Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Large Sample Size Effects on Maximum Likelihood Estimation in Multinomial Logistic Regression

Abubakar Abdullahi

abubakar.abdullahi2@fubk.edu.ng

Federal University Birnin Kebbi

A. W. Babayemi

Abdullahi Fadio Abdullahi Fodio University of Science and Technology, Nigeria

Research Article

Keywords: multinomial logistic regression, Monte Carlo simulation, sample size, coefficient estimation, bias, mean squared error

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11464640/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Large Sample Size Effects on Maximum Likelihood Estimation in Multinomial Logistic Regression

Abubakar Abdullahi[1] and Babayemi, A. W.[2]
1 Department of Mathematics, Federal University Birnin Kebbi
2 Department of Mathematics, Abdullahi Fodio University of Science and Technology, Aliero

Abstract

Multinomial logistic regression is used for outcomes with more than two categories, but evidence on how large sample sizes affect coefficient-estimation performance remains limited for specific model structures. Previous studies have often focused on prediction, sample-size planning, sparse-data methods, or other estimation settings. This study assessed coefficient performance in a five-category multinomial logistic regression model with ten binary explanatory variables, using category 5 as the reference. Monte Carlo simulation was conducted at sample sizes of 1,000, 3,000, 5,000, and 10,000, with 10,000 datasets generated at each size. The model was fitted to each dataset, and the mean estimates, bias, empirical standard errors, and mean squared errors (MSEs) were calculated for 44 coefficients. All 10,000 model fits succeeded at each sample size. Across the 44 coefficients, mean absolute bias decreased from 0.0453 at $n = 1{,}000$ to 0.0038 at $n = 10{,}000$. Median coefficient-level MSE decreased from 0.0598 to 0.0055, while mean empirical standard error declined from 0.3216 to 0.0922. MSE decreased for every coefficient at each successive sample size. The study fills a focused evidence gap by comparing coefficient-level estimation measures across several large sample sizes within the same five-category model. Under the specified simulation conditions, larger sample sizes were associated with estimates that were, overall, less biased and more precise.

Keywords: multinomial logistic regression; Monte Carlo simulation; sample size; coefficient estimation; bias; mean squared error

1 Introduction

Multinomial logistic regression is used when the outcome variable has more than two categories and the categories have no natural order. It estimates the probability of each outcome category from a set of explanatory variables, using one outcome category as the reference. The model is used in many fields, including health and social science, where researchers need to examine several possible outcomes at once. The quality of its estimates depends in part on the amount and structure of the data. Sample size affects the information available to estimate model coefficients. When the sample is limited, estimates may be imprecise or biased, especially when outcome categories are uncommon or the model contains many parameters.

Increasing the sample size can reduce variability, but it does not by itself guarantee a well-specified model or valid conclusions. A larger sample can also make small effects statistically significant (Abdullahi *et al.*, 2023). Researchers therefore need to consider estimate size and precision, not significance alone. Recent studies have examined sample size and multinomial logistic regression from different perspectives. Hamid *et al.* (2016) used Monte Carlo simulation to study sample size and covariate type in a three-category model. They found that estimation patterns depended on the covariate distribution as well as the sample size. De Jong *et al.* (2019) examined how the sizes of outcome categories, the number of predictors, and sample size affected the predictive performance of multinomial logistic models. In binary logistic regression, van Smeden *et al.* (2016) showed that coefficient bias and mean squared error can depend on both the total sample size and the number of events per variable. Pate *et al.* (2023) proposed criteria for planning sample sizes for multinomial prediction models, including criteria related to overfitting and the precision of predicted risks. Tan and Bellec (2023) studied inference for multinomial logistic regression in a high-dimensional setting. Gehringer *et al.* (2024) provided guidance on developing and validating multinomial prediction models and identified further sample-size research as an important need. Together, these studies show that sample-size requirements depend on the model structure, outcome categories, predictors, and the performance measure of interest. However, these studies do not all address the same question. Some focus on prediction and overfitting, while others examine sparse data, bias reduction, or high-dimensional inference. Findings from one setting cannot automatically be applied to another. In particular, the reviewed studies do not provide a single assessment of how mean coefficient estimates, bias, empirical standard errors, and mean squared error change across large sample sizes in a five-category model with ten binary predictors. The present study uses Monte Carlo simulation to examine this question. It generates data under a specified five-category multinomial logistic regression model, treats category 5 as the reference, and includes

ten binary explanatory variables. The model is fitted at sample sizes of 1,000, 3,000, 5,000, and 10,000. For each sample size, 10,000 datasets are generated and analysed. The mean coefficient estimates, bias, empirical standard errors, and mean squared errors are compared with the known coefficient values. The study therefore provides a focused assessment of coefficient-estimation performance under the specified simulation conditions.

2 Literature Review

Research on sample size in multinomial logistic regression has considered coefficient estimation, sparse data, prediction, and statistical inference. These studies provide useful evidence, but their results apply to different models and research questions. Bull *et al.* (2002) examined coefficient estimation in small-sample multinomial logistic regression. Their simulation studies compared ordinary maximum-likelihood estimates with bias-reducing methods. In the settings examined, the modified-score estimator reduced bias and mean squared error. It could also produce finite estimates in cases where ordinary maximum-likelihood estimates did not exist. This study is directly relevant to finite-sample coefficient estimation, although it focused on comparing estimation methods rather than on the large sample sizes used in the present study. Bull *et al.* (2007) studied confidence intervals for multinomial logistic regression with sparse data. They considered settings in which separation could cause ordinary maximum-likelihood estimates to become very large or non-finite. Their results supported penalized profile-likelihood intervals over conventional Wald intervals in the settings studied. This work shows that both the sample and the estimation method can affect inference when data are sparse. The present study has a different focus: it compares coefficient-estimation measures across increasing sample sizes under a specified data-generating model. Hamid *et al.* (2016) conducted a Monte Carlo study of sample size and covariate type in a three-category multinomial logistic regression model. Their design used one explanatory

variable and considered continuous and categorical covariates. They found that estimates generally became less variable and moved closer to their generating values as the sample size increased. The pattern also depended on the covariate distribution. The sample-size values suggested by that study apply to its one-covariate design and should not be treated as universal thresholds. The present study extends this line of work by examining ten binary predictors and four non-reference outcome categories. De Jong *et al.* (2019) examined sample size and predictive performance in multinomial logistic prediction models. Their simulation varied the number of predictors, outcome-category frequencies, and the size of the smallest category relative to the number of model parameters. They found that predictive performance improved as sample size and information in the smallest outcome category increased. Small samples could produce overfitting and poor calibration. The study's main outcomes were predictive performance measures, rather than coefficient-level bias, empirical standard errors, or mean squared error. It therefore provides relevant background but does not answer the estimation question addressed here. Pate *et al.* (2023) proposed minimum-sample-size criteria for developing multinomial logistic prediction models. The criteria address overfitting, optimism in model performance, and the precision of predicted outcome risks. Their simulation assessment supported the proposed overfitting criterion in the settings examined. This work is important for prediction-model planning, but it does not compare coefficient bias, empirical standard errors, and mean squared error across a fixed set of large sample sizes. Tan and Bellec (2023) studied the asymptotic behaviour of maximum-likelihood estimates in multinomial logistic regression when the number of dimensions and the sample size grow at a similar rate. They developed results for inference on null covariates and used simulations to assess proposed significance tests. Their study focuses on high-dimensional inference. The present study instead holds the number of predictors fixed and examines coefficient-estimation measures as the sample size increases. Gehringer *et al.* (2024) provided guidance on developing, validating,

and updating clinical prediction models using multinomial logistic regression. Their article discusses model development and evaluation, including calibration and validation. They also identify further methodological work on sample-size criteria, particularly for external validation, as an area for future research. This supports the continuing importance of sample-size research in multinomial models. The present study focuses on estimation performance when fitting a model, not on external validation. Evidence from binary logistic regression is also relevant as background, although it should not be treated as direct evidence for multinomial outcomes. Van Smeden *et al.* (2016) used Monte Carlo simulation to examine the "one variable per ten events" rule in binary logistic regression. They assessed coefficient bias, confidence-interval behaviour, and mean squared error across different event counts and sample sizes. Their findings showed that performance depended on several factors, including total sample size, effect size, and separation. They cautioned against applying a single events-per-variable rule in every setting. Nemes *et al.* (2009) used simulation to examine the effect of sample size on odds-ratio estimates in binary logistic regression. They found that estimates moved towards their generating values as sample size increased, while smaller samples could produce greater bias and more extreme estimates. Their findings support the general need to examine finite-sample performance, but their model had a binary outcome and a specific predictor structure. Hsieh *et al.* (1998) developed methods for calculating sample size and power in linear and binary logistic regression. Their methods show how factors such as effect size and correlation between predictors can affect the sample size needed to test a coefficient. This work is useful background for sample-size planning, but it does not examine a five-category outcome or compare Monte Carlo bias and mean squared error across the sample sizes used in the present study. Abdullahi *et al.* (2024) compared parametric and semiparametric estimators of the long-memory parameter in an ARFIMA time-series model. The study used Monte Carlo simulation and compared measures such as bias and mean squared error across

simulated settings. It is relevant as an example of simulation-based estimator comparison, but it did not study multinomial logistic regression or categorical outcomes. Its relevance to the present study is therefore methodological rather than direct.

3 Monte Carlo Simulation Method

Monte Carlo simulation was used to assess the performance of the regression coefficient estimates. In particular, the simulation was used to examine whether the maximum likelihood estimates for the multinomial logistic regression model approach their true values as the sample size increases. The outcome variable *Y* has five possible categories: 1, 2, 3, 4, and 5. Category 5 was chosen as the reference category. The model includes binary explanatory variables, each coded as 0 or 1. Using $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$, and x_{10}. the model is written as:

$$\log\left(\frac{P(Y=1)}{P(Y=5)}\right) = \beta_{10} + \sum_{j=1}^{10} \beta_{1j} x_j \qquad 1$$

$$\log\left(\frac{P(Y=2)}{P(Y=5)}\right) = \beta_{20} + \sum_{j=1}^{10} \beta_{2j} x_j \qquad 2$$

$$\log\left(\frac{P(Y=3)}{P(Y=5)}\right) = \beta_{30} + \sum_{j=1}^{10} \beta_{3j} x_j \qquad 3$$

$$\log\left(\frac{P(Y=4)}{P(Y=5)}\right) = \beta_{40} + \sum_{j=1}^{10} \beta_{4j} x_j \qquad 4$$

Under the multinomial logistic regression model, with five response categories $Y = 1,2,3,4,5$ and category 5 as the reference category, the response probabilities are given by:

$$P(Y=1) = \frac{\exp\left(\beta_{10} + \sum_{j=1}^{10} \beta_{1j} x_j\right)}{1 + \sum_{k=1}^{4} \exp\left(\beta_{k0} + \sum_{j=1}^{10} \beta_{kj} x_j\right)} \qquad 5$$

$$P(Y=2)=\frac{\exp\left(\beta_{20}+\sum_{j=1}^{10}\beta_{1j}x_j\right)}{1+\sum_{k=1}^{4}\exp\left(\beta_{k0}+\sum_{j=1}^{10}\beta_{kj}x_j\right)} \quad 6$$

$$P(Y=3)=\frac{\exp\left(\beta_{30}+\sum_{j=1}^{10}\beta_{1j}x_j\right)}{1+\sum_{k=1}^{4}\exp\left(\beta_{k0}+\sum_{j=1}^{10}\beta_{kj}x_j\right)} \quad 7$$

$$P(Y=4)=\frac{\exp\left(\beta_{40}+\sum_{j=1}^{10}\beta_{1j}x_j\right)}{1+\sum_{k=1}^{4}\exp\left(\beta_{k0}+\sum_{j=1}^{10}\beta_{kj}x_j\right)} \quad 8$$

The probability of the reference category, $Y=5$. is

$$P(Y=5)=\frac{1}{1+\sum_{k=1}^{4}\exp\left(\beta_{k0}+\sum_{j=1}^{10}\beta_{kj}x_j\right)} \quad 9$$

These probabilities satisfy the condition below and each probability lies between 0 and 1.

$$\sum_{k=1}^{5}P(Y=k)=1 \quad 10$$

For the simulation, ten binary explanatory variables, $x_1, x_2, \ldots, x_{10}$, were generated using binomial distributions. Category 5 was treated as the reference category. Fixed values were assigned to the regression coefficients for each of the other four categories. For Category 1, the beta coefficients are $\beta_{10}=-4.2$, $\beta_{11}=1.3$, $\beta_{12}=-0.7$, $\beta_{13}=0.9$, $\beta_{14}=1.8$, $\beta_{15}=-1.1$, $\beta_{16}=0.5$, $\beta_{17}=1.2$, $\beta_{18}=-0.4$, $\beta_{19}=0.8$, and $\beta_{1,10}=1.5$. For Category 2, the beta coefficients are $\beta_{20}=-3.6$, $\beta_{21}=0.6$, $\beta_{22}=1.4$, $\beta_{23}=-1.2$, $\beta_{24}=0.7$, $\beta_{25}=1.1$, $\beta_{26}=-0.5$, $\beta_{27}=1.6$, $\beta_{28}=0.9$, $\beta_{29}=-0.8$, and $\beta_{2,10}=1.2$. For Category 3, the beta coefficients are $\beta_{30}=-4.7$, $\beta_{31}=-0.9$, $\beta_{32}=1.1$, $\beta_{33}=0.6$, $\beta_{34}=-1.3$, $\beta_{35}=1.5$, $\beta_{36}=0.8$, $\beta_{37}=-0.6$, $\beta_{38}=1.4$, $\beta_{39}=0.5$, and $\beta_{3,10}=-1.0$. For Category 4, the beta coefficients are $\beta_{40}=-3.9$, $\beta_{41}=1.0$, $\beta_{42}=-0.5$, $\beta_{43}=1.3$, $\beta_{44}=-0.7$, $\beta_{45}=0.9$, $\beta_{46}=1.6$, $\beta_{47}=-1.1$,

$\beta_{48} = 0.6$, $\beta_{49} = 1.2$, and $\beta_{4,10} = -0.4$. These coefficient values are assumed for the simulation. The simulation was conducted using four sample sizes: $n = 1{,}000,\ 3{,}000,\ 5{,}000$ and $10{,}000$. For each sample size, 10,000 independent datasets were generated. The multinomial logistic regression model was fitted separately to each dataset, and the parameter estimates were recorded. For each parameter, the mean estimate, bias, standard error, and mean squared error (MSE) were calculated across the 10,000 simulation runs to assess the performance of the estimates.

4 Evaluation of Parameter Performance

The performance of each regression coefficient was evaluated using the mean estimate, bias, empirical standard error, and mean squared error (MSE). These measures were calculated from R successful simulation runs, where $\hat{\beta}_r$ represents the estimated coefficient in run r, and β is the true coefficient value. The measures are defined as follows:

$$\text{Mean Estimate} = \bar{\hat{\beta}} = \frac{1}{R}\sum_{r=1}^{R}\hat{\beta}_r \qquad 11$$

$$\text{Bias} = \bar{\hat{\beta}} - \beta \qquad 12$$

$$\text{Standard Error} = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\hat{\beta}})^2} \qquad 13$$

$$\text{MSE} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\beta}_r - \beta)^2 \qquad 14$$

Generally, an estimation method performs better when it has lower absolute bias, lower standard error, and lower MSE.

5 Analysis and Results

Table 1. Mean Estimates and Empirical Standard Errors of the Multinomial Logistic Regression Coefficients by Sample Size

		$n = 1,000$		$n = 3,000$		$n = 5,000$		$n = 10,000$	
Parameter	**True value**	**Mean**	**Std. error**	**Mean**	**Std. error**	**Mean**	**Std. error**	**Mean**	**Std. error**
β_{10}	-4.2000	-4.2882	0.4132	-4.2314	0.2305	-4.2177	0.1771	-4.2091	0.1253
β_{11}	1.3000	1.3240	0.2152	1.3097	0.1216	1.3054	0.0942	1.3028	0.0670
β_{12}	-0.7000	-0.7135	0.2187	-0.7049	0.1220	-0.7035	0.0949	-0.7008	0.0663
β_{13}	0.9000	0.9183	0.2158	0.9060	0.1232	0.9036	0.0956	0.9022	0.0674
β_{14}	1.8000	1.8360	0.2330	1.8138	0.1298	1.8071	0.0998	1.8031	0.0709
β_{15}	-1.1000	-1.1267	0.2302	-1.1088	0.1287	-1.1054	0.0976	-1.1025	0.0695
β_{16}	0.5000	0.5120	0.2096	0.5032	0.1180	0.5034	0.0923	0.5010	0.0648
β_{17}	1.2000	1.2250	0.2186	1.2092	0.1228	1.2036	0.0944	1.2025	0.0670
β_{18}	-0.4000	-0.4085	0.2127	-0.4046	0.1218	-0.4016	0.0937	-0.4009	0.0649
β_{19}	0.8000	0.8138	0.2114	0.8055	0.1211	0.8029	0.0935	0.8010	0.0655
$\beta_{1,10}$	1.5000	1.5320	0.2227	1.5086	0.1221	1.5065	0.0963	1.5035	0.0688
β_{20}	-3.6000	-3.6700	0.3779	-3.6223	0.2100	-3.6125	0.1638	-3.6052	0.1145
β_{21}	0.6000	0.6076	0.2003	0.6020	0.1140	0.6020	0.0872	0.6018	0.0614
β_{22}	1.4000	1.4259	0.2145	1.4076	0.1220	1.4062	0.0933	1.4019	0.0655
β_{23}	-1.2000	-1.2226	0.2158	-1.2082	0.1225	-1.2061	0.0945	-1.2027	0.0654
β_{24}	0.7000	0.7108	0.2057	0.7051	0.1151	0.7020	0.0884	0.7006	0.0620
β_{25}	1.1000	1.1218	0.2093	1.1084	0.1178	1.1045	0.0908	1.1019	0.0634
β_{26}	-0.5000	-0.5103	0.2069	-0.5016	0.1143	-0.5021	0.0882	-0.5007	0.0628
β_{27}	1.6000	1.6343	0.2136	1.6100	0.1202	1.6048	0.0942	1.6022	0.0658
β_{28}	0.9000	0.9164	0.2043	0.9043	0.1139	0.9027	0.0887	0.9010	0.0630
β_{29}	-0.8000	-0.8212	0.2076	-0.8052	0.1176	-0.8043	0.0914	-0.8016	0.0634
$\beta_{2,10}$	1.2000	1.2240	0.2051	1.2072	0.1173	1.2052	0.0888	1.2015	0.0634
β_{30}	-4.7000	-5.0327	0.9216	-4.7944	0.4406	-4.7563	0.3465	-4.7275	0.2356
β_{31}	-0.9000	-0.9948	0.5978	-0.9277	0.2757	-0.9188	0.2089	-0.9070	0.1470
β_{32}	1.1000	1.1749	0.4944	1.1203	0.2541	1.1083	0.1934	1.1057	0.1352
β_{33}	0.6000	0.6428	0.4445	0.6081	0.2394	0.6089	0.1839	0.6039	0.1272
β_{34}	-1.3000	-1.4222	0.6779	-1.3324	0.2812	-1.3144	0.2110	-1.3083	0.1466
β_{35}	1.5000	1.5925	0.5650	1.5256	0.2620	1.5176	0.2005	1.5055	0.1382
β_{36}	0.8000	0.8539	0.4427	0.8131	0.2411	0.8094	0.1835	0.8041	0.1278
β_{37}	-0.6000	-0.6571	0.4693	-0.6100	0.2495	-0.6127	0.1906	-0.6069	0.1332
β_{38}	1.4000	1.4752	0.5013	1.4213	0.2534	1.4139	0.1985	1.4067	0.1366
β_{39}	0.5000	0.5277	0.4347	0.5057	0.2343	0.5043	0.1807	0.5029	0.1277
$\beta_{3,10}$	-1.0000	-1.0859	0.5717	-1.0224	0.2678	-1.0146	0.2055	-1.0073	0.1445
β_{40}	-3.9000	-4.0032	0.4515	-3.9305	0.2542	-3.9153	0.1956	-3.9089	0.1387
β_{41}	1.0000	1.0226	0.2419	1.0061	0.1370	1.0036	0.1051	1.0023	0.0738
β_{42}	-0.5000	-0.5160	0.2477	-0.5042	0.1369	-0.5045	0.1055	-0.5011	0.0752
β_{43}	1.3000	1.3393	0.2584	1.3091	0.1448	1.3045	0.1099	1.3029	0.0789
β_{44}	-0.7000	-0.7217	0.2555	-0.7049	0.1428	-0.7058	0.1108	-0.7028	0.0774
β_{45}	0.9000	0.9215	0.2480	0.9064	0.1374	0.9040	0.1061	0.9032	0.0756
β_{46}	1.6000	1.6445	0.2627	1.6122	0.1471	1.6074	0.1134	1.6023	0.0798
β_{47}	-1.1000	-1.1374	0.2706	-1.1107	0.1520	-1.1080	0.1159	-1.1024	0.0834
β_{48}	0.6000	0.6149	0.2382	0.6035	0.1347	0.6033	0.1041	0.6013	0.0737
β_{49}	1.2000	1.2300	0.2495	1.2098	0.1419	1.2067	0.1074	1.2029	0.0772
$\beta_{4,10}$	-0.4000	-0.4174	0.2457	-0.4076	0.1396	-0.4032	0.1079	-0.4019	0.0753

Table 1showed the mean estimates and empirical standard errors for each coefficient at all four sample sizes. The sample size increases from 1,000 to 10,000, the mean estimates become

closer to the corresponding true values overall. The average absolute difference between the mean estimates and true values decreases from 0.0453 at $n = 1{,}000$ to 0.0130 at $n = 3{,}000$, 0.0081 at $n = 5{,}000$, and 0.0038 at $n = 10{,}000$. The empirical standard errors also decrease for all 44 coefficients at each successive sample size. Their average falls from 0.3216 at $n = 1{,}000$ to 0.0922 at $n = 10{,}000$.

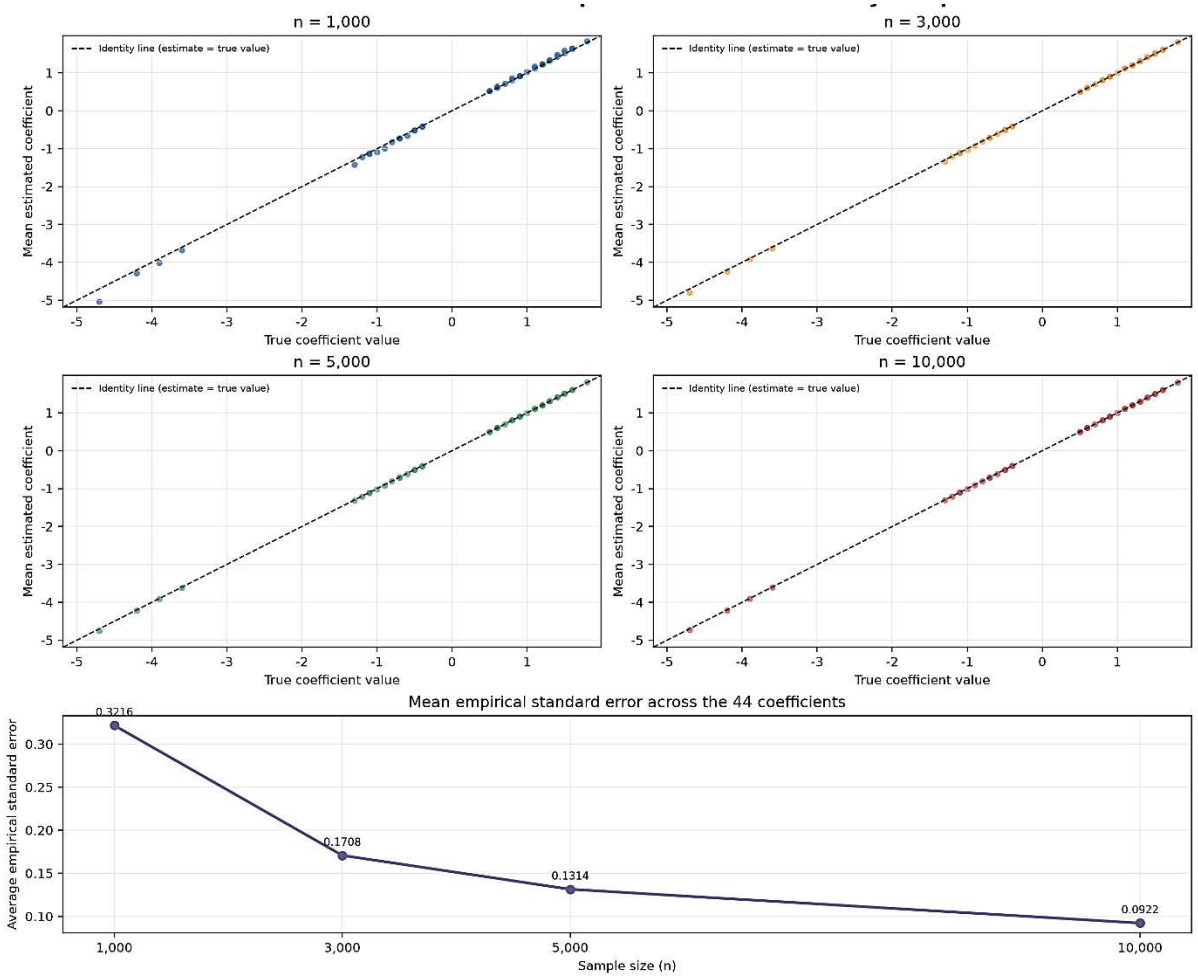


Figure 1: Mean Coefficient Estimates and Empirical Standard Errors by Sample Size

Figure 1 compared the mean estimated coefficients with their known true values. The dashed line marks where an estimate equals its true value. As sample size increases, the plotted estimates lie closer to this line. The lower panel shows that the average empirical standard error across the 44 coefficients decreases from 0.3216 at $n = 1{,}000$ to 0.0922 at $n = 10{,}000$.

Table 2. Bias and Mean Squared Error of the Multinomial Logistic Regression Coefficient Estimates by Sample Size

		$n = 1,000$		$n = 3,000$		$n = 5,000$		$n = 10,000$	
Parameter	**True value**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
$\boldsymbol{\beta_{10}}$	-4.2000	-0.0882	0.1785	-0.0314	0.0541	-0.0177	0.0317	-0.0091	0.0158
$\boldsymbol{\beta_{11}}$	1.3000	0.0240	0.0469	0.0097	0.0149	0.0054	0.0089	0.0028	0.0045
$\boldsymbol{\beta_{12}}$	-0.7000	-0.0135	0.0480	-0.0049	0.0149	-0.0035	0.0090	-0.0008	0.0044
$\boldsymbol{\beta_{13}}$	0.9000	0.0183	0.0469	0.0060	0.0152	0.0036	0.0092	0.0022	0.0045
$\boldsymbol{\beta_{14}}$	1.8000	0.0360	0.0556	0.0138	0.0170	0.0071	0.0100	0.0031	0.0050
$\boldsymbol{\beta_{15}}$	-1.1000	-0.0267	0.0537	-0.0088	0.0166	-0.0054	0.0096	-0.0025	0.0048
$\boldsymbol{\beta_{16}}$	0.5000	0.0120	0.0441	0.0032	0.0139	0.0034	0.0085	0.0010	0.0042
$\boldsymbol{\beta_{17}}$	1.2000	0.0250	0.0484	0.0092	0.0152	0.0036	0.0089	0.0025	0.0045
$\boldsymbol{\beta_{18}}$	-0.4000	-0.0085	0.0453	-0.0046	0.0149	-0.0016	0.0088	-0.0009	0.0042
$\boldsymbol{\beta_{19}}$	0.8000	0.0138	0.0449	0.0055	0.0147	0.0029	0.0088	0.0010	0.0043
$\boldsymbol{\beta_{1,10}}$	1.5000	0.0320	0.0506	0.0086	0.0150	0.0065	0.0093	0.0035	0.0047
$\boldsymbol{\beta_{20}}$	-3.6000	-0.0700	0.1477	-0.0223	0.0446	-0.0125	0.0270	-0.0052	0.0131
$\boldsymbol{\beta_{21}}$	0.6000	0.0076	0.0402	0.0020	0.0130	0.0020	0.0076	0.0018	0.0038
$\boldsymbol{\beta_{22}}$	1.4000	0.0259	0.0467	0.0076	0.0149	0.0062	0.0087	0.0019	0.0043
$\boldsymbol{\beta_{23}}$	-1.2000	-0.0226	0.0471	-0.0082	0.0151	-0.0061	0.0090	-0.0027	0.0043
$\boldsymbol{\beta_{24}}$	0.7000	0.0108	0.0424	0.0051	0.0133	0.0020	0.0078	0.0006	0.0038
$\boldsymbol{\beta_{25}}$	1.1000	0.0218	0.0443	0.0084	0.0139	0.0045	0.0083	0.0019	0.0040
$\boldsymbol{\beta_{26}}$	-0.5000	-0.0103	0.0429	-0.0016	0.0131	-0.0021	0.0078	-0.0007	0.0039
$\boldsymbol{\beta_{27}}$	1.6000	0.0343	0.0468	0.0100	0.0145	0.0048	0.0089	0.0022	0.0043
$\boldsymbol{\beta_{28}}$	0.9000	0.0164	0.0420	0.0043	0.0130	0.0027	0.0079	0.0010	0.0040
$\boldsymbol{\beta_{29}}$	-0.8000	-0.0212	0.0436	-0.0052	0.0139	-0.0043	0.0084	-0.0016	0.0040
$\boldsymbol{\beta_{2,10}}$	1.2000	0.0240	0.0426	0.0072	0.0138	0.0052	0.0079	0.0015	0.0040
$\boldsymbol{\beta_{30}}$	-4.7000	-0.3327	0.9600	-0.0944	0.2031	-0.0563	0.1233	-0.0275	0.0563
$\boldsymbol{\beta_{31}}$	-0.9000	-0.0948	0.3663	-0.0277	0.0768	-0.0188	0.0440	-0.0070	0.0216
$\boldsymbol{\beta_{32}}$	1.1000	0.0749	0.2500	0.0203	0.0650	0.0083	0.0375	0.0057	0.0183
$\boldsymbol{\beta_{33}}$	0.6000	0.0428	0.1994	0.0081	0.0574	0.0089	0.0339	0.0039	0.0162
$\boldsymbol{\beta_{34}}$	-1.3000	-0.1222	0.4745	-0.0324	0.0801	-0.0144	0.0447	-0.0083	0.0216
$\boldsymbol{\beta_{35}}$	1.5000	0.0925	0.3277	0.0256	0.0693	0.0176	0.0405	0.0055	0.0191
$\boldsymbol{\beta_{36}}$	0.8000	0.0539	0.1989	0.0131	0.0583	0.0094	0.0337	0.0041	0.0163
$\boldsymbol{\beta_{37}}$	-0.6000	-0.0571	0.2235	-0.0100	0.0623	-0.0127	0.0365	-0.0069	0.0178
$\boldsymbol{\beta_{38}}$	1.4000	0.0752	0.2569	0.0213	0.0647	0.0139	0.0396	0.0067	0.0187
$\boldsymbol{\beta_{39}}$	0.5000	0.0277	0.1897	0.0057	0.0549	0.0043	0.0327	0.0029	0.0163
$\boldsymbol{\beta_{3,10}}$	-1.0000	-0.0859	0.3342	-0.0224	0.0722	-0.0146	0.0424	-0.0073	0.0209
$\boldsymbol{\beta_{40}}$	-3.9000	-0.1032	0.2145	-0.0305	0.0655	-0.0153	0.0385	-0.0089	0.0193
$\boldsymbol{\beta_{41}}$	1.0000	0.0226	0.0590	0.0061	0.0188	0.0036	0.0111	0.0023	0.0054
$\boldsymbol{\beta_{42}}$	-0.5000	-0.0160	0.0616	-0.0042	0.0188	-0.0045	0.0111	-0.0011	0.0056
$\boldsymbol{\beta_{43}}$	1.3000	0.0393	0.0683	0.0091	0.0210	0.0045	0.0121	0.0029	0.0062
$\boldsymbol{\beta_{44}}$	-0.7000	-0.0217	0.0658	-0.0049	0.0204	-0.0058	0.0123	-0.0028	0.0060
$\boldsymbol{\beta_{45}}$	0.9000	0.0215	0.0620	0.0064	0.0189	0.0040	0.0113	0.0032	0.0057
$\boldsymbol{\beta_{46}}$	1.6000	0.0445	0.0710	0.0122	0.0218	0.0074	0.0129	0.0023	0.0064
$\boldsymbol{\beta_{47}}$	-1.1000	-0.0374	0.0746	-0.0107	0.0232	-0.0080	0.0135	-0.0024	0.0070
$\boldsymbol{\beta_{48}}$	0.6000	0.0149	0.0570	0.0035	0.0182	0.0033	0.0108	0.0013	0.0054
$\boldsymbol{\beta_{49}}$	1.2000	0.0300	0.0632	0.0098	0.0202	0.0067	0.0116	0.0029	0.0060
$\boldsymbol{\beta_{4,10}}$	-0.4000	-0.0174	0.0607	-0.0076	0.0195	-0.0032	0.0117	-0.0019	0.0057

Table 2 revealed the bias and MSE for each coefficient at all four sample sizes. The average absolute bias decreases from 0.0453 at $n = 1{,}000$ to 0.0130 at $n = 3{,}000$, 0.0081 at $n =$

5,000, and 0.0038 at $n = 10{,}000$. The mean MSE also decreases across the sample sizes, from 0.1338 to 0.0345, 0.0204, and 0.0099, respectively. MSE is lower for all 44 coefficients at each successive sample size. Absolute bias is lower for all 44 coefficients from 1,000 to 3,000, for 37 from 3,000 to 5,000, and for all 44 from 5,000 to 10,000.

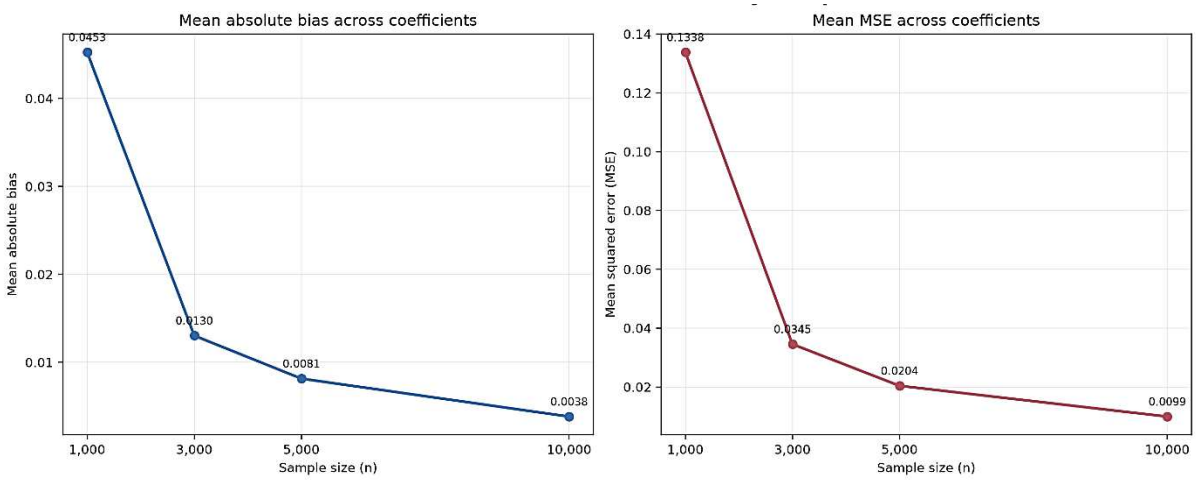


Figure 2: Title: Mean Absolute Bias and Mean Squared Error by Sample Size

Figure 2 showed the mean absolute bias and MSE across the coefficients, and. Both measures decrease across the sample sizes shown. Mean absolute bias falls from 0.0453 at $n = 1{,}000$ to 0.0038 at $n = 10{,}000$; mean MSE falls from 0.1338 to 0.0099. These summaries are calculated across the 44 coefficients using the known true values as the comparison baseline.

6. Conclusion

This study addressed the limited evidence on how increasing sample size affects coefficient-estimation performance in a five-category multinomial logistic regression model with ten binary explanatory variables. Using 10,000 Monte Carlo replications at each of four sample sizes—1,000, 3,000, 5,000, and 10,000—the model was fitted and 44 coefficients were assessed using their mean estimates, bias, empirical standard errors, and MSE. All model fits succeeded. As sample size increased, mean absolute bias, median coefficient-level MSE, and mean empirical standard error decreased. MSE also decreased for all 44 coefficients at each successive sample size. The study therefore fills a focused gap by comparing these coefficient-

level measures across several large sample sizes under one clearly defined multinomial model. The results indicate improved estimation accuracy and precision in the simulated setting, but they should not be treated as a universal sample-size rule. They apply to the specified model, predictors, coefficient values, and outcome structure.

Declarations

Funding: The authors received no funding for this research.

Clinical Trial Number: Not applicable.

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Abdullahi, A., James, T. O., Uzaomaka, A., Ezafulukwe, I. T. V., Ogundeji, A. A., Usman, F., George, G. G., Agboluaje, A. A., Shamsudden, A. S., & Ogini, S. I. (2023, December 13–15). Monte Carlo simulation techniques on evaluating the effect of sample size on multiple linear regression [Conference presentation]. *11th International Mardin Artuklu Scientific Research Conference, Mardin, Türkiye*. https://epale.ec.europa.eu/en/content/11th-international-mardin-artuklu-scientific-researches-conference

Abdullahi, A., Hussaini, J., Babayemi, A. W., Onwuka, G. I., & James, T. O. (2024). Monte Carlo simulation for comparison of parametric and semiparametric estimation methods of long-memory parameter: An ARFIMA model approach. *In R. Baltezarević (Ed.), 6th International Black Sea Modern Scientific Research Congress* (pp. 1371–1382). İKSAD Yayınları.

Bull, S. B., Mak, C., & Greenwood, C. M. T. (2002). A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics & Data Analysis,* 39(1), 57–74. https://doi.org/10.1016/S0167-9473(01)00048-2

Bull, S. B., Lewinger, J. P., & Lee, S. S. F. (2007). Confidence intervals for multinomial logistic regression in sparse data. *Statistics in Medicine*, 26(4), 903–918. https://doi.org/10.1002/sim.2518

de Jong, V. M. T., Eijkemans, M. J. C., van Calster, B., Timmerman, D., Moons, K. G. M., Steyerberg, E. W., & van Smeden, M. (2019). Sample size considerations and predictive performance of multinomial logistic prediction models. *Statistics in Medicine*, 38(9), 1601–1619. https://doi.org/10.1002/sim.8063

Gehringer, C. K., Martin, G. P., Van Calster, B., Hyrich, K. L., Verstappen, S. M. M., & Sergeant, J. C. (2024). How to develop, validate, and update clinical prediction models

using multinomial logistic regression. *Journal of Clinical Epidemiology*, 174, Article 111481. https://doi.org/10.1016/j.jclinepi.2024.111481

Hamid, H. A., Wah, Y. B., & Xie, X.-J. (2016). Effects of different type of covariates and sample size on parameter estimation for multinomial logistic regression model. *Jurnal Teknologi (Sciences & Engineering*), 78(12–3), 155–161.https://doi.org/10.11113/jt.v78.10036

Hsieh, F. Y., Bloch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, 17(14), 1623–1634. https://doi.org/10.1002/(SICI)1097-0258(19980730)17:143.0.CO;2-S

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, Article 56. https://doi.org/10.1186/1471-2288-9-56

Pate, A., Riley, R. D., Collins, G. S., van Smeden, M., Van Calster, B., Ensor, J., & Martin, G. P. (2023). Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research*, 32(3), 555–571. https://doi.org/10.1177/09622802231151220

Tan, K., & Bellec, P. C. (2023). Multinomial logistic regression: Asymptotic normality on null covariates in high-dimensions. *Advances in Neural Information Processing System*s, 36. https://doi.org/10.52202/075280-3106

van Smeden, M., de Groot, J. A. H., Moons, K. G. M., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, 16, Article 163. https://doi.org/10.1186/s12874-016-0267-3

R code for the Study

```
library(nnet)
library(parallel)

set.seed(12345)
sample_sizes <- c(1000, 3000, 5000, 10000)
B <- 10000L
predictor_probability <- 0.5
n_cores <- max(1L, detectCores() - 1L)
predictor_names <- paste0('x', 1:10)

beta <- rbind(
  '1' = c(-4.2, 1.3, -0.7, 0.9, 1.8, -1.1, 0.5, 1.2, -0.4, 0.8, 1.5),
  '2' = c(-3.6, 0.6, 1.4, -1.2, 0.7, 1.1, -0.5, 1.6, 0.9, -0.8, 1.2),
  '3' = c(-4.7, -0.9, 1.1, 0.6, -1.3, 1.5, 0.8, -0.6, 1.4, 0.5, -1.0),
  '4' = c(-3.9, 1.0, -0.5, 1.3, -0.7, 0.9, 1.6, -1.1, 0.6, 1.2, -0.4)
)
colnames(beta) <- c('(Intercept)', predictor_names)

one_run <- function(n, predictor_probability, beta, predictor_names) {
  X <- matrix(rbinom(n * 10L, 1L, predictor_probability), nrow=n, ncol=10L)
  colnames(X) <- predictor_names
  design <- cbind('(Intercept)'=1, X)
  eta <- design %*% t(beta)
  eta_all <- cbind(eta, category_5=0)
```

```
  eta_all <- eta_all - apply(eta_all, 1L, max)
  ex <- exp(eta_all)
  probs <- ex / rowSums(ex)
  cumulative <- t(apply(probs, 1L, cumsum))
  u <- runif(n)
  y <- 1L + (u  cumulative[,1L]) + (u  cumulative[,2L]) +
    (u  cumulative[,3L]) + (u  cumulative[,4L])
  dat <- data.frame(Y=factor(y, levels=c(5,1,2,3,4)), X, check.names=FALSE)
  fit <- tryCatch(
    suppressWarnings(multinom(Y ~ ., data=dat, trace=FALSE, maxit=200,
      MaxNWts=1000)),
    error=function(e) NULL
  )
  if (is.null(fit) || fit$convergence != 0L) return(NULL)
  est <- coef(fit)
  if (is.null(dim(est)) || !all(rownames(beta) %in% rownames(est)) ||
      !all(colnames(beta) %in% colnames(est))) return(NULL)
  est[rownames(beta), colnames(beta), drop=FALSE]
}

run_n <- function(n) {
  message(format(Sys.time()), ' starting n=', n, ' with ', B, ' repetitions on ', n_cores, '
      cores')
  cl <- makeCluster(n_cores)
  on.exit(stopCluster(cl))
  clusterEvalQ(cl, library(nnet))
  clusterExport(cl, c('one_run','beta','predictor_names','predictor_probability','n'),
      envir=environment())
  clusterSetRNGStream(cl, iseed=12345 + n)
  fits <- parLapply(cl, seq_len(B), function(i) one_run(n, predictor_probability, beta,
      predictor_names))
  good <- Filter(Negate(is.null), fits)
  R <- length(good)
  if (R == 0L) stop('No successful fits for n=', n)
  rows <- vector('list', nrow(beta)*ncol(beta)); idx <- 1L
  for (cat in rownames(beta)) for (coef in colnames(beta)) {
    vals <- vapply(good, function(z) z[cat,coef], numeric(1))
    truth <- beta[cat,coef]; avg <- mean(vals)
    rows[[idx]] <- data.frame(sample_size=n, outcome_category=cat, coefficient=coef,
      true_value=truth, mean_estimate=avg, bias=avg-truth,
      empirical_standard_error=sd(vals), MSE=mean((vals-truth)^2),
      successful_runs=R, failed_runs=B-R)
    idx <- idx+1L
  }
  out <- do.call(rbind, rows)
  write.csv(out, paste0('/home/ubuntu/monte_carlo/results_n_', n, '.csv'),
      row.names=FALSE)
  message(format(Sys.time()), ' finished n=', n, '; successful=',R,'/',B)
  out
}
```

```
all <- lapply(sample_sizes, run_n)
results <- do.call(rbind, all)
write.csv(results, '/home/ubuntu/monte_carlo/monte_carlo_results_all.csv',
    row.names=FALSE)
write.csv(unique(results[,c('sample_size','successful_runs','failed_runs')]),
        '/home/ubuntu/monte_carlo/convergence_summary.csv', row.names=FALSE)
message(format(Sys.time()), ' all sample sizes complete')
```

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.jpg
- Table2.jpg
- Table1.jpg
- Table2.jpg