

Monte Carlo Evaluation of Consistency and Normality of Dichotomous Logistic and Multinomial Logistic Regression Models

Naima Shifa¹ & Mamunur Rashid²

¹Department of Mathematics, DePauw University
Greencastle, IN USA 46135, naimashifa@depauw.edu

²Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis
Indianapolis, IN USA 46202, mrashid@iupui.edu

Abstract

The dichotomous logistic regression model is one of the popular mathematical models for the analysis of binary data with applications in physical, biomedical, and behavioral sciences, among others. The feature of this model is to quantify the effects of several explanatory variables on one dichotomous outcome variable. Multinomial logistic regression model, on the other hand, handles the categorical dependent outcome variable with more than two levels. Normally, the asymptotic properties of the maximum likelihood estimates for the model parameters are used for statistical inference, for example, normality allows one to compute the confidence interval and perform statistical tests in a manner analogous to the analysis of linear multiple regression models, provided the sample size is large. However, asymptotic properties of the maximum likelihood (ML) estimator in logistic models had been studied earlier, see, for example, Gourieroux and Monfort (1981) and Amemiya (1985), and different results have been established. As none of the authors verified their work via the Monte Carlo simulation study, this research article performs an extensive Monte Carlo simulation study to examine consistency and normality of the maximum likelihood estimators for parameters of both the dichotomous logistic and multinomial logistic regression models.

Key words

Logistic regression; Consistency; Maximum Likelihood Estimator; Monte Carlo simulation.

1. Introduction

Logistic regression analysis is a statistical modeling method for analyzing a categorical outcome variable. This statistical model describes the relationship between a categorical response variable and a set of explanatory variables. The response variable in logistic regression model is usually dichotomous, but more than two response options can be modeled using multinomial or polytomous logistic regression model. Cramer (2003) discussed an overview of the development of the logistic regression model, and he identifies three sources that had a profound impact on the

model: applied mathematics, experimental statistics, and economic theory. Agresti (2002) also provided in detail of the development on logistic regression in different areas. The maximum likelihood estimation (MLE) is the most widely-used general method of estimation procedures and is treated as a standard approach to parameter estimation and inference in statistics (van der Vaart, 1998). The MLE has good asymptotic (large sample) properties for the estimates. Under very general conditions, maximum likelihood estimates are consistent, asymptotically efficient, and asymptotically-normally distributed. Notice that this normality allows one to compute the confidence interval and perform statistical tests in a manner analogous to the analysis of linear multiple regression models, provided the sample size is large. However, asymptotic properties of the maximum likelihood (ML) estimator in logistic models had been studied earlier, see, for example, Gourieroux and Monfort (1981) and Amemiya (1985), and different results have been established. For example, different proofs of consistency can be found in the literature such as Beer (2005), Gourieroux and Monfort (1981), and Amemiya (1985). All of them are based upon the fact that the probability of the existence of the estimators approaches one as sample size tends to infinity. Furthermore, they proceed on the assumption that the number of explanatory variables is fixed. In other words, the number of variables is compelled to remain constant while the sample size increases. Another result presented by Beer (2001) enables us to relax the former condition. It allows for any number of variables, but depends on sample size, and examines the relationship between the number of variables and sample size that is necessary to preserve the consistency of the estimators. However, it needs to be pointed out that none of the authors cited above verified their work via the Monte Carlo simulation study. Gourieroux and Monfort (1981) note, “it should be stressed that all these asymptotic results give little indication on the properties of the estimators in finite sample, and it would be interesting to clarify this point by means of Monte Carlo studies.” The purpose of this paper is to provide an extensive standard Monte Carlo simulation study to show the consistency and asymptotic normality of the ML estimators of the logistic and multinomial logistic regression models.

In section 2 we describe the models and estimation methods in the binary logistic and polytomous logistic regression models. Section 3 provides the simulation results for both models, and section 4 discusses a summary.

2. The models and estimation methods for the binary and multinomial logistic regression models

2.1 The binary logistic regression model

Suppose a binary random variable y follows a Bernoulli distribution, that is, y takes either the value 1 or the value 0 with probabilities $\pi(x)$ or $1 - \pi(x)$ respectively, where

$x = (x_1, x_2, \dots, x_p) \in \mathfrak{R}^p$ is a vector of p explanatory variables. In fact, $\pi(x)$ represents the conditional probability $P(y=1|x)$ of $y=1$, given x . Based on the binary outcome variable, we use the logistic distribution (see, for example, Cox and Snell, 1989; Hosmer and Lameshow, 2000). The specific form of the logistic regression model with unknown parameters $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ and $x_0 = 1$ becomes

$$\pi(x) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}} \quad (1)$$

A transformation of $\pi(x)$ is called the *logit transformation*, and is given by

$$\text{logit } \pi(x) = \ln \frac{\pi(x)}{1-\pi(x)}$$

Thus equation (1) can be rewritten as

$$\text{logit } \pi(x) = x^T \beta \quad (2)$$

Suppose we have a sample of n independent observations $\{(y_i, x_i)\}_{i \in \{1, 2, \dots, n\}} \in (\{0, 1\} \times \mathfrak{R}^{p+1})^n$, where y_i denotes the value of a dichotomous outcome variable, and x_i is the value of the explanatory variables for the i th subject. The likelihood function to find the ML estimator of β is

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Therefore, the log-likelihood function yields, $\ell(\beta) = \sum_{i=1}^n y_i x_i^T \beta - \sum_{i=1}^n \ln(1 + e^{x_i^T \beta})$

The first derivative of the log-likelihood function gives the gradient as follows

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^n (y_i - \mu_i) x_{ij}, \text{ where } \mu_i = E(y_i) = \pi_i$$

The second and third derivatives are, respectively, as follows

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n \pi_i (1 - \pi_i) x_{ij} x_{ik}$$

$$\frac{\partial^3 \ell(\beta)}{\partial \beta_j \partial \beta_k \partial \beta_r} = - \sum_{i=1}^n \pi_i (1 - \pi_i) (1 - 2\pi_i) x_{ij} x_{ik} x_{ir}$$

$$\text{Let } y = \begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}, \text{ and } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$$

Notice that y and μ are $n \times 1$, X is $n \times (p+1)$, and the elements of μ are non-linear functions of an assumed value for β . Also, we define

$$W = \text{diag}(\pi_i(1 - \pi_i))$$

which is $n \times n$. Then, we can write the gradient

$$\ell'(\beta) = \frac{\partial \ell(\beta)}{\partial \beta_j} = X^T (y - \mu)$$

and the Hessian matrix

$$\ell''(\beta) = \frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_k} = -X^T W X.$$

Since $\ell''(\beta)$ is negative semi-definite, the log-likelihood, ℓ , is a concave function of the parameter β ; several optimization techniques are available for finding the maximizing

parameters (see, for example, Mak, 1993; Givens and Hoeting, 2005). We use the Newton-Raphson algorithm for maximizing ℓ . For one step of the Newton-Raphson, we use $\beta^{(t)}$, the current estimate of β , to calculate $\mu^{(t)}$ and $W^{(t)}$. The new estimate of β is then

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W^{(t)} X)^{-1} X^T (y - \mu^{(t)}).$$

This process is repeated until the estimates stop changing, that is, until $\beta^{(t+1)}$ is sufficiently close to $\beta^{(t)}$, then we say the Newton-Raphson method converges. The value at which the Newton-Raphson method converges is the estimate of parameter vector $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$. To better understand what ensures convergence, we must carefully analyze the errors at successive steps. This can be shown by using the following theorem, a notation and terminology that differs slightly from that of the theorem discussed by Givens and Hoeting (2005).

Theorem 1.1 If $l'''(\beta)$ is continuous and β^* is a simple root of $\ell'(\beta)$, then there exists a neighborhood of β^* for which Newton-Raphson method converges to β^* when started from any $\beta^{(t)}$, $t = 0, 1, 2, \dots$ in that neighborhood.

2.2 The multinomial logistic regression model

According to Hosmer and Lameshow (2000), the polytomous logistic model could be extended by any number of levels (or categories) of the outcome variable, but the details of the model would be most understandable if the outcome variable has three categories. This is because the generalization to more than three categories is a problem more of notation than of concept. Following Hosmer and Lameshow (2000), in this article we consider only the situation where the outcome variable has three levels. Let Y be a categorical response variable with three categories, codes as 1, 2, or 3. Since the outcome variable has three categories, we need two logit models as the logistic regression model uses for a binary outcome variable which parameterizes in terms of the logit $Y = 1$ versus $Y = 0$. We assume there are p explanatory variables, $x = (x_1, x_2, \dots, x_p)$, in the model. The logit models for nominal responses pair each response category to a baseline category and the choice is arbitrary. If the last category is the baseline, then the baseline-category logits are

$$\begin{aligned} \ln \left[\frac{P(Y = 1 | x)}{P(Y = 3 | x)} \right] &= \beta_{10} + \beta_{11}x_1 + \dots + \beta_{1p}x_p = \beta_1'x \\ \ln \left[\frac{P(Y = 2 | x)}{P(Y = 3 | x)} \right] &= \beta_{20} + \beta_{21}x_1 + \dots + \beta_{2p}x_p = \beta_2'x \end{aligned}$$

The conditional likelihood function given the covariates for sample of n independent observations is

$$L(\beta) = \prod_{i=1}^n \left[\left(\frac{e^{\beta_1'x_i}}{1 + e^{\beta_1'x_i} + e^{\beta_2'x_i}} \right)^{y_{1i}} \left(\frac{e^{\beta_2'x_i}}{1 + e^{\beta_1'x_i} + e^{\beta_2'x_i}} \right)^{y_{2i}} \left(\frac{1}{1 + e^{\beta_1'x_i} + e^{\beta_2'x_i}} \right)^{y_{3i}} \right]$$

Taking *log* on both sides we have,

$$\ell(\beta) = \sum_{i=1}^n \left[y_{1i} \beta'_1 x_i + y_{2i} \beta'_2 x_i - \ln(1 + e^{\beta'_1 x_i} + e^{\beta'_2 x_i}) \right] \text{ as } \sum_{j=1}^3 y_{ji} = 1 \text{ for each } i,$$

The maximum likelihood estimators are obtained by taking the first partial derivatives of $\ell(\beta)$ with respect to each of the unknown parameters and setting these equations equal to zero. As nonlinear equations, we use similar iterative procedures like Newton-Raphson method. The Hessian matrix is calculated to obtain the estimator of the covariance matrix of the ML estimator, which is the inverse of the observed information matrix. Again, the estimates of the parameters and variance covariance matrix can be obtained by any standard statistical computer packages like SAS, SPSS, and R (nnet package).

3. The simulations results for consistency and normality of the binary logistic and multinomial logistic regression models

3.1.1 The binary logistic regression model: Consistency of the ML estimators

We now assess, via standard Monte Carlo simulation, the finite sample performance of consistency of the maximum likelihood estimators of the logistic regression model. In our simulation study, we consider four explanatory variables $x_1, x_2, x_3,$ and $x_4,$ which are fixed and the binary response variable $y,$ which is treated as a random variable in the logistic model. For fixed values of the intercept parameter β_0 and four other parameters $\beta_1, \beta_2, \beta_3,$ and $\beta_4,$ our aim is to compare the performance of the values of parameters and their standard errors when sample size increases. For fixed values of $\beta_0 = 0.7, \beta_1 = 1.0, \beta_2 = 1.3, \beta_3 = 0.25,$ and $\beta_4 = 0.05,$ the logistic regression model becomes

$$\pi(x) = \frac{e^{0.7 + 1.0 x_1 + 1.3 x_2 + 0.25 x_3 + 0.05 x_4}}{1 + e^{0.7 + 1.0 x_1 + 1.3 x_2 + 0.25 x_3 + 0.05 x_4}}$$

In the simulation, we consider sample sizes of $n = 50, 100, 150,$ and 200 and generate 1,000 independent sets of random samples for each different sample size. For each set of random sample with a particular sample size, we estimate $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 and their standard errors based on the logistic regression model. The final estimates and standard errors of $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 are the average of 1,000 estimates of $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 for that particular sample size. The following table gives the results of the simulation study for different sample sizes.

Table 1. Estimated parameter values and their standard errors using the logistic regression model for different sample sizes of 50, 100, 150, and 200.

Parameters	$n = 50$		$n = 100$		$n = 150$		$n = 200$	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	1.236	0.132	0.864	0.043	0.736	0.017	0.747	0.015
β_1	2.644	0.184	1.263	0.058	1.084	0.026	1.082	0.025
β_2	4.143	0.225	1.759	0.081	1.461	0.041	1.382	0.025
β_3	1.030	0.159	0.320	0.041	0.252	0.017	0.263	0.015
β_4	0.380	0.147	0.016	0.044	0.060	0.017	0.045	0.015

SE=Simulation standard error

As seen in the above table, for sample size $n = 50$, the estimated values of parameters are different from the true values ($\beta_0 = 0.7$, $\beta_1 = 1.0$, $\beta_2 = 1.3$, $\beta_3 = 0.25$, and $\beta_4 = 0.05$), and also the standard errors become larger. However, when the sample size increases from $n = 50$ to $n = 200$, the estimated values of the parameters β_0 , β_1 , β_2 , β_3 , and β_4 are very close to the true values, and the standard errors of the estimates are noticeably smaller.

3.1.2 The binary logistic regression model: Normality of the ML estimators

In this section, we illustrate the large sample behavior of the estimated parameters

$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)^T$. Specifically, we want to show

$$\sqrt{n} (\hat{\beta} - \beta) \xrightarrow{L} N(0, [I(\beta)]^{-1}) \quad (3)$$

Where, $I(\beta) = -E_{\beta}$

$$\begin{bmatrix} \frac{\partial^2 \log L}{\partial \beta_0^2} & \frac{\partial^2 \log L}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \log L}{\partial \beta_0 \partial \beta_2} & \frac{\partial^2 \log L}{\partial \beta_0 \partial \beta_3} & \frac{\partial^2 \log L}{\partial \beta_0 \partial \beta_4} \\ \frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \log L}{\partial \beta_1^2} & \frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_3} & \frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_4} \\ \frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_0} & \frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 \log L}{\partial \beta_2^2} & \frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_3} & \frac{\partial^2 \log L}{\partial \beta_2 \partial \beta_4} \\ \frac{\partial^2 \log L}{\partial \beta_3 \partial \beta_0} & \frac{\partial^2 \log L}{\partial \beta_3 \partial \beta_1} & \frac{\partial^2 \log L}{\partial \beta_3 \partial \beta_2} & \frac{\partial^2 \log L}{\partial \beta_3^2} & \frac{\partial^2 \log L}{\partial \beta_3 \partial \beta_4} \\ \frac{\partial^2 \log L}{\partial \beta_4 \partial \beta_0} & \frac{\partial^2 \log L}{\partial \beta_4 \partial \beta_1} & \frac{\partial^2 \log L}{\partial \beta_4 \partial \beta_2} & \frac{\partial^2 \log L}{\partial \beta_4 \partial \beta_3} & \frac{\partial^2 \log L}{\partial \beta_4^2} \end{bmatrix}$$

For different sample sizes of $n = 100, 250, 500$, we calculate the equation (3) and repeat it 1,000 times. The results are presented below in Figure 1 through the quantile-normal graphs of $\hat{\beta}$.

Figure 1. Monte Carlo simulation of finite sample behavior of normality for the parameters $\beta_1, \beta_2, \beta_4$ and β_4 based on simulation size 1000 and sample sizes 100, 250, and 500.

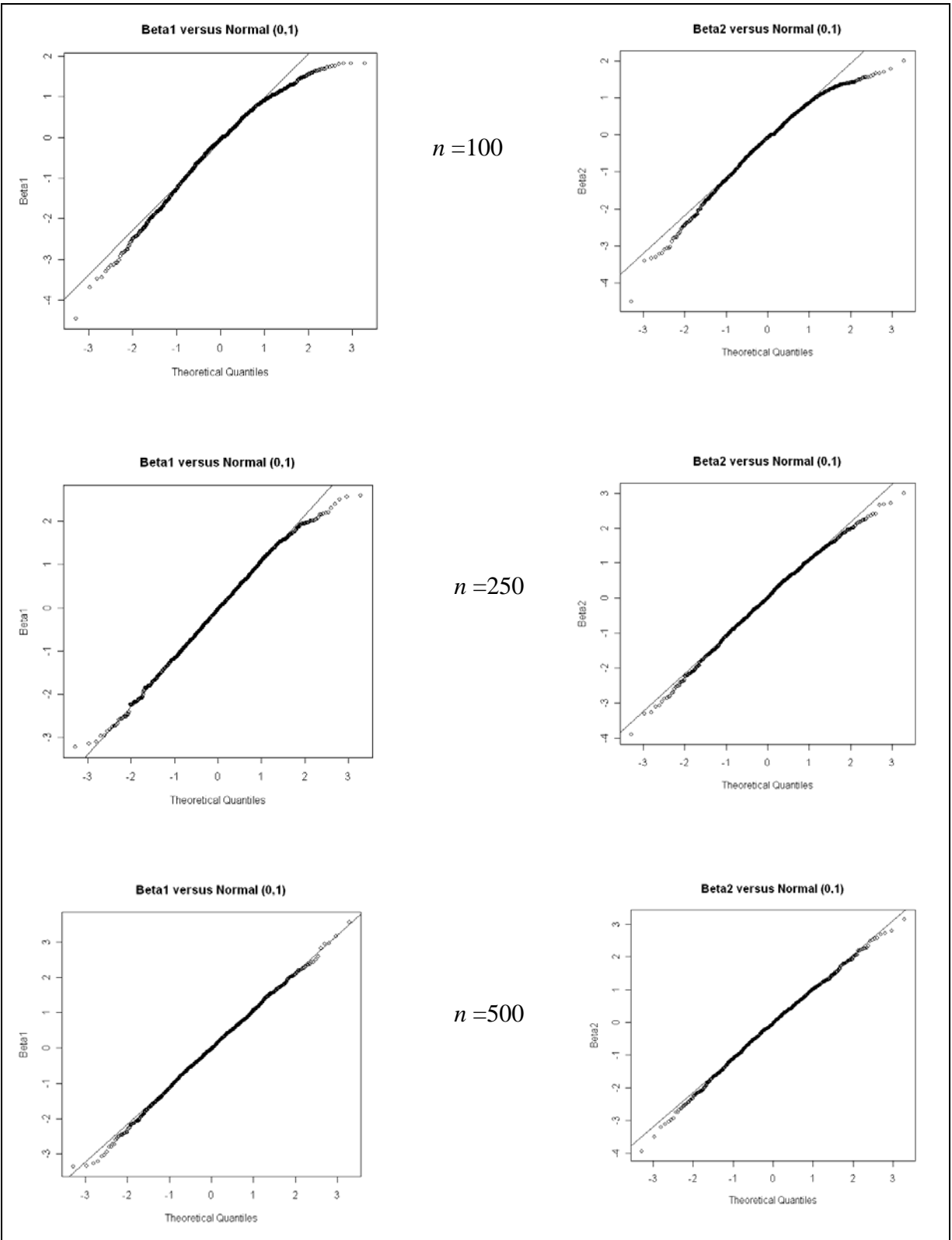
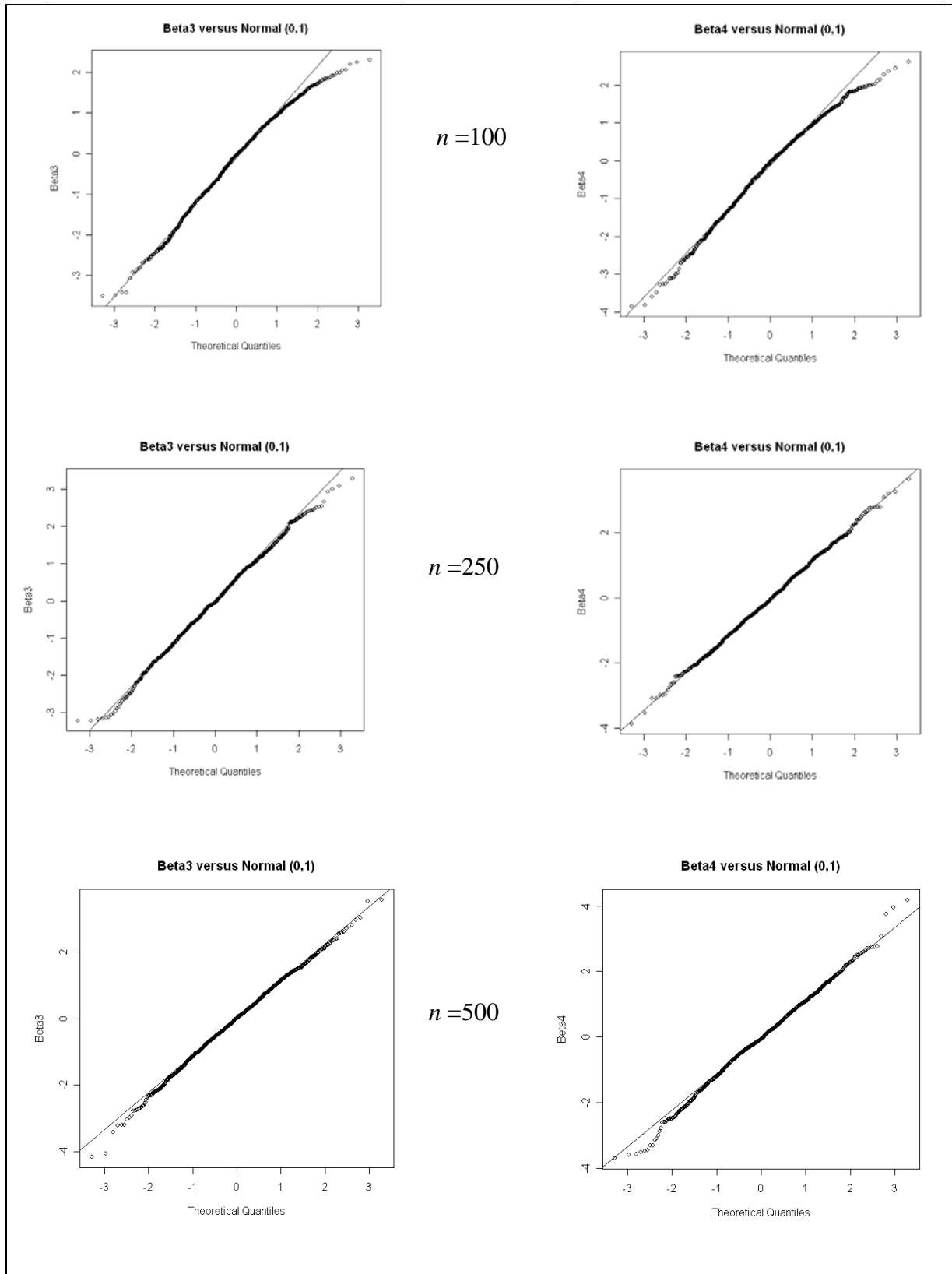


Figure 1 (Continued)



3.2.1 The multinomial logistic regression model: Consistency of the ML estimators

In this section, we show the consistency of the ML estimators for the multinomial logistic regression model via standard Monte Carlo simulation. In this case, we consider the outcome variable Y is random and has three categories, that is, Y takes values coded as 1, 2, and 3. We assume that there are four explanatory variables x_1 , x_2 , x_3 , and x_4 in the model, where each of them is a vector and takes two possible values coded as 0 or 1. If we treat the last category of the outcome variable as the baseline, then the multinomial logistic regression model can be written as

$$\ln \left[\frac{P(Y = 1 | x_1, x_2, x_3, x_4)}{P(Y = 3 | x_1, x_2, x_3, x_4)} \right] = \beta_{01} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \beta_{14}x_4$$

$$\ln \left[\frac{P(Y = 2 | x_1, x_2, x_3, x_4)}{P(Y = 3 | x_1, x_2, x_3, x_4)} \right] = \beta_{02} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \beta_{24}x_4$$

Under these models, the response probabilities are

$$P(Y = 1 | x_1, x_2, x_3, x_4) = \frac{e^{\beta_{01} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \beta_{14}x_4}}{1 + e^{\beta_{01} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \beta_{14}x_4} + e^{\beta_{02} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \beta_{24}x_4}} \quad (4)$$

$$P(Y = 2 | x_1, x_2, x_3, x_4) = \frac{e^{\beta_{02} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \beta_{24}x_4}}{1 + e^{\beta_{01} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \beta_{14}x_4} + e^{\beta_{02} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \beta_{24}x_4}} \quad (5)$$

$$P(Y = 3 | x_1, x_2, x_3, x_4) = \frac{1}{1 + e^{\beta_{01} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \beta_{14}x_4} + e^{\beta_{02} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \beta_{24}x_4}} \quad (6)$$

For the above models, we estimate the unknown parameters $\beta_{01}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{02}, \beta_{21}, \beta_{22}, \beta_{23}$ and β_{24} . The purpose is to show that if the number of observations $(y_i, x_{1i}, x_{2i}, x_{3i}, x_{4i})$, $i = 1, 2, \dots, n$ increases, then the estimates of the parameters converge to their true values. Now, we simulate the values of the outcome and explanatory variables. As the explanatory variables are fixed, the variables x_1, x_2, x_3 , and x_4 are created based on the binomial distribution for arbitrary number of sample size. Once the variables x_1, x_2, x_3 , and x_4 are in hand, we calculate probabilities for the outcome variable based on the above equations (4), (5), and (6). These probabilities are used to simulate the data for Y from the multinomial distribution as Y exceeds more than two categories (actually, in this case it would be trinomial since Y has only three categories). For standard Monte Carlo simulation, we consider sample sizes of $n = 200, 500$, and $1,000$. For the arbitrary fixed values of $\beta_{10} = 0.4, \beta_{11} = 0.8, \beta_{12} = 1.3, \beta_{13} = -0.5, \beta_{14} = 1.1, \beta_{20} = 1.2, \beta_{21} = 1.5, \beta_{22} = 0.9, \beta_{23} = 0.2$, and $\beta_{24} = -0.5$, we generate 1,000 independent sets of random samples for each different sample sizes. Then we estimate $\beta_{01}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{02}, \beta_{21}, \beta_{22}, \beta_{23}$, and β_{24} based on the average of 1,000 estimates of $\beta_{01}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{02}, \beta_{21}, \beta_{22}, \beta_{23}$,

and β_{24} , which are estimated from the simultaneously fitted multinomial logistic regression model, and so are their standard errors for each estimated parameter. The results of the simulation study are provided in the following table.

Table 2. Estimated parameter values and their standard errors using the multinomial logistic regression model for different sample sizes of 200, 500, and 1,000.

Estimated parameter	$n = 200$		$n = 500$		$n = 1,000$	
	Estimate	SE	Estimate	SE	Estimate	SE
$\hat{\beta}_{10}$	0.462	0.026	0.437	0.013	0.420	0.010
$\hat{\beta}_{11}$	0.884	0.027	0.804	0.013	0.811	0.009
$\hat{\beta}_{12}$	2.403	0.095	1.365	0.018	1.344	0.011
$\hat{\beta}_{13}$	-0.494	0.026	-0.497	0.0130	-0.505	0.010
$\hat{\beta}_{14}$	1.171	0.028	1.113	0.013	1.109	0.010
$\hat{\beta}_{20}$	1.263	0.024	1.235	0.012	1.231	0.010
$\hat{\beta}_{21}$	1.637	0.026	1.520	0.013	1.507	0.009
$\hat{\beta}_{22}$	2.012	0.095	0.967	0.017	0.942	0.012
$\hat{\beta}_{23}$	0.233	0.025	0.204	0.012	0.200	0.009
$\hat{\beta}_{24}$	-0.475	0.027	-0.506	0.013	-0.504	0.009

SE: Simulated standard error

3.2.2 The multinomial logistic regression model: Normality of the ML estimators

In this section, we show the large sample behavior of ML estimators of the parameters for the multinomial logistic regression model; that is, we show that the ML estimators of the parameters follow approximately normal distribution as sample size increases. This idea is similar to what we demonstrated in Section 3.1.2. The result of the simulation study is provided below in Figure 2 for the sample of sizes 750, 1,500, and 3,000. For each of the sample sizes, we replicate 1,000 times, and then results of the estimated parameters are provided through Q-Q plots.

Figure 2. Monte Carlo simulation of finite sample behavior of normality for the parameters $\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{21}, \beta_{22}, \beta_{23},$ and β_{24} based on simulation size 1,000 and sample sizes 750, 1,500, and 3,000.

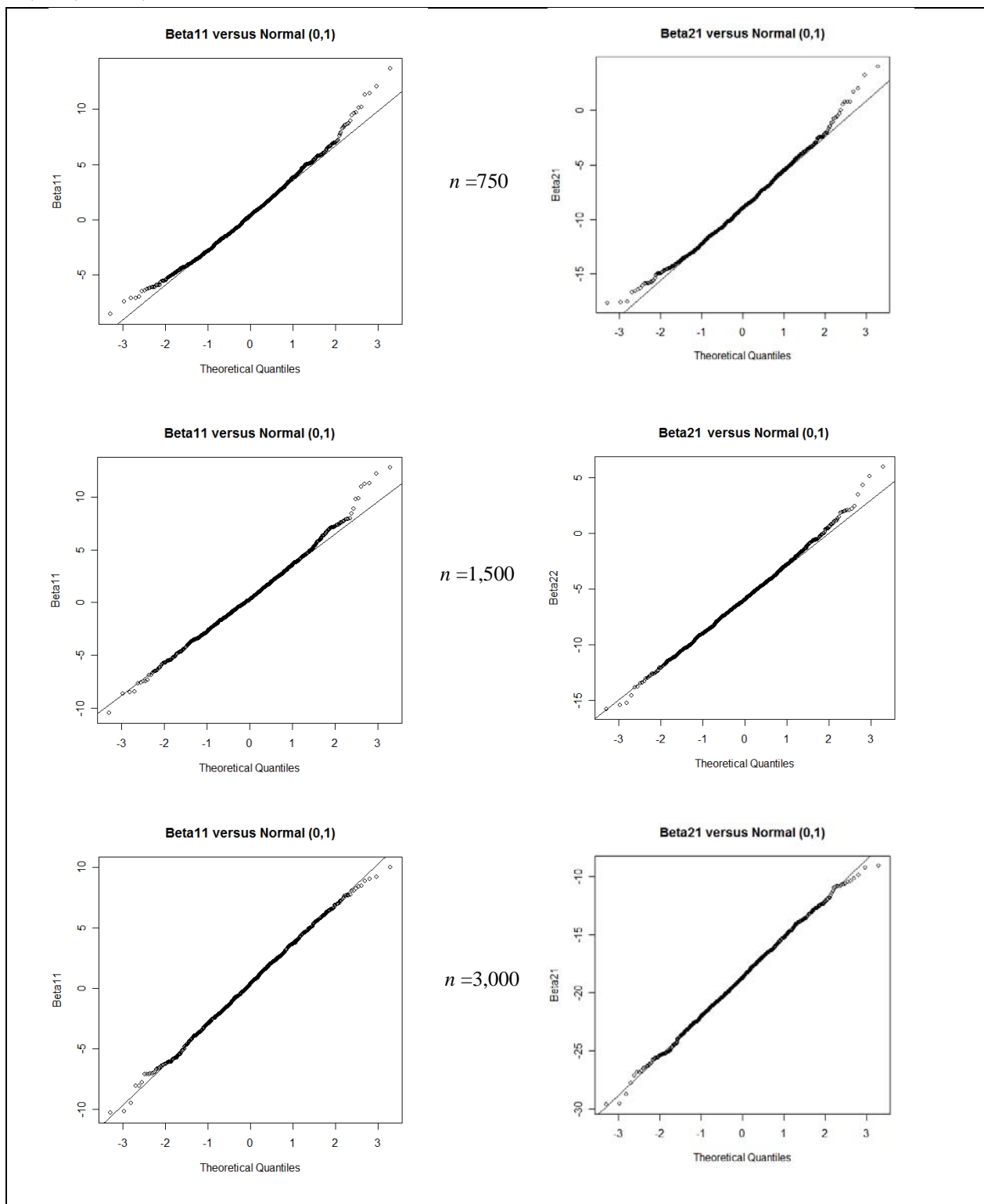


Figure 2 (continued)

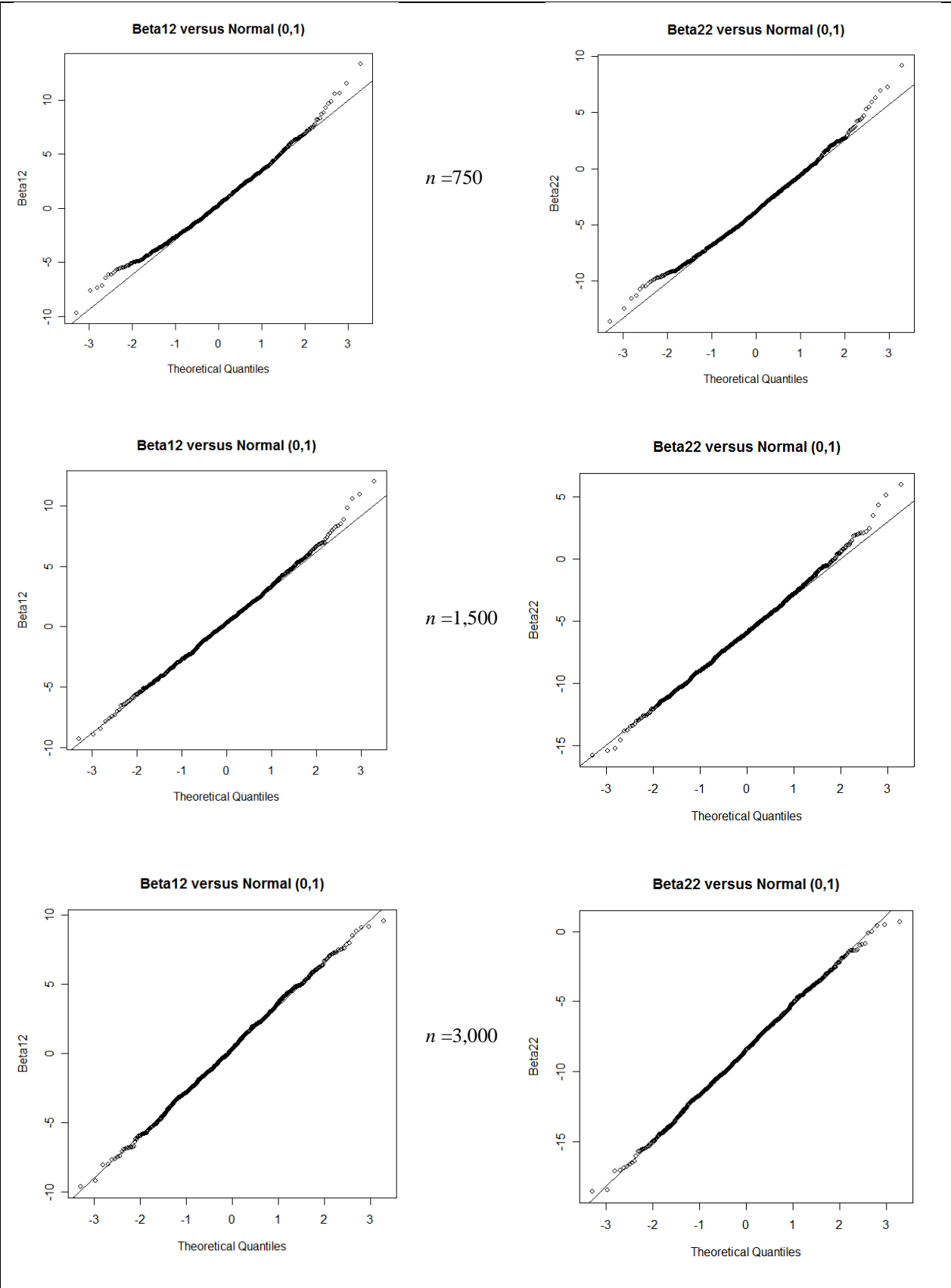


Figure 2 (continued)

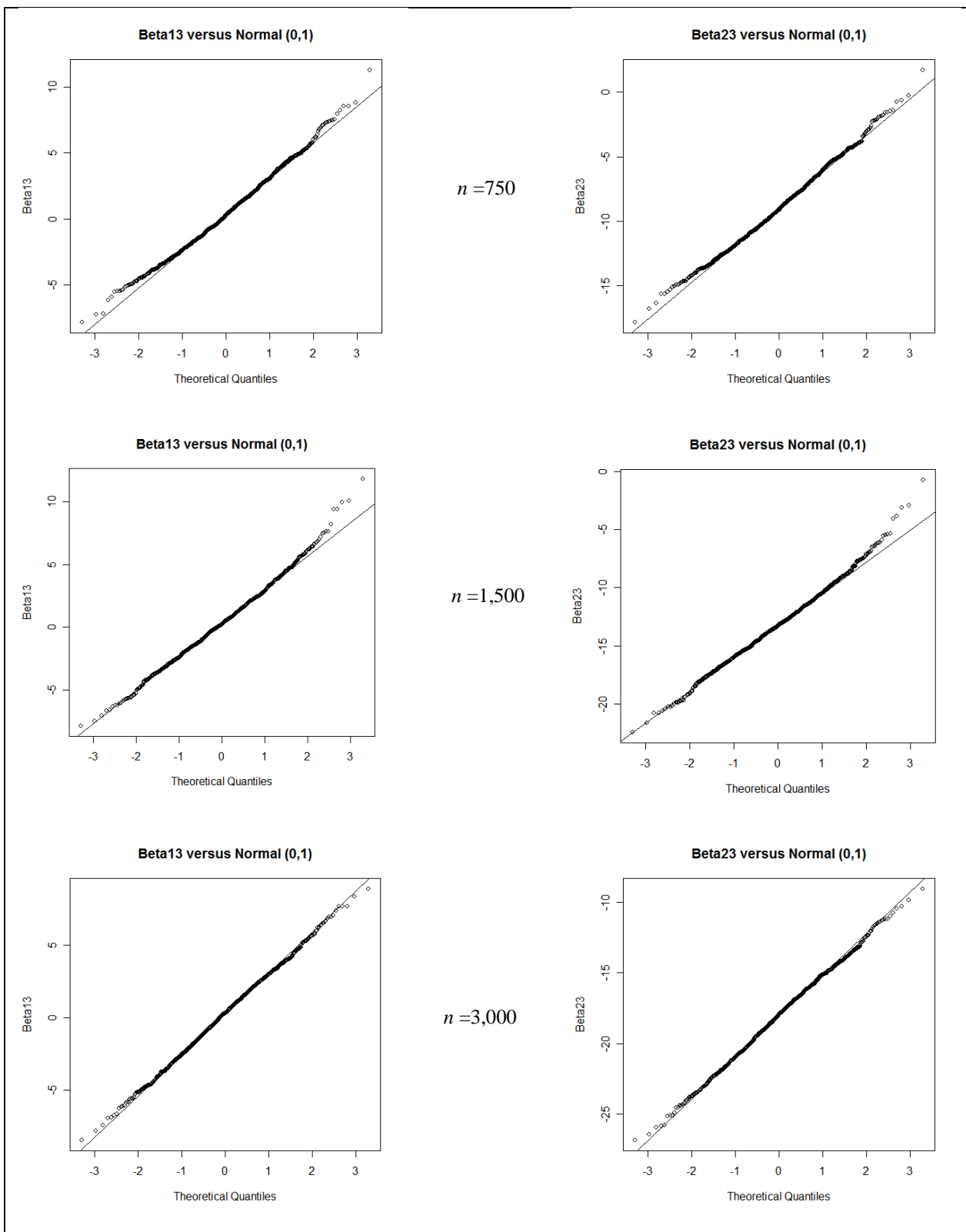
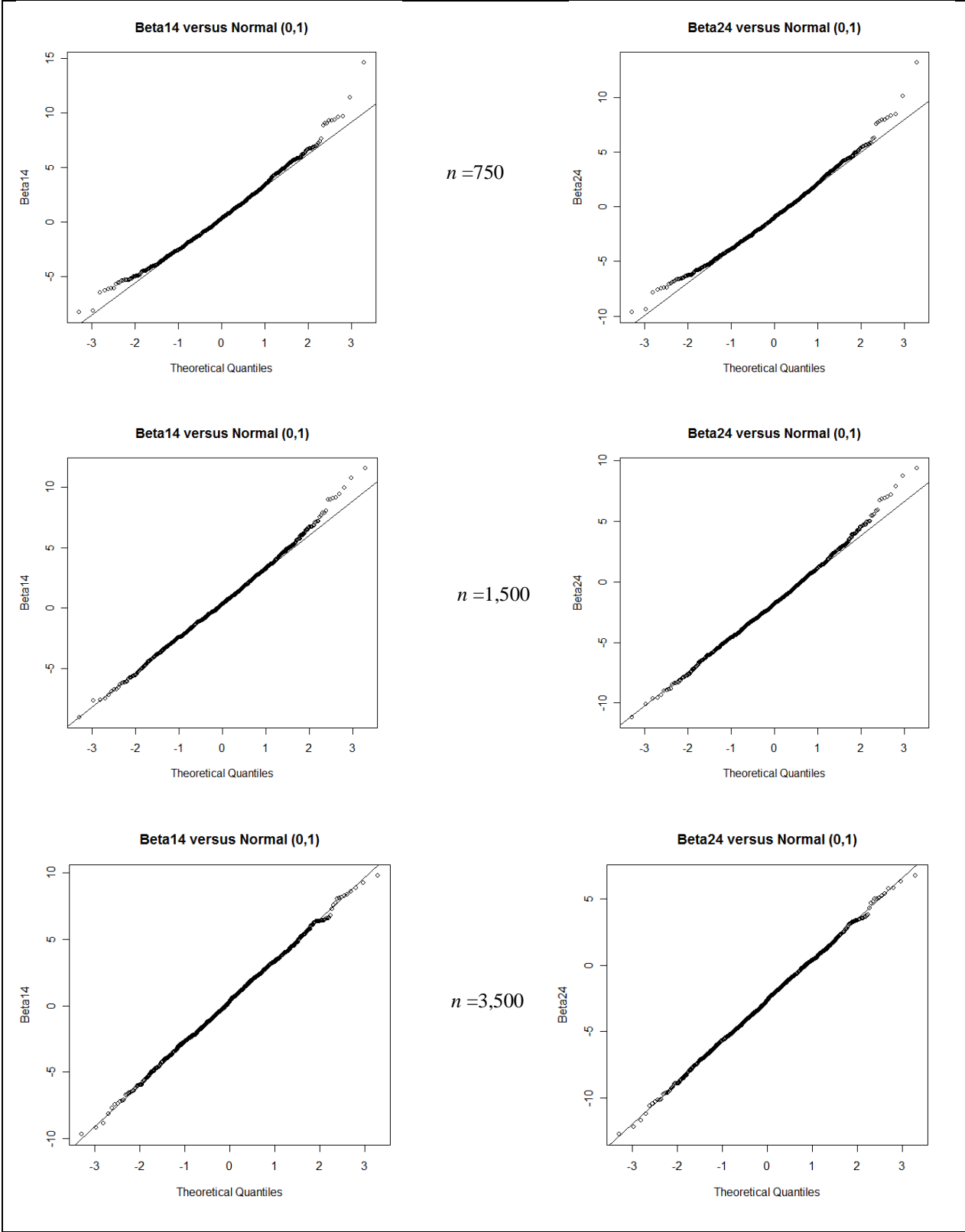


Figure 2 (continued)



4. Discussion

In this article we have shown simulation studies for the consistency and normality for both binary and multinomial logistic regression. The results for both binary and multinomial logistic regression indicate that the simulation study performs well in showing the consistency of the maximum likelihood estimators for parameters of the models. We also believe that the computation results show the distribution of parameters approximates normal distribution as sample size increases; however, it takes quite a large set of data for such results in the case of multinomial logistic regression models. Therefore it is very important to check the model assumptions before applying these results in a real life situation.

References

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Inc.
- Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Harvard University Press.
- Beer, M. (2001). Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logistic Regression Models. Diploma Thesis, *University of Fribourg Switzerland*.
- Cox, D. R., and Snell, E. J. (1989). *Analysis of Binary Data*. London: Chapman and Hall.
- Cramer, J.S. (2003). *Logit Models from Economics and Other Fields*. Cambridge University Press.
- Givens, G.H. and Hoeting, J.A. (2005). *Computational Statistics*, John Wiley & Sons, Inc.
- Gourieroux, C. and Monfort, A. (1981). Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logit Models. *Journal of Econometrics*, **17**, 83-97.
- Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression*. Second Edition, John Wiley & Sons Inc., New York.
- Mak, K. Tak (1993). Solving Non-Linear Estimation Equations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 945-955.
- van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.