$\inf_{\boldsymbol{\beta} \in \mathbf{B} \setminus \mathcal{N}(\boldsymbol{\beta}_0, \epsilon)} \| E[u^w(\boldsymbol{\beta})] \| > 0$$

where $\mathcal{N}(\boldsymbol{\beta}_0, \epsilon) = \{ \boldsymbol{\beta} \in \mathbb{R}^l \big| \, \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \| < \epsilon \}$ is an open ϵ-neighborhood of a point $\boldsymbol{\beta}_0$.

Assumption F

F1. Let $u^w(\boldsymbol{\beta})$ be continuous in $\boldsymbol{\beta} \in \mathcal{B}$, and be twice differentiable in $\boldsymbol{\beta}$ on $\mathcal{N}(\boldsymbol{\beta}_0, \epsilon)$ almost surely.

F2. Expectation $E \sup_{\boldsymbol{\beta} \in \mathcal{B}} \| u^w(\boldsymbol{\beta}) \|$, $E \sup_{\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0, \epsilon)} \| \partial u^w(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}_k \|$, and $E \sup_{\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0, \epsilon)} \left\| \partial^2 u^w(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_l \right\|$ exists and are finite for $k, l = 1, ..., p$.

Assumption W

W1. $\lim_{\epsilon \to 0} \sup_{\| \Delta \| \leq \epsilon} | w(\boldsymbol{\beta} + \Delta) - w(\boldsymbol{\beta}) | = 0$.

W2. $\lim_{\epsilon \to 0} \sup_{\| \Delta \| \leq \epsilon} | \partial w(\boldsymbol{\beta} + \Delta) / \partial \boldsymbol{\beta} - \partial w(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} | = 0$.

When the above assumptions are met, we can prove that $\hat{\boldsymbol{\beta}}_w$ is consistent for $\boldsymbol{\beta}_0$. We begin with studying the behavior of the weighted moment equations in a neighborhood of $\boldsymbol{\beta}_0$. And proving their asymptotic linearity is followed. The linearity ensures the applicability of the central limit theorem for the asymptotic normality of GMWM.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.299v1 | CC-BY 3.0 Open Access | received: 23 Mar 2014, published: 23 Mar 2014

Theorem 1. *Let the Assumption F and I hold, then the GMWM estimator* $\hat{\boldsymbol{\beta}}_w$ *is asymptotically normal, that is,* $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\right) \xrightarrow{F} N(0, M^T S^w M)$ *as* $n \to \infty$*, where*

$$M = \left((V^w)^T \Sigma V^w\right)^{-1} (V^w)^T \Sigma,$$

$$S^w = E\left[\frac{1}{n}\sum_{i=1}^{n} u^w(\hat{\boldsymbol{\beta}}) u^w(\hat{\boldsymbol{\beta}})^T\right].$$

with $V^w = E\left[\frac{\partial U^w(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}\right]$.

We start with proving two lemmas before we present the proof of Theorem 1.

Lemma 1. *Let the assumption F, I and W hold, and let* $U_r^w(\boldsymbol{\beta})$ *be the* r*th element of the vector* $U^w(\boldsymbol{\beta})$*,* $r = 1, ..., p$*. Then, for* $0 < s < 1$*,*

$$\sup_{\|t\| \le C} \left| \frac{1}{n} \sum_i \sum_{l=1}^{p} t_l \left\{ (\partial/\partial\beta_l) U_{i,r}^w \left(\boldsymbol{\beta} + \frac{st}{\sqrt{n}}\right) - (\partial/\partial\beta_l) U_{i,r}^w(\boldsymbol{\beta}) \right\} \right| = o_p(1). \tag{14}$$

Proof. For $l, r = 1, ..., p$, by differentiating the ith component of $U_r^w(\boldsymbol{\beta})$, we get

$$\frac{\partial U_{i,r}^w(\boldsymbol{\beta})}{\partial \beta_l} = -w(\boldsymbol{\beta}) X_i \frac{\partial \pi_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_l} + \frac{\partial w_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_l} X_i (y_i - \pi_i(\boldsymbol{\beta}))$$

Then,

$$\begin{aligned} &\sup_{\|t\| \le C} \left| \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{p} t_l \left\{ (\partial/\partial\beta_l) U_{i,r}^w \left(\boldsymbol{\beta} + \frac{st}{\sqrt{n}}\right) - (\partial/\partial\beta_l) U_{i,r}^w(\boldsymbol{\beta}) \right\} \right| \\ &\le \frac{C}{n} \sum_{i=1}^{n} \sum_{l=1}^{p} t_l \sup_{\|t\| \le C} \left| (\partial/\partial\beta_l) U_{i,r}^w \left(\boldsymbol{\beta} + \frac{st}{\sqrt{n}}\right) - (\partial/\partial\beta_l) U_{i,r}^w(\boldsymbol{\beta}) \right| \end{aligned}$$

and

$$\begin{aligned}
&\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)U^w_{i,r}\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)U^w_{i,r}(\boldsymbol{\beta})\right| \\
&\le \sup_{\|t\|\le C}\left\{\left|w_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-w_i\left(\boldsymbol{\beta}\right)\right|\left|\left(\partial/\partial\beta_l\right)\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right|\right\} \\
&+\sup_{\|t\|\le C}\left\{\left|(\partial/\partial\beta_l)\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)\pi_i\left(\boldsymbol{\beta}\right)\right||X_i w_i\left(\boldsymbol{\beta}\right)|\right\} \\
&+\sup_{\|t\|\le C}\left\{\left|(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}\right)\right|\left|X_i\left(y_i-\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right)\right|\right\} \\
&+\sup_{\|t\|\le C}\left\{\left|\left(y_i-\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right)-\left(y_i-\pi_i\left(\boldsymbol{\beta}\right)\right)\right|\left|(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}\right)\right|\right\}.
\end{aligned}$$

Then, by taking expectation at both sides,

$$\begin{aligned}
&E\left\{\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)U^w_{i,r}\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)U^w_{i,r}(\boldsymbol{\beta})\right|\right\} \\
&\le \sup_{\|t\|\le C}\left|w_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-w_i\left(\boldsymbol{\beta}\right)\right|\sup_{\|t\|\le C}\left|\left(\partial/\partial\beta_l\right)\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right| \\
&+\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)\pi_i\left(\boldsymbol{\beta}\right)\right|\sup_{\|t\|\le C}|X_i w_i\left(\boldsymbol{\beta}\right)| \\
&+\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}\right)\right|E\left\{\sup_{\|t\|\le C}\left|X_i\left(y_i-\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right)\right|\right\} \\
&+\sup_{\|t\|\le C}\left|\left(y_i-\pi_i\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)\right)-\left(y_i-\pi_i\left(\boldsymbol{\beta}\right)\right)\right|\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)w_i\left(\boldsymbol{\beta}\right)\right|.
\end{aligned}$$

Thus, by conditions F and W, we have

$$E\left\{\sup_{\|t\|\le C}\left|(\partial/\partial\beta_l)U^w_{i,r}\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)U^w_{i,r}(\boldsymbol{\beta})\right|\right\}\longrightarrow 0,\quad \forall i.$$

and

$$E\left\{\sup_{\|t\|\le C}\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{p}t_l\left\{(\partial/\partial\beta_l)U^w_{i,r}\left(\boldsymbol{\beta}+\frac{st}{\sqrt{n}}\right)-(\partial/\partial\beta_l)U^w_{i,r}(\boldsymbol{\beta})\right\}\right|\right\}\longrightarrow 0,\quad \forall i.$$

Therefore, we have the results in (14). □

Lemma 2. *Let the Assumption F, I and W hold, it holds that*

$$\frac{1}{\sqrt{n}}\sup_{\|t\|\le C}\left\|U^w_n(\boldsymbol{\beta}_0+n^{-\frac{1}{2}}t)-U^w_n(\boldsymbol{\beta}_0)+V^w n^{-\frac{1}{2}}t\right\|=o_p(1), \tag{15}$$

as $n \to \infty$, *where* $V^w = E\left[\frac{\partial U^w(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right]$.

Proof. Write

$$U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - U_n^w(\boldsymbol{\beta}_0) = \sum_{i=1}^n w_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t)u_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - \sum_{i=1}^n w_i(\boldsymbol{\beta}_0)u_i(\boldsymbol{\beta}_0)$$

By the Taylor expansion, $u_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) = u_i(\boldsymbol{\beta}_0) + n^{-\frac{1}{2}}t\left[\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0 + \frac{t}{\sqrt{n}})\right]$, where $0 < s < 1$. Then, we can write

$$\begin{aligned}
&U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - U_n^w(\boldsymbol{\beta}_0) \\
&= \sum_{i=1}^n u_i(\boldsymbol{\beta}_0)\left\{w_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - w_i(\boldsymbol{\beta}_0)\right\} && (16)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i(\boldsymbol{\beta}_0)t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0) && (17)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{w_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - w_i(\boldsymbol{\beta}_0)\right\}t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0) && (18)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i\left(\boldsymbol{\beta}_0 + \frac{t}{\sqrt{n}}\right)t\left\{\frac{\partial u_i\left(\boldsymbol{\beta}_0 + \frac{t}{\sqrt{n}}\right)}{\partial \boldsymbol{\beta}} - \frac{\partial u_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right\} && (19)
\end{aligned}$$

We will now show that terms (16), (18) and (19) are asymptotically negligible. As to the term (16), By Assumption W1, $\{w_i(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}}t) - w_i(\boldsymbol{\beta}_0)\} \to 0$, and $u_i(\boldsymbol{\beta}_0)$ is independent of $\boldsymbol{\beta}$. So we have the term (16) tends to zero. Similarly, $\partial/\partial\boldsymbol{\beta}u_i(\boldsymbol{\beta}_0)$ is independent of $\boldsymbol{\beta}$ and t is bounded. Hence, the term (18) tends to zero. Lemma 1 implies $\frac{\partial u_i\left(\boldsymbol{\beta}_0 + \frac{t}{\sqrt{n}}\right)}{\partial \boldsymbol{\beta}} - \frac{\partial u_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \to 0$, as $n \to \infty$. So the term (19) can be neglect too.

Now, let us analyze the term (17). Let $w_i^*(\boldsymbol{\beta}_0)$ be the limit of $w_i(\boldsymbol{\beta}_0)$. Rewrite (17) as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i(\boldsymbol{\beta}_0)t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[w_i(\boldsymbol{\beta}_0) - w_i^*(\boldsymbol{\beta}_0)\right]t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0) && (20)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{w_i^*(\boldsymbol{\beta}_0)t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0) - E\left[w_i^*(\boldsymbol{\beta}_0)t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0)\right]\right\} && (21)\\
&+ \frac{1}{\sqrt{n}}E\left[w_i^*(\boldsymbol{\beta}_0)t\frac{\partial}{\partial \boldsymbol{\beta}}u_i(\boldsymbol{\beta}_0)\right] && (22)
\end{aligned}$$

The first term (20) is negligible because $\partial/\partial\boldsymbol{\beta}u_i(\boldsymbol{\beta}_0)$ is independent of $\boldsymbol{\beta}$, t is bounded, and $[w_i(\boldsymbol{\beta}_0) - w_i^*(\boldsymbol{\beta}_0)] \to$

0. By the central limit theorem, each element of vector (21) converges in distribution to a normally distribution random variable with zero mean and a finite variance which is uniformly bounded by t. Hence, (21) is bounded in probability. The last term (20) is

$$\frac{1}{\sqrt{n}} E\left[w_i^*(\boldsymbol{\beta}_0) t \frac{\partial}{\partial \boldsymbol{\beta}} u_i(\boldsymbol{\beta}_0)\right] = \frac{t}{\sqrt{n}} V^w$$

This proves the lemma. □

Proof of Theorem 1. Since $t_n = \sqrt{n} = \mathcal{O}_p(1)$ as $n \to \infty$ by Lemma 2, we can write (15) as

$$U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}} t_n) - U_n^w(\boldsymbol{\beta}_0) + V^w n^{-\frac{1}{2}} t_n = o_p(n^{-\frac{1}{2}}). \tag{23}$$

with a probability arbitrarily close to one uniformly in $t_n \in \{t : \|t\| \leq C\}$. Moreover, with $n^{-\frac{1}{2}} t_n = o_p(1)$, $\partial U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}} t_n)/\partial \boldsymbol{\beta} \to V^w$ in probability as $n \to \infty$.

Note that the first order conditions of GMWM equal to 0, that is,

$$\frac{\partial Q_n^w(\boldsymbol{\beta}_w)}{\partial \boldsymbol{\beta}} = \left[\frac{\partial U_n^w(\boldsymbol{\beta}_w)}{\partial \boldsymbol{\beta}}\right]^T \Sigma(\boldsymbol{\beta}_w) U_n^w(\boldsymbol{\beta}_w) = 0,$$

Replace $U_n^w(\boldsymbol{\beta}_w)$ with $U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}} t_n)$ from Equation (23),

$$\begin{aligned} & \left[\frac{\partial U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}} t_n)}{\partial \boldsymbol{\beta}}\right]^T \Sigma(\boldsymbol{\beta}_w) U_n^w(\boldsymbol{\beta}_0 + n^{-\frac{1}{2}} t_n) \\ = \; & \left[V^w + o_p(1)\right]^T \Sigma(\boldsymbol{\beta}_w) \left[U_n^w(\boldsymbol{\beta}_0) - V^w n^{-\frac{1}{2}} t_n\right] = 0. \end{aligned}$$

Then we have

$$t_n = \sqrt{n}(\boldsymbol{\beta}_w - \boldsymbol{\beta}_0) = \sqrt{n}\left[(V^w)^T \Sigma(\boldsymbol{\beta}_w) V^w\right]^{-1} (V^w)^T \Sigma(\boldsymbol{\beta}_w) U_n^w(\boldsymbol{\beta}_0) + o_p(1). \tag{24}$$

Next we examine the behavior of $\sqrt{n}U_n^w(\boldsymbol{\beta}_0)$, which can be written as

$$\begin{aligned}\sqrt{n}U_n^w(\boldsymbol{\beta}_0) &= n^{-\frac{1}{2}}\sum_{i=1}^{n} u_i(\boldsymbol{\beta}_0)w_i(\boldsymbol{\beta}_0) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n} u_i(\boldsymbol{\beta}_0)\left\{w_i(\boldsymbol{\beta}_0) - w_i^*(\boldsymbol{\beta}_0)\right\} \quad (25) \\ &+ n^{-\frac{1}{2}}\sum_{i=1}^{n} u_i(\boldsymbol{\beta}_0)w_i^*(\boldsymbol{\beta}_0). \quad (26)\end{aligned}$$

Note that the term (25) is asymptotically negligible in probability due the the triangle inequality and Assumption W1. The term (26) is stationary sequence of absolutely random variables. By Assumption I3 and F2, (26) have zero mean and finite second moments. So the central limit theorem can be applied on (26), giving $\sqrt{n}U_n^w(\boldsymbol{\beta}_0) \sim N(0, S^w)$ [28, Section 25.3].

With Equation (24), we have asymptotic normality of $\boldsymbol{\beta}^w$, and its asymptotic variance is given by $M^T S^w M$ [28]. □

References

1. McCullagh P, Nelder JA (1989) Generalized Linear Models, Second Edition (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). Chapman and Hall/CRC.

2. Liu I, Agresti A (2005) The analysis of ordered categorical data: An overview and a survey of recent developments. TEST: An Official Journal of the Spanish Society of Statistics and Operations Research 14: 1-73.

3. Pregibon D (1982) Resistant fits for some commonly used logistic models with medical applications. Biometrics 38: 485-498.

4. Copas JB (1988) Binary regression models for contaminated data. Journal of the Royal Statistical Society Series B (Methodological) 50: 225-265.

5. Victoria-Feser M, Ronchetti E (1997) Robust estimation for grouped data. Journal of the American Statistical Association 92: 333–340.

6. Gupta A, Kasturiratna D, Nguyen T, Pardo L (2006) A new family of ban estimators for polytomous logistic regression models based on ϕ-divergence measures. Statistical Methods & Applications 15: 159-176.

7. Flores, Garrido (2001) Robust logistic regression for insurance risk classification. Working Paper 01-64, Universidad Carlos III de Madrid.

8. Mebane J W R, Sekhon JS (2004) Robust estimation and outlier detection for overdispersed multinomial models of count data. American Journal of Political Science 48: 392–411.

9. Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50: 1029-1054.

10. Newey WK, West KD (1987) Hypothesis testing with efficient method of moments estimation. International Economic Review 28: 777-787.

11. Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. Econometrica 57: 1027-1057.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.299v1 | CC-BY 3.0 Open Access | received: 23 Mar 2014, published: 23 Mar 2014

12. Hansen LP, Heaton J, Yaron A (1996) Finite-sample properties of some alternative gmm estimators. Journal of Business & Economic Statistics 14: 262-280.

13. Newey WK, McFadden D (1986) Large sample estimation and hypothesis testing, In: Handbook of Econometrics: vol. 4, 1st edn. Elsevier.

14. Hayashi F (2000) Econometrics. Princeton University Press.

15. Baum CF, Schaffer ME, Stillman S (2002) Instrumental variables and gmm: Estimation and testing. Boston College Working Papers in Economics 545, Boston College Department of Economics.

16. Ronchetti E, Trojani F (2001) Robust inference with gmm estimators. Journal of Econometrics 101: 37-69.

17. Huber PJ (1981) Robust Statistics (Wiley Series in Probability & Mathematical Statistics). Wiley-Interscience.

18. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust Statistics: The Approach Based on Influence Functions (Wiley Series in Probability and Statistics). Wiley.

19. Rousseeuw PJ, Leroy AM (1987) Robust Regression and Outlier Detection (Wiley Series in Probability and Statistics). Wiley.

20. Kuersteiner GM (2012) Kernel-weighted GMM estimators for linear time series models. Journal of Econometrics 170: 399-421.

21. Institute of Medicine (2005) Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. The National Academies Press.

22. He FJ, MacGregor GA (2004) Effect of longer-term modest salt reduction on blood pressure. Cochrane database of systematic reviews Online 3: CD004937.

23. Agresti A (2002) Categorical Data Analysis (Wiley Series in Probability and Statistics). Wiley-Interscience.

24. Stokes ME, Davis CS, Koch GG (2001) Categorical Data Analysis Using The SAS System. WA (Wiley-SAS).

25. Hansen LP, Heaton J, Yaron A (1996) Finite-sample properties of some alternative gmm estimators. Journal of Business & Economic Statistics 14: 262-280.

26. Lesaffre E, Albert A (1989) Multiple-group logistic regression diagnostics. Journal of the Royal Statistical Society Series C (Applied Statistics) 38: 425-440.

27. Heritier S, Cantoni E, Copt S, Victoria-Feser MP (2009) Robust Methods in Biostatistics (Wiley Series in Probability and Statistics). Wiley.

28. Davidson J (1994) Stochastic Limit Theory: An Introduction for Econometricicans (Advanced Texts in Econometrics). Oxford University Press, USA.