Sample Size and Robustness of Inferences from Logistic Regression in the Presence of Nonlinearity and Multicollinearity

Jason S. Bergtold[1], Elizabeth A. Yeager[2] and Allen Featherstone[3]

[1] Assistant Professor, Department of Agricultural Economics, Kansas State University, 304G Waters Hall, Manhattan, KS 66506-4011, (785) 532-0984, bergtold@ksu.edu (contacting author).

[2] Graduate Research Assistant, Department of Agricultural Economics, Kansas State University, 304G Waters Hall, Manhattan, KS 66506-4011

[3] Professor, Department of Agricultural Economics, Kansas State University, 304G Waters Hall, Manhattan, KS 66506-4011

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011.

Sample Size and Robustness of Inferences from Logistic Regression in the Presence of Nonlinearity and Multicollinearity

Abstract

The logistic regression models has been widely used in the social and natural sciences and results from studies using this model can have significant impact. Thus, confidence in the reliability of inferences drawn from these models is essential. The robustness of such inferences is dependent on sample size. The purpose of this study is to examine the impact of sample size on the mean estimated bias and efficiency of parameter estimation and inference for the logistic regression model. A number of simulations are conducted examining the impact of sample size, nonlinear predictors, and multicollinearity on substantive inferences (e.g. odds ratios, marginal effects) and goodness of fit (e.g. pseudo-R^2, predictability) of logistic regression models. Findings suggest that sample size can affect parameter estimates and inferences in the presence of multicollinearity and nonlinear predictor functions, but marginal effects estimates are relatively robust to sample size.

Keywords: Logistic Regression Model, Multicollinearity, Nonlinearity, Robustness, Small Sample Bias,

Sample Size and Robustness of Inferences from Logistic Regression in the Presence of Nonlinearity and Multicollinearity

Introduction

A quick literature search identifies that logistic regression models have been widely used in the social and natural sciences to examine a myriad of problems. The results of these studies can have tremendous impacts (e.g. prediction of default or treatment effects), thus confidence in the reliability of inferences drawn from these models is essential. The robustness of such inferences is dependent on sample size. For example, in medical and experimental research selecting a sample size is crucial in that too small a sample could affect the meaningfulness of results obtained, but too large of sample could expose an excess number of individuals to study treatments (Biau et al., 2008).

Hosmer and Lemeshow (2000) identify two sample size issues: how many observations do I need and do I have enough data to fit my model? Despite these two issues, the authors are surpised by the “surprisingly little work on sample size for logistic regression” (Hosmer and Lemeshow 2000, p. 339). A number of quantitative measures for estimating the number of observations that may be appropriate have been suggested, but these measures provide no clear indication of the effect on the robustness of inferences made using the model. This is particularly an issue for stated choice methods in environmental and natural resource economics, especially in estimating willingness-to-pay (WTP) and other similar measures.

Greenland (2000) and Greenland et al. (2000) have addressed the impact of small sample bias examining studies on electrical wiring, as well as childhood cancer and caloric intake. These examples were chosen because unusually high odds ratios led to suspicion of the results.

A concern arises in the plausibility of small sample bias going unnoticed due to plausible results that do not trigger an issue. A number of small-sample corrections were compared to see how well they performed. In addition, small perturbations to the data were made to observe how large an impact these changes have on estimation results. Both studies argue that small-sample bias needs to be checked more often.

Nemes et al. (2009) show that as sample size increases the size of bias in logistic regression parameter estimates approaches zero. They don't offer recommendations for correcting for small sample sizes, but instead provide guidelines for when larger samples may be more appropriate. Researchers have tried to show the magnitude of small-sample bias through various means of bias correction (Carroll and Pederson 1990; Firth 1993; MacKinnon and Smith Jr. 1997; Bull et al. 2007). However, it is difficult to know when to correct for bias and how concerned individuals should be with bias due to small samples.

The purpose of this study is to examine the impact of sample size on the bias and efficiency of parameter estimation in the logistic regression model. A number of simulations are conducted examining the impact of sample size, nonlinear predictors, and multicollinearity on substantive inferences (e.g. odds ratios, marginal effects) and goodness of fit (e.g. pseudo-R^2, predictability) of logistic regression models.

Simulation Methods

Simulations were conducted following the specification of the logistic regression model in Bergtold et al. (2010), which approaches the specification of the index or predictor function of the logit model using the inverse conditional distribution. That is,

$$Y_i = h(\mathbf{X}_i;\beta) + u_i = \left[1 + \exp\{-\eta(\mathbf{X}_i;\beta)\}\right]^{-1} + u_i\,, \tag{1}$$

where Y_i is a binary dependent variable, $\mathbf{X}_i$ is a vector of explanatory variables, β is a vector of parameters to be estimated, u_i is an IID stochastic error term, and:

$$\eta(\mathbf{X}_i;\beta) = \ln\left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i;\theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i;\theta_0)}\right) + \kappa, \qquad (2)$$

where $\kappa = \ln\left(\frac{p}{1-p}\right)$ and $f_{\mathbf{X}|Y}(\mathbf{X}_i;\theta_j)$ is the inverse conditional distribution of $\mathbf{X}_i$ on Y_i. The model specification arises from the link between the logistic regression model and discriminant analysis. That is, from the relationship between the conditional distribution of Y_i on $\mathbf{X}_i$ and the inverse conditional distribution of $\mathbf{X}_i$ on Y_i, which is a requirement for the existence of an underlying joint distribution from which the model arises (Bergtold et al., 2010). This model specification approach provides a more parsimonious description of the model and allows for the specification and simulation of models with multicollinearity and nonlinear index/predictor functions.

Using this modeling approach, logistic regression models can be simulated using a two-step procedure, which involves the inverse conditional distribution(s). First, a realization of the dependent Bernoulli random variable is generated. Using this variable as a conditioning factor, the explanatory variables are generated using the inverse conditional distribution. With multiple explanatory variables the modeler is dealing with a multivariate inverse conditional distribution. To make the data generation more tractable, these distributions can be decomposed into products of conditional distributions, allowing one to generate realizations of the multivariate distribution in a sequential manner. For example, the inverse conditional distribution $f_{\mathbf{X}|Y}(X_1, X_2;\theta_j)$ can be simulated by first simulating $f_{X_1|Y}(X_1;\theta_{1,j})$ and then using the results to simulate $f_{X_1|X_2,Y}(X_2 \mid X_1;\theta_{2,j})$. The advantage of this data generation approach is that it allows for a purely statistical method to generate binary choice process data without any *a priori* theoretical assumptions. Furthermore, exact formulas for the parameters, β, can be derived as functions of the parameters of the inverse conditional distributions (Bergtold et al. 2010).

To examine potential bias in the presence of multicollinearity and nonlinear predictor functions, four cases are examined: (1) a base case or bivariate inverse conditional normal distribution with no correlation; (2) bivariate inverse conditional normal distribution with correlation between the explanatory variables equal to 0.95; (3) a mixture inverse conditional distribution between gamma and Bernoulli distributed random variables; and (4) a mixture inverse conditional distribution between exponential and Bernoulli random variables. The resulting four logistic regression models with inverse conditional distribution assumptions are presented in Table 1. Of interest is that the two inverse conditional mixture distributions give rise to models that are nonlinear in the variables and include interaction terms as a result of the distributional assumptions (see Bergtold et al. 2010).

Using the models in Table 1, Monte Carlo simulations were conducted for differing sample sizes of N = 50; 100; 250; 500; 1000; 2,500; 5,000; 10,000; 20,000; 30,000; 40,000; and 50,000 to examine estimated bias as sample size increases. Each Monte Carlo simulation was conducted using 1,000 runs. For each run, logistic regression models were estimated, along with marginal effects, odds ratios, pseudo-R^2, and within-sample prediction percentage for each sample realization generated. From these statistics, estimated bias was calculated as the difference between the statistic and the “true value”. While the true values for the β parameters and odds ratios are exact, for marginal effects, pseudo-R^2 values and within sample prediction, these statistics are all functions of $\mathbf{X}_i$. Thus, the “true values” for these latter statistics represent the means of the calculated “true” measures across all runs using the randomly generated data and true values of the β parameters for each model. Data, simulation and model estimation were all carried out in MATLAB.

Results

Results for the Monte Carlo simulations are reported in Tables 2 to 5 for each of the cases examined in Table 1. Mean estimated bias in coefficient estimates (β) are as high as 300 percent above and below the "true value" of the coefficient for small samples up to 100 observations. This bias becomes significantly less with sample sizes above 250. Estimated biases in coefficient estimates are the most pronounced for case (iii) with the mixture inverse conditional distribution of the gamma and Bernoulli random variables. While many researchers utilize odds ratios for inference, given the limited interpretability of the coefficient estimates, the bias from the coefficient estimates, results in the same significant bias in these measures, as well.

A significant and unexpected result in all the models estimated is that marginal effects were relatively robust to sample size. That is, the mean estimated bias in marginal effects estimates was on an order of magnitude of 1 x 10^{-3} or less for many of the simulation runs, even at small sample sizes. The exception is that in the models with nonlinear predictor functions the confidence intervals for the estimated bias are quite wide. For example, for case (iii) the "true" marginal effect for N = 100 for X_1 is 0.11 with mean estimated bias of 0.02. The 95 percent confidence interval around the bias estimate is -0.22 to 0.32, giving an approximate range for the marginal effect of -0.11 to 0.44, which is quite large and changes in sign. These results can have significant bearing on inferences obtained from the logistic regression model. It seems to suggest, that marginal effects provide a more accurate and robust measure of the impact of a variable on the probability that Y_i = 1, than inference from using odds ratios, which is more likely to be biased in small samples. This is in realization that the confidence interval around (or statistical significance) of the marginal effects may be large (or insignificant). Thus, small sample size does play a role on the variation in marginal effect estimates.

The estimated mean bias in measures of fit (e.g. McFadden Pseudo-R^2 and within sample predictive ability) was not greater than 0.08 for any of the models or sample sizes examined. The only significant difference was the result for cases (iii) and (iv), where predictive ability was approximately lower by 34 and 30 percent for all sample sizes, respectively.

The presence of multicollinearity, surprisingly, seems to have no significant impact on the estimated bias results when comparing them to the bivariate normal case (i) where the two explanatory variables are independent. Given the numerical issues that can be caused by the presence of multicollinearity (Greene, 2003), evidence from this study suggests that in some cases multicollinearity may not result in any significant bias beyond that caused by a small sample size.

The two cases of nonlinear index or predictor functions ((iii) and (iv)) did result in greater bias in coefficient estimates and other statistics. This result is to be expected with a more highly nonlinear function, as more data or degrees of freedom (i.e. information) may be needed to accurately estimate the parameters of the model. Increasing accuracy with increased data is the case for all the parameters and statistics estimated. As more information or data is provided, the estimated bias declines and the confidence intervals around the bias tighten.

Conclusion

This study re-examines the robustness and efficiency of inferences from logitistic regression models and finds that the impact of small sample bias is dependent on the type of model estimated, nature of the data and inference being conducted. Furthermore, given the close relationship between the logistic regression and probit models, these results should extend to those models, as well. If the objective of the study is to obtain meaningful and interpretable marginal effects, sample size, while still an important consideration, may not be as large of an issue as previously thought.

References

Bergtold, J.S. A. Spanos and E. Onukwugha. “Bernoulli Regression Models: Revisiting the Specification of Statistical Models with Binary Dependent Variables.” *Journal of Choice Modeling*, 3(2010): 1 – 28.

Biau, David Jean, Solen Kerneis, and Raphael Porcher. "Statistics in Brief: The Importance of Sample Size in the Planning and Interpretation of Medical Research." *Clinical Orthopaedics and Related Research* 466, no. 9 (2008): 2282-2288.

Bull, Shelley B., Juan Pablo Lewinger, and Sophia S. F. Lee. “Confidence intervals for multinomial logistic regression in sparse data.” *Statistics in Medicine* 26 (2007): 903-918.

Carroll, R.J., and Shane Pederson. *Further Remarks On Robustness In The Logistic Regression Model.* Technical Report #90-43C, Purdue University: Department of Statistics, 1990.

Greenland, Sander. "Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators." *Biostatistics* 1, no. 1 (2000): 113-122.

Greene W.H. (2003). *Econometric Analysis*, 5th ed., Prentice Hall: Upper Saddle River, NJ.

Greenland, Sander, Judith A. Schwartzbaum, and William D. Finkle. "Problems due to Small Samples and Sparse Data in Conditional Logistic Regression Analysis." *American Journal of Epidemiology* 151, no. 5 (2000): 531-539.

Firth, David. “Bias reduction of maximum likelihood estimates.” *Biometrika* 80, no. 1 (1993): 27-38.

Hosmer, David W., and Stanley Lemeshow. *Applied Logistic Regression.* Second Edition. John Wiley & Sons, Inc., 2000.

MacKinnon, James G., and Anthony A. Smith Jr. "Approximate Bias Correction in Econometrics." 1997.

Nemes, Szilard, Junmei Miao Jonasson, Anna Genell, and Gunnar Steineck. "Bias in odds ratios by logistic regression modelling and sample size." *BMC Medical Research Methodology* 9, no. 56 (July 2009): 1-5.

Table 1: Inverse Conditional Distribution Assumptions for Monte Carlo Simulations

Inverse Conditional Distributional Assumptions	$\boldsymbol{P}(Y_i = 1) =$	$h(\boldsymbol{X_i};\boldsymbol{\beta})$ and $f_{X\mid Y=i}(\boldsymbol{X};\theta_{\boldsymbol{i}})$
Bivariate Normal	0.50	$h(\boldsymbol{X_i};\boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big\vert Y = 0 \sim N\left(\begin{bmatrix} 1 \\ 1.5 \end{bmatrix}\begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}\right);\ \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big\vert Y = 1 \sim N\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}\right);$
Bivariate Normal with Multicollinearity	0.50	$h(\boldsymbol{X_i};\boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \vert Y = 0 \sim N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 1.25 & \rho(1.25)^2 \\ \rho(1.25)^2 & 1.25 \end{bmatrix}\right)$ $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \vert Y = 1 \sim N\left(\begin{bmatrix} 1.5 \\ 3 \end{bmatrix}\begin{bmatrix} 1.25 & \rho(1.25)^2 \\ \rho(1.25)^2 & 1.25 \end{bmatrix}\right)$ and $\rho = 0.95$
Gamma and Bernoulli Mixture	0.50	$h(\boldsymbol{X_i};\boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln(X_2) + \beta_4 X_1 X_2 + \beta_5 X_1 \ln(X_2)$ $X_1 \vert Y = 0 \sim Bernoulli(0.25);\ X_1 \vert Y = 1 \sim Bernoulli(0.75);$ $X_2 \vert X_1 = 0, Y = 0 \sim Gamma(1,1);\ X_2 \vert X_1 = 0, Y = 1 \sim Gamma(1,2) \vert$ $X_2 \vert X_1 = 1, Y = 0 \sim Gamma(1.5,3);$ and $X_2 \vert X_1 = 1, Y = 1 \sim Gamma(2,4);$
Exponential and Bernoulli Mixture	0.50	$h(\boldsymbol{X_i};\boldsymbol{\beta}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$ $X_1 \vert Y = 0 \sim Bernoulli(0.8);\ X_1 \vert Y = 1 \sim Bernoulli(0.2);$ $X_2 \vert X_1 = 0, Y = 0 \sim Exponential(1);\ X_2 \vert X_1 = 0, Y = 1 \sim Exponential(1.5) \vert$ $X_2 \vert X_1 = 1, Y = 0 \sim Exponential(2);$ and $X_2 \vert X_1 = 1, Y = 1 \sim Exponential(2.7);$

Table 2: Monte Carlo Simulation Results for the Normal Bivariate Inverse Condition Distribution with Zero Correlation Case

N	**Beta0**		**Beta1**		**Beta2**		**Odds Ratio 1**		**Odds Ratio 2**	
	True Values									
	-3.12		0.64		0.96		1.8965		2.6117	
	Estimated Bias									
50	-0.4393		0.09432		0.13056		0.3571		0.76702	
	-0.51896	-0.35964	0.071434	0.11721	0.10354	0.15758	0.28647	0.42772	0.58074	0.9533
100	-0.1759		0.039811		0.053572		0.13426		0.23672	
	-0.22256	-0.12923	0.025248	0.054375	0.038164	0.068981	0.10249	0.16603	0.18604	0.2874
250	-0.064475		0.0064747		0.02204		0.031924		0.089043	
	-0.092943	-0.036008	-0.0023194	0.015269	0.01269	0.031389	0.014456	0.049392	0.063232	0.11485
500	-0.041715		0.0035476		0.016556		0.016848		0.058703	
	-0.061985	-0.021445	-0.0028054	0.0099006	0.0099936	0.023119	0.0044873	0.029209	0.040778	0.076628
1000	-0.019176		0.004123		0.0040572		0.012163		0.017292	
	-0.032416	-0.0059354	-4.73E-05	0.0082932	-0.00035687	0.0084713	0.0041474	0.020179	0.0056132	0.028971
2500	-0.0020245		0.0006251		0.00025679		0.00291		0.0034608	
	-0.010517	0.0064676	-0.0020155	0.0032657	-0.0026049	0.0031185	-0.0021201	0.00794	-0.0040549	0.010977
5000	-0.0069958		0.0011127		0.0027249		0.0029373		0.0084445	
	-0.01298	-0.0010114	-0.00071486	0.0029402	0.00075836	0.0046914	-0.00054245	0.0064171	0.0032842	0.013605
10000	-0.0018396		0.00072342		0.00065828		0.0018001		0.0023702	
	-0.0060833	0.0024041	-0.00059224	0.0020391	-0.00072452	0.0020411	-0.00070145	0.0043016	-0.0012495	0.0059898
20000	-0.0044734		0.00080104		0.0010745		0.0017406		0.0031821	
	-0.0075821	-0.0013648	-0.0001446	0.0017467	2.52E-05	0.0021239	-5.62E-05	0.0035375	0.00043925	0.005925
30000	-0.0013061		-0.00023356		0.000648		-0.00030035		0.0019087	
	-0.0037395	0.0011273	-0.00099388	0.00052676	-0.00014857	0.0014446	-0.0017422	0.0011415	-0.00017485	0.0039922
40000	0.0002796		-0.00068615		0.00042796		-0.0011912		0.0012785	
	-0.0018244	0.0023836	-0.001353	-1.93E-05	-0.0002593	0.0011152	-0.0024551	7.26E-05	-0.00051891	0.0030759
50000	0.0010452		-8.71E-05		-0.00080541		-7.55E-05		-0.0019731	
	-0.0008914	0.0029817	-0.00068996	0.00051577	-0.0014232	-0.0001876	-0.001219	0.0010679	-0.0035858	-0.00036046

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 2 continued.

N	**Marginal Effect 1**			**Marginal Effect 2**			**Predictive Ability**			**McFadden Pseudo R^2**		
	Estimated True Value	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***	
50	0.11199	0.0011784		0.16799	-0.00059905		0.7642	0.01372		0.29771	0.034554	
		-0.0014114	0.0037682		-0.0027552	0.0015571		0.011314	0.016126		0.032322	0.036785
100	0.10132	7.17E-05		0.15198	-0.00072551		0.76557	0.00587		0.29948	0.016701	
		-0.0017302	0.0018736		-0.0021589	0.00070792		0.0044502	0.0072898		0.015663	0.017739
250	0.10346	-0.0011079		0.15519	0.00022394		0.76433	0.003196		0.29826	0.0069371	
		-0.0022509	3.50E-05		-0.00069417	0.0011421		0.0024401	0.0039519		0.006514	0.0073601
500	0.10407	-0.00079147		0.1561	0.00064511		0.76488	0.001404		0.30041	0.0035399	
		-0.0016163	3.33E-05		1.50E-05	0.0012752		0.0009703	0.0018377		0.0033221	0.0037577
1000	0.10229	0.00011516		0.15343	-0.00016348		0.76456	0.0009		0.29956	0.0016411	
		-0.00044126	0.00067157		-0.00060416	0.00027719		0.00063579	0.0011642		0.0015348	0.0017474
2500	0.10103	-2.95E-05		0.15154	-0.00015578		0.76411	0.000176		0.2988	0.00067032	
		-0.00038404	0.00032504		-0.00044222	0.00013066		4.47E-05	0.00030726		0.00062864	0.00071199
5000	0.10291	-1.60E-05		0.15437	0.00014073		0.7649	0.0001134		0.30018	0.00033706	
		-0.00025824	0.00022624		-5.37E-05	0.0003352		3.04E-05	0.00019638		0.00031674	0.00035739
10000	0.10253	4.14E-05		0.1538	-4.24E-06		0.76457	6.83E-05		0.29984	0.00016457	
		-0.00013044	0.00021332		-0.00013874	0.00013025		2.18E-05	0.00011478		0.00015424	0.00017491
20000	0.10264	4.68E-05		0.15397	4.89E-05		0.76465	3.37E-05		0.30002	8.75E-05	
		-7.86E-05	0.00017218		-5.41E-05	0.00015199		5.12E-06	6.23E-05		8.21E-05	9.28E-05
30000	0.10249	-6.66E-05		0.15373	6.05E-05		0.76468	1.07E-05		0.29983	5.36E-05	
		-0.0001675	3.44E-05		-1.84E-05	0.00013948		-9.03E-06	3.05E-05		5.05E-05	5.68E-05
40000	0.10231	-0.00011467		0.15347	6.10E-05		0.7646	6.30E-06		0.29978	4.29E-05	
		-0.00020452	-2.48E-05		-8.45E-06	0.00013055		-1.12E-05	2.38E-05		4.03E-05	4.55E-05
50000	0.10259	1.68E-05		0.15388	-8.19E-05		0.76449	1.01E-05		0.29963	3.37E-05	
		-6.18E-05	9.54E-05		-0.00014205	-2.17E-05		-3.79E-06	2.40E-05		3.18E-05	3.56E-05

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 3: Monte Carlo Simulation Results for the Normal Bivariate Inverse Condition Distribution with Correlation Equal to 0.95.

N	**Beta 0**		**Beta 1**		**Beta 2**		**Odds Ratio 1**		**Odds Ratio 2**	
	True Value									
	-2.7789		-1.1789		2.0211		0.3076		7.5463	
	Estimated Bias									
50	-0.37802		-0.14708		0.26402		0.048851		16.884	
	-0.44914	-0.3069	-0.19902	-0.095151	0.20601	0.32204	0.031215	0.066488	5.261	28.506
100	-0.15122		-0.0583		0.10791		0.020541		2.4709	
	-0.19393	-0.10851	-0.090191	-0.026408	0.074075	0.14175	0.01053	0.030552	1.9377	3.0042
250	-0.055032		-0.034407		0.045127		0.0042862		0.80612	
	-0.081153	-0.028912	-0.053801	-0.015014	0.024504	0.06575	-0.0017863	0.010359	0.62319	0.98904
500	-0.035825		-0.027769		0.034773		-0.0013338		0.48929	
	-0.054498	-0.017153	-0.041307	-0.014232	0.020293	0.049253	-0.0054249	0.0027572	0.36565	0.61294
1000	-0.017227		-0.0031052		0.0079787		0.0024662		0.15515	
	-0.029376	-0.0050779	-0.012369	0.006159	-0.001744	0.017701	-0.00039545	0.0053279	0.079035	0.23126
2500	-0.0013356		0.00045241		7.57E-05		0.0015647		0.039862	
	-0.0091336	0.0064625	-0.0055186	0.0064234	-0.0062266	0.0063779	-0.00027729	0.0034066	-0.0084712	0.088195
5000	-0.006065		-0.0040919		0.0057555		-0.0005963		0.062178	
	-0.01156	-0.0005701	-0.0081643	-1.96E-05	0.0014175	0.010093	-0.0018436	0.00065102	0.02906	0.095297
10000	-0.001512		-0.00052415		0.0013422		0.00015526		0.019318	
	-0.0054182	0.0023942	-0.0033376	0.0022893	-0.0017111	0.0043956	-0.00071044	0.001021	-0.0038578	0.042494
20000	-0.0040521		-0.0013583		0.0023501		-0.00023274		0.023021	
	-0.006908	-0.0011963	-0.0035084	0.00079184	3.57E-05	0.0046644	-0.00089424	0.00042875	0.0055158	0.040527
30000	-0.0011303		-0.0014951		0.0014121		-0.00035072		0.013708	
	-0.0033722	0.0011117	-0.0031456	0.00015539	-0.00034707	0.0031713	-0.00085807	0.00015663	0.00038649	0.02703
40000	0.00030869		-0.0016012		0.0010224		-0.00040748		0.0099782	
	-0.0016296	0.002247	-0.0030577	-0.00014472	-0.00049312	0.0025379	-0.00085451	3.96E-05	-0.0014957	0.021452
50000	0.00075692		0.0014526		-0.0016882		0.0005105		-0.0109	
	-0.0010294	0.0025433	0.00019503	0.0027102	-0.0030544	-0.000322	0.00012301	0.00089799	-0.021199	-0.0006008

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 3 continued.

N	**Marginal Effect 1**			**Marginal Effect 2**			**Predictive Ability**			**McFadded Pseudo R^2**		
	Estimated True Value	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***	
50	-0.2128	0.0022831		0.3648	-0.0012638		0.751	0.01512		0.27099	0.034043	
		-0.0040419	0.008608		-0.0065476	0.0040199		0.01269	0.01755		0.031835	0.036252
100	-0.19503	0.001744		0.33434	-0.0018563		0.75295	0.00618		0.27271	0.01655	
		-0.0026795	0.0061675		-0.0053767	0.0016642		0.0047449	0.0076151		0.01552	0.01758
250	-0.19975	-0.0015962		0.34243	0.00050943		0.75144	0.003432		0.27141	0.0068499	
		-0.0044152	0.0012228		-0.0017597	0.0027786		0.0026512	0.0042128		0.0064293	0.0072705
500	-0.20039	-0.0021903		0.34353	0.0015695		0.75252	0.001296		0.27346	0.0035022	
		-0.0041638	-0.00021674		9.87E-06	0.0031291		0.00085128	0.0017407		0.003286	0.0037184
1000	-0.19715	0.00049468		0.33796	-0.00041284		0.75197	0.000761		0.27276	0.0016204	
		-0.00087573	0.0018651		-0.0015004	0.00067471		0.0004906	0.0010314		0.001515	0.0017257
2500	-0.19478	0.0003172		0.33392	-0.00040065		0.7514	0.000296		0.27199	0.00066176	
		-0.00056681	0.0012012		-0.0011058	0.00030448		0.00015963	0.00043237		0.00062055	0.00070297
5000	-0.19801	-0.00032163		0.33945	0.00034538		0.75223	0.0001972		0.27333	0.00033204	
		-0.00092206	0.00027881		-0.00013574	0.0008265		0.0001134	0.000281		0.00031192	0.00035216
10000	-0.19742	4.65E-05		0.33843	-7.42E-06		0.75189	6.77E-05		0.27302	0.00016253	
		-0.00036802	0.00046105		-0.00034046	0.00032562		2.13E-05	0.00011408		0.00015228	0.00017278
20000	-0.19761	-7.53E-05		0.33877	0.00013365		0.75195	3.44E-05		0.27318	8.66E-05	
		-0.00039002	0.00023947		-0.00012129	0.0003886		4.50E-06	6.42E-05		8.13E-05	9.19E-05
30000	-0.19734	-0.0001988		0.3383	0.00014753		0.75201	8.47E-06		0.27299	5.31E-05	
		-0.00044377	4.62E-05		-4.74E-05	0.00034249		-1.27E-05	2.97E-05		4.99E-05	5.62E-05
40000	-0.19705	-0.00025907		0.33779	0.000156		0.75196	1.69E-05		0.27295	4.22E-05	
		-0.00047751	-4.06E-05		-1.53E-05	0.00032726		-1.77E-08	3.38E-05		3.97E-05	4.48E-05
50000	-0.19747	0.00019177		0.33852	-0.0001955		0.75186	1.04E-05		0.27283	3.34E-05	
		5.36E-06	0.00037818		-0.00034457	-4.64E-05		-4.53E-06	2.54E-05		3.15E-05	3.53E-05

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 4: Monte Carlo Simulation Results for a Mixture Inverse Condition Distribution Between Gamma and Bernoulli Distributed Random Variables.

N	**Beta 0**		**Beta 1**		**Beta 2**		**Beta 3**		**Beta 4**		**Beta 5**	
	True Values											
	2.62574848		1.64491923		0.66666667		-0.4166667		0.5		0.5	
	Estimated Bias											
50	-6.728523894		5.446008218		2.228236875		-1.661193211		0.139926713		-0.234565971	
	-119.80584	2.56373358	-8.7737448	117.674902	-3.8039074	43.5263565	-43.01947	7.87242512	-21.704779	10.7075909	-15.307275	23.308977
100	-0.584880815		0.048508522		0.050036919		0.144990752		0.505315336		-0.452233232	
	-3.6589348	1.44233659	-3.8957519	3.62276112	-1.292285	1.62441509	-1.679957	2.63028236	-1.4409128	4.51130038	-5.6923227	3.31578489
250	-0.133226846		-0.042807105		0.016502363		0.037919721		0.13713916		-0.087576362	
	-1.5253116	0.97656051	-1.6718764	1.77295266	-0.6909212	0.87386773	-0.9364058	1.01352702	-0.8886392	1.71977127	-2.1910497	1.85099491
500	-0.063574268		-0.04187495		0.008588658		0.016458231		0.06330618		-0.032324493	
	-1.0084145	0.76307032	-1.145314	1.09074812	-0.4962604	0.59223522	-0.6703569	0.62142777	-0.6656807	1.16313975	-1.3374729	1.22037311
1000	-0.04366258		0.006140461		0.007867736		0.001708516		0.027203051		-0.011572763	
	-0.7127055	0.52954444	-0.7743222	0.82613306	-0.3515444	0.3969059	-0.4309039	0.44455164	-0.5188683	0.76466259	-0.920347	0.82539998
2500	0.001215997		-0.018709627		-0.001188087		0.008147906		0.012571602		-0.012669222	
	-0.4036259	0.3852039	-0.5298013	0.48073709	-0.2315735	0.24138413	-0.2729959	0.28931863	-0.3501233	0.43270868	-0.6200041	0.5614764
5000	-0.005141992		-0.005717442		-6.84E-05		0.002542473		0.006568034		-0.002396295	
	-0.28252	0.2505945	-0.3306515	0.32709773	-0.1645771	0.16227077	-0.1837817	0.19361023	-0.2481631	0.29547656	-0.4116538	0.40237235
10000	-0.006103698		0.002792809		0.002788147		-0.00191837		-0.002041385		0.002140586	
	-0.2089194	0.1752754	-0.2347098	0.26151414	-0.1095641	0.1129043	-0.1390995	0.13563651	-0.1734113	0.18513523	-0.275272	0.28312551
20000	0.000301382		-0.002005706		-0.000760672		0.002860551		0.002357807		-0.007651579	
	-0.1412841	0.13293404	-0.169569	0.17070442	-0.0773387	0.08053514	-0.0963666	0.09719118	-0.1279681	0.13759485	-0.2189022	0.19952046
30000	-0.00258092		0.003404174		0.001044626		-0.001306721		0.000597763		-0.000308217	
	-0.1181133	0.10996141	-0.1362082	0.14091062	-0.0638419	0.07134054	-0.0839338	0.07510753	-0.1153316	0.11375535	-0.172404	0.16898047
40000	-0.002109787		0.001013691		0.00073992		-0.000552287		0.00022482		0.001344101	
	-0.1023506	0.09888677	-0.1237936	0.13022158	-0.0544192	0.05610819	-0.0691111	0.0655653	-0.0873223	0.10169789	-0.1472932	0.14710332
50000	-8.38E-05		-8.81E-05		-0.000475466		-1.96E-05		0.001543953		-0.000203391	
	-0.0867573	0.08768357	-0.1071106	0.10789061	-0.051578	0.05071426	-0.0629728	0.05803796	-0.0807532	0.08198454	-0.1232747	0.12683537

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 4 continued.

N	**Odds Ratio1**		**Odds Ratio 2**		**Odds Ratio 3**		**Odds Ratio 4**		**Odds Ratio 5**	
	True Value									
	5.18059146		1.94773404		0.65924063		1.64872127		1.64872127	
	Estimated Bias									
50	7.095E+175		4.59E+85		9.60E+19		2.97E+55		3.43E+61	
	-5.1797898	6.61E+51	-1.9043317	1.56E+19	-0.6592406	1729.13623	-1.6487213	73686.0339	-1.6487209	2.1883E+10
100	4.05E+40		5541222.478		832.7629199		159240.294		5175.52114	
	-5.07528	188.784916	-1.4128035	7.93789169	-0.5363701	8.48949504	-1.2584498	148.4510749	-1.6431622	43.7636649
250	2.249939896		0.198335008		0.11176582		0.982323471		0.856810577	
	-4.2071885	25.3238378	-0.9716968	2.71934966	-0.4007957	1.15716633	-0.9707425	7.556503869	-1.4643953	8.84728585
500	0.667586837		0.094127763		0.045426028		0.327300263		0.348024641	
	-3.532521	10.2394391	-0.7619475	1.57381827	-0.3220219	0.56799122	-0.8014046	3.627128429	-1.2159184	3.9378912
1000	0.473443551		0.053374727		0.018730204		0.146594933		0.167457605	
	-2.7922552	6.65430264	-0.5773071	0.94896715	-0.2307861	0.36903677	-0.6674127	1.893176198	-0.9919024	2.11496919
2500	0.066596303		0.012273108		0.012505039		0.056364125		0.052493294	
	-2.1306638	3.19780316	-0.402627	0.53175075	-0.1574953	0.22118657	-0.4870303	0.892662392	-0.7618045	1.24191431
5000	0.047677454		0.006666203		0.005052516		0.027864223		0.032470615	
	-1.4585672	2.00456261	-0.295564	0.34314919	-0.110675	0.14082906	-0.3623351	0.566775293	-0.556355	0.8167238
10000	0.054721477		0.008700768		0.00035446		0.004031106		0.020776356	
	-1.0837782	1.54845448	-0.2021266	0.2328025	-0.085608	0.09576494	-0.26249	0.335318842	-0.3967392	0.53957987
20000	0.008056382		0.000165657		0.002703499		0.00781118		-0.003398239	
	-0.8080247	0.96431581	-0.1449575	0.16335048	-0.0605638	0.06728938	-0.198042	0.243203769	-0.3241381	0.36406599
30000	0.030950857		0.003187693		-0.000315568		0.003738444		0.005551831	
	-0.6596915	0.7839361	-0.1204609	0.14402887	-0.0530741	0.05142081	-0.1795942	0.198634592	-0.2610931	0.30352467
40000	0.015676194		0.002310103		4.84E-05		0.002393717		0.006573731	
	-0.6032168	0.72052054	-0.1031616	0.11240785	-0.0440221	0.04467177	-0.1378633	0.176493908	-0.2258075	0.26127888
50000	0.007226607		-0.000236544		0.000301553		0.004114077		0.003012978	
	-0.5262117	0.59020347	-0.0979134	0.1013255	-0.0402341	0.03939307	-0.1279056	0.14086514	-0.1912173	0.22295676

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 4 continued

N	**Marginal Effect 1**			**Marginal Effect 2**			**Predictive Ability**			**McFadden Pseudo R^2**		
	Estimated True Value	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***	
50	0.1098566	0.018654352		0.11823813	-0.038907646		0.84482	-0.34362		0.5653085	0.075680281	
		-0.2162105	0.32059509		-0.2788967	0.07534172		-0.52	-0.18		0.0104064	0.1889523
100	0.11003332	0.00543655		0.11863512	-0.02262214		0.84895	-0.34894		0.3778433	0.039708793	
		-0.1517403	0.21230351		-0.2264235	0.03657325		-0.46	-0.22		0.0059191	0.1016726
250	0.109606	0.001404061		0.11854331	-0.021021971		0.849896	-0.349416		0.4778066	0.015723877	
		-0.0969144	0.10475964		-0.0786433	0.02255537		-0.424	-0.272		0.0022546	0.0367766
500	0.10985986	-0.001699259		0.11821589	-0.004974642		0.849518	-0.349664		0.4182241	0.007772272	
		-0.0697269	0.06716602		-0.0385977	0.01757352		-0.402	-0.296		0.0008912	0.0189213
1000	0.10983199	0.000701905		0.11841929	-0.002456445		0.849103	-0.348946		0.4585243	0.003927761	
		-0.0489234	0.04715294		-0.0184173	0.01142491		-0.386	-0.313		0.0004813	0.0093843
2500	0.10986121	-0.000319478		0.11850305	-6.50E-05		0.848674	-0.3491944		0.4945821	0.001584473	
		-0.031083	0.03063506		-0.0085509	0.00737015		-0.372	-0.3256		0.0002036	0.0039072
5000	0.10986848	-7.29E-05		0.11839841	-0.000141761		0.8488742	-0.348758		0.4976316	0.000798157	
		-0.0217904	0.02178182		-0.0050084	0.00497842		-0.3652	-0.3328		7.42E-05	0.0020061
10000	0.10993617	-2.77E-06		0.11843617	-4.81E-05		0.8487941	-0.3487364		0.4881336	0.000384867	
		-0.0154425	0.01775937		-0.0034684	0.00352922		-0.36	-0.3373		5.65E-05	0.0009947
20000	0.10995552	0.0001682		0.11854212	1.68E-05		0.84868925	-0.3485632		0.4886731	0.000192186	
		-0.0103527	0.01045643		-0.0023229	0.00243759		-0.35645	-0.34105		2.45E-05	0.00046
30000	0.10988014	3.40E-05		0.11848012	-4.85E-06		0.84894073	-0.348903567		0.5032357	0.000131745	
		-0.0087168	0.00948667		-0.0019988	0.00213229		-0.3556333	-0.3422		1.75E-05	0.000319
40000	0.10989595	4.16E-05		0.11847509	-1.80E-05		0.84900485	-0.348989275		0.4938374	9.75E-05	
		-0.0074905	0.00766787		-0.0018009	0.00167678		-0.355175	-0.343		1.14E-05	0.000258
50000	0.10992241	-3.32E-05		0.11845834	-1.35E-05		0.84883304	-0.34875966		0.5003647	7.77E-05	
		-0.0068374	0.00694927		-0.0016168	0.00164039		-0.3538	-0.34326		1.09E-05	0.0001964

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 5: Monte Carlo Simulation Results for a Mixture Inverse Condition Distribution Between Exponential and Bernoulli Distributed Random Variables.

N	**Beta 0**	**Beta 1**	**Beta 2**	**Beta 3**	**Odds Ratio 1**	**Odds Ratio 2**	**Odds Ratio 3**
				True Values			
	0.69315	-2.6672	0.5	-0.2037	0.069444	1.6487	0.8157
				Estimated Bias			
50	-0.18954	-1.2095	0.54534	-0.35856	3.17E+09	1.05E+15	5.44E+63
	-2.3619 1.6084	-3.7204 2.596	-0.69116 3.9426	-4.8004 1.2553	-0.067762 0.86176	-0.82272 83.343	-0.80899 2.0465
100	-0.052295	-0.071775	0.16222	-0.13385	0.021065	3.6986	-0.0071298
	-1.1599 1.0828	-1.8928 1.4443	-0.45321 1.3483	-1.3546 0.6824	-0.058982 0.22493	-0.60082 4.7002	-0.60521 0.79827
250	-0.019866	-0.035164	0.051011	-0.037312	0.0060998	0.14689	-0.0014173
	-0.64994 0.6882	-1.0514 0.9031	-0.34255 0.63148	-0.63984 0.43173	-0.045177 0.10189	-0.4782 1.4515	-0.38552 0.44042
500	-0.021264	-0.007836	0.033886	-0.024815	0.003671	0.080845	-0.0069858
	-0.5174 0.43572	-0.71293 0.6161	-0.24739 0.37481	-0.40435 0.30231	-0.035402 0.059146	-0.36134 0.74969	-0.2713 0.28793
1000	-2.85E-05	-0.0087034	0.010634	-0.008538	0.0012131	0.02807	-0.00037484
	-0.29951 0.31695	-0.45648 0.42984	-0.17843 0.26012	-0.27593 0.231	-0.025451 0.037293	-0.26943 0.48981	-0.1967 0.21197
2500	-0.00015599	0.00073091	0.0035135	-0.0038808	0.00084462	0.0096113	-0.00078785
	-0.21374 0.1961	-0.3135 0.28513	-0.12346 0.14551	-0.15222 0.1444	-0.018688 0.022912	-0.19148 0.25824	-0.11518 0.12672
5000	-0.0017953	0.0038418	0.0024805	-0.0030562	0.00067023	0.0060494	-0.0012753
	-0.14717 0.13335	-0.20502 0.22496	-0.089093 0.098654	-0.1135 0.097592	-0.012873 0.017519	-0.14054 0.17095	-0.087519 0.08362
10000	-0.00095628	-0.0012287	0.0026861	-0.0014328	0.00011303	0.0054562	-0.00052192
	-0.10459 0.10188	-0.15608 0.13628	-0.065319 0.073607	-0.081926 0.074894	-0.010035 0.010139	-0.10425 0.12594	-0.064163 0.063437
20000	-0.00021355	-0.0021623	0.00065721	-0.00017113	-4.64E-05	0.00158	0.00017468
	-0.074122 0.069926	-0.10554 0.1043	-0.043106 0.052776	-0.057117 0.051937	-0.0069556 0.0076339	-0.069559 0.08935	-0.045285 0.043485
30000	0.00059249	-0.002462	0.00038121	-7.94E-05	-9.91E-05	0.00094289	0.00013685
	-0.058395 0.057829	-0.091641 0.09136	-0.035561 0.040594	-0.04534 0.043843	-0.0060811 0.0066433	-0.057599 0.068305	-0.036158 0.036559
40000	-0.0012882	0.0013877	0.0010451	-0.001345	0.00014621	0.0019582	-0.00094399
	-0.053764 0.051593	-0.071668 0.070794	-0.032556 0.031878	-0.03726 0.039188	-0.0048028 0.0050944	-0.052812 0.053404	-0.029834 0.0326
50000	-0.00036999	-3.73E-05	-2.10E-05	0.00041416	3.78E-05	0.00015024	0.00046075
	-0.049253 0.045912	-0.066106 0.065749	-0.029622 0.029646	-0.034625 0.032912	-0.0044423 0.0047193	-0.048123 0.04961	-0.027761 0.027293

Note: The confidence intervals for each calculated bias are given below the bias estimate

Table 5 continued

N	**Marginal Effect 1**			**Marginal Effect 2**			**Predictive Ability**			**McFadden Pseudo R^2**		
	Estimated True Value	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***		***Estimated True Value***	***Estimated Bias***	
50	-0.59665	-0.0043061		0.059928	0.011965		0.80576	-0.3058		0.30796	0.05493	
		-0.21033	0.22326		-0.093984	0.14511		-0.48	-0.12		-0.0033824	0.1697
100	-0.59665	-0.00094674		0.059935	0.00659		0.80455	-0.30571		0.33193	0.024836	
		-0.16292	0.15669		-0.046365	0.077992		-0.43	-0.18		-0.0013492	0.073541
250	-0.59672	-0.00063074		0.059937	-0.00063074		0.80394	-0.30352		0.30183	0.0097132	
		-0.099558	0.098883		-0.030834	0.04254		-0.38	-0.224		-0.0006031	0.031238
500	-0.59673	0.00046189		0.059894	0.0021134		0.80347	-0.30339		0.30404	0.0049392	
		-0.072399	0.074406		-0.022424	0.027759		-0.36	-0.246		-0.0004012	0.015074
1000	-0.59676	-0.00045135		0.059939	0.00041428		0.80394	-0.30424		0.25728	0.0022873	
		-0.047709	0.044087		-0.015643	0.017358		-0.345	-0.266		-0.0003095	0.006832
2500	-0.59668	0.00017202		0.059904	7.25E-05		0.80363	-0.30377		0.30114	0.00096593	
		-0.030861	0.029834		-0.010891	0.011487		-0.3288	-0.2784		-0.0001033	0.0029436
5000	-0.59674	0.00037502		0.059921	8.92E-05		0.80354	-0.3036		0.32418	0.00048102	
		-0.021182	0.022729		-0.0073228	0.0078557		-0.3222	-0.2848		-3.73E-05	0.0015131
10000	-0.59673	-1.51E-05		0.05992	0.00021806		0.80379	-0.30381		0.32199	0.00023928	
		-0.015233	0.015726		-0.0052545	0.0056533		-0.3166	-0.2901		-1.57E-05	0.0006995
20000	-0.5967	-0.0001752		0.059924	3.97E-05		0.80386	-0.30388		0.31314	0.00012232	
		-0.010965	0.010635		-0.0036399	0.0041125		-0.31315	-0.2952		-5.97E-06	0.0003851
30000	-0.59672	-0.0002688		0.059915	4.84E-06		0.8039	-0.30373		0.33092	8.04E-05	
		-0.009397	0.008929		-0.003003	0.0032509		-0.31083	-0.29647		-2.54E-06	0.0002328
40000	-0.59673	-8.96E-06		0.059919	4.64E-05		0.80378	-0.30367		0.32375	6.02E-05	
		-0.0083003	0.0071868		-0.0028126	0.0025585		-0.31042	-0.29685		-8.13E-06	0.0001698
50000	-0.59671	0.0001662		0.059919	3.30E-05		0.80371	-0.30375		0.32033	4.86E-05	
		-0.0068571	0.0068886		-0.0023079	0.0025878		-0.30962	-0.2978		-3.46E-06	0.0001488

Note: The confidence intervals for each calculated bias are given below the bias estimate