

## ABSTRACT

Title of Dissertation:       Diagnostics for Nonlinear Mixed-Effects Models

Mohamed O. Nagem, Doctor of Philosophy, 2009

Dissertation directed by: Professor Benjamin Kedem  
Department of Mathematics

The estimation methods in Nonlinear Mixed-Effects Models (NLMM) still largely rely on numerical approximation of the likelihood function and the properties of these methods are yet to be characterized. These methods are available in most statistical software packages, such as S-plus and SAS; However approaches on how to assess the reliability of these estimation methods are still open to debate. Moreover, the lack of a common measure to capture the best fitted model is still an open area of research. Common Software packages such as SAS and S-plus do not provide a specific method for computing such a measure other than the traditional Akaike's Information Criterion (AIC) Akaike [2],

Bayesian Information Criterion (BIC) Schwarz [38], or the likelihood ratio. These methods are comparative in nature and are very hard to interpret in this context due to the complex structure and dependent nature of the populations that they were intended to analyze.

This dissertation focuses on approximate methods of estimating parameters of NLMM. In chapter 1, the general form of a NLMM is introduced and real data examples are presented to illustrate the usefulness of NLMM where a standard regression model is not appropriate. A general review of the approximation methods of the log-likelihood function is described. In chapter 2, we compared three approximation techniques, which are widely used in the estimation of NLMM, based on simulation studies. In this chapter we first compared these approximation methods through extensive simulation studies motivated by two widely used data sets. We compared the empirical estimates from three different approximations of the log-likelihood function and their bias, precision, convergence rate, and the 95% confidence interval coverage probability. We compared the First Order approximation (FO) of Beal and Sheiner [5], the Laplace approximation (LP) of Wolfinger

[49], and the Gaussian Quadrature (GQ) of Davidian and Gallant [10]. We also compared these approaches under different sample size configurations and analyzed their effects on both fixed effects estimates and the precision measures. The question of which approximation yields the best estimates and the degree of precision associated with it seems to depend greatly on many aspects. We explored some of these aspects such as the magnitude of variability among the random effects, the random parameters covariance structure, and the way in which such random parameters enter the model as well as the “linearity” or the “close to linearity” of the model as a function of these random parameters. We concluded that, while no method outperformed the others on a consistent basis, both the GQ and LP methods provided the most accurate estimates. The FO method has the advantage that it is exact when the model is linear in the random effects. It also has the advantage of being computationally simple and provides reasonable convergence rates.

In chapter 3 we investigated the robustness and sensitivity of the three approximation techniques to the structure of the random

effect parameters, the dimension of these parameters, and the correlation structure of the covariance matrix. We expanded the work of Hartford and Davidian [18] to assess the robustness of these approximation methods under different scenarios (models) of random effect covariance structures:(1) Under the assumption of single random effect models;(2) under the assumption of correlated random effect models;(3) under the assumption of non-correlated random effect models. We showed that the LP and GQ methods are very similar and provided the most accurate estimates. Even though the LP is fairly robust to mild deviations, the LP estimates can be extremely biased due to the difficulty of achieving convergence. The LP method is sensitive to misspecification of the inter-individual model.

In chapter 4 we evaluated the Goodness of Fit measure (GOF) of Hosmer et. al. [20] and Sturdivant and Hosmer [43] to a class of NLMM and evaluated the asymptotic sum of residual squares statistics as a measure of goodness of fit by conditioning the response on the random effect parameter and using Taylor series approximations in the estimation technique. Simulations of different mixed logistic regression

models were evaluated, as well as the effect of the sample size on such statistics. We showed that the proposed sum of squares residual statistics works well for a class of mixed logistic regression models with the presence of continuous covariates with a modest sample size dataset. However, the same statistics failed to provide an adequate power to detect the correct model in the presence of binary covariates.

Diagnostics for Nonlinear Mixed-Effects Models

by

Mohamed O. Nagem

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2009

Advisory Committee:

Professor Benjamin Kedem, Chairman/Advisor  
Professor Frank Alt  
Professor Bilal Ayyub  
Professor Doron Levy  
Professor Udaya Shankar

© Copyright by

Mohamed O. Nagem

2009

# Acknowledgements

Professor Benjamin Kedem, my advisor, has provided me with all the necessary resources and guidance to complete this dissertation.

I will always be appreciative of the time and energy that he spent working with me. I thank all of my committee members for their time and support. I owe special thanks to my committee members, Professors Benjamin Kedem, Frank Alt, Bilal Ayyub, Doron Levy, and Udaya Shankar.

I dedicate this work to my uncle (truly my only father and mentor) who raised me to be a man and dedicated his life to support me to accomplish all that I have become today. His support was both moral and financial and was always ready when needed. Without him, I would not have found the path to this journey and achieved the many goals that I never imagined possible. So, at the completion of this one goal, I say "say Thank you Mohamed Yahya O. Ahmed O. Nagem."



## TABLE OF CONTENTS

<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>x</b>
<b>1 An Overview of Nonlinear Mixed-Effects Models</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.1.1 Application of NLMM in Medical Studies . . . . .	6
1.1.2 PK Model . . . . .	8
1.2 General Nonlinear Mixed-Effects Model . . . . .	11
1.3 Examples of Data following NLMM . . . . .	13
1.3.1 Theophylline Data . . . . .	13
1.3.2 Carbon Dioxide Data . . . . .	17
1.3.3 Orange Tree Data . . . . .	23
1.3.4 Binary Data Model . . . . .	25

1.4	Problem of Estimation of NLMM . . . . .	26
1.5	Standard Nonlinear Regression Model . . . . .	28
1.6	Approximation Methods of the Likelihood . . . . .	32
1.6.1	First Order Approximation Method (FO) . . . . .	33
1.6.2	Laplace Approximation (LP) . . . . .	36
1.6.3	Gaussian Quadrature (GQ) . . . . .	39
<b>2</b>	<b>Comparison of the Approximation Methods</b>	<b>43</b>
2.1	Simulation Methodology . . . . .	44
2.2	Theophylline data . . . . .	44
2.3	Carbon Dioxide Data . . . . .	49
2.4	Simulation Results . . . . .	50
2.4.1	Theophylline Simulation Results: ( $N = 120$ ) . . . . .	51
2.4.2	Theophylline Simulation Results: ( $N = 240$ ) . . . . .	58
2.4.3	Carbon Dioxide Simulation Results: ( $N = 84$ ) . . . . .	65
2.4.4	Carbon Dioxide Simulation Results: ( $N = 168$ ) . . . . .	76
2.5	Conclusion . . . . .	86
<b>3</b>	<b>Diagnostics of the Random Effects</b>	<b>89</b>

3.1	Simulation . . . . .	90
3.2	Single Random Component Model . . . . .	92
3.2.1	Model 1: Only $\beta_1$ Has an Associated Random Effect . . . . .	93
3.2.2	Model 2: Only $\beta_2$ Has an Associated Random Effect . . . . .	96
3.2.3	Model 3: Only $\beta_3$ Has an Associated Random Effect . . . . .	99
3.3	Correlated Random Parameters Model . . . . .	102
3.4	NonCorrelated Random Parameters Model . . . . .	116
3.5	Conclusion . . . . .	127
<b>4</b>	<b>Goodness of Fit</b>	<b>129</b>
4.1	Motivation . . . . .	129
4.2	Mixed Logistic Model . . . . .	134
4.2.1	Standard Logistic Model . . . . .	136
4.2.2	Mixed Logistic Model . . . . .	136
4.3	Sum of Squares Residual Statistics (SSR) . . . . .	140
4.4	Simulations Studies . . . . .	148

4.4.1	Results . . . . .	150
4.5	Conclusions . . . . .	166
<b>5</b>	<b>Conclusions and Future work</b>	<b>168</b>
5.1	Recommendations for the Use of NLMM . . . . .	171
5.2	Future work . . . . .	173
	<b>Appendix A</b>	<b>177</b>
	<b>Bibliography</b>	<b>201</b>

## LIST OF TABLES

1.1	Subject's Estimate of $\beta_1$ (True $\beta_1 = -3.23$ ) . . . . .	30
1.2	Subject's Estimate of $\beta_2$ (True $\beta_2 = 0.48$ ) . . . . .	30
1.3	Subject's Estimate of $\beta_3$ (True $\beta_3 = -2.46$ ) . . . . .	31
2.1	Empirical Estimates for Theophylline Model (N = 120). . . . .	52
2.2	Empirical Estimates for Theophylline Model (N = 240). . . . .	60
2.3	Empirical Estimates for Carbon Dioxide Model (N = 84). . . . .	67
2.4	Empirical Estimates for Carbon Dioxide Model (N = 168). . . . .	78
3.1	Single Random Component Models . . . . .	93
3.2	Empirical Estimates: $\beta_1$ with a Random Component. . . . .	94
3.3	Empirical Estimates: $\beta_2$ with a Random Component. . . . .	97
3.4	Empirical Estimates: $\beta_3$ with a Random Component. . . . .	100

3.5	Correlated Random Parameters Models . . . . .	104
3.6	Correlated Random Effects: Model 1 ( $\rho = 0.00$ ). . . . .	105
3.7	Correlated Random Effects: Model 2 ( $\rho = 0.25$ ). . . . .	106
3.8	Correlated Random Effects: Model 3 ( $\rho = 0.50$ ). . . . .	107
3.9	Correlated Random Effects: Model 4 ( $\rho = 0.75$ ). . . . .	108
3.10	Correlated Model Estimates for $\beta_1$ (True $\beta_1=-3.23$ ). . . . .	110
3.11	Correlated Model Estimates for $\beta_2$ (True $\beta_2=0.48$ ). . . . .	111
3.12	Correlated Model Estimates for $\beta_3$ (True $\beta_3=-2.46$ ). . . . .	112
3.13	NonCorrelated Random Effects: Model 1 ( $\rho = 0.00$ ). . . . .	117
3.14	NonCorrelated Random Effects: Model 2 ( $\rho = 0.25$ ). . . . .	118
3.15	NonCorrelated Random Effects: Model 3 ( $\rho = 0.50$ ). . . . .	119
3.16	NonCorrelated Random Effects: Model 4 ( $\rho = 0.75$ ). . . . .	120
3.17	NonCorrelated Model Estimates for $\beta_1$ (True $\beta_1=-3.23$ ). . . . .	121
3.18	NonCorrelated Model Estimates for $\beta_2$ (True $\beta_2=0.48$ ). . . . .	122
3.19	NonCorrelated Model Estimates for $\beta_3$ (True $\beta_3=-2.46$ ). . . . .	123
4.1	Skin Irritation Scoring Scale . . . . .	131
4.2	Testing for Normality . . . . .	132
4.3	Range of Dataset: Frequencies of the Mean Irritation Scores . . . . .	133

4.4	Power Estimate with 95% CI Model 1: $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 U_i[-1, 1] +$ $\beta_2 U_{ij}[-2, 2] + b_i$ . . . . .	151
4.5	Power Estimate with 95% CI Model 2: $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 B_i(0.5) +$ $\beta_2 U_{ij}[-1, 1] + b_i$ . . . . .	157
4.6	Power Estimate with 95% CI Model 3: $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 B_i(0.5) +$ $\beta_2 B_{ij}(0.25) + b_i$ . . . . .	162

## LIST OF FIGURES

1.1	Average Curve of Theophylline Concentration . . . . .	16
1.2	Carbon Dioxide Uptake . . . . .	20
1.3	Average Carbon Dioxide Uptake Per Group . . . . .	22
1.4	Orange Tree Data . . . . .	24
2.1	Theophylline Concentration in Serum From 12 individuals . . . . .	47
2.2	$\beta_1$ Estimate Ranges from Model 2.1 with (N = 120) . . . . .	54
2.3	$\beta_2$ Estimate Ranges from Model 2.1 with (N = 120) . . . . .	55
2.4	$\beta_3$ Estimate Ranges from Model 2.1 with (N = 120) . . . . .	57
2.5	$\beta_1$ Estimate Ranges from Model 2.1 with (N = 240) . . . . .	61
2.6	$\beta_2$ Estimate Ranges from Model 2.1 with (N = 240) . . . . .	63
2.7	$\beta_3$ Estimate Ranges from Model 2.1 with (N = 240) . . . . .	64
2.8	$\beta_0$ Estimate Ranges from Model 2.2 with (N = 84) . . . . .	69
2.9	$\beta_1$ Estimate Ranges from Model 2.2 with (N = 84) . . . . .	70



2.10	$\beta_2$ Estimate Ranges from Model 2.2 with (N = 84)	. . .	72
2.11	$\beta_3$ Estimate Ranges from Model 2.2 with (N = 84)	. . .	73
2.12	$\beta_4$ Estimate Ranges from Model 2.2 with (N = 84)	. . .	74
2.13	$\beta_5$ Estimate Ranges from Model 2.2 with (N = 84)	. . .	75
2.14	$\beta_0$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	80
2.15	$\beta_1$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	81
2.16	$\beta_2$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	82
2.17	$\beta_3$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	83
2.18	$\beta_4$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	84
2.19	$\beta_5$ Estimate Ranges from Model 2.2 with (N = 168)	. . .	85
3.1	Empirical Estimates: $\beta_1$ with a Random Component	. . .	95
3.2	Empirical Estimates: $\beta_2$ with a Random Component	. . .	98
3.3	Empirical Estimates: $\beta_3$ with a Random Component	. . .	101
3.4	The Relative Bias Plot for $\beta_1(\rho)$	. . . . .	113
3.5	The Relative Bias Plot for $\beta_2(\rho)$	. . . . .	114
3.6	The Relative Bias Plot for $\beta_3(\rho)$	. . . . .	115
3.7	The Relative Bias Plot for $\beta_1(\rho)$	. . . . .	124
3.8	The Relative Bias Plot for $\beta_2(\rho)$	. . . . .	125

3.9	The Relative Bias Plot for $\beta_3(\rho)$	126
4.1	The SSR Histogram for Model 1 with $N = 20$ and $M = 10$	152
4.2	Percentiles Plot for Model 1 with $N = 20$ and $M = 10$	153
4.3	The SSR Histogram for Model 1 with $N = 20$ and $M = 20$	154
4.4	Percentiles Plot for Model 1 with $N = 20$ and $M = 20$	155
4.5	The SSR Histogram for Model 2 with $N = 20$ and $M = 10$	158
4.6	Percentiles Plot for Model 2 with $N = 20$ and $M = 10$	159
4.7	The SSR Histogram for Model 2 with $N = 20$ and $M = 20$	160
4.8	Percentiles Plot for Model 2 with $N = 20$ and $M = 20$	161
4.9	The SSR Histogram for Model 3 with $N = 20$ and $M = 10$	163
4.10	Percentiles: Model 3 with $N = 20$ and $M = 10$	164
4.11	The SSR Histogram for Model 3 with $N = 20$ and $M = 20$	165
4.12	Percentiles: Model 3 with $N = 20$ and $M = 20$	166
A-1	The Mean Square Error Plot for $\beta_1$ against $\rho$ values.	177
A-2	The Mean Square Error Plot for $\beta_2$ against $\rho$ values.	178
A-3	The Mean Square Error Plot for $\beta_3$ against $\rho$ values.	179
A-4	The Coefficient of Variation Plot for $\beta_1$ against $\rho$ values.	180
A-5	The Coefficient of Variation Plot for $\beta_2$ against $\rho$ values.	181

A-6	The Coefficient of Variation Plot for $\beta_3$ against $\rho$ values.	182
A-7	The Coverage Probability Plot for $\beta_1$ against $\rho$ values.	183
A-8	The Coverage Probability Plot for $\beta_2$ against $\rho$ values.	184
A-9	The Coverage Probability Plot for $\beta_3$ against $\rho$ values.	185
A-10	The Mean Square Error Plot for $\beta_1$ against $\rho$ values.	186
A-11	The Mean Square Error Plot for $\beta_2$ against $\rho$ values.	187
A-12	The Mean Square Error Plot for $\beta_3$ against $\rho$ values.	188
A-13	The Coefficient of Variation Plot for $\beta_1$ against $\rho$ values.	189
A-14	The Coefficient of Variation Plot for $\beta_2$ against $\rho$ values.	190
A-15	The Coefficient of Variation Plot for $\beta_3$ against $\rho$ values.	191
A-16	The Coverage Probability Plot for $\beta_1$ against $\rho$ values.	192
A-17	The Coverage Probability Plot for $\beta_2$ against $\rho$ values.	193
A-18	The Coverage Probability Plot for $\beta_3$ against $\rho$ values.	194

# Chapter 1

## An Overview of Nonlinear Mixed-Effects Models

### 1.1 Introduction

Methodological issues with respect to the application of Nonlinear Mixed-Effects Models (NLMM) are of current interest in the statistics community. Bates and Watt [3], Pinheiro and Bates [27] and Davidian and Giltinan [12]. Linear Mixed-Effects Models (LMM) have been with us for decades (e.g. Searle [39]), whereas NLMM have received considerably more attention only in the last decade. There are many areas of research, including, but not limited to agricultural, ecological, pharmacokinetics, and bioassay studies that require NLMM techniques. Studies with longitudinal data often are easier to interpret using NLMM techniques for repeated measure designs with a continuous response variable than the more traditional linear models. The

general context in which one might be interested in using a Mixed Effects Model (MM) is for multivariate correlated data or longitudinal data. The most attractive feature of the MM is its flexibility. In particular, this type of modeling allows for variation across individuals and it also strengthens each individual's data by "borrowing" from the ensemble. For example, subject specific parameter estimates for an individual with very few data points may still be estimated by "borrowing" information, or data from the group. In addition, the ability to model both within and between subject heterogeneity is also the strength of MM. There are some limitations for the use of NLMM, when the distance between observations are too far, the correlation between the corresponding responses are much less related and therefore a NLMM may not be the best approach to model these responses.

The most common NLMMs are those in which individual response curves share the same functional form, but some particular characteristics vary across individuals. For example, each individual's data are modeled as an exponential, but the rate of increase could vary across

individuals. The general idea is to specify the functional form that each individual's data follow and allow some of those parameters to vary among subjects. Methods for point estimation of population parameters have been addressed for the LMM (e.g., Gallant [14], Sheiner and Beal [40] and Racine-Poon [30]).

The focus of this research is centered on the application of NLMM in the regulatory setting, for example, the application of NLMM in bioequivalence studies, dose response relationship and dose finding studies, as well as toxicologically-based quantitative risk assessment studies.

The Food and Drug Administration (FDA) Guidance for Industry (Statistical Approaches to Establishing Bioequivalence, 2001, and Guidance for Industry Population Pharmacokinetics, 1999) attempts to standardize the techniques used for the assessment of bioequivalence studies. The Guidance is based on the application of the products for various time periods (dose durations), where the responses are measured over a period of time and the areas under the curves ( $AUC$ ) are calculated as a function of the exposure time period to obtain

dose-response relationships. Extrapolation of the *AUC* is needed to compute the limits when time is set to infinity. In the USA, generic drug companies have to demonstrate bioequivalence of a product to the original through comparative clinical trials with a bioequivalence study. Under current FDA standards, bioequivalence is achieved in cases where the innovator and test product differ in terms of their rate and extent of absorption by pre-specified margin (e.g.  $-20/+25$  percent or less).

Another example of an application of NLMM is the toxicologically based quantitative risk assessment. In this area we are concerned with estimating human risks based upon experimental data linking an environmental hazard to a known outcome (tumor incidence, acute toxicity, etc.). Risk, the probability of some adverse response, is often derived from dose-response models which parameterize risk as a function of dose. The NLMM is commonly used to fit the dose response relationship. For example, a logistic mixed model can be appropriate to model the probability of some adverse response as a function of the dose level.

The estimation methods in NLMM largely rely on numerical approximations of the likelihood function, and their properties are yet to be characterized. These methods are available in most statistical software packages such as S-plus and SAS. However, approaches on how to assess the reliability of these estimation methods are still open to debate. Moreover, the lack of a common measure to capture the best fitted model is still an open area of research. Common software packages such as SAS and S-plus do not provide specific methods for computing such measures other than the traditional likelihood based Akaike's Information Criterion (AIC) Akaike [2], Bayesian Information Criterion (BIC) Schwarz [38], or the likelihood ratio that are comparative in nature and are very hard to interpret in this context due to the complex structure and dependent nature of these populations.

This dissertation focuses on approximate methods of estimating parameters of NLMM. Previous works to assess these approximate methods have been mainly focussed on the comparisons of the parameters estimates, both fixed and random effects. The goal of this dissertation is to evaluate and investigate through extensive simula-



tion studies the properties, the robustness, and sensitivity of these methods. We use simulation studies on a number of different models to investigate the strength and weakness of these approximation methods and to give a general comparison of the methods. The evaluation will not only be based on the fixed parameters estimates, but on measures such as mean square error, bias, and coverage probability as well as other measures which are defined in a later chapter. In this chapter, we define the general form of a NLMM and then present real data examples to illustrate the usefulness of NLMMs and the regulatory applications in the drug development process from dose findings to safety studies. We also show that the nonlinear regression model lacks the efficacy and accuracy for these types of studies. We also give a general review of the approximation methods of the log-likelihood function of NLMM.

### **1.1.1 Application of NLMM in Medical Studies**

Population Pharmacokinetics (PK) is the study of variability in plasma concentrations between and within individuals. This type of

study helps in identifying the demographic, pathophysiological, environmental, and drug related factors that contribute to the variability observed in the safety and efficacy of a drug. PK studies offer an advantage of estimating PK parameters in target patient population with sparse sampling methodology. PK uses a NLMM approach. Variability in NLMM is characterized in terms of fixed and random effects. The fixed effects are population average values of pharmacokinetic parameters, such as, clearance, and volume of distribution, that may in turn be a function of patient characteristics. The random effects quantify the variability that is not explained by the fixed effects. These random effects include inter-subject, intra-subject, and residual variability. This methodology has also been extended to model population pharmacokinetic/pharmacodynamics (PK/PD) relationships. This NLMM approach was introduced by Sheiner and Beal [40] approximately 30 years ago.

Therapeutic Equivalence (Bioavailability: BA) for invivo studies focuses on determining the process by which a drug is released from the dosage form and moves to the site of action. BA data provides

an estimate of the fraction of the drug absorbed, as well as its subsequent distribution and elimination. BA can be generally documented by a systemic exposure profile obtained by measuring drug and/or metabolite concentration in the systemic circulation over time.

Bioequivalence (BE) studies establish therapeutic equivalence between two products, a test product (T) and a reference product (R). The FDA has defined bioequivalence between two products as “the absence of a significant difference in the rate and extent to which the active ingredient becomes available at the site of drug action when administered at the same dose under similar conditions in an appropriately designed study, FDA [13].

### **1.1.2 PK Model**

Pharmacokinetics (PK) is the quantitative relationship between the administered dose and the observed concentration of a drug and its metabolites in the body(plasma, tissue) over time. Absorbtion, distribution metabolism, and elimination rates are the most relevant pharmacokinetics measures. PK models are a way to describe the

concentration-time profile of a drug in mathematical terms. For most drugs, the concentration  $C(t)$  can be described as follows:

$$C(t) = \sum_{i=1}^N C_i \exp(-\lambda_i t) \quad (1.1)$$

where  $C_i$ , and  $\lambda_i$  are constant rates, and it is also assumed that  $N \leq 3$ . In this work, we will assume a one compartment model ( $N = 2$ ).

The FDA's guidelines recommend that reliance on systemic exposure measures reflect comparable rate and extent of absorption, which in turn would achieve the underlying statutory and regulatory objective of ensuring comparable therapeutic effects. Bioequivalence studies attempt to gain insight on formulation "switchability" (i.e., the ability to substitute one formulation for another without concern for the potential of reduced effectiveness or increased probability of adverse effects). A key assumption is that switchability may be inferred from plasma concentration vs. time data and metrics reflecting the rate and extent of drug absorption. The area under the plasma concentration vs. time curve ( $AUC$ ) is commonly employed as the

metric describing the extent of drug absorption, while the maximal concentration observed following drug administration ( $C_{max}$ ) is the metric recommended by the FDA to evaluate the rate of drug absorption.  $AUC$  analysis, often used as a measure of drug exposure, plays many important roles in pharmacokinetics.  $AUC$  provides a measure of how much and how long a drug stays in a body.

Exposure measures are defined in relation to early, peak, and total portions of the plasma, serum, or blood concentration-time profile as follows:

- $C_{max}$ : is the maximum drug concentration obtained directly from the data without interpolation (Peak Exposure)
- $AUC_{\infty}$ : is the area under the plasma/serum/blood concentration-time curve from time zero to infinity.

Bioequivalence of the two drugs require that all three PK endpoints should be equivalent in term of bioequivalence intervals using the ratio  $\frac{\mu_T}{\mu_R}$ ; that is the two-sided 90% confidence intervals of the geometric mean ratios need to fall within the bioequivalence limits [80%, 125%]. For this, the estimates of  $C_{max}$  and  $AUC_{\infty}$  need to be

calculated.  $C_{max}$  can be directly computed from the collected data; however for the  $AUC_{\infty}$ , only observation at time  $T = T_{max}$  are available.

One question is: How can we compute the total area  $AUC_{\infty}$  for time  $T = \infty$ ? Extrapolation of the function  $C_t$ , and therefore the estimation of its parameters  $\lambda_1$  and  $\lambda_2$  are needed.

## 1.2 General Nonlinear Mixed-Effects Model

Davidian and Giltinan [12] suggested a two-stage NLMM. This is the most common model in use in statistical literature. For the  $j^{th}$  observation on the  $i^{th}$  cluster (subject) the model response is related to the covariate  $x_{ij}$ , a fixed parameter effects  $\beta$ , and a random parameter effects  $b_i$  as follows (Pinheiro and Bates [27]):

$$y_{ij} = f(\phi_{ij}, x_{ij}) + \epsilon_{ij}, \text{ and } \phi_{ij} = A_{ij}\beta + B_{ij}b_i \quad (1.2)$$

where  $A_{ij}$  and  $B_{ij}$  are design matrices for the fixed and random effects respectively.  $i = 1 \dots N$ ,  $j = 1 \dots n_i$ .

$N$  is the total number of subjects.

$y_{ij}$  is a scalar response of the  $i^{th}$  subject at the  $j^{th}$  measurement.

$x_{ij}$  is a known vector (covariate).

$f$  is nonlinear function of cluster-specific parameters  $\beta$  and  $b_i$ .

$\beta$  is a  $p$ -dimensional vector of unknown parameters (fixed-effects).

$b_i$  is a  $q$ -dimensional vector of random effects associated with the  $i^{th}$  subject, and  $\epsilon_{ij}$  is a random error normally distributed.

The following are common distributional assumptions:

$$b_i \sim N(0, \sigma^2 \Delta), \epsilon_{ij} \sim N(0, \sigma^2)$$

where  $\Delta$  is a positive definite symmetric variance-covariance matrix.

It is further assumed that the observations made on different subjects are independent and that  $\epsilon_{ij}$  and  $b_i$  are independent. It should be noted that the covariance of  $b_i$  is modeled as a multiple of the variance of  $\epsilon_{ij}$  in order to simplify the likelihood function.

The simple case of model( 1.2) would be when  $f$  is a linear function, and the linear mixed model of Verbeke and Molenberghs [44] follows:

$$y_{ij} = X_{ij}\beta + Z_{ij}b_i + \epsilon_{ij}, \text{ and } \phi_{ij} = X_{ij}\beta + Z_{ij}b_i \quad (1.3)$$

where  $X_{ij}$  and  $Z_{ij}$  are known covariates such as weight and treatment type, for example.

### 1.3 Examples of Data following NLMM

We list a few examples of real data, where the use of NLMM is warranted, and we give especial attention to the first two examples: 1.3.1 and 1.3.2. These two examples will be used as a basis for our simulations in both chapters 2 and 3.

#### 1.3.1 Theophylline Data

In this example we consider the Theophylline dataset where 12 subjects were orally administered the Theophylline dose. The serum concentrations were measured at 11 time points over 25 consecutive hours. A common model for such data is a First-Order Open-Compartment model. Gibaldi and Perrier [15].

Let  $y_{ij}$  denote the  $i^{th}$  subject's observed serum concentration at time  $t_{ij}$  and let  $D_i$  be the initial dose for the  $i^{th}$  subject. Pinheiro and Bates [27], First-Order Open-Compartment model that describe the



scalar response  $y_{ij}$  is given as follows:

$$y_{ij} = D_i \frac{\exp[-\phi_{i1} + \phi_{i2} + \phi_{i3}]}{\exp(\phi_{i2}) - \exp(\phi_{i3})} (\exp[-t_{ij} \exp(\phi_{i3})] - \exp[-t_{ij} \exp(\phi_{i2})]) + \epsilon_{ij} \quad (1.4)$$

$$\phi_{i1} = \beta_1 + b_{i1},$$

$$\phi_{i2} = \beta_2 + b_{i2},$$

$$\phi_{i3} = \beta_3,$$

$$\beta = (\beta_1, \beta_2, \beta_3),$$

$$b_i = (b_{i1}, b_{i2}, b_{i3}),$$

$$i = 1, \dots, 12, j = 1, \dots, 11.$$

The design matrices  $A_{ij}$  and  $B_{ij}$  in model 1.2 are given in this model as follows:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

In addition, the following functions of the parameters are widely used for the first-order open-compartment model:

$Cl_i = \exp(\phi_{i1})$ : is the clearance rate.

$Ka_i = \exp(\phi_{i2})$ : is the absorption rate.

$Ke_i = \exp(\phi_{i3})$ : is the elimination rate.

With The following parameters values widely published in the NLMM literature, Pinheiro and Bates [27] and simple calculation from model 1.4

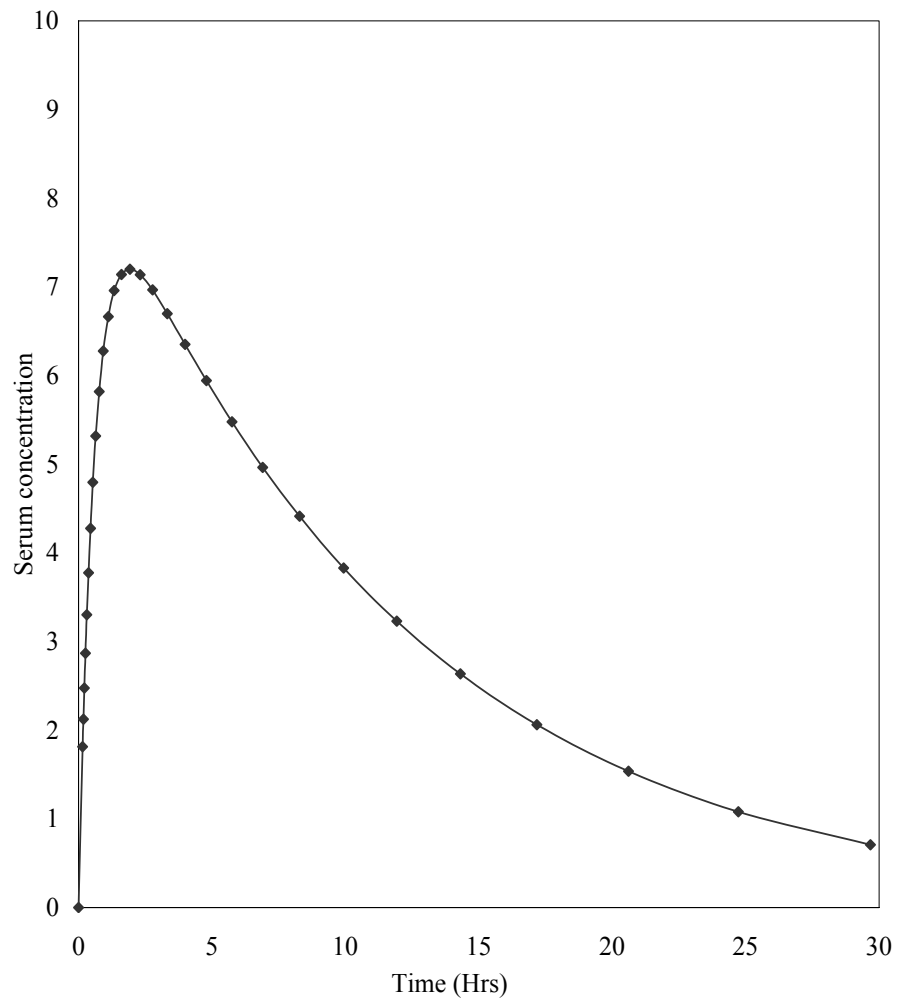
$$\beta_1 = -3.23, \beta_2 = 0.48,$$

$$\beta_3 = -2.46, b_{i1} = b_{i2} = \epsilon_{ij} = 0$$

The typical response curve evaluated at the above parameters is as follows:

$$\hat{y}_{ij} = 2.275 \times D_j [\exp(-0.085 \times t_{ij}) - \exp(-1.613 \times t_{ij})]$$

The corresponding curve is shown below:



**Figure 1.1:** Average Curve of Theophylline Concentration

Obtained from many individuals.

### 1.3.2 Carbon Dioxide Data

The second dataset is the Carbon Dioxide (CO<sub>2</sub>) data reported in Vonesh and Chinchilli [45], on the results of a study on cold tolerance. Twelve plants (six from Quebec and six from Mississippi) were divided into two groups: control plants that were kept at 26<sup>0</sup> and chilled plants that were subject to 14 hours of chilling at 7<sup>0</sup>. After 10 hours of recovery at 20<sup>0</sup>C, Carbon Dioxide Uptake rate was measured for each plant at seven concentrations of ambient CO<sub>2</sub>. The goal of the study was to evaluate the effect of plant type and/or chilling treatment on the CO<sub>2</sub> Uptake. The expected uptake rate  $y_{ij}$  is a function of the Carbon Dioxide concentration  $x_{ij}$  (e.g., figure 1.2 that shows Carbon Dioxide Uptake per subject) which can be expressed as an asymptotic regression with an offset, as in the model of Pinheiro Bates [27] and as follows:

$$y_{ij} = (\alpha + b_{1i}) \times [1 - \exp(-\exp(\beta_4 + b_{2i})) \times (x_{ij} - \beta_5)] + \epsilon_{ij} \quad (1.5)$$

It should be noted that only  $\alpha$  varies among groups. Since it is the

only fixed-effect parameter that is dependant on the covariates (plant type and treatment), it can be expressed as follows:

$$\alpha = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{1i} \times x_{2i}$$

We defined  $\phi_1$  and  $\phi_2$  as the two mixed parameters as follows:

$$\phi_1 = \alpha + b_{1i} \text{ and,}$$

$$\phi_2 = \beta_4 + b_{2i}$$

Where  $\beta_i$ s are the fixed-effect parameters representing the population average of the individual parameters. The  $b_{ij}$ s are the random effect parameters representing the deviations of the Carbon Dioxide uptake from the population averages. The random effects  $(b_{1i}, b_{2i})$  are assumed to have a bivariate normal distribution with mean  $(0, 0)$  and covariance  $\Psi$  defined as follows:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \sim \mathcal{N}(\varphi, \Psi) \varphi = \begin{bmatrix} \alpha \\ \beta_4 \end{bmatrix}, \Psi = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

where

$x_{1i} = -1$ , if the  $i$ th plant type is Quebec;

$x_{1i} = -1$ , if the  $i$ th plant type is Mississippi;

$x_{2i} = -1$ , if treatment of the  $i$ th plant is non-chilled and

$x_{2i} = 1$ , if treatment of the  $i$ th plant is chilled.

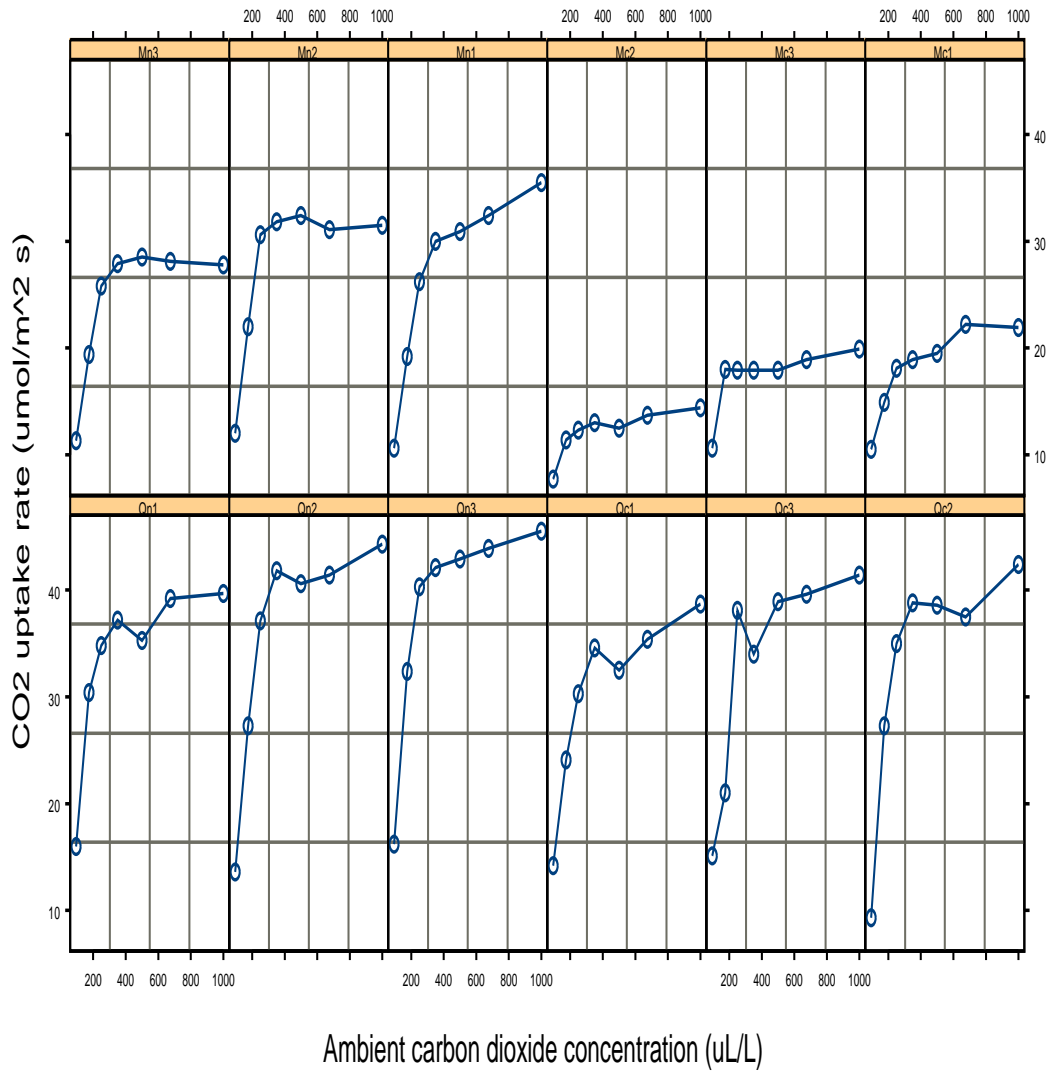
The data consists of four groups of plants:

QC: Quebec chilled plant,

QN: Quebec non-chilled,

MC: Mississippi chilled, and

MN: Mississippi non-chilled.



**Figure 1.2:** Carbon Dioxide Uptake

Each group has three plants measured at seven different levels of ambient CO<sub>2</sub>. We considered the model from Pinheiro and Bates

[27] where they used plant type and treatment as covariates to describe the variability within and amongst groups. Such covariates, when available, could lead to decreasing the number of random effects and therefore the dimension of the objective function (likelihood function). As in the first dataset, we simulate 1000 replicates (datasets) from the uptake model 2.2 with dependent random coefficients using the Cholesky decomposition William et al. [28] to simulate those parameters from a bivariate normal distribution with mean  $(0, 0)$ , and covariance  $\Psi$ .

The mean models (where the random effects parameters are set to zero) for the corresponding four groups are given as follows:

$$U_{QN} = 43.14 \times [1 - \exp(-0.009307 \times (CO2 - 48.73))]$$

$$U_{QC} = 37.80 \times [1 - \exp(-0.009307 \times (CO2 - 48.73))]$$

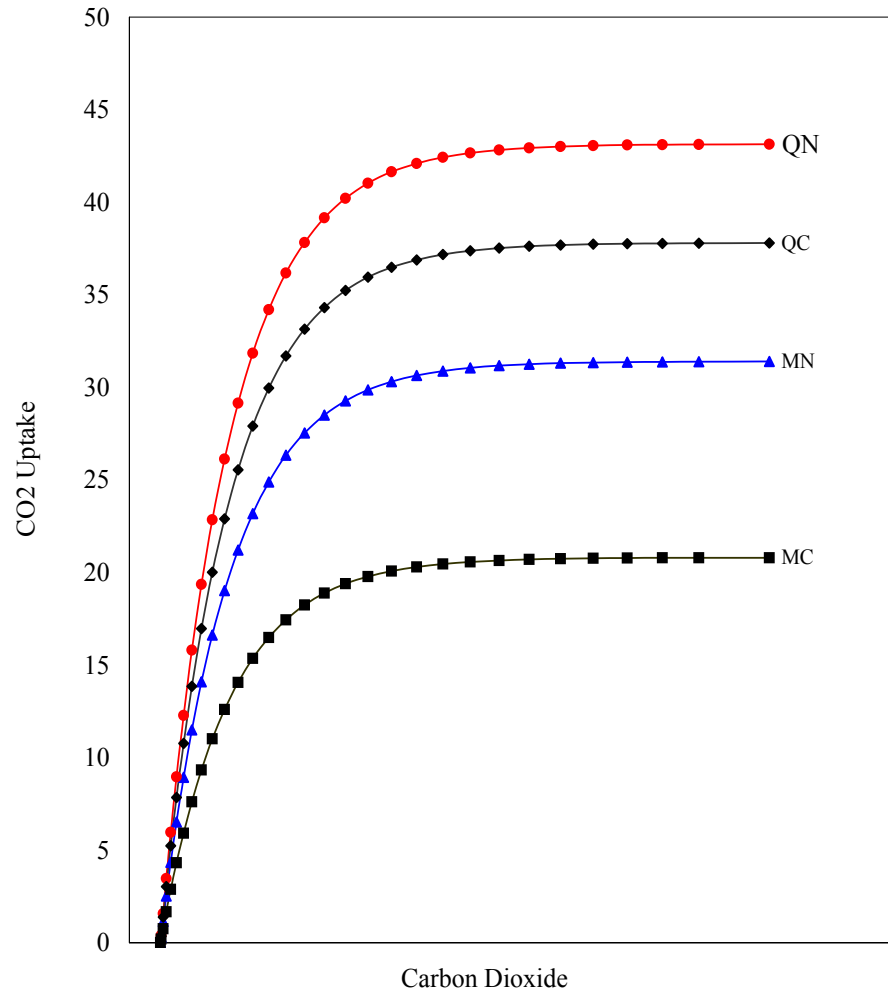
$$U_{MN} = 31.40 \times [1 - \exp(-0.009307 \times (CO2 - 48.73))]$$

$$U_{MC} = 20.80 \times [1 - \exp(-0.009307 \times (CO2 - 48.73))]$$

As shown in the above equations only  $\alpha$  or the linear function of  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$  varies among groups since it is the only fixed-effect parameter to be expressed as a function of the covariates (plant type,



treatments) in the model.



**Figure 1.3:** Average Carbon Dioxide Uptake Per Group

### 1.3.3 Orange Tree Data

Our second set of real data is the Orange data from Pinheiro and Bates [27](see figure 1.4). The example describes the growth of trunk circumferences (in millimeters) of the orange trees. Seven sets of measurements of trunk circumferences,  $y_{ij}$ , were taken on five trees. The NLMM corresponds to the following logistic model:

$$y_{ij} = \frac{\phi_{1i}}{1 + \exp[-(x_{ij} - \phi_{2i}) / \phi_{3i}]} + \epsilon_{ij}$$

$i = 1, \dots, 5$  and  $j = 1, \dots, 7$

$\phi_{ki} = \beta_k + b_{ki}$ ,  $k = 1, \dots, 3$  and  $i = 1, \dots, 5$

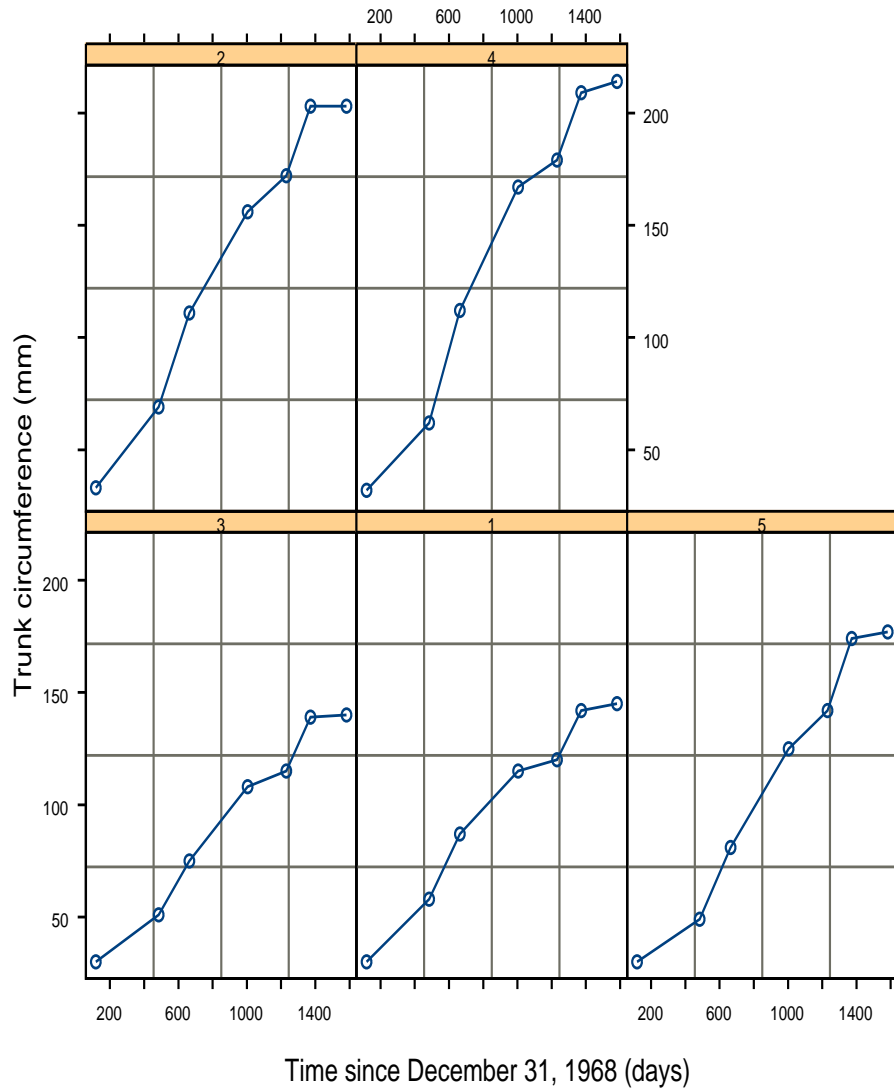
$\phi_1$  : The the asymptotic trunk circumference

$\phi_2$  : The age at which the tree attains half of its asymptotic trunk circumference

$\phi_3$  : The growth scale

$x_{ij}$  : The age of the  $i^{th}$  tree at the  $j^{th}$  measurement

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



**Figure 1.4:** Orange Tree Data

### 1.3.4 Binary Data Model

Our third example concerns the data from a toxicology study. Price et. al. [29] investigated the birth defects or malformations effects of a chemical compound (ethylene glycol also known as monoethylene glycol: MEG) where pregnant mice were exposed to the compound at one of four different dose levels: 0, 0.75, 1.5, and  $3mg/kg$ . The fetal binary malformation indicator for each fetus within a litter and litter size were recorded. It was of interest to estimate the dose effect on adverse outcomes malformation as well as the effect of litter size on this malformation.

Let  $y_{ij}$  denote the response (malformation indicator)  $y_{ij} = 1$  if the  $j^{th}$  fetus in the  $i^{th}$  litter is malformed and  $y_{ij} = 0$  otherwise, ( $i = 1, \dots, I, j = 1, \dots, n_i$ ). Let  $x_i$  be the dose level and  $z_i$  be the litter's weight. Then a reasonable model for this dataset is the following logistic model with two random effects to account for the litter-specific and fetus-specific interclass correlations :

$$y_{ij}|(b_i, c_j) \sim \text{Bernoulli}(p_{ij}) \text{ and}$$

$$E[y_{ij}|(b_i, c_j)] = \eta_{ij} = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_0 + \beta_1 \times x_i + \beta_2 \times z_i + b_i + c_j \quad (1.6)$$

$b_i$  is the litter-specific random effect independent of the fetus-specific random effect  $c_i$ , and both are assumed to have normal distribution  $N(0, \sigma_b^2)$  and  $N(0, \sigma_c^2)$  respectively.

## 1.4 Problem of Estimation of NLMM

Let  $y_i$  be a  $n_i$ -dimensional response vector for the  $i^{th}$  subject  $i = 1, \dots, s$ . We assume that the  $y_{is}$  are independent responses across subjects, but that the within-subject covariance is likely to exist because each of the elements of  $y_{is}$  are measured on the same subject  $i^{th}$ . As a statistical mechanism for modeling this within-subject covariance, assume that there exist unobserved (latent) random-effect vectors  $b_i$  of small dimension (typically one to three) that are also independent across  $i$ . We also assume that an appropriate model linking  $y_i$  and  $b_i$  exists, and note the joint distribution of  $(y_i, b_i)$  as  $p(y, b)$ .

Since the response variable  $y$  is observed, and the random effects  $b$  are unobserved, we consider a two-stage model in which the random components are drawn from a density  $p(b)$ , and  $y$  is drawn from a

density conditional on  $b$ ,  $p(y|b)$ :

- Choose  $b$  from  $p(b)$
- Given  $b$ , choose  $y$  from  $p(y|b)$
- $p(y|b) = \int p(y, b)db$

The marginal density function of  $y_i$  given  $b$  is given as follows:

$$L(\theta) = p(y|\beta, \Delta, \sigma^2) = \int p(y|b, \beta, \Delta, \sigma^2)p(b|\Delta, \sigma^2)db \quad (1.7)$$

Where  $\theta = (\beta, \Delta, \sigma^2)$

There is no general closed form for the integral and therefore numerical approximations are needed in order to estimate  $\beta, \Delta$  and  $\sigma^2$ . The inverse Hessian (second derivative) matrix at the estimates provides an approximate variance-covariance matrix for the parameters estimates. The function  $-\log(L(\theta))$  is referred to both as the negative log-likelihood function and as the objective function for optimization.

The likelihood cannot be expressed as a closed-form integral and therefore numerical approximations are needed to have a quasi-likelihood function in order to compute the estimates of the model parameters. Different methods have been proposed to estimate the parameters in

the NLMM and they can be divided into two categories: “exact” methods and approximate methods. The “exact” method to compute the maximum likelihood estimates using the EM algorithm was introduced by Walker [48] for a special case where only random coefficients are present in the model. Davidian and Gallant [10] considered a maximum likelihood estimation using the Gaussian Quadrature (GQ) approach allowing the random effects to be smooth, but not necessarily normal. A Bayesian approach was also discussed by Racine-Poon [30].

## 1.5 Standard Nonlinear Regression Model

A simple and intuitive way to estimate parameters of model 1.7 in general and model 1.4 in particular is to ignore the random effects and use a simple nonlinear regression model for each individual without the random effects to get the model parameters. Since each subject in the study will have a separate model, this implies that for example in model 1.4, the  $i^{th}$  individual model will be written as follows:

$$y_{ij} = D_i \frac{\exp[-\beta_{1i} + \beta_{2i} + \beta_{3i}]}{\exp(\beta_{i2}) - \exp(\beta_{i3})} (\exp[-t_{ij} \exp(\beta_{i3})] - \exp[-t_{ij} \exp(\beta_{i2})]) + \epsilon_{ij}$$

$i = 1, \dots, 12, j = 1, \dots, 11$ . Therefore there are  $12 \times 4 = 48$  parameters that need to be estimated. In other words, for subject  $i^{th}$  model,  $\beta_{1i}, \beta_{2i}, \beta_{3i}$ , and  $\sigma_i^2$  are unknown and must be estimated from a total of 11 observations, that is less than three observations per unknown parameter.

The Gauss-Newton method can be used to solve for the parameters by performing a Taylor series expansion on  $f(\beta)$ . Then we approximated the nonlinear model with linear terms and employ ordinary least squares to estimate the parameters. This procedure is performed in iterative nature and it generally leads to a solution of the nonlinear problem.



Subject	Estimate	MSE	CV	Bias	Is True $\beta_1$ in the 95% CI?
1	-3.92	0.491	3.25%	21.35%	N
2	-3.11	0.045	5.65%	-3.74%	Y
3	-3.23	0.002	1.32%	0.10%	Y
4	-3.29	0.027	4.71%	1.83%	Y
5	-3.13	0.047	6.22%	-2.92%	Y
6	-2.97	0.081	4.33%	-7.86%	Y
7	-2.96	0.075	2.57%	-8.14%	N
8	-3.07	0.043	4.39%	-4.89%	Y
9	-3.42	0.049	3.08%	6.01%	Y
10	-3.43	0.045	2.04%	6.24%	N
11	-2.86	0.136	1.49%	-11.36%	N
12	-3.17	0.011	2.78%	-1.76%	Y

**Table 1.1:** Subject's Estimate of  $\beta_1$ (True  $\beta_1 = -3.23$ )

Model without Random Components  
CI: The 95% Confidence Interval  
 $\beta_{10}$  falls outside the 95% CI for 4 subjects out of 12  
The estimated coverage probability is about 67%.

Subject	Estimate	MSE	CV	Bias	Is True $\beta_2$ in the 95% CI?
1	0.58	0.039	29.99%	20.28%	Y
2	0.66	0.135	47.75%	38.87%	Y
3	0.90	0.181	8.26%	87.68%	N
4	0.16	0.163	155.15%	-66.90%	Y
5	0.39	0.109	82.01%	-19.22%	Y
6	0.15	0.159	150.86%	-68.30%	Y
7	-0.39	0.767	36.37%	-180.72%	N
8	0.32	0.077	71.20%	-33.33%	Y
9	2.18	3.124	21.50%	356.34%	N
10	-0.36	0.718	28.50%	-175.93%	N
11	1.35	0.764	6.43%	181.85%	N
12	-0.18	0.463	88.57%	-138.23%	N

**Table 1.2:** Subject's Estimate of  $\beta_2$ (True  $\beta_2 = 0.48$ )

Model without Random Components  
CI: The 95% Confidence Interval  
 $\beta_{20}$  falls outside the 95% CI for 6 subjects out of 12  
The estimated coverage probability is about 50%

Subject	Estimate	MSE	CV	Bias	Is True $\beta_3$ in the 95% CI?
1	-2.92	0.241	5.84%	18.71%	N
2	-2.29	0.101	11.65%	-7.05%	Y
3	-2.51	0.006	2.39%	1.98%	Y
4	-2.44	0.059	9.90%	-0.93%	Y
5	-2.43	0.089	12.22%	-1.38%	Y
6	-2.31	0.067	9.08%	-6.18%	Y
7	-2.28	0.052	6.22%	-7.28%	Y
8	-2.39	0.049	8.78%	-2.97%	Y
9	-2.45	0.018	5.50%	-0.54%	Y
10	-2.60	0.034	4.38%	5.88%	Y
11	-2.32	0.023	2.55%	-5.61%	Y
12	-2.25	0.070	7.03%	-8.58%	Y

**Table 1.3:** Subject's Estimate of  $\beta_3$ (True  $\beta_3 = -2.46$ )

Model without Random Components  
CI: The 95% Confidence Interval  
 $\beta_{30}$  falls outside the 95% CI for 1 subjects out of 12.  
The estimated coverage probability is about 92%.

While it can be concluded that omitting the random effects from the nonlinear model provided reasonable estimates for most of  $\beta_1$ , the estimates have a poor probability coverage of 67%. In addition, the same type of model failed to provide good estimates for  $\beta_2$  and  $\beta_3$ . For  $\beta_2$  the absolute relative bias ranged from 19% to 356% for subject number 5 and 9 respectively. The probability coverage averaged is about 50%. Therefore due to the large number of parameters in mostly small to moderate sample size data, as well as the lack of accuracy due to simple nonlinear standard regression models, the alternative is to

find a way to approximate the likelihood function in 1.7 to a simpler form and estimate the parameters in the presence of random effects.

## 1.6 Approximation Methods of the Likelihood

An alternative approach consists of finding an approximation to the marginal likelihood function, Assuming that the conditional distribution of the response within individual is normal (e.g., the use of the first order Taylor expansion to approximate the log-likelihood). Kedem and Fokianos [23] used partial likelihood to overcome this problem in the context of logistic time series model. Beal and Sheiner [5] proposed an expansion around the means of the random effects often referred to as First order approximation (FO); Wolfinger [49] showed an alternative approximation using a modified version of the Laplace approximation (LP); and Davidian and Gallant [10] considered a maximum likelihood estimation using the Gaussian-Hermit Quadrature approach (GQ).

### 1.6.1 First Order Approximation Method (FO)

The method of Beal and Sheiner in [4] and in [5] and Sheiner and Beal [40] is known in the field of pharmacokinetics as the First order method (FO) and is based on using a Taylor expansion to approximate the model in 1.2, or its marginal distribution of 1.7, in order to have a closed-form likelihood function that can be maximized. The approximation is used only in the case where  $p(y_i|x_{ij}, \phi_{ij})$  is normal. Using the model given in 1.2, the first order Taylor expansion of  $f$  around  $b_i = E(b_i) = b_0 = 0$  gives the following approximate model:

$$y_{ij} \approx f(x_{ij}, \beta, b_0) + F(x_{ij}, \beta, b_0)b_i + \epsilon_{ij} \quad (1.8)$$

where  $F(x_{ij}, \beta, b_0) = \partial f / \partial b_i(x_{ij}, \beta, b_0)$  is the partial derivative of  $f(x_{ij}, \beta, b_i)$  with respect to  $b_i$  evaluated at  $b_i = b_0 = 0$ . The approximate marginal distribution of  $y_i$ , then is normal with mean  $E$  and covariance  $\Sigma$  given by:

$$E_i = E(y_i) \approx f(x_{ij}, \beta, b_0)$$

$$\Sigma = \sigma^2 I_{n_i} + \sigma^2 f'(x_{ij}, \beta, b_0) \Delta f'(x_{ij}, \beta, b_0)$$

The covariance-weighted least squares estimates (LSE) of the

fixed effects  $\beta$  and the covariance matrix  $\Sigma$  are obtained by minimizing the objective function. The First Order approximation (FO) of Beal and Sheiner [5] used to approximate the likelihood of model 1.2 is reduced to the objective function:

$$\ell_{FO}(\beta, \Delta) = \sum_{i=1}^N \{ [y_i - f(x_{ij}, \beta, b_0)]^T \Sigma^{-1} [y_i - f(x_{ij}, \beta, b_0)] + \log |\Sigma| \} \quad (1.9)$$

Under the normality assumption, the LSE is equivalent to the joint maximum likelihood estimation. The minimizers of 1.9 in  $\beta$  and  $\Sigma$  are the Beal and Sheiner [5] estimates. Since  $\Sigma$  is a function of  $\beta$  the minimization of 1.9 can be very complex and not straightforward optimization. Beal and Sheiner [5] used a derivative-free Quasi-Newton algorithm William et al. [28] to obtain the estimates of  $\beta$  and  $\Sigma$ . One disadvantage of the LSE is that when the random effects are misspecified, the assumed covariance matrix  $\Sigma$  can be wrong; and this can produce inefficient and possibly inconsistent estimates of  $\beta$  Carroll and Ruppert [8]. The FO method performance will be assessed in later simulations and compared with other methods of approximation

described in this section.

Pinheiro and Bates [27], using two data sets with small sample sizes, compared three methods for estimating the NLMM parameters: GQ, LP, and Monte Carlo method (MC). They concluded that the FO and the LP were the most efficient methods, while the GQ performed poorly when the numbers of abscissas used to approximate the integral, were small. Also the MC method was very inefficient computationally and gave results quite similar to GQ.

The presence of the random coefficients in a nonlinear fashion makes it impossible to explicitly have a closed-form of the likelihood function and therefore all the estimation methods rely on the Taylor expansion or other forms of approximations, such as the GQ, the LP, the or FO. Numerical optimization methods such as the Quasi-Newton are then employed for the nonlinear objective function.

To show the effects of the random effects on the FO approximation, let  $y_{ij}$  be a simple exponential decay model as follows:

$$y_{ij} = e^{(\beta+b_i)t_{ij}} + \epsilon_{ij}, \text{ where: } b_i \sim \mathcal{N}(0, \sigma_b^2), \epsilon_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

The true marginal mean and variance of  $y_{ij}$  are given as follows:

$$E(y_{ij}) \approx e^{(\beta t_{ij} + \frac{1}{2} \sigma_b^2 t_{ij}^2)}, \quad Var(y_{ij}) \approx e^{(2\beta t_{ij})} (e^{(2\sigma_b^2 t_{ij}^2)} - e^{(\sigma_b^2 t_{ij}^2)}) + \sigma^2$$

The first order Taylor approximation around  $b_i = 0$  is given as follows:

$$y_{ij} \approx e^{(\beta t_{ij})} + t_{ij} e^{(\beta t_{ij})} b_i + \epsilon_{ij}$$

The approximate mean and variance under the FO method are given as follows:

$$E(y_{ij}) \approx e^{(\beta t_{ij})}, \quad Var(y_{ij}) \approx e^{(2\beta t_{ij})} \sigma_b^2 t_{ij}^2 + \sigma^2$$

It should be noted that the difference between the true marginal mean and the FO approximate mean is an offset in the exponent of  $\frac{1}{2} \sigma_b^2 t_{ij}^2$ . Therefore the accuracy of the FO approximation will greatly depend on the magnitude of the random effects covariance (Variance). The smaller the random effect variance, the closer the approximate mean is to the true marginal mean.

### 1.6.2 Laplace Approximation (LP)

We first define the laplacian approximations in a general context and then define it for the integral in 1.7 and give the LP approximation of Wolfinger [49]:

Let  $\mathcal{I} = \int g(b) db$  where  $g(b)$  is a positive function. Let  $l(b) =$

$\log(g(b))$ , and  $\hat{b} = \operatorname{argmax}(g(b))$ . Then  $l'(\hat{b}) = 0$  and  $l(b) \simeq l(\hat{b}) + 1/2(b - \hat{b})'l''(\hat{b})(b - \hat{b})$

Using this approximation,

$$\mathcal{I} \approx \exp(l(\hat{b})) \int \exp(1/2(b - \hat{b})'l''(\hat{b})(b - \hat{b}))db$$

The integrand can be written as the density of a normal random variable to obtain the following:

$$\mathcal{I} \approx (2\pi)^{q/2} \exp(l(\hat{b})) | -l''(\hat{b}) |^{-1/2}$$

This is the first order Laplace approximation to the integral  $\mathcal{I}$ . There is a natural extension to higher order approximations. The advantage of the laplace approximation is that Monte Carlo (MC) integration is replaced by maximization.

Let us define the Laplace approximation for joint density of  $y$  and  $b$ . Let  $\gamma$  be a row vector of all parameters to be estimated and let  $d$  be the length of  $\gamma$ . Let  $p^*(y, b)$  be the Laplace approximation of the integral  $\int p(y, b)db$  and let  $l^*(\gamma) = \log p^*(y|\gamma)$ . Then:

$$p^*(\gamma) = l(y, \hat{b}(\gamma)|\gamma) - 1/2 \log \left| -l''_{22}(y, \hat{b}(\gamma) |\gamma) \right|$$

where  $l(y, b|\gamma) = \log p(y, b|\gamma)$  and  $\hat{b}(\gamma)$  is the value of  $b$  that maximizes  $p(y, b|\gamma)$ .



The integral that we want to estimate for the marginal distribution of  $y_i$  in 1.7 can be written as follows:

$$p(y|\beta, \Delta, \sigma^2) = (2\pi\sigma^2)^{-(n_i+q)/2} |D|^{-1/2} \exp[-g(\beta, \Delta, y_i, b_i)/2\sigma^2]$$

We defined the first and second partial derivative as follows:

$$\begin{aligned} g(\beta, \Delta, y_i, b_i) &= \|y_i - f_i(\beta, b_i)\|^2 + b_i' \Delta^{-1} b_i \hat{b}_i \\ &= b_i(\beta, \hat{\Delta}, y_i) = \operatorname{argmin}(g(\beta, \Delta, y_i, b_i)) \\ g'(\beta, \Delta, y_i, b_i) &= \frac{\partial g(\beta, \Delta, y_i, b_i)}{\partial b} \\ g''(\beta, \Delta, y_i, b_i) &= \frac{\partial^2 g(\beta, \Delta, y_i, b_i)}{\partial b_i \partial b_i^T} \end{aligned}$$

The Laplace approximation to the log-likelihood of model 1.2 is defined as follows:

$$\begin{aligned} p_{lp}(y|\beta, \Delta, \sigma^2) &\approx (2\pi\sigma^2)^{-N/2} |D|^{-M/2} \exp[-1/(2\sigma^2) \sum_{i=1}^N g(\beta, \Delta, y_i, \hat{b}_i)] \\ &\times \int (2\pi\sigma^2)^{q/2} \exp -1/(2\sigma^2) \sum_{i=1}^N [b_i - \hat{b}_i]^T g''(\beta, \Delta, y_i, \hat{b}_i) [b_i - \hat{b}_i] db_i \end{aligned}$$

One can rewrite the formula as follows:

$$\begin{aligned} p_{lp}(y|\beta, \Delta, \sigma^2) &= (2\pi\sigma^2)^{-N/2} |D|^{-M/2} \prod_{i=1}^N |g''(\beta, \Delta, y_i, \hat{b}_i)|^{-1/2} \times \\ &\exp[-g(\beta, \Delta, y_i, \hat{b}_i)/2\sigma^2] \end{aligned}$$

Denote  $G$  as the following:

$$g''(\beta, \Delta, y_i, \hat{b}_i) \approx G(\beta, \Delta, y_i) = \frac{\partial f(\beta, b_i)}{\partial b^T} \Big|_{b_i=\hat{b}_i} + \frac{\partial f(\beta, b_i)}{\partial b} \Big|_{b_i=\hat{b}_i} + \Delta^{-1}$$

By eliminating the small terms in the approximation and using the notation above, the laplace approximation (LP) of Wolfinger [49] used to approximate the likelihood of model 1.2 is reduced to the following objective function:

$$\begin{aligned} \ell_{LP} = & -\frac{1}{2} \{ N[1 + \log(2\pi) + \log(\hat{\sigma}^2)] + \\ & N \log(|\Delta|) + \sum_{i=1}^N \log(G(\beta, \Delta, y_i)) \} \end{aligned} \quad (1.10)$$

This laplace approximation the likelihood of model 1.2 is known as the first order Laplace approximation.

### 1.6.3 Gaussian Quadrature (GQ)

Gaussian quadrature is widely used to approximate integrals of functions with respect to a given Kernel by a weighted average of the integrand evaluated at pre-determined abscissas. The weights and abscissas used in Gaussian quadrature rules for the most common Ker-

nels can be obtained from the table of Abramowitz and Stegun [1]. Gaussian quadrature for multiple integrals are known to be numerically complex, Davis and Rabinowitz [9]; but by using the structure of the integrand in the NLMM we can transform the problem into successive applications of simple one-dimensional Gaussian quadrature rules. The critical condition for success is the choice of an importance distribution that approximates the integrand.

In numerical analysis, a quadrature rule is an approximation of the definite integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration. The standard formula for the approximation of integral is defined as follows:

$$\int_a^b f(x) dx \approx \sum_{i=1}^{N_{GQ}} w_i f(x_i)$$

where  $f$  is the function for which the integral needs to be evaluated and  $w_i$  is a weight function evaluated at the abscissas points  $x_i$ . The above Gaussian quadrature will produce accurate results if the function  $f(x)$  is well approximated by a polynomial function within

the integral limits. A particular weight function is the Gauss-Hermite weight defined as  $w(x) = e^{-x^2}$ .

In the case of NLMM models the marginal distribution of the random effects has an integrand that is proportional to  $\exp[-g(\beta, \Delta, y_i, b_i)/2\sigma^2]$ .

Let  $z_j, w_j, j = 1, \dots, N_{GQ}$  denote respectively the abscissas and the weights for the (one-dimensional) Gaussian quadrature rule with  $N_{GQ}$  points based on the standard normal distribution Kernel,  $N(0, 1)$ . Then the Gaussian quadrature method (Using Gauss-Hermite weight function) for the likelihood function  $\mathcal{K}$  is defined as follows:

$\mathcal{K} = \int_{-\infty}^{\infty} \exp[-g(\beta, \Delta, y_i, b_i)/2\sigma^2] db_i$ , is given by:

$$\mathcal{K} \approx C \times \sum_{j_1=1}^{N_{GQ}} \dots \sum_{j_q=1}^{N_{GQ}} [\exp[-g(\beta, \Delta, y_i, \hat{b}_i + \sigma G^{-1/2}(\beta, \Delta, y_i) z_j)/2\sigma^2 + \frac{||z_j||^2}{2}] \prod_{k=1}^q w_{j_k}]$$

where  $z_j = (z_{j1}, \dots, z_{jq})$ ,  $C = \sigma^q |G(\beta, \Delta, y_i)|^{-1/2}$ , and  $G(\beta, \Delta, y_i) = g''(\beta, \Delta, y_i, \hat{b}_i)$

The Gaussian Quadrature approximation (GQ) of Davidian and Gallant [10] used to approximate the likelihood of model 1.2 is reduced to the following objective function:

$$\begin{aligned}
\ell_{GQ} = & -[N \log(2\pi\sigma^2) + N \log |\Delta| + \sum_{i=1}^N |G(\beta, \Delta, y_i)|]/2 + \\
& \sum_{i=1}^N \left[ \sum_{j=1}^{N_{GQ}} \exp[-g(\beta, \Delta, y_i, \hat{b}_i + \sigma[G(\beta, \Delta, y_i)]^{-1/2} z_j)/2\sigma^2 + \right. \\
& \left. \|z_j\|^2/2\right] \prod_{k=1}^q w_{jk} \tag{1.11}
\end{aligned}$$

## Chapter 2

### Comparison of the Approximation Methods

In this chapter we will compare three techniques used to approximate the likelihood function of model 1.2, which are widely used in the estimation of Nonlinear Mixed-Effect Models (NLMM), based on simulation studies. We first compare these approximation methods through extensive simulation studies motivated by two widely used datasets. We compare the empirical estimates from the three different approximations of the log-likelihood function and its bias, precision, convergence rate, and coverage probability. We compare the First order approximation (FO) of Beal and Sheiner [5], Laplace approximation (LP) of Wolfinger [49] and Gaussian Quadrature (GQ) of Davidian and Gallant [10].

We also compare these approaches under different sample size configurations by increasing the number of observations per subject,

and analyzing their effects on both fixed-effects estimates and the precision measures. We compare these methods of estimation based on the model parameters' empirical estimate, precision and bias as well as their 95% confidence interval coverage probability.

## **2.1 Simulation Methodology**

## **2.2 Theophylline data**

We first consider the Theophylline data described in example 1.3.1 and obtained from a study of the kinetics of the anti-asthmatic agent (Theophylline) reported and analyzed in Davidian and Giltinan [12], Vonesh and Chichilli [45], and Pinheiro and Bates [27]. In this experiment, 12 volunteers received oral doses of the Theophylline drug and each of the 12 subjects' serum concentration was measured at eleven time points over the subsequent 25 hours. Davidian and Giltinan [12] used the one-compartment open model to fit the Theophylline data, and Pinheiro and Bates [27] suggested a refined version of such a model and concluded that a diagonal covariance matrix is as good as the un-

constrained one. In the first simulation, we assumed that the Pinheiro and Bates ([27] model version (as in model 2.1 in the following section) to avoid the boundary issue, where the covariance component of the two random coefficients is very small and can lead to an ill-conditioned matrix. On one hand, this assumption makes the optimization very difficult and sometimes getting satisfactory convergence is impossible. On the other hand, this issue has been of large interest in the statistics literature and still constitutes great computational challenges.

A first-order open-compartment model of Davidian and Giltinan [12] is used to express the serum concentration  $y_{ij}$  for patients  $i^{th}$  at observation  $j^{th}$ , after an initial dose  $D_j$  as follows:

$$y_{ij} = D_j \frac{\exp[-(\beta_1 + b_{i1}) + (\beta_2 + b_{i2}) + \beta_3]}{\exp(\beta_2 + b_{i2}) - \exp(\beta_3)} \times (\exp[-t_{ij} \exp(\beta_3)] - \exp[-t_{ij} \exp(\beta_2 + b_{i2})]) + \epsilon_{ij} \quad (2.1)$$

$$\beta = (\beta_1, \beta_2, \beta_3), b_i = (b_{i1}, b_{i2}, b_{i3}) \quad i = 1, \dots, 12, j = 1, \dots, 11$$

$y_{ij}$  is the observed concentration for occasion  $j$  and individual  $i$

$D_j$ : is the initial dose level

$Cl_i = \exp(\beta_1 + b_{i1})$ : is the clearance rate for individual  $i$

$Ka_i = \exp(\beta_2 + b_{i2})$ : is the absorption rate for individual  $i$



$Ke_i = \exp(\beta_3)$ : is the elimination rate for individual  $i$

We assume that only  $Ka_i$  and  $Cl_i$  are random (i.e., they vary by subject). Further, they follow normal distribution, that is:

$$\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right), i = 1, \dots, 12, \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

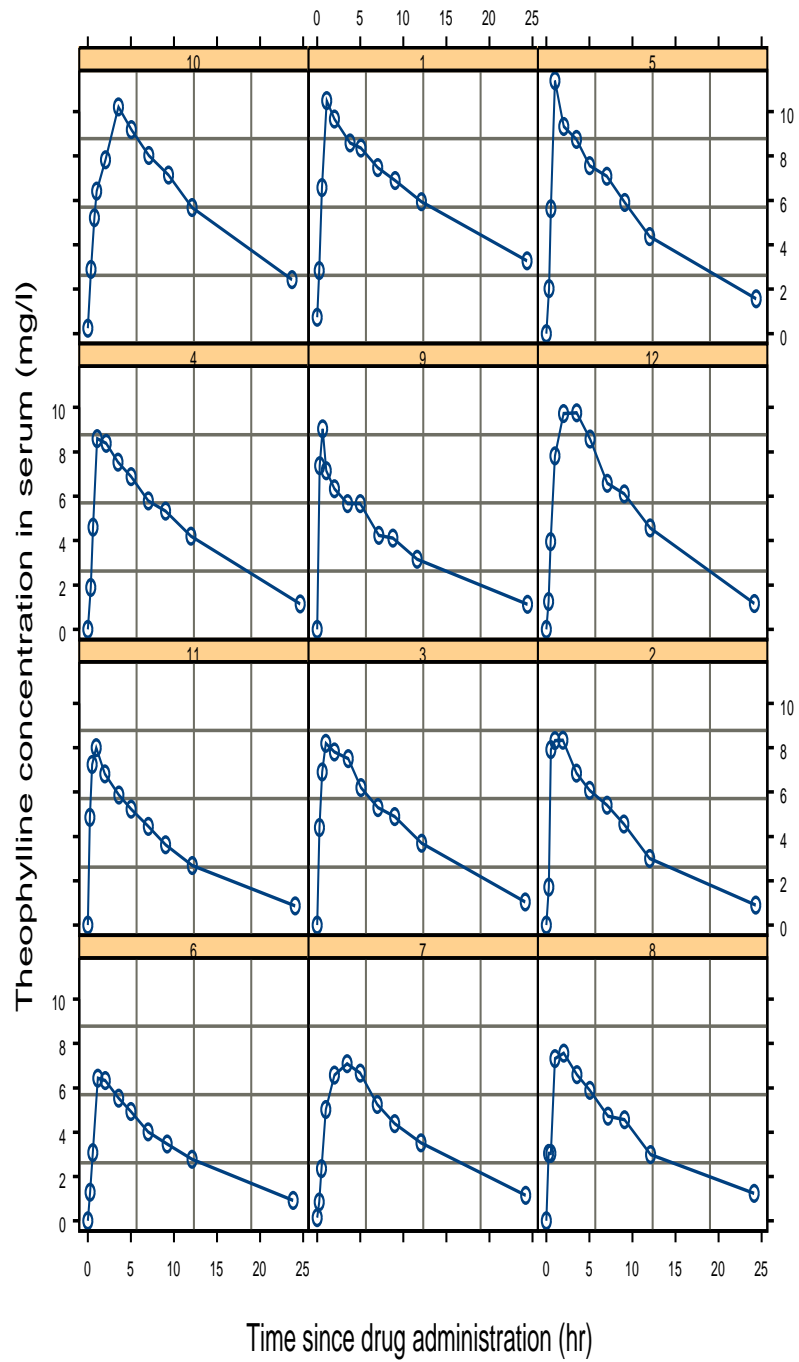
We assume that the following are the true parameter values:

$$\beta_1 = -3.2269, \beta_2 = 0.4782,$$

$$\beta_3 = -2.4594, \sigma^2 = 0.5581,$$

$$\sigma_1^2 = 0.02787, \sigma_2^2 = 0.4239$$

The following graph illustrates the 12 individual's curves of the observed Theophylline concentration in the blood.



**Figure 2.1:** Theophylline Concentration in Serum  
From 12 individuals

We use the above parameters estimates, widely used and published in the literature, Pinheiro and Bates [27], to generate 1000 datasets from these parameters and form model 2.1 to evaluate the accuracy of the parameter estimates using the three approximation methods, FO, GQ, and LP and the sensitivity of these approximation techniques.

Davidian and Giltinan [12] suggested that the response at time zero should be excluded from the analysis, since all but three data points collected at time zero are zeroes and the presence of such zeroes might complicate the analysis. In their simulation of a similar dataset from the one-compartment model, Vonesh and Chinchilli [45], overcame the problems of negative responses and zeroes concentrations by excluding the time zero from their simulation and only simulating the regression coefficients and not the error term. Roe [34], in an extensive Pharmacokinetics simulation study, excluded the initial time at zero and started the simulated data at time =0.1 hour with an experiment time of 12 hours. Since we want to be consistent with these previous efforts, and compare the capability and consistency of these likelihood approximations the serum concentrations at time zero were deleted

from the dataset and therefore from all the simulated replicates.

## 2.3 Carbon Dioxide Data

The second dataset is the Carbon Dioxide (CO<sub>2</sub>) data reported in Vonesh and Chinchilli [45], and described in example 1.3.2, model 2.2 of Pinheiro Bates [27] is as follows:

$$y_{ij} = (\alpha + b_{1i}) \times [1 - \exp(-\exp(\beta_4 + b_{2i})) \times (x_{ij} - \beta_5)] + \epsilon_{ij} \quad (2.2)$$

In this dataset the following true parameter values will be used to simulate 1000 replicates of the dataset:

$$\beta_0 = 32.3580, \beta_1 = -7.1815,$$

$$\beta_2 = -3.9845, \beta_3 = -1.3124,$$

$$\beta_4 = -4.6133, \beta_5 = 48.7323,$$

$$\sigma^2 = 3.5299$$

$$\Psi = \begin{bmatrix} 8.1028 & -0.3435 \\ -0.3435 & 0.01814 \end{bmatrix}$$

## 2.4 Simulation Results

In our simulations, methods of approximation of the likelihood function and the estimates were obtained using SAS procedure nlmixed [35] and the empirical estimates were compared based on model convergence rate, relative bias, coverage probability of the 95% confidence interval (CI), the coefficient of variation also known as “relative variability” and the mean square error (MSE).

Assuming that  $\theta$  is the parameter to be empirically estimated,  $\hat{\theta}$  its estimator, and  $\theta_0$  its true value of  $\theta$ , we define the following performance measures, to evaluate the approximation methods:

$$\text{Estimate of } \beta_i = \frac{\text{Sum of all } \beta_i \text{ Estimates}}{\text{Number of datasets where the model converged}}$$

$$\text{Convergence Rate (CR)} = \frac{\text{Number of datasets where the model converged}}{\text{Number of datasets}} \times 100$$

$$\text{Coverage Probability (CP)} = \frac{\text{The number of } \theta_0 \text{ that lies in the 95\% CI of } \hat{\theta}}{\text{Number of simulation}} \times 100$$

$$\text{The Coefficient of Variation (CV)} = \frac{\text{Standard deviation}}{\text{The mean}} \times 100$$

$$\text{Relative Bias (Bias)} = \left| \frac{\hat{\theta} - \theta_0}{\theta_0} \times 100 \right|$$

$$\text{Mean Square Error (MSE)} = (\hat{\theta} - \theta_0)^2$$

### 2.4.1 Theophylline Simulation Results: ( $N = 120$ )

In the Theophylline dataset where the random effects are assumed independent as in model 2.1 (i.e., random effects are normally distributed with zero correlation), the LP and GQ methods were close, giving very similar estimates and coverage probability and seeming to give better results than the FO method. For the parameter with the smallest magnitude,  $\beta_2$  the LP and GQ methods have poor precisions with biases above 10% and coverage probabilities below 90%.

Of all four methods of approximation (estimations), the LP and GQ seem to provide the most accurate estimation. The relative biases are 0.74% and 1.05% for  $\beta_1$ , 14.22% and 14.93% for  $\beta_2$ , and 1.68% and 2.50% for  $\beta_3$  for the GQ and PL method respectively. It should be noted that  $\beta_2$  is the parameter with the smallest magnitude (true value = 0.48). The convergence rates for both GQ and LP were 92% and 93% respectively. The coverage probabilities for the GQ method

were 90%, 92%, and 88% for  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  respectively. The same coverage probabilities for the LP method were 94%, 92%, and 92% for  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  respectively.

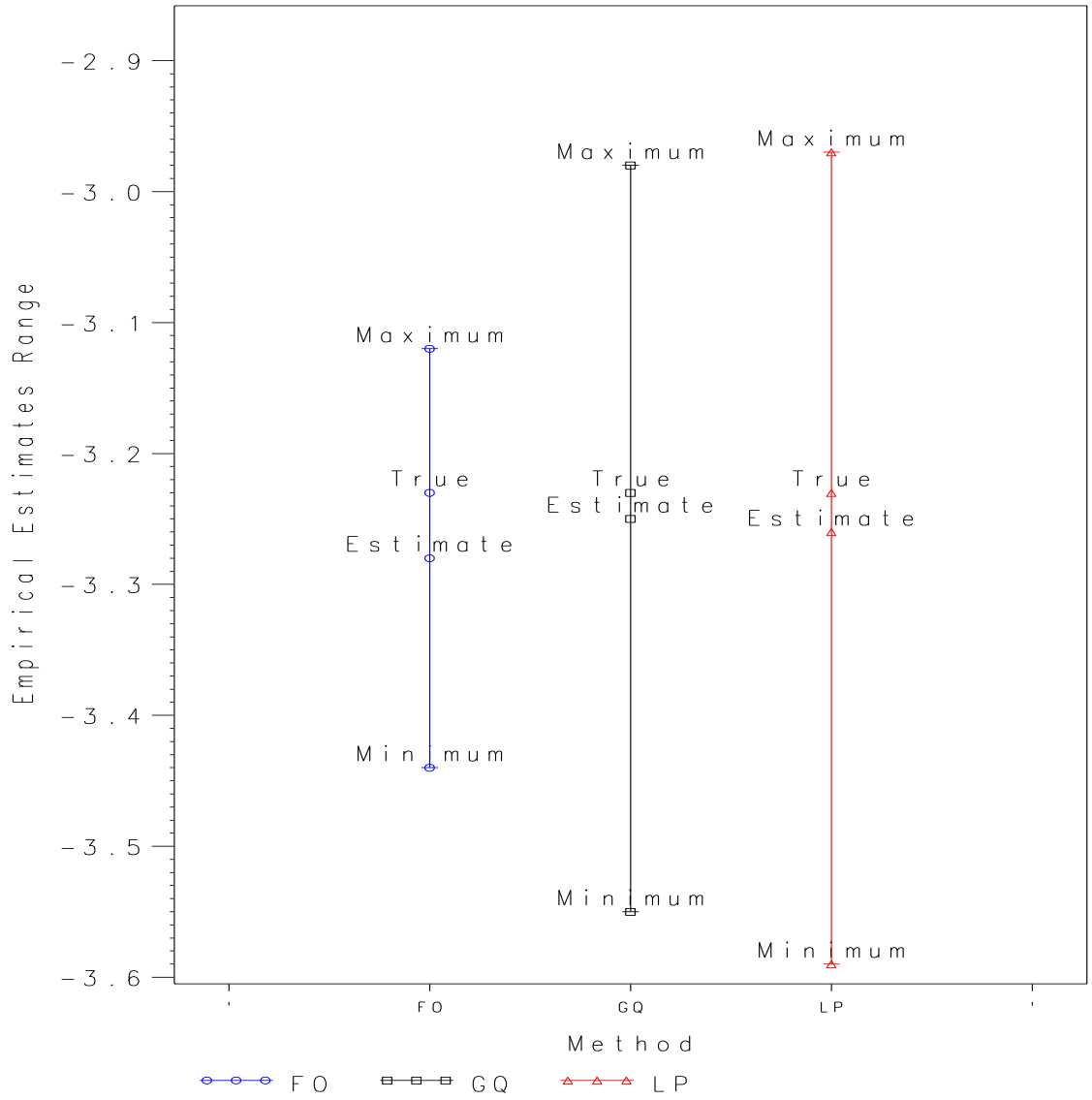
The FO method's precision and accuracy seem to depend on the degree of linearity at which the parameter is in the model. For example, the FO method seems to give good estimation for  $\beta_1$  and less accurate estimation for  $\beta_2$  and  $\beta_3$ . Table 2.1 summarizes the simulation results for all four methods used for the Theophylline dataset model.

Parameter (True Value)	Method	Estimate	Min	Max	MSE	CV	Bias	CR	CP
$\beta_1 (= -3.23)$	GQ	-3.25	-3.55	-2.98	0.009	2.86%	0.73%	92%	90%
	LP	-3.26	-3.59	-2.97	0.007	2.41%	1.05%	93%	94%
	FO	-3.28	-3.44	-3.12	0.005	1.38%	1.62%	70%	99%
$\beta_2 (= 0.48)$	GQ	0.55	-0.03	1.23	0.044	36.38%	14.22%	92%	92%
	LP	0.55	0.03	1.27	0.043	35.46%	14.93%	93%	92%
	FO	0.56	0.14	1.83	0.040	33.15%	16.14%	70%	92%
$\beta_3 (= -2.46)$	GQ	-2.50	-2.90	-2.03	0.025	6.05%	1.68%	92%	88%
	LP	-2.52	-3.06	-2.03	0.020	4.99%	2.50%	93%	92%
	FO	-2.55	-2.83	-2.26	0.015	3.25%	3.68%	70%	94%

**Table 2.1:** Empirical Estimates for Theophylline Model (N = 120). Estimates from converging models among 1000 cases.

For all three parameters,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  the estimates ranges, the difference between the maximum estimate value and the minimum estimate value were higher for both the GQ and the LP approximation techniques, with the exception of  $\beta_2$  where the magnitude of such parameter is the lowest. In this case the FO approximation method has the largest estimate ranges. It is also worth noting that the GQ and LP methods' estimate values were the closest to the true parameter values, whereas for the FO method the distance between the estimates and the true values was the largest. Figures 2.2, 2.3, and 2.4 display the estimates ranges as well as the true parameter value for  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  respectively.



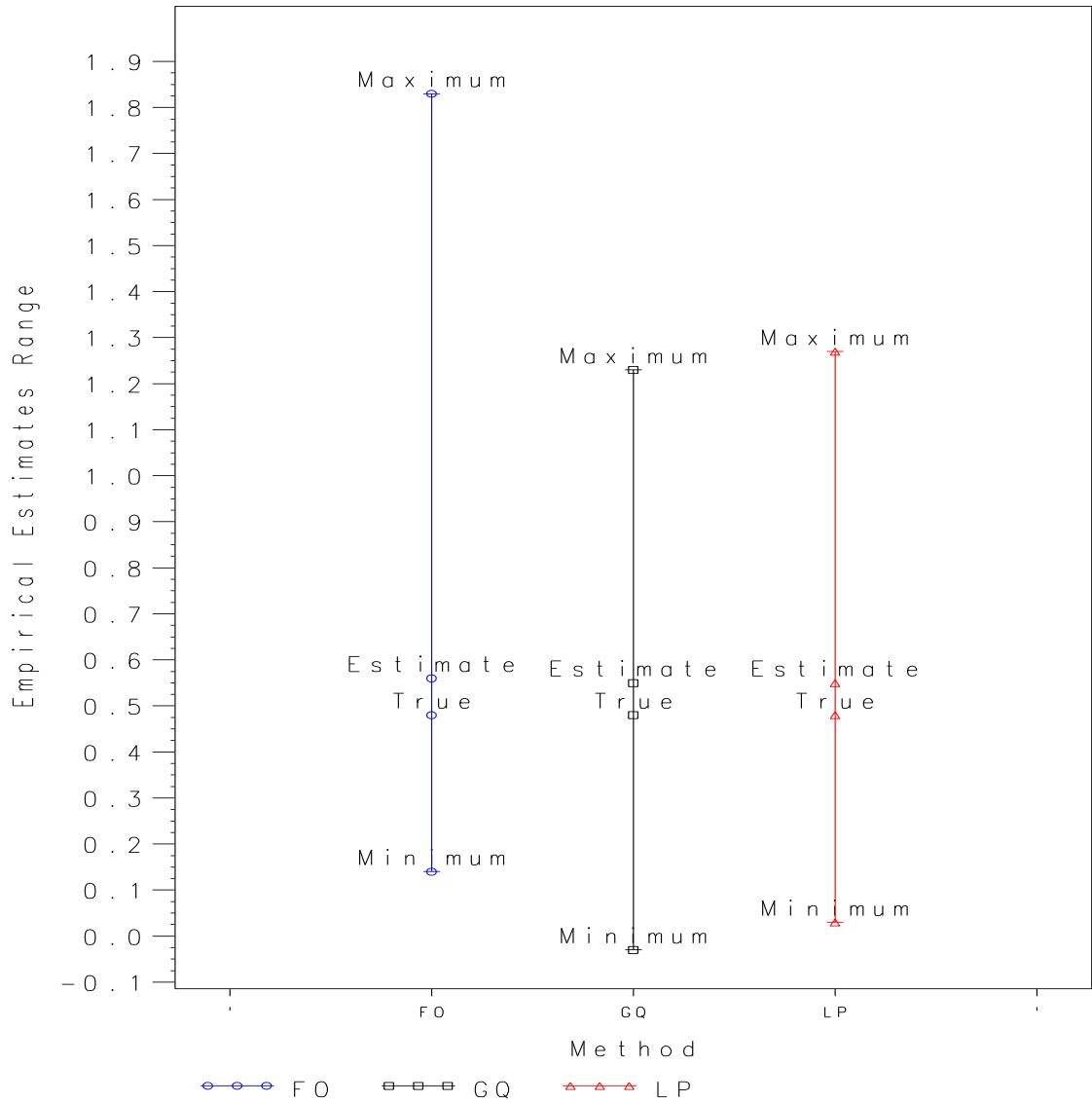


**Figure 2.2:**  $\beta_1$  Estimate Ranges from Model 2.1 with ( $N = 120$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

It is worth observing that the true parameter value of  $\beta_1$  and its estimate are the closest for the GQ method. On the other hand it is also worth noting the wide estimate ranges of LP and GQ methods.

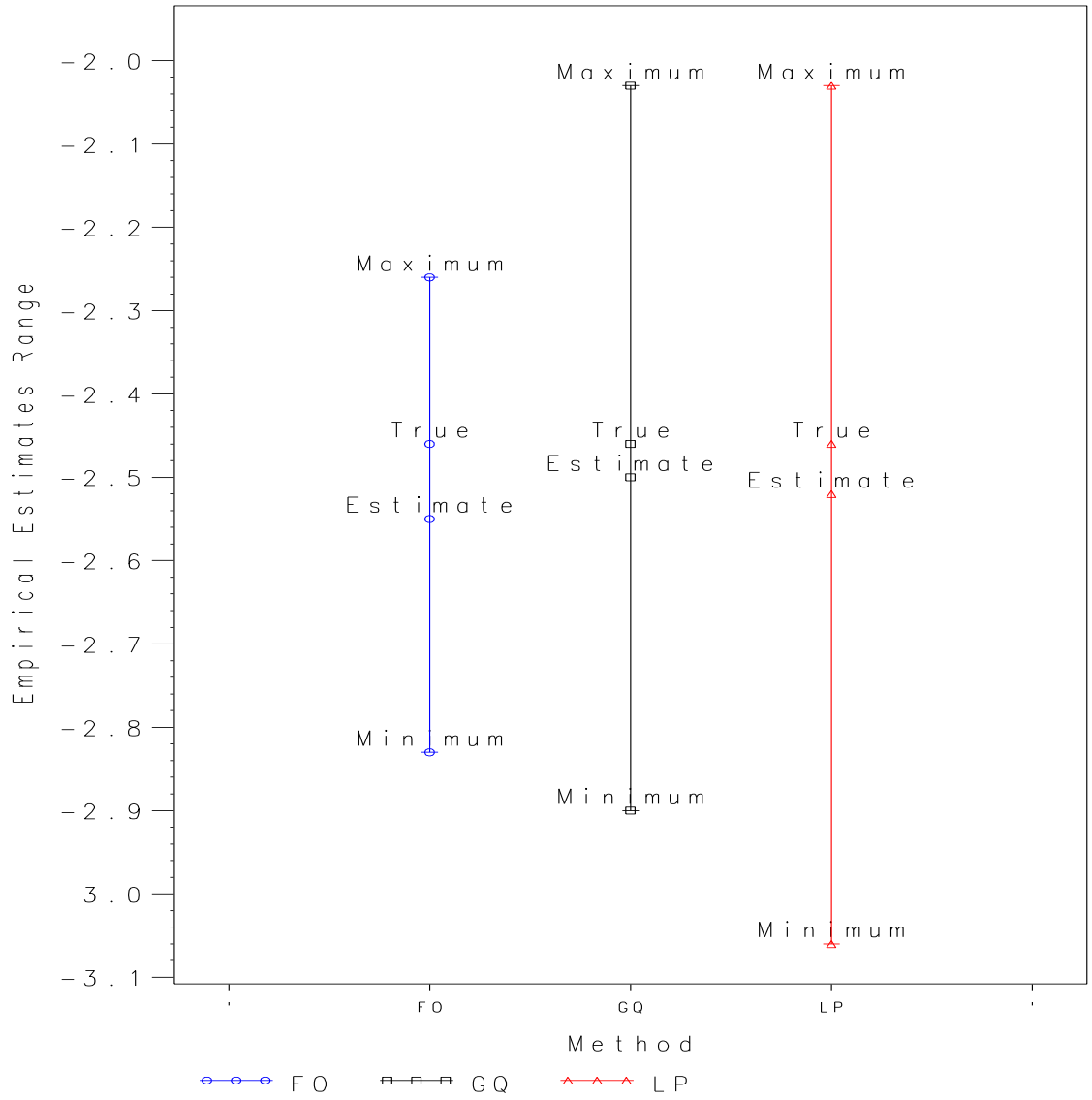
The wide estimate ranges is due to some of the simulated models resulting in  $\beta_1$ 's estimates that are too far from the true value. The LP and GQ methods seem more likely to results in extreme estimates.



**Figure 2.3:**  $\beta_2$  Estimate Ranges from Model 2.1 with (N = 120) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

It is worth noting that the true parameter value of  $\beta_2$  and its estimate are the closest for the GQ and LP methods. On the other hand, it is also worth noting the wider  $\beta_2$ 's estimate ranges from the FO method. The wide estimate ranges is due to some of the simulated models resulting in  $\beta_2$ 's estimates that are too far from the true value. The FO method seems to be the one that gives the extreme estimates.



**Figure 2.4:**  $\beta_3$  Estimate Ranges from Model 2.1 with ( $N = 120$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The true parameter value of  $\beta_3$  and its estimate are the closest for the GQ and LP methods. On the other hand we observed that the width of  $\beta_3$ 's estimate ranges of the LP method is too large. The

wide estimate ranges is due to some of the simulated models giving estimates that result in  $\beta_3$ 's estimates that are too far from the true value (much lower or/and much higher than the true value). The LP method seems to be the one with the worst dispersed estimates.

#### **2.4.2 Theophylline Simulation Results: ( $N = 240$ )**

In this second simulation where the Theophylline dataset's sample sizes were doubled by replicating each observation twice, results confirmed our initial finding that the GQ and LP methods give the best mix of efficiency and accuracy; and when the sample size increased such accuracy increased as well. The convergence rates improved from 92% to 97% for the GQ method and from 93% to 97% for the LP method. The coverage probability also improved and ranged from 95% for  $\beta_1$  and  $\beta_2$  to 97% for  $\beta_3$  for both the GQ and the LP methods. One can observe that the GQ and LP methods gave almost similar estimation and accuracy when the Theophylline replicates' sample sizes were doubled. The FO method also improved for the efficiency and accuracy; the relative bias, convergence rate and

coverage probability improved for all three parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  (see table 2.2).

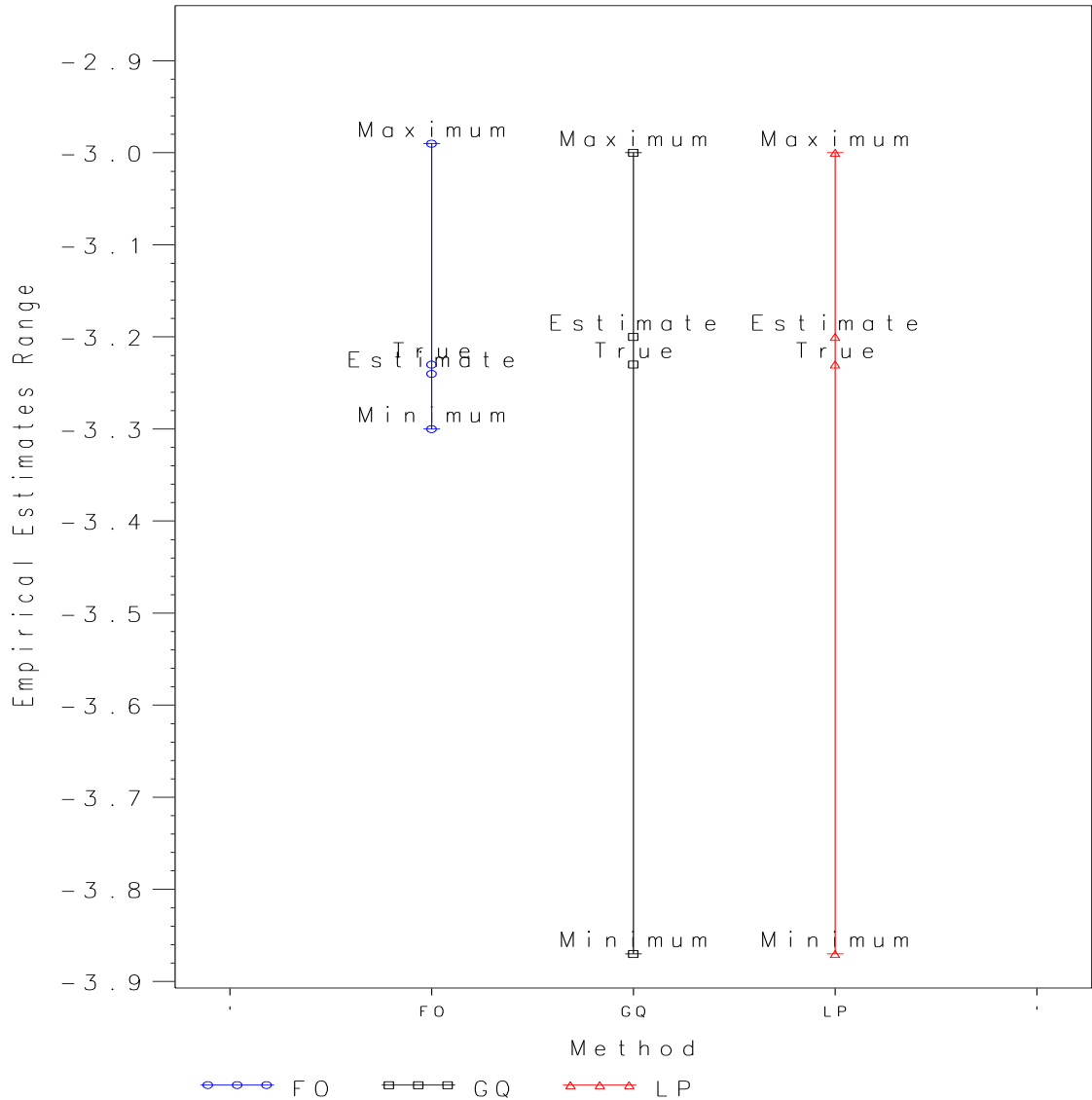
In this simulation we were also able to improve the results, when we increased the sample size from  $N = 120$  observations to  $N = 240$  observations, by replicating each subject in the Theophylline data twice. Bias improvement was between 2% for  $\beta_3$  to 30% for  $\beta_2$ . In addition, the coverage probability rates have increased by at least 7%. The minimum coverage probability rate for the first simulation ( $N = 120$ ) was 88%; however, the minimum coverage probability rate in the second simulation ( $N = 240$ ) was about 95%, a net gain of 7%. It should be noted that the increase in the sample size seemed to have the effect of underestimating most of the parameters estimates  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ . Table 2.2 summarizes the simulation results for all four methods used for the Theophylline dataset model with  $N = 240$ .

Parameter (True Value)	Method	Estimate	Min	Max	MSE	CV	Bias	CR	CP
$\beta_1 (= -3.23)$	GQ	-3.20	-3.87	-3.00	0.003	1.58%	-0.87%	97%	95%
	LP	-3.20	-3.87	-3.00	0.003	1.58%	-0.87%	97%	95%
	FO	-3.24	-3.30	-2.99	0.001	0.81%	0.31%	100%	100%
$\beta_2 (= 0.48)$	GQ	0.31	-0.34	0.89	0.041	37.12%	-34.82%	97%	95%
	LP	0.31	-0.34	0.89	0.041	37.12%	-34.82%	97%	95%
	FO	0.37	0.04	0.65	0.018	20.72%	-23.02%	100%	99%
$\beta_3 (= -2.46)$	GQ	-2.44	-2.72	-2.04	0.006	2.98%	-0.89%	97%	97%
	LP	-2.44	-2.72	-2.04	0.006	2.98%	-0.89%	97%	97%
	FO	-2.50	-2.62	-2.11	0.003	1.57%	1.53%	100%	100%

**Table 2.2:** Empirical Estimates for Theophylline Model ( $N = 240$ ). Estimates from converging models among 1000 cases.

For all three parameters,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  the estimate ranges, the difference between the maximum estimate value and the minimum estimate value were higher for both the GQ and LP approximation techniques. The FO approximation method has the largest estimate ranges. The GQ and LP approximation methods seem to be the more accurate approximation technique for  $\beta_2$ , the parameter with the smallest magnitude for which the distance between the estimates and the true values was the smallest. Figures 2.5, 2.6, and 2.7 display

the estimate ranges as well as the true parameter value for  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  respectively.



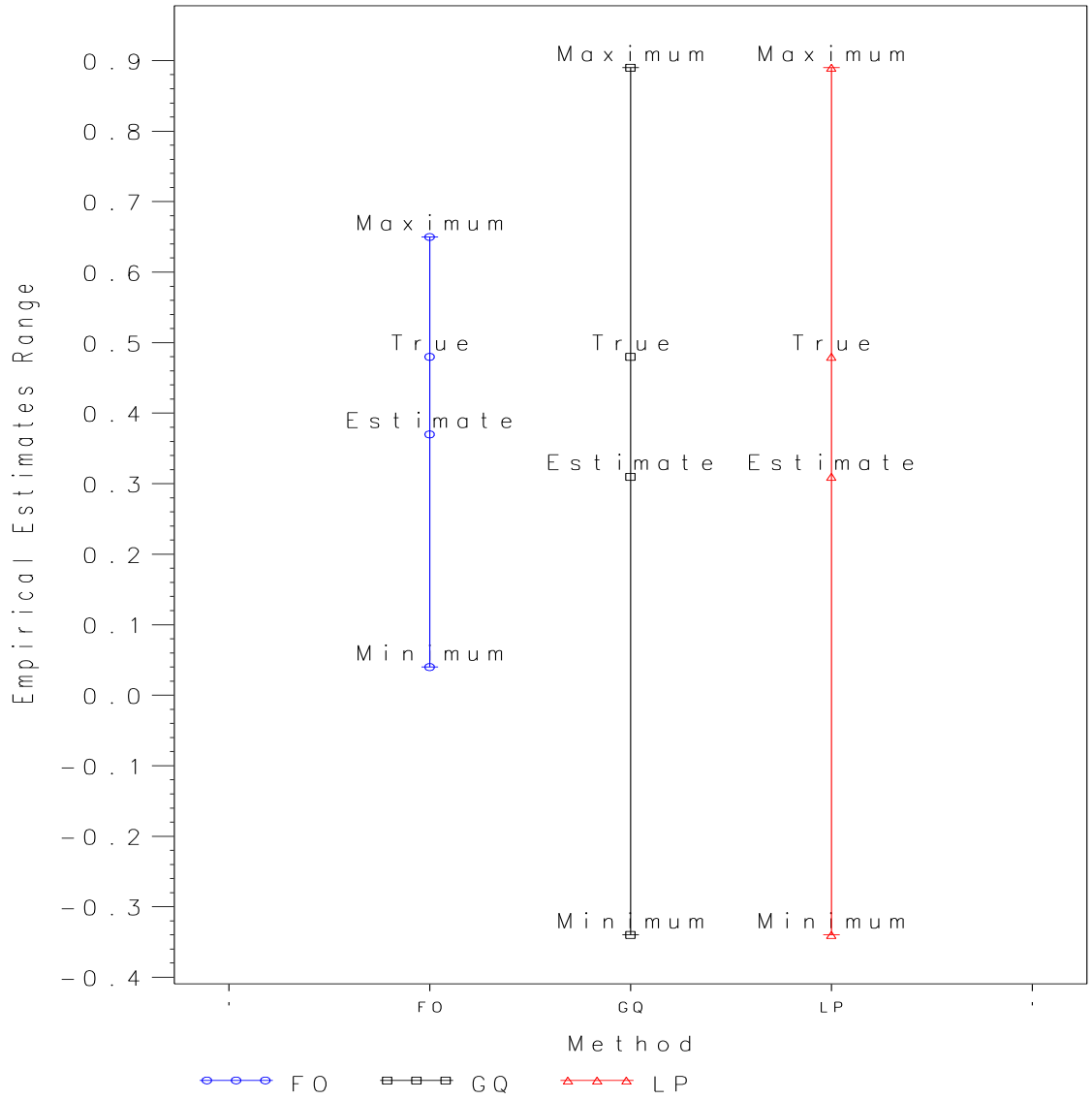
**Figure 2.5:**  $\beta_1$  Estimate Ranges from Model 2.1 with ( $N = 240$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The sample size increase seems to especially benefit the FO method



for estimation of  $\beta_1$ . The distance between the true parameter value of  $\beta_1$  and its estimate are the closest for the FO. The PL and GQ methods are consistent, in that they have the most dispersed estimate ranges. These two methods seem to underestimate  $\beta_1$  by relative biases of more than 20% .

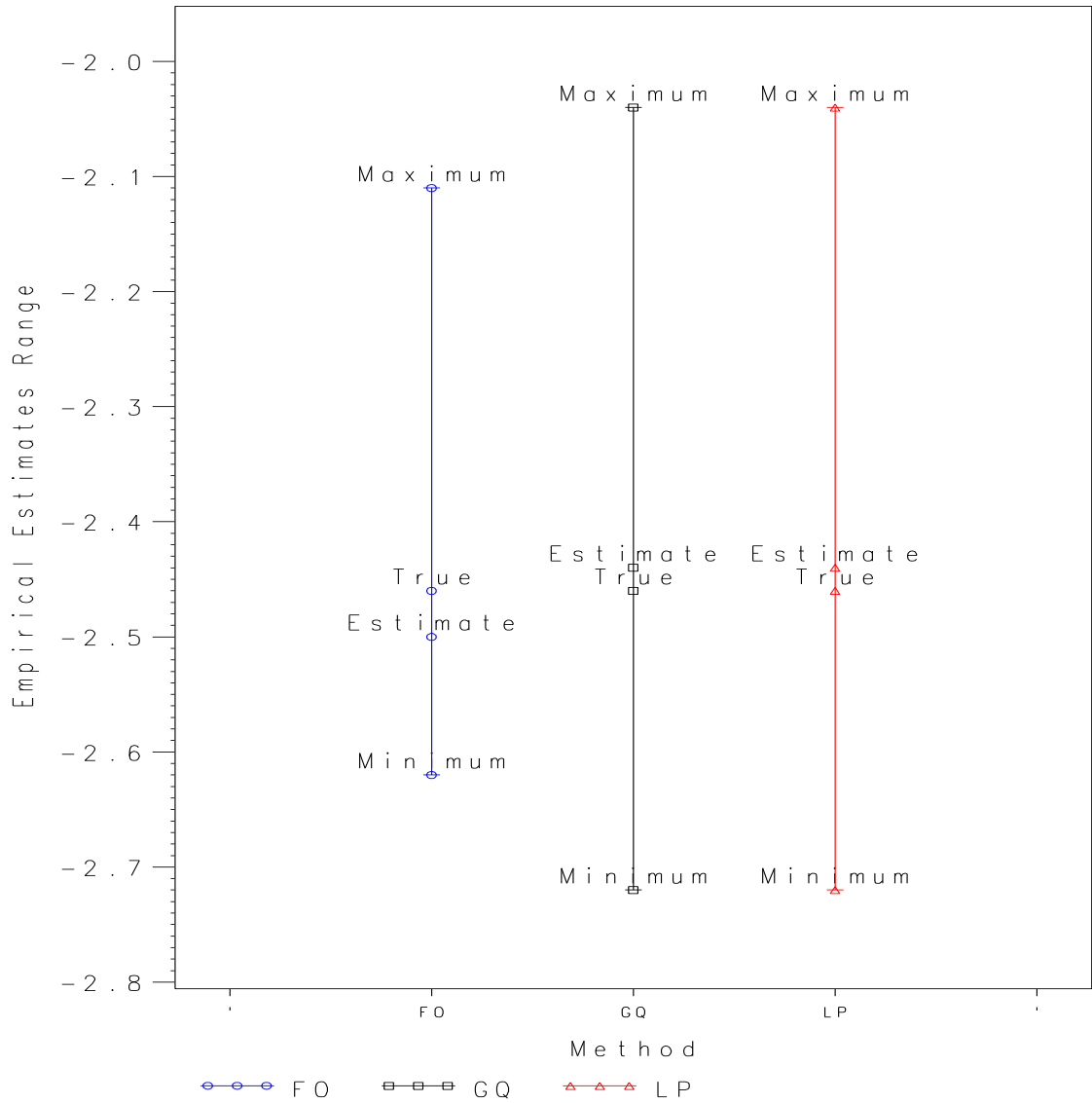


**Figure 2.6:**  $\beta_2$  Estimate Ranges from Model 2.1 with ( $N = 240$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_2$  the FO method seems to have the least dispersed estimate ranges. The distance between the true parameter value of  $\beta_2$  and its estimates are the closest for the FO. The PL and GQ methods

seem to have larger estimate ranges. These two methods sometimes underestimate  $\beta_2$  by relative biases of about  $-170\%$ , and in other times overestimate  $\beta_2$  by relative bias of about  $80\%$ .



**Figure 2.7:**  $\beta_3$  Estimate Ranges from Model 2.1 with ( $N = 240$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_3$  the FO method seems to be consistent in that they have the least dispersed estimate ranges. The LP and GQ methods have the least biases given that the distance between the true parameter value of  $\beta_3$  and its estimate are the closest. However, the PL and GQ methods seem to have larger estimate ranges. These two methods, even with estimates that are nearest to the true value of  $\beta_3$ , they sometimes tend to underestimate  $\beta_3$  by a relative bias of at least 10%. In other times they tend to overestimate it by a relative bias of more than 15%.

### 2.4.3 Carbon Dioxide Simulation Results: ( $N = 84$ )

In the Carbon Dioxide data, where an unrestricted variance-covariance matrix ( $\Psi$ ) was assumed, ten parameters in the model were to be estimated. The GQ and LP methods have the worst convergence rates and the worst coverage probability rates. Both methods failed to converge in three out of four cases (convergence rate is about 25%) and their coverage probabilities ranged from 67% for  $\beta_5$  to 86% for  $\beta_1$ . The coefficient of variation varied from 1.94% for  $\beta_4$  to 22.49% for  $\beta_3$ .

Even when these methods did converge they gave the largest MSEs and their 95% confidence intervals missed the true values 25% of the time.

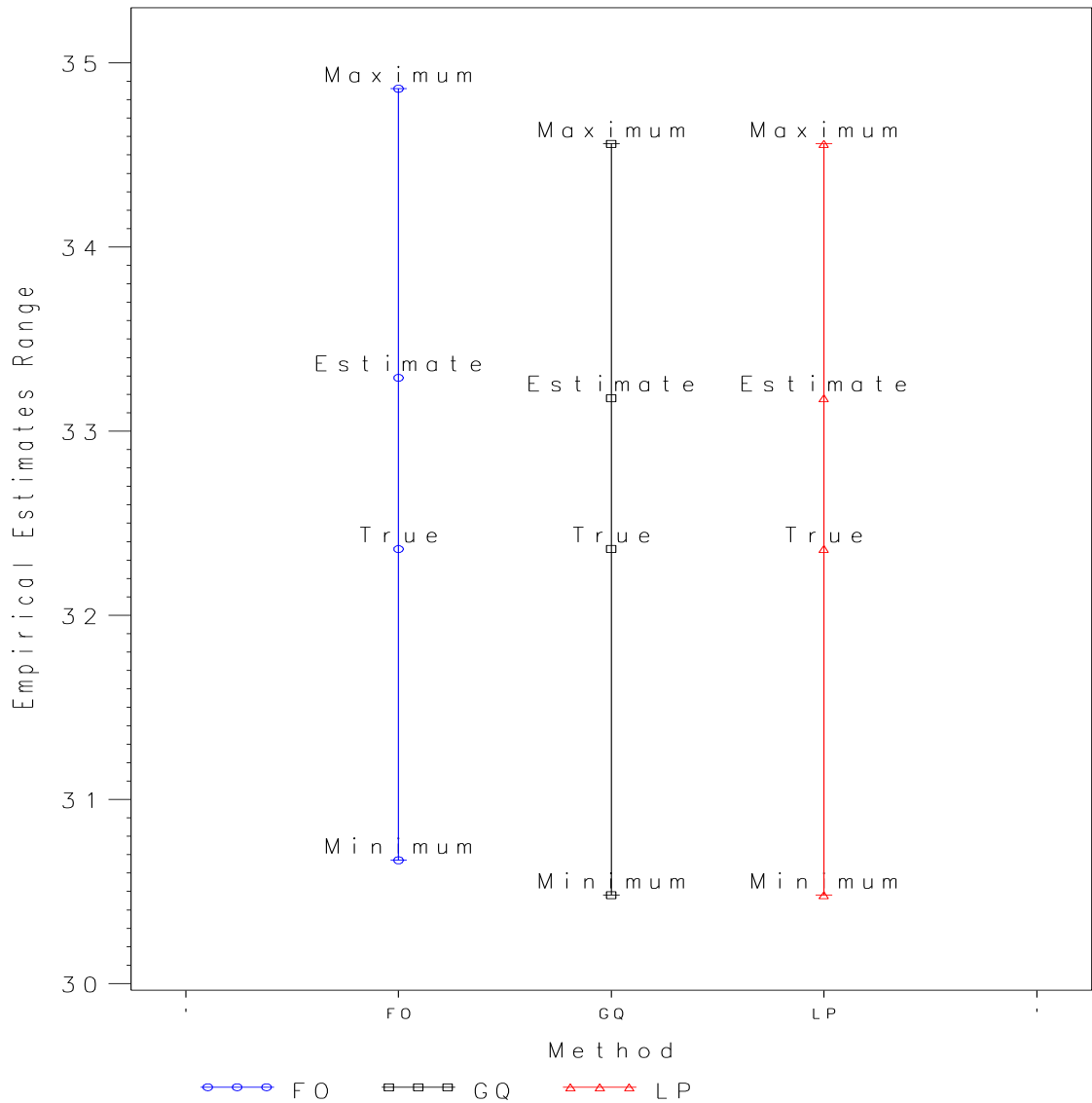
In contrast, the FO method provided the best convergence rates of almost 100%. In addition the FO approximation method gave the smallest coefficient of variation and the smallest MSE when compared to those of the GQ and LP approximations. Table 2.3 summarizes the simulation results for all four methods used for the Carbon Dioxide model with  $N = 84$ .

Parameter (True Value)	Method	Estimate	Min	Max	MSE	CV	Bias	CR	CP
$\beta_0 (= 32.36)$	GQ	33.18	30.48	34.56	1.094	1.97%	2.53%	25%	73%
	LP	33.18	30.48	34.56	1.094	1.97%	2.53%	25%	73%
	FO	33.29	30.67	34.86	1.011	1.13%	2.88%	100%	77%
$\beta_1 (= -7.18)$	GQ	-7.19	-8.12	-6.11	0.081	3.96%	0.13%	25%	86%
	LP	-7.19	-8.12	-6.11	0.081	3.96%	0.13%	25%	86%
	FO	-7.18	-8.27	-5.90	0.071	3.71%	0.03%	100%	80%
$\beta_2 (= -3.98)$	GQ	-4.02	-5.39	-3.11	0.104	8.00%	0.77%	25%	82%
	LP	-4.02	-5.39	-3.11	0.104	8.00%	0.77%	25%	82%
	FO	-3.98	-5.03	-3.18	0.075	6.88%	-0.05%	100%	78%
$\beta_3 (= -1.31)$	GQ	-1.36	-2.39	0.14	0.095	22.49%	3.30%	25%	79%
	LP	-1.36	-2.39	0.14	0.095	22.49%	3.30%	25%	79%
	FO	-1.33	-2.42	-0.45	0.073	20.34%	1.23%	100%	77%
$\beta_4 (= -4.68)$	GQ	-4.67	-4.88	-4.31	0.008	1.94%	-0.10%	25%	79%
	LP	-4.67	-4.88	-4.31	0.008	1.94%	-0.10%	25%	79%
	FO	-4.68	-5.15	-4.44	0.005	1.46%	0.11%	100%	85%
$\beta_5 (= 48.73)$	GQ	46.76	35.46	61.34	31.948	11.33%	-4.06%	25%	67%
	LP	46.76	35.46	61.34	31.948	11.33%	-4.06%	25%	67%
	FO	47.96	32.07	62.50	22.958	9.86%	-1.59%	100%	86%

**Table 2.3:** Empirical Estimates for Carbon Dioxide Model (N = 84).

Estimates from converging models among 1000 cases.

For all six parameters, the range of estimates of  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  (the difference between the maximum estimate value and the minimum estimate value) were higher for the FO,GQ and LP approximation techniques. The FO approximation method has the smallest estimate ranges. The FO approximation method seems to be the more accurate approximation technique for  $\beta_3$ , the parameter with the smallest magnitude for which the distance between the estimates and the true values were the smallest. Figures 2.8, 2.9, 2.10, 2.11, 2.12, and 2.13 display the range of estimates as well as the true parameter values for  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  respectively.



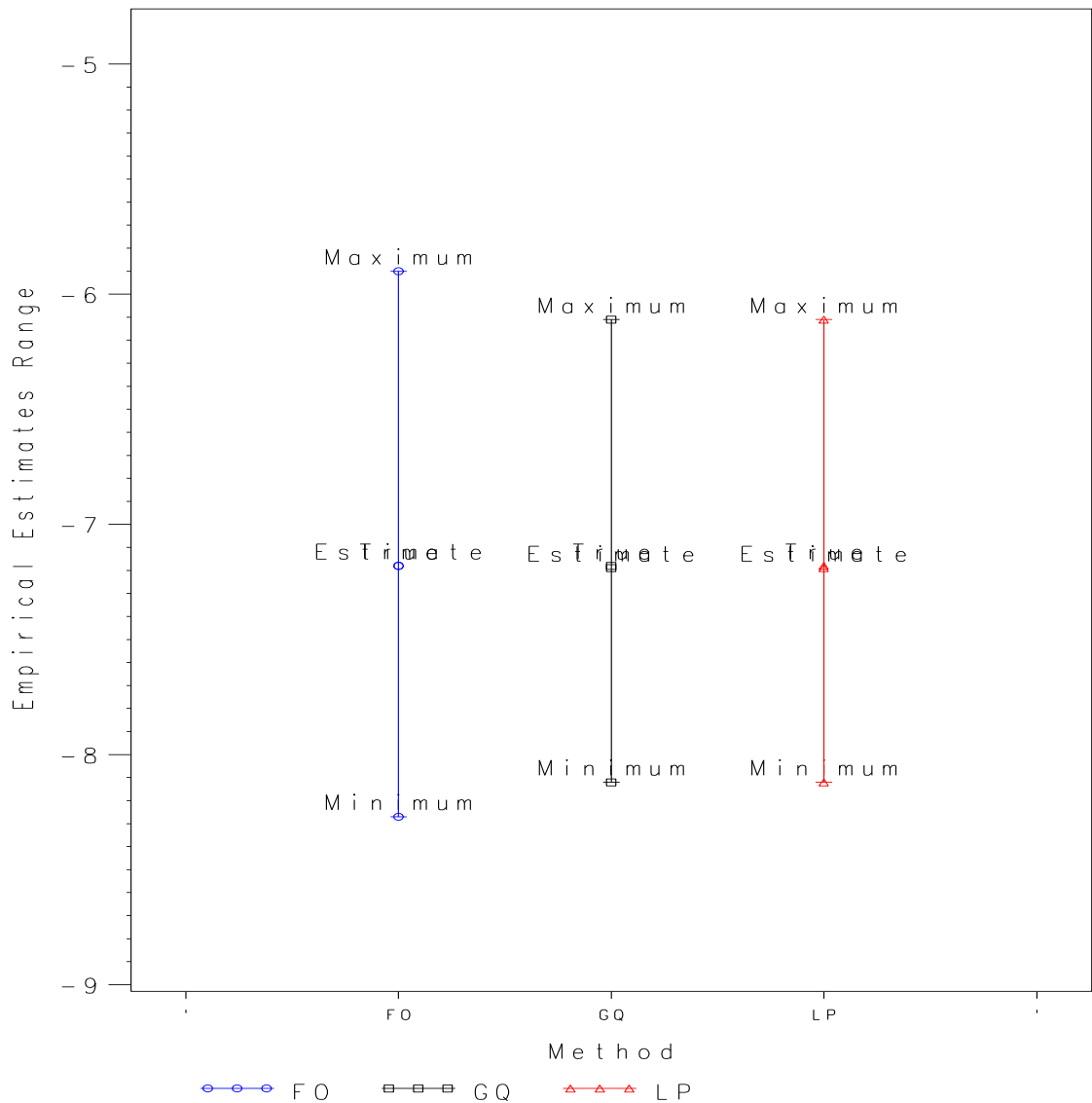
**Figure 2.8:**  $\beta_0$  Estimate Ranges from Model 2.2 with ( $N = 84$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_0$  all four methods seem to have similar biases. The FO, GQ and LP methods have about similar estimate range widths for this parameter. In addition, the GQ and LP methods have almost identical



estimate ranges.



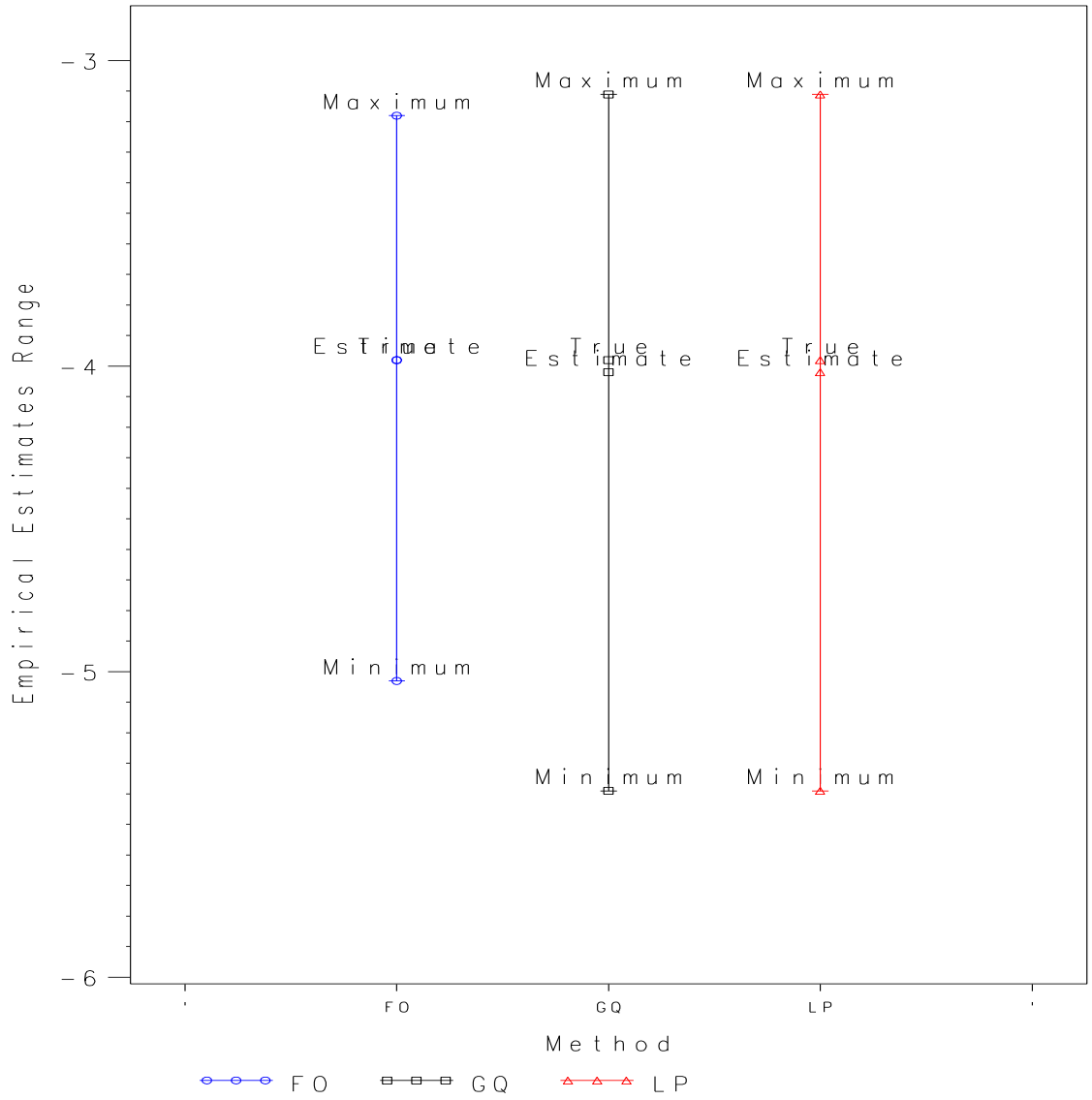
**Figure 2.9:**  $\beta_1$  Estimate Ranges from Model 2.2 with ( $N = 84$ )

Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The  $\beta_1$ 's estimate from the FO method is very much identical to the true parameter value. This can be explained by the fact that

the marginal distribution is a linear function of  $\beta_1$ . The other three methods also have relatively small biases. However, the FO method seems to have more dispersed estimate ranges. The GQ and LP methods have about the same estimate range widths and they have almost identical estimate ranges.

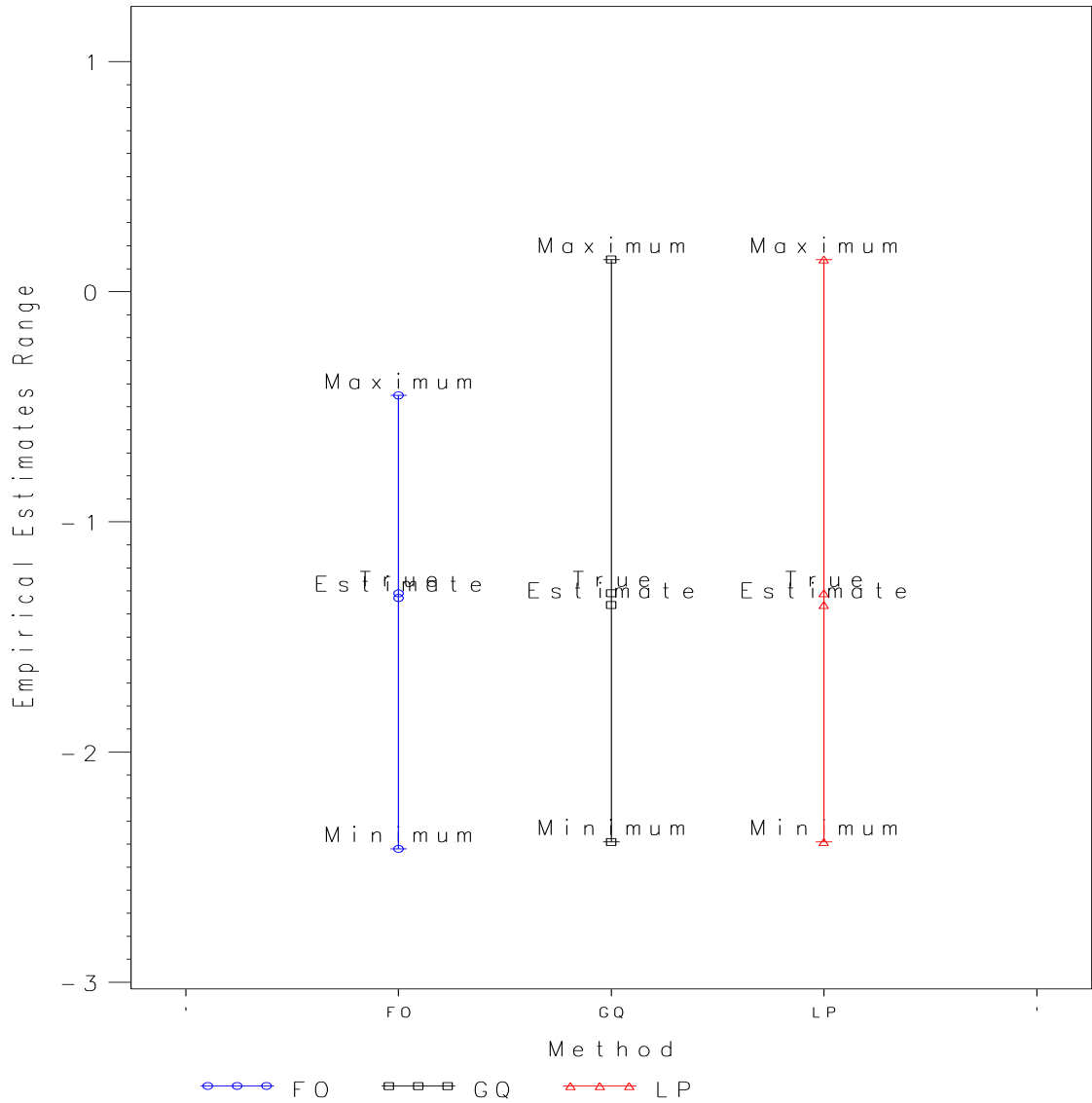


**Figure 2.10:**  $\beta_2$  Estimate Ranges from Model 2.2 with ( $N = 84$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The  $\beta_2$ 's estimate from the FO method is very much identical to the true parameter value. This is also due to the fact that the marginal distribution of  $y_{ij}$  is a linear function of  $\beta_2$ . The other three

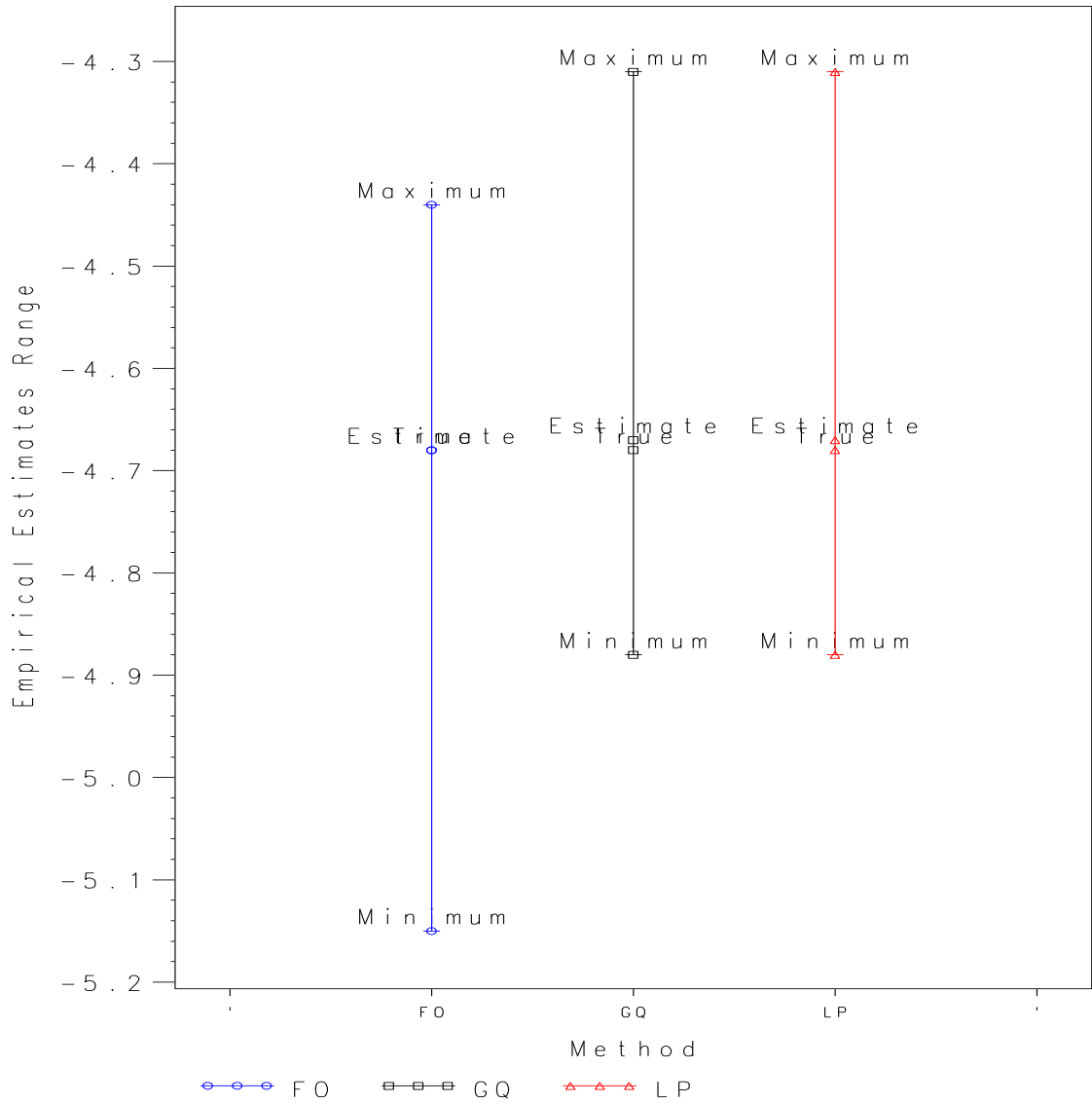
methods also have relatively small biases. However, the FO method seems to have the least dispersed estimate ranges. The GQ and LP methods have about the same estimate range width and they have almost identical estimate ranges.



**Figure 2.11:**  $\beta_3$  Estimate Ranges from Model 2.2 with ( $N = 84$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

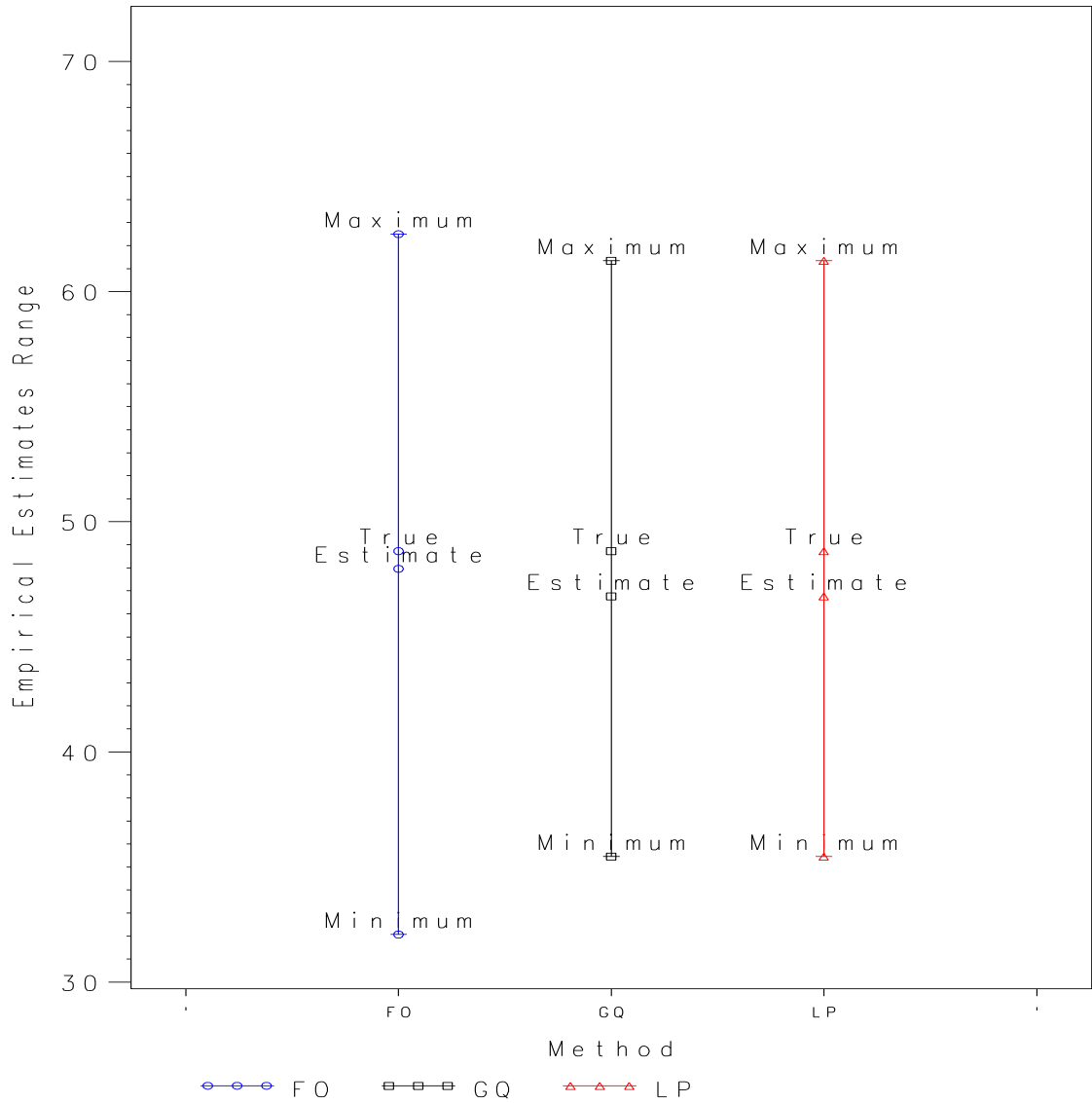
The  $\beta_3$ 's estimate from the FO method has the smallest bias. The FO method has the least dispersed estimate ranges. The GQ and LP methods have almost identical estimate ranges.



**Figure 2.12:**  $\beta_4$  Estimate Ranges from Model 2.2 with (N = 84) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The  $\beta_4$ 's estimates from the FO, LP and GQ methods have the smallest biases. However, the FO method has the widest estimate range. The GQ and LP methods have almost identical estimate ranges.



**Figure 2.13:**  $\beta_5$  Estimate Ranges from Model 2.2 with (N = 84) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The  $\beta_5$ 's estimate from the FO method has the smallest bias. However, the FO method has the widest estimate range. The GQ and LP methods have almost identical estimate ranges.

#### **2.4.4 Carbon Dioxide Simulation Results: ( $N = 168$ )**

In the second Carbon Dioxide model, when we increased the sample from  $N = 84$  observations to  $N = 168$  observations by replicating each subject in the dataset twice, there was a small improvement in the convergence rates and the coverage probabilities; but there was no consistent improvement in the precision of the estimation compared to the original dataset with a sample size  $N = 84$ . For the LP method the convergence rate increased from 25% to 32%. However the coverage probability decreased on average by 20%. For the GQ method the estimation parameters were almost unchanged. The FO approximation method was relatively better when the sample size increased; however, the improvement may not be worth the cost to double the sample size.

This issue illustrates the difficulty we face when trying to select

an approximation method to the log-likelihood in order to estimate the NLMM's parameters. Such selection hinges around the purpose of the study and on the method which is best in keeping with that purpose. In addition the amount of data available per subject will determine to some extent what method is preferred. Table 2.4 summarizes the simulation results for all four methods for the Carbon Dioxide model with  $N = 168$ .

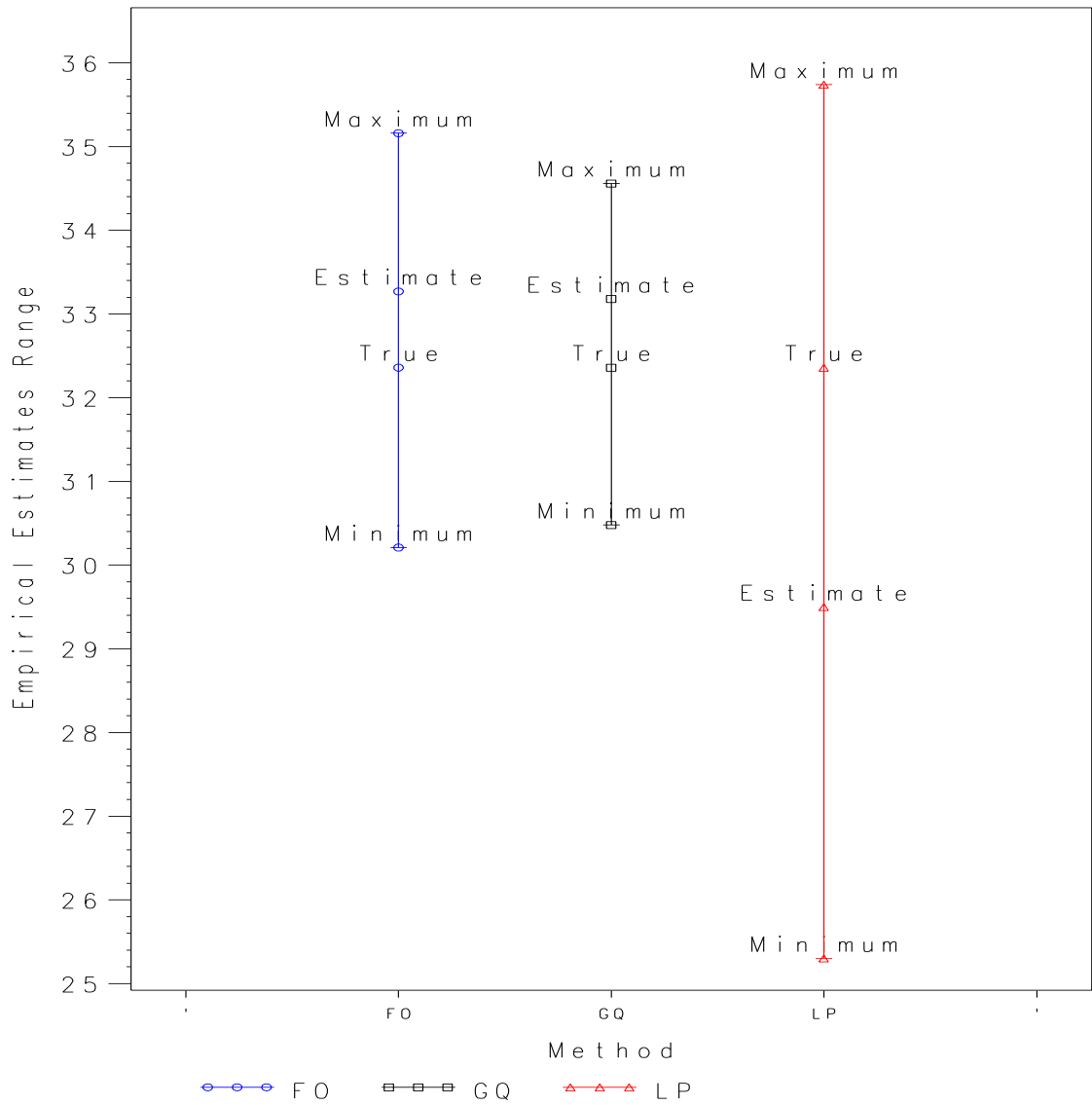


Parameter (True Value)	Method	Estimate	Min	Max	MSE	CV	Bias	CR	CP
$\beta_0(= 32.36)$	GQ	33.18	30.48	34.56	1.094	1.97%	2.53%	25%	73%
	LP	29.50	25.30	35.74	8.764	2.66%	-8.82%	32%	53%
	FO	33.27	30.21	35.16	0.937	1.01%	2.81%	100%	84%
$\beta_1(= -7.18)$	GQ	-7.19	-8.12	-6.11	0.081	3.96%	0.13%	25%	86%
	LP	-6.33	-8.39	-6.30	0.982	8.11%	-11.80%	32%	57%
	FO	-7.18	-8.20	-6.08	0.040	2.80%	-0.04%	100%	82%
$\beta_2(= -3.98)$	GQ	-4.02	-5.39	-3.11	0.104	8.00%	0.77%	25%	82%
	LP	-2.38	-4.72	1.70	2.922	24.40%	-40.36%	32%	52%
	FO	-3.99	-4.62	-3.20	0.038	4.89%	0.03%	100%	82%
$\beta_3(= -1.31)$	GQ	-1.36	-2.39	0.14	0.095	22.49%	3.30%	25%	79%
	LP	-1.64	-5.70	0.21	0.328	28.34%	25.31%	32%	57%
	FO	-1.31	-2.01	-0.57	0.040	15.20%	0.16%	100%	81%
$\beta_4(= -4.68)$	GQ	-4.67	-4.88	-4.31	0.008	1.94%	-0.10%	25%	79%
	LP	-4.43	-9.10	-3.88	0.207	8.64%	-5.24%	32%	57%
	FO	-4.68	-5.17	-4.42	0.003	1.15%	-0.01%	100%	87%
$\beta_5(= 48.73)$	GQ	46.76	35.46	61.34	31.948	11.33%	-4.06%	25%	67%
	LP	39.93	17.52	49.82	95.830	10.72%	-18.06%	32%	49%
	FO	48.54	36.62	59.01	10.935	6.80%	-0.39%	100%	85%

**Table 2.4:** Empirical Estimates for Carbon Dioxide Model(N = 168).

Estimates from converging models among 1000 cases.

For all six parameters the sample size increase did not change the overall conclusions. The FO, GQ, and LP approximation techniques have higher estimate ranges. The FO approximation method seems to be the more accurate approximation technique for  $\beta_3$ , the parameter with the smallest magnitude for which the distance between the estimate and the true values was the smallest. Figures 2.14, 2.15, 2.16, 2.17, 2.18, and 2.19 display the estimate ranges as well as the true parameter values for  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  respectively.

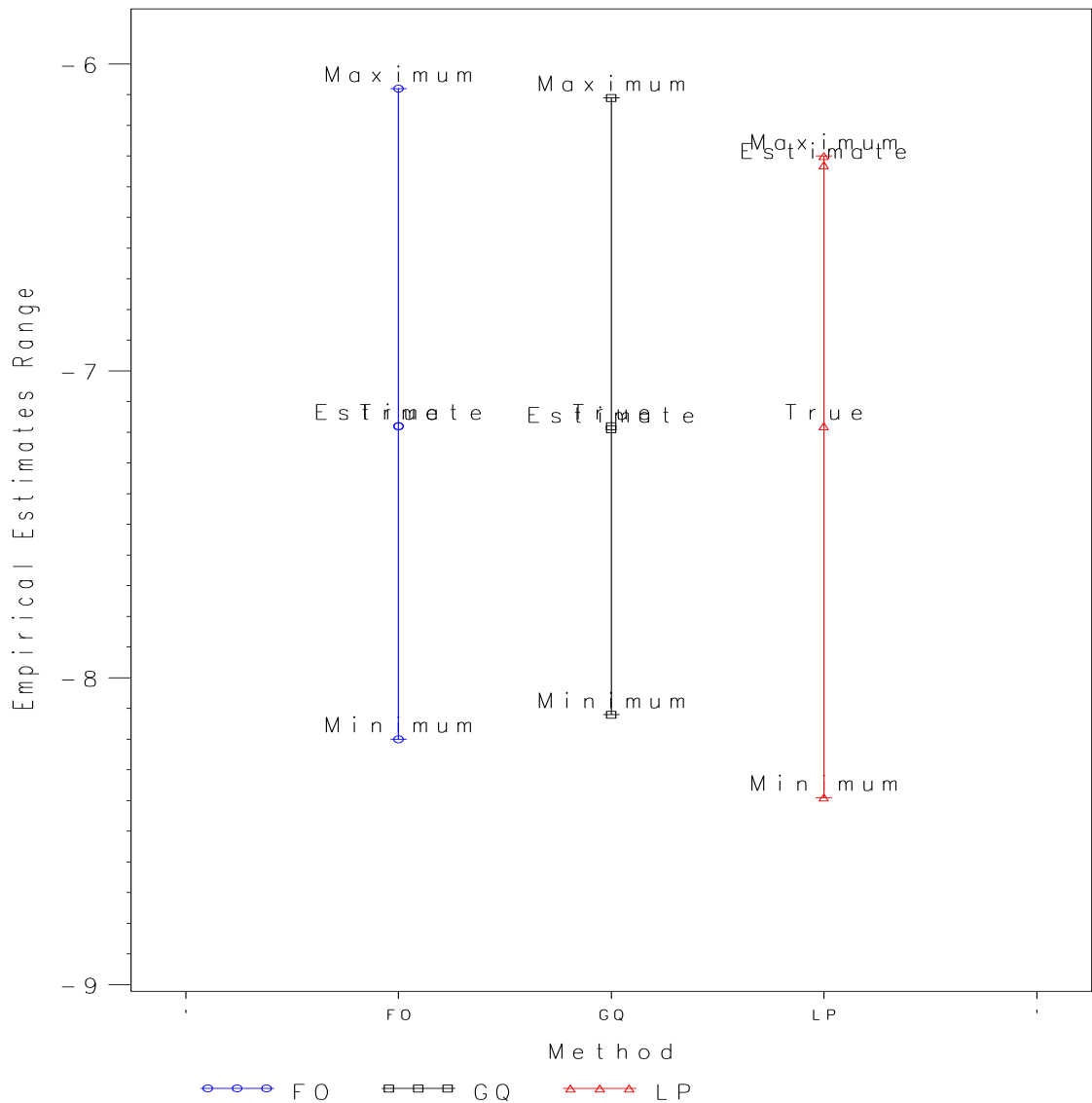


**Figure 2.14:**  $\beta_0$  Estimate Ranges from Model 2.2 with ( $N = 168$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_0$  the FO and GQ methods seem to have similar and unchanged biases with respect to sample size increase. The GQ method has the least dispersed estimate ranges. However, the PL method has

the largest bias and the widest estimate range.

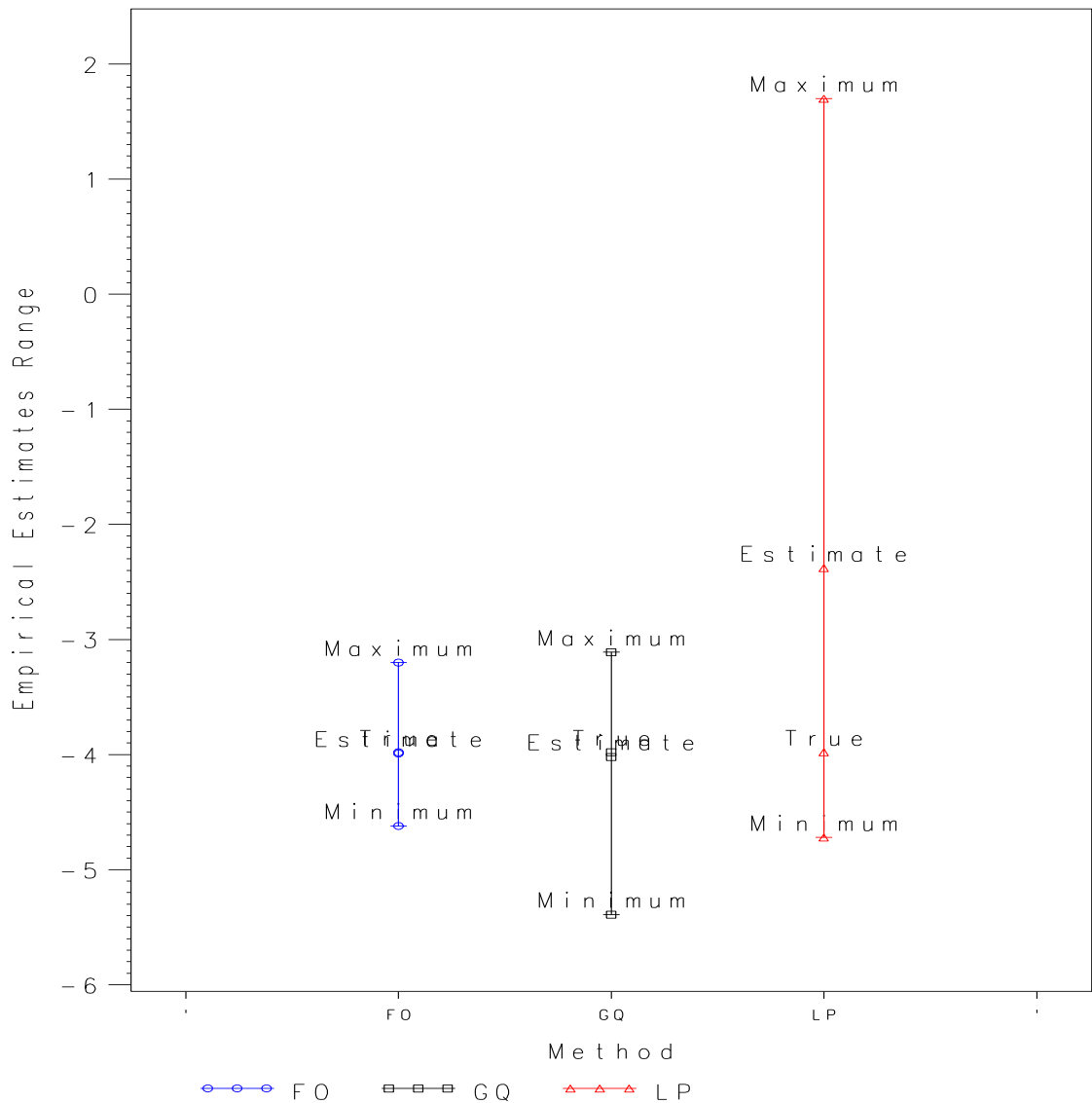


**Figure 2.15:**  $\beta_1$  Estimate Ranges from Model 2.2 with ( $N = 168$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

The sample size increase seems to especially benefit the FO method for estimating  $\beta_1$  with almost no bias. The FO, PL, and GQ methods

have similar estimate ranges.

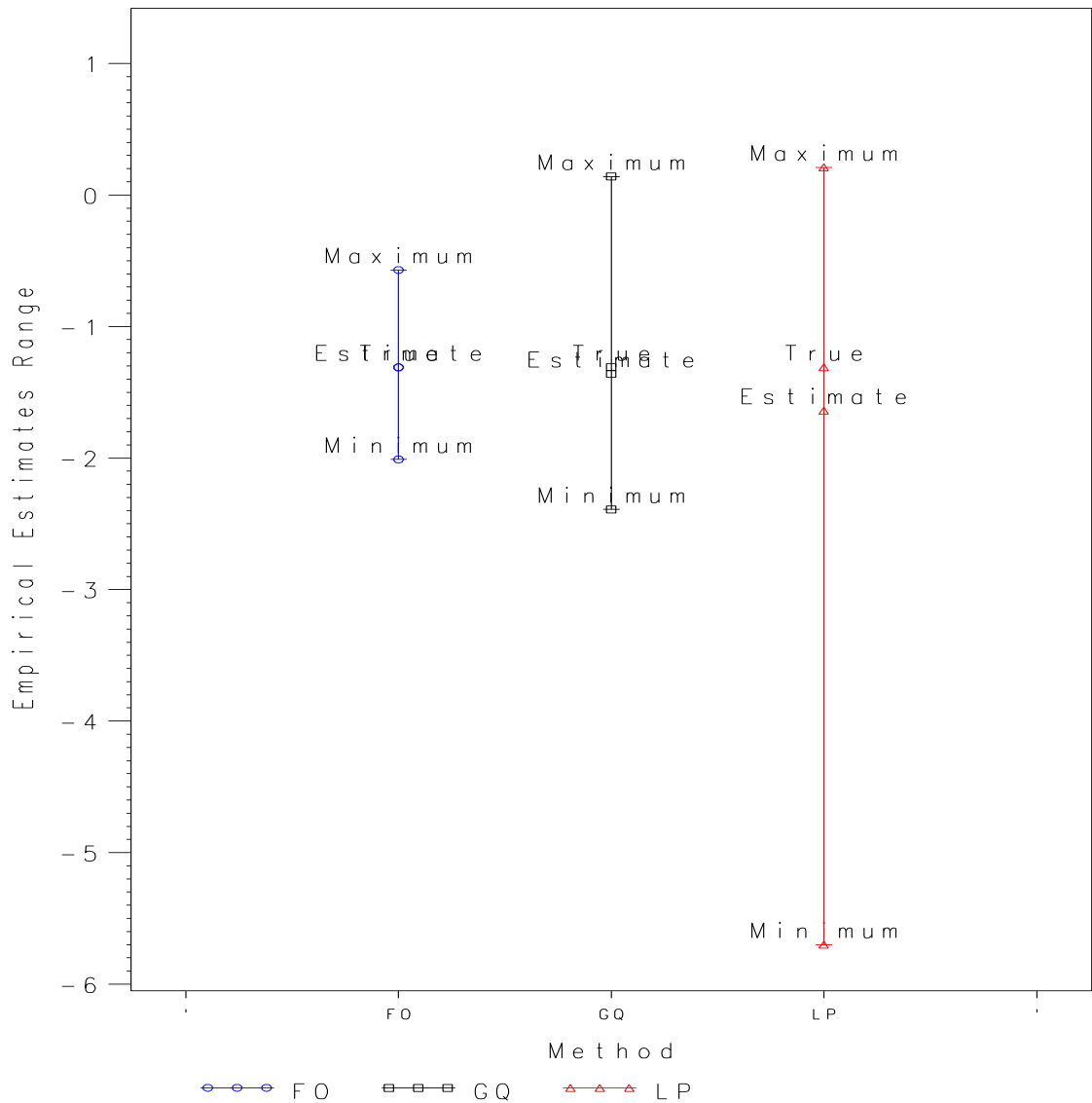


**Figure 2.16:**  $\beta_2$  Estimate Ranges from Model 2.2 with (N = 168) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_2$  the FO method has the smallest bias and about the same estimate range. The PL method has the most dispersed estimate

range, is the most overestimating of  $\beta_2$ , and has the largest relative bias of about  $-40\%$ .

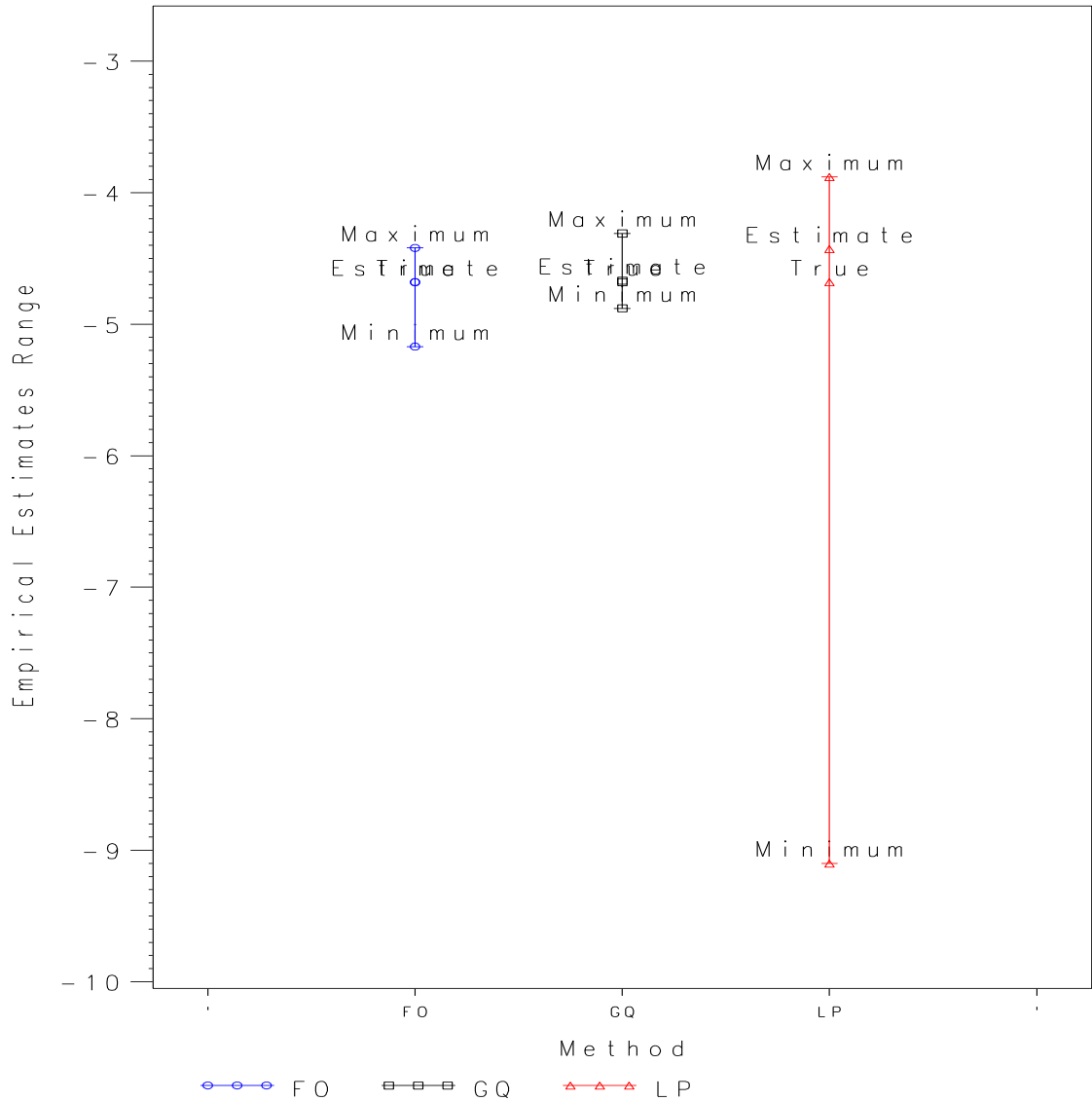


**Figure 2.17:**  $\beta_3$  Estimate Ranges from Model 2.2 with ( $N = 168$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_3$  the FO method has the smallest bias and about the same

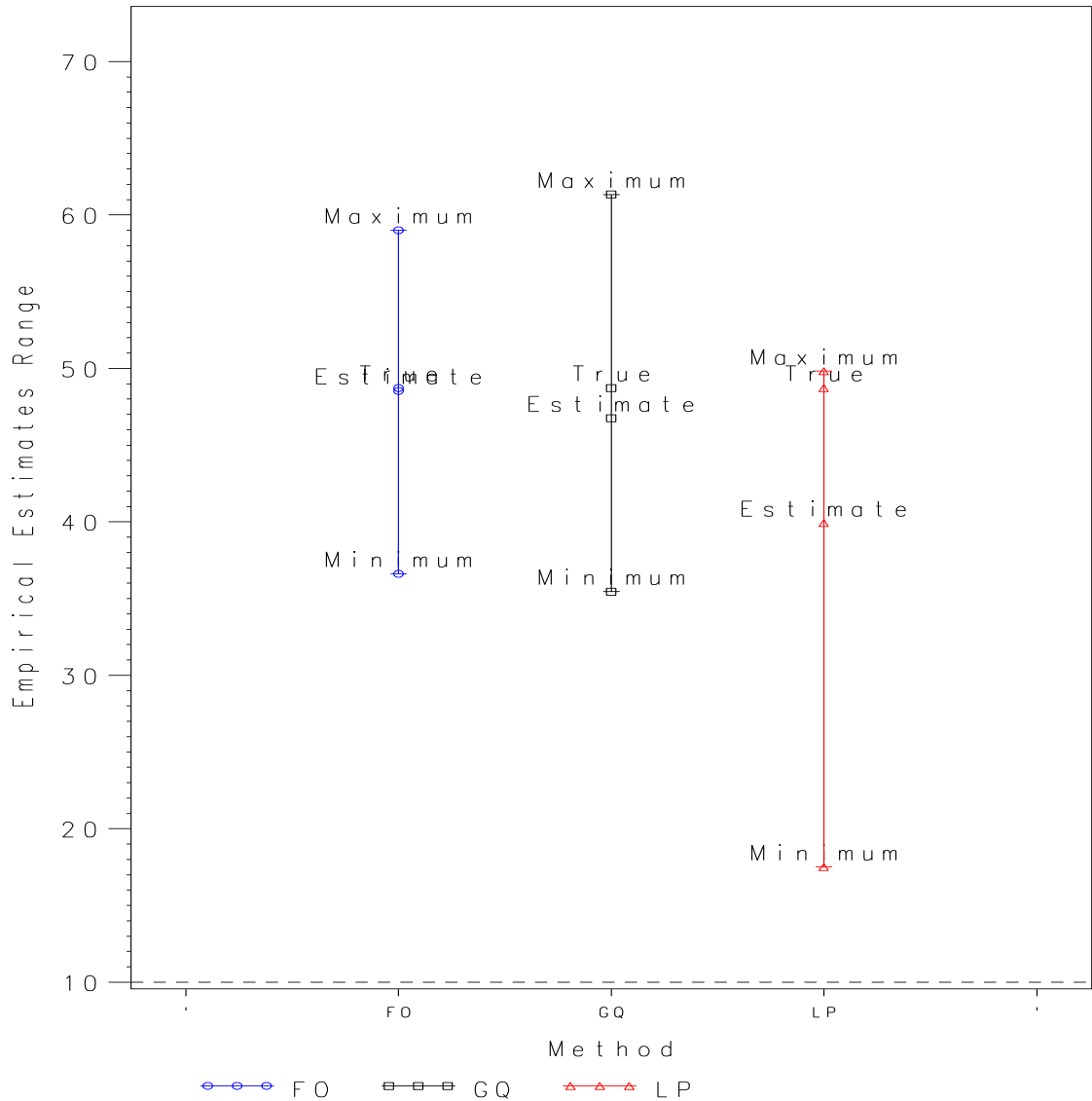
estimate range. The PL method has the most dispersed estimate range, is the most underestimating of  $\beta_3$ , and has the largest relative bias of about 25%.



**Figure 2.18:**  $\beta_4$  Estimate Ranges from Model 2.2 with ( $N = 168$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.

For  $\beta_4$  the FO and GQ methods have relative biases of less than 1% and both methods have similar estimate ranges. The PL method has the most dispersed estimate range, is the most underestimating of  $\beta_4$ , and has the largest relative bias among all methods of about 5%.



**Figure 2.19:**  $\beta_5$  Estimate Ranges from Model 2.2 with ( $N = 168$ ) Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and FO = First Order Method.



For  $\beta_5$  the FO and GQ methods have relative biases of less than 5% and both methods have about similar estimate ranges. The PL method has the most dispersed estimate range, is the most underestimating of  $\beta_5$ , and has the largest relative bias among all methods of about  $-18\%$ . In addition it is worth observing that the maximum of the estimates is very close to the true value. This observance can explain the low coverage probability of  $\beta_5$ .

## 2.5 Conclusion

The question of which approximation yields the best estimates and the degree of precision associated with it seems to depend greatly on many factors. We tried in this work to address and explore some of these factors, the magnitude of variability among the random effects, the random parameters covariance structure, and the way in which such random parameters enter the model as well as the “linearity” or the “close to linearity” of the model as a function of these random parameters.

The coefficients of variation (CV) provides a measure of such mag-

nitude. A number of simulations have been carried out to compare the performance of the various Taylor approximations Beal and Sheiner [5], Lindstrom and Bates [25] , and Davidian and Giltinan [11]. While no method outperformed the other on a consistent basis, both the bias and efficiency of the variance estimates were found to depend on the magnitude of the inter- and intra- subject coefficients of variation. In particular, as the coefficient of variation between and within subject variation increased, the bias in the FO estimates increased. The FO method has been widely used and there are many examples for which it performs well. It has the advantage that it is exact when the model is linear in the random effects. It also has the advantage of being computationally simple and provide a reasonable convergence rates. However as the probability of  $b_i$  being zero increases (The Variance of the random effects much much greater than zero), the accuracy of the the FO approximation decreases. Vonesh [47] showed some improvement of the LP approximation under the assumption that the number of subject and observations per subject were sufficiently large. Both the LP and GQ methods were very similar and provided similar

estimates and coverage probability and seem to give better coverage than the FO method. For the parameter with the smallest magnitude,  $\beta_2$  the LP and GQ methods have poor precisions with biases above 10% and coverage probabilities below 90%.

The Theophylline dataset's model provided an acceptable confidence interval coverage when the sample size was increased, similar increase in the Carbon Dioxide did not provide the expected 95% coverage.

## Chapter 3

### Diagnostics of the Random Effects

This chapter discusses our investigation into the robustness and sensitivity of the three approximation techniques used to approximate the likelihood function of model 1.2, in terms of the structure of the random effect parameters, the dimension of these parameters, and the correlation structure of the covariance matrix. In this section we expand the work of Hartford and Davidian [18] to assess the robustness of the approximation methods under different scenarios (models) of random effect covariance structures:

- (1) under assumption of single univariate random effect models;
- (2) under assumption of correlated multivariate random effects models; and
- (3) under assumption of Non-correlated multivariate random effects

models.

This different covariance structure models allows us to better assess the effect of the number of random effects in the model and the correlation level of these random effects on the estimation methods.

### 3.1 Simulation

In these simulations we investigate the robustness and the sensitivity of the three approximation techniques in terms of the random effects correlation and the level of that correlation, as well as the effects of the number of random effects in the model, on the fixed effect parameters' estimation and their precisions. The three methods of approximation of FO, GQ, and LP were put under the same conditions and their estimation and precision under these conditions were compared. In the simulation, we assumed a more complex model version of Davidian and Giltinan [12], model 2.1. The first-order open-compartment model is used to express the serum concentration  $y_{ij}$  for patients  $i^{th}$  at observation  $j^{th}$  after an initial dose  $D_j$  as follows:

$$y_{ij} = D_j \frac{\exp[-(\alpha_1 + \alpha_2 + \alpha_3)]}{\exp(\alpha_2) - \exp(\alpha_3)} \{ \exp[-t_{ij} \exp(\alpha_3)] - \exp[-t_{ij} \exp(\alpha_2)] \} + \epsilon_{ij} \quad (3.1)$$

$$\alpha_1 = \beta_1 + b_{i1},$$

$$\alpha_2 = \beta_2 + b_{i2},$$

$$\alpha_3 = \beta_3 + b_{i3}$$

In this section we define the general form of the model's random effects covariance that we will be using in the entire chapter. Further we define the special case in each corresponding section.

The general form of the random effects structure is given as follows:

$$\begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \right)$$

$$b_i = (b_{i1}, b_{i2}, b_{i3}), \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), i = 1, \dots, 12 \text{ and } j = 1, \dots, 11$$

We simulate 1000 replicates of the Theophylline dataset using the following true parameter values:

$$\beta_1 = -3.2269, \beta_2 = 0.4782, \beta_3 = -2.4594 \text{ and } \sigma^2 = 0.5581$$

### 3.2 Single Random Component Model

We first consider a single random effect model in which there is only one single random effect parameter in the model 3.1. Doing so allows us to better understand the effects of the dimension of the random effect parameters, as well as the effect of the way in which the random effect parameters enter the model, on the estimations. We will assume three cases of a single random effect parameter model for these simulation as listed in table 3.1.

Given that each of these cases correspond to a different level of linearity and complexity level in the mixed model, we will compare the single random component parameter model with the multiple random component parameters model that we will describe in the next following sections. Additionally, we will analyze the effect of the number of

random component parameters on the estimation methods as well as the way in which the random effect enters the model. We will consider a model with only one random component parameter associated with one of the three parameters,  $\beta_1$ ,  $\beta_2$ , or  $\beta_3$  respectively.

Case Model	Random Effect	Variance	Assumption
Model 1	$b_{i1}$	$\sigma_1^2 = 0.02787$	$\sigma_2^2 = \sigma_3^2 = \sigma_{12} = \sigma_{13} = \sigma_{23} = 0$
Model 2	$b_{i2}$	$\sigma_2^2 = 0.4239$	$\sigma_1^2 = \sigma_3^2 = \sigma_{12} = \sigma_{13} = \sigma_{23} = 0$
Model 3	$b_{i3}$	$\sigma_3^2 = 0.0458$	$\sigma_1^2 = \sigma_2^2 = \sigma_{12} = \sigma_{13} = \sigma_{23} = 0$

**Table 3.1:** Single Random Component Models

### 3.2.1 Model 1: Only $\beta_1$ Has an Associated Random Effect

In this section, we explore the effect of the number of random effects on the estimation methods as well as the way in which the random effect enters the model. We will consider a model with only one random component parameter associated with the parameter  $\beta_1$  in order to evaluate the effect of the way in which the random effect parameter enters the model on the approximation methods.

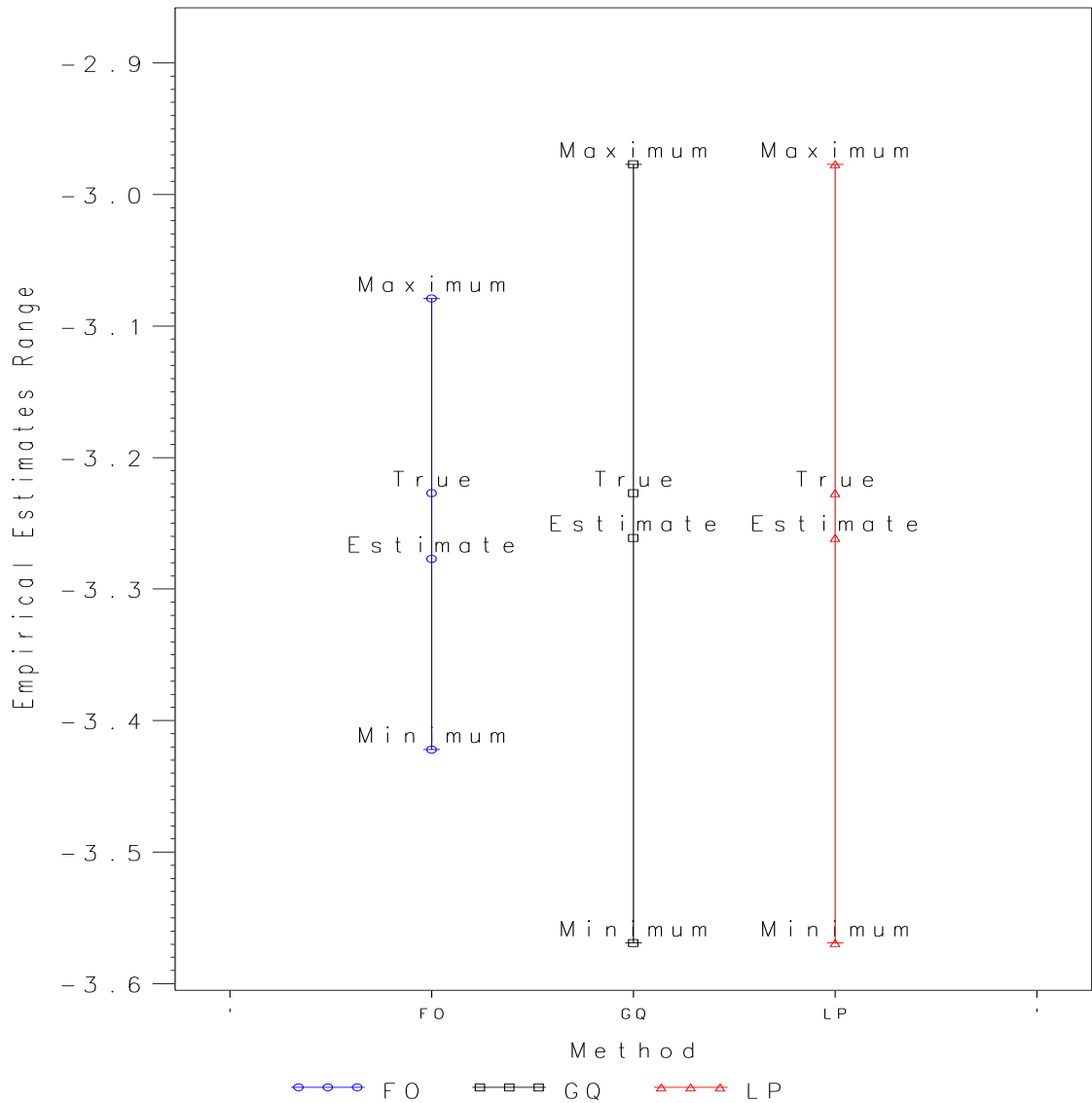


Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1 (= -3.23)$	FO	-3.28	1.56%	0.004	1.32%	98%	97%
	GQ	-3.26	1.04%	0.008	2.46%	100%	95%
	LP	-3.26	1.04%	0.008	2.46%	100%	95%
$\beta_2 (= 0.48)$	FO	0.56	16.39%	0.033	29.36%	98%	83%
	GQ	0.55	14.89%	0.042	34.97%	100%	85%
	LP	0.55	14.89%	0.042	34.97%	100%	85%
$\beta_3 (= -2.46)$	FO	-2.55	3.51%	0.014	3.11%	98%	94%
	GQ	-2.53	2.71%	0.019	4.77%	100%	92%
	LP	-2.53	2.71%	0.019	4.77%	100%	92%

**Table 3.2:** Empirical Estimates:  $\beta_1$  with a Random Component. Estimates from converging models among 1000 cases.

All the estimates of both  $\beta_1$  and  $\beta_3$  are similar across methods FO, GQ, and LP. The relative bias, convergence rate, coefficient of variation, and the coverage probability are within similar range. The FO seems to perform worse than the GQ and LP methods. The FO method has shorter estimates range (Maximum value - Minimum value). The coverage probability is significantly small at 83% for the

FO and 85% for both the GQ and LP methods. The FO method has the smallest MSE. Table 3.2 lists the results for this model simulation.



**Figure 3.1:** Empirical Estimates:  $\beta_1$  with a Random Component  
Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and  
FO = First Order Method.

For  $\beta_1$  the FO, GQ, and LP methods seem to have similar bias ranges. The GQ and LP methods have the most dispersed estimates range. However, these two methods's estimates are centered between the minimum and the maximum estimates of  $\beta_1$ , resulting in good coverage probabilities for both of 95%.

### **3.2.2 Model 2: Only $\beta_2$ Has an Associated Random Effect**

In this model case, we explore the effect of the number of random effects on the estimation methods as well as the way in which the random effect enters the model. We consider a model with only one random effect component associated with  $\beta_2$  in order to further evaluate the effect of the way in which the random effect enters the model on the approximation methods.

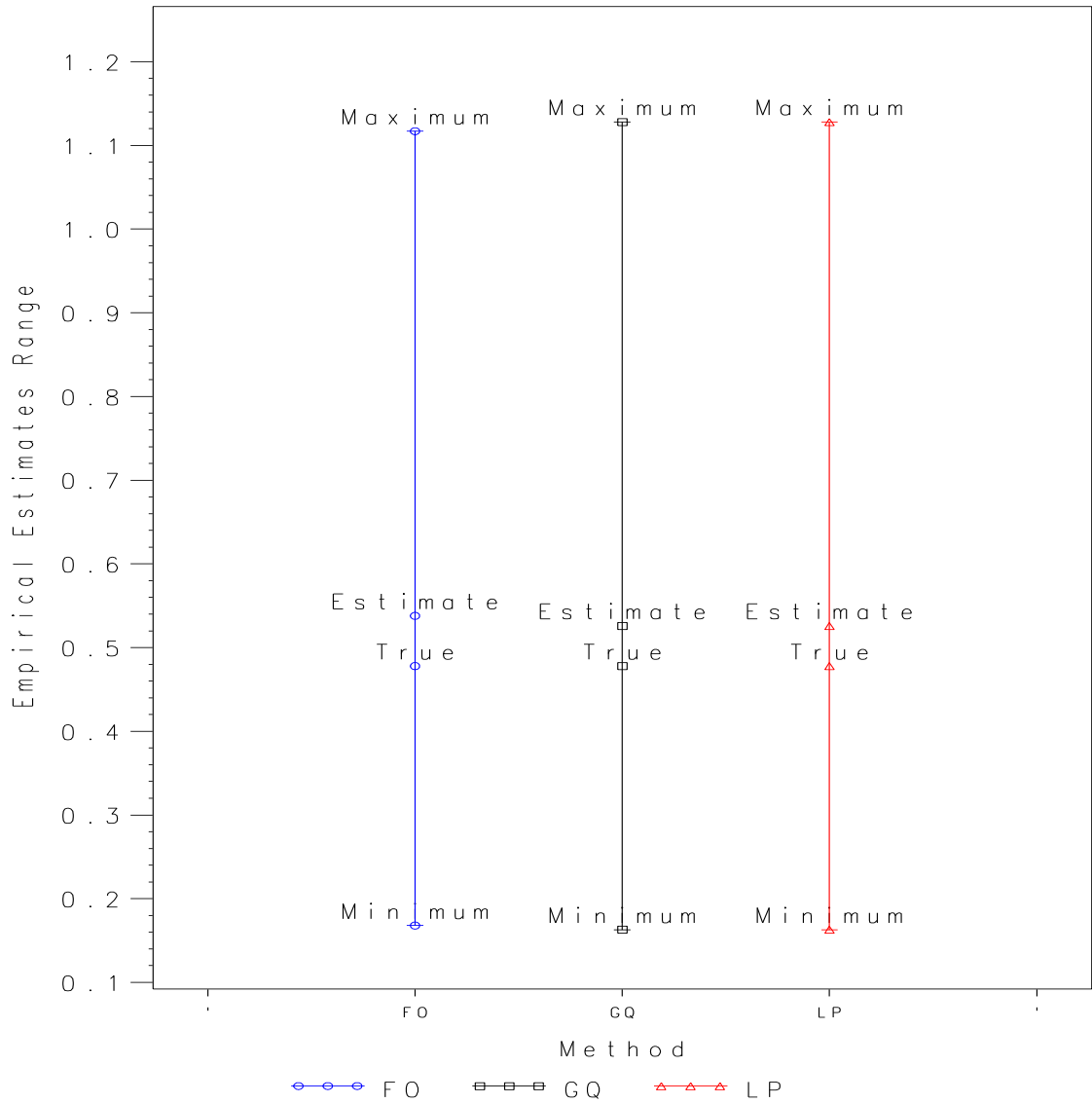
Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.66%	0.005	1.29%	100%	98%
	GQ	-3.27	1.40%	0.004	1.36%	100%	99%
	LP	-3.27	1.40%	0.004	1.36%	100%	99%
$\beta_2(= 0.48)$	FO	0.54	12.48%	0.028	29.24%	100%	92%
	GQ	0.53	9.95%	0.025	28.74%	100%	94%
	LP	0.53	9.95%	0.025	28.74%	100%	94%
$\beta_3(= -2.46)$	FO	-2.55	3.77%	0.014	2.97%	100%	92%
	GQ	-2.54	3.08%	0.012	3.08%	100%	96%
	LP	-2.54	3.08%	0.012	3.08%	100%	96%

**Table 3.3:** Empirical Estimates:  $\beta_2$  with a Random Component. Estimates from converging models among 1000 cases.

For this model where  $\beta_2$ , the parameter that has the smallest magnitude scale (0.48), is the only random component in the model, the estimates of  $\beta_1$  and  $\beta_3$  were unchanged. The relative bias, convergence rate, coefficient of variation, as well as the coverage probability improved compared to the first case model where only  $\beta_1$  has a random component. This improvement is worth noting and might be explained by the fact that, by having a random component, the

parameter's precisions were estimated using data from all individual subjects. This is one of the core advantages of Mixed-Effect Models.

Table 3.3 lists the results for this model simulation.



**Figure 3.2:** Empirical Estimates:  $\beta_2$  with a Random Component  
Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and  
FO = First Order Method.

For  $\beta_2$  the FO, GQ, and LP methods seem to have similar biases, estimates range. However, the methods estimates seem to underestimate  $\beta_2$  resulting in a lack of symmetry and therefore a poor coverage probabilities in the range of 85%.

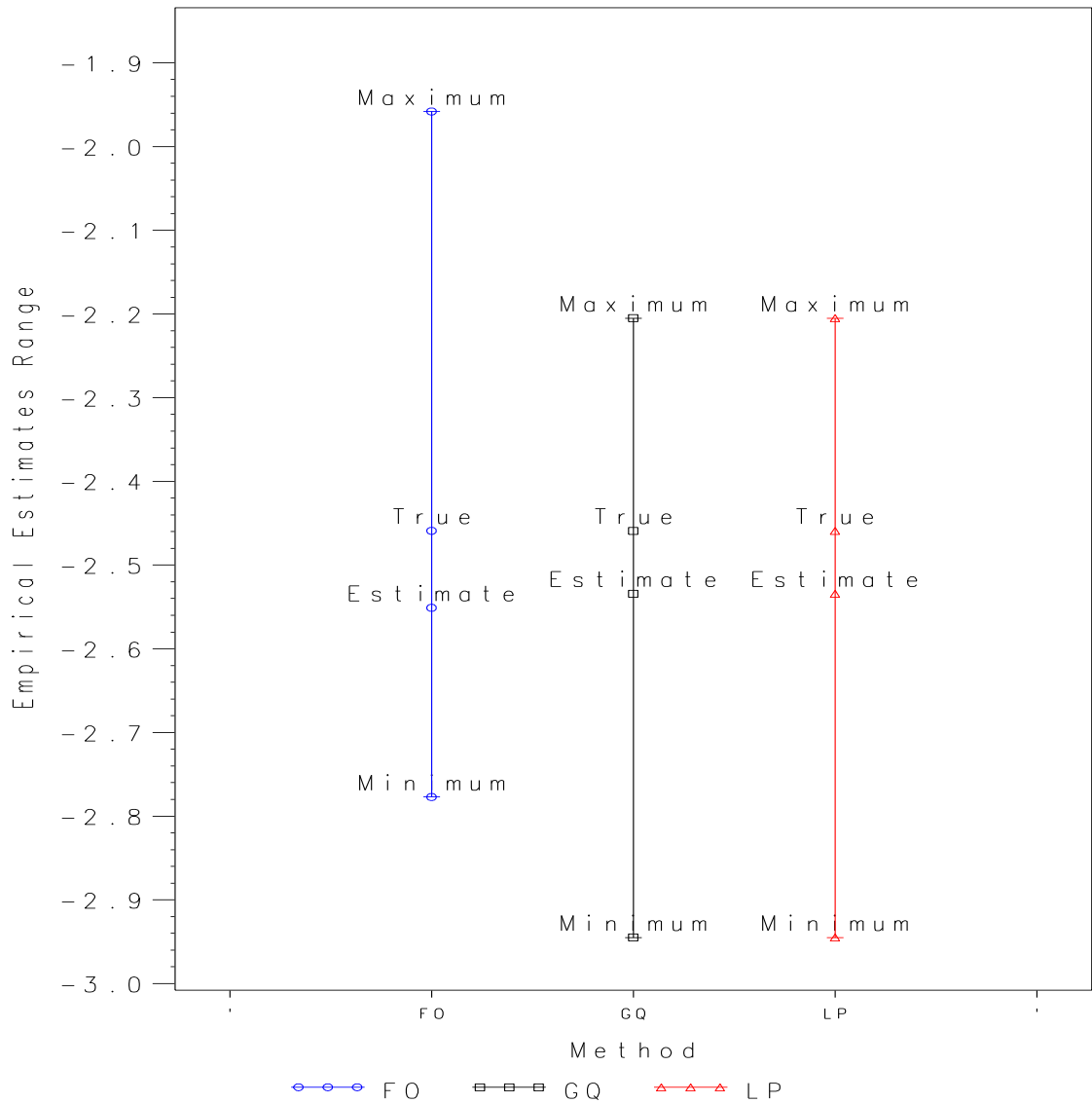
### **3.2.3 Model 3: Only $\beta_3$ Has an Associated Random Effect**

In this section, we explore the effect the way in which the random effect enters the model. We consider a model with only one random effect component associated with  $\beta_3$  in order to further evaluate the effect of the way in which the random effect enters the model on the approximation methods in order to further evaluate the effect of the way in which the random effect enters the model on the approximation methods.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.63%	0.005	1.28%	99%	96%
	GQ	-3.27	1.36%	0.006	1.97%	100%	93%
	LP	-3.27	1.36%	0.006	1.97%	100%	93%
$\beta_2(= 0.48)$	FO	0.51	6.94%	0.018	25.64%	99%	90%
	GQ	0.52	8.51%	0.021	26.67%	100%	89%
	LP	0.52	8.51%	0.021	26.67%	100%	89%
$\beta_3(= -2.46)$	FO	-2.55	3.73%	0.014	2.83%	99%	93%
	GQ	-2.53	3.03%	0.015	3.81%	100%	91%
	LP	-2.53	3.03%	0.015	3.81%	100%	91%

**Table 3.4:** Empirical Estimates:  $\beta_3$  with a Random Component. Estimates from converging models among 1000 cases.

The same findings regarding the first case model where only  $\beta_1$  has a random effect component can be extended to this model where  $\beta_3$  is the only parameter with a random effect component in the model. One should note that the scale magnitude of the true values of parameters  $\beta_1$  and  $\beta_3$  are close to each other. Table 3.4 lists the results for this model simulation.



**Figure 3.3:** Empirical Estimates:  $\beta_3$  with a Random Component  
Estimates from converging models among 1000 cases.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and  
FO = First Order Method.

For  $\beta_3$  the FO, GQ, and LP methods seem to have similar biases  
ranges. However, the methods seem to overestimate  $\beta_3$  resulting in  
a lack of perfect symmetry and therefore a lack of perfect coverage



probabilities in the range of 92%.

### 3.3 Correlated Random Parameters Model

In this simulation we assume that there are two of random components  $b_{i1}$  and  $b_{i2}$  in this model, and they are correlated, and we assume that the correlation coefficient ( $\rho$ ) of these two random components varied and took one of the four values  $\rho = 0.0, 0.25, 0.50$  and  $0.75$  in order to evaluate the effects of the correlation level in the model on the approximation techniques as well as their robustness and performances. The datasets used in this model were assumed to be simulated from the correct model: that is the dataset replicates were generated using the same model. the true parameter values, and a correlation coefficient that took the values  $\rho = 0.0, 0.25, 0.50$  and  $0.75$  respectively.

The general form of the model is given as follows:

$$y_{ij} = D_j \frac{\exp[-(\alpha_1 + \alpha_2 + \alpha_3)]}{\exp(\alpha_2) - \exp(\alpha_3)} \{\exp[-t_{ij} \exp(\alpha_3)] - \exp[-t_{ij} \exp(\alpha_2)]\} + \epsilon_{ij} \quad (3.2)$$

where:

$$\alpha_1 = \beta_1 + b_{i1},$$

$$\alpha_2 = \beta_2 + b_{i2},$$

$$\alpha_3 = \beta_3 (b_{i3} = 0)$$

In this section, we will assume that the variance-covariance matrix of the random effects  $b_i = (b_{1i}, b_{2i})$  is given as follows:

$$\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right), i = 1, \dots, 12$$

where  $\rho$  is the correlation coefficient between  $(b_{1i}, b_{2i})$  and  $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ . The parameter values to be used to generate the replicates are as follow:

$$\beta_1 = -3.2269, \beta_2 = 0.4782$$

$$\beta_3 = -2.4594, \sigma^2 = 0.5581$$

$$\sigma_1^2 = 0.02787, \sigma_2^2 = 0.4239$$

In order to better understand the effects of the random parameters correlation level as well as the way in which the random parame-

ter enters the model on the estimations, we consider four cases of the model 3.2.

Model	correlation coefficient ( $\rho$ )
Model 1	$\rho = 0.00$
Model 2	$\rho = 0.25$
Model 3	$\rho = 0.50$
Model 4	$\rho = 0.75$

**Table 3.5:** Correlated Random Parameters Models

When the random effects in the model are correlated, the most significant effect observed on the overall methods of approximation is that the estimates of the parameter with the smallest magnitude, in this case  $\beta_2 = 0.48$ , have the worst performance. The relative bias and the coefficient of variation were the largest and the the coverage probability was the lowest across methods. The GQ and the PL performed better than the FO method. Tables 3.5, 3.6, 3.7, 3.8, 3.9 list the results for the model simulation for  $\rho = 0.0, 0.25, 0.50$  and  $0.75$  respectively.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1 (= -3.23)$	FO	-3.28	1.58%	0.005	1.48%	96%	76%
	GQ	-3.27	1.30%	0.004	1.36%	62%	97%
	LP	-3.26	1.09%	0.003	1.43%	68%	96%
$\beta_2 (= 0.48)$	FO	0.57	18.67%	0.047	34.59%	96%	69%
	GQ	0.56	17.57%	0.038	31.32%	62%	91%
	LP	0.56	17.11%	0.037	31.30%	68%	84%
$\beta_3 (= -2.46)$	FO	-2.55	3.68%	0.015	3.34%	96%	73%
	GQ	-2.53	2.72%	0.011	3.18%	62%	96%
	LP	-2.52	2.48%	0.011	3.29%	68%	94%

**Table 3.6:** Correlated Random Effects: Model 1 ( $\rho = 0.00$ ).

Estimates from converging models among 1000 cases.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.64%	0.005	1.47%	98%	77%
	GQ	-3.27	1.35%	0.004	1.33%	64%	98%
	LP	-3.26	1.17%	0.004	1.41%	66%	97%
$\beta_2(= 0.48)$	FO	0.54	12.80%	0.030	30.09%	98%	72%
	GQ	0.53	10.69%	0.026	28.70%	64%	93%
	LP	0.53	11.44%	0.026	28.24%	66%	87%
$\beta_3(= -2.46)$	FO	-2.55	3.83%	0.015	3.13%	98%	73%
	GQ	-2.53	2.96%	0.011	2.98%	64%	95%
	LP	-2.53	2.78%	0.011	3.12%	66%	94%

**Table 3.7:** Correlated Random Effects: Model 2 ( $\rho = 0.25$ ).

Estimates from converging models among 1000 cases.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.66%	0.005	1.41%	96%	77%
	GQ	-3.27	1.46%	0.004	1.28%	68%	95%
	LP	-3.26	1.16%	0.004	1.64%	67%	94%
$\beta_2(= 0.48)$	FO	0.52	8.16%	0.020	26.35%	96%	74%
	GQ	0.51	5.96%	0.018	25.99%	68%	92%
	LP	0.51	7.15%	0.019	25.84%	67%	89%
$\beta_3(= -2.46)$	FO	-2.55	3.87%	0.014	2.85%	96%	73%
	GQ	-2.54	3.32%	0.011	2.73%	68%	92%
	LP	-2.53	2.88%	0.013	3.55%	67%	92%

**Table 3.8:** Correlated Random Effects: Model 3 ( $\rho = 0.50$ ).

Estimates from converging models among 1000 cases.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.76%	0.005	1.36%	96%	74%
	GQ	-3.28	1.56%	0.004	1.29%	68%	96%
	LP	-3.27	1.26%	0.005	1.69%	68%	94%
$\beta_2(= 0.48)$	FO	0.49	3.41%	0.014	23.62%	96%	72%
	GQ	0.49	1.53%	0.013	22.99%	68%	93%
	LP	0.48	1.19%	0.013	23.23%	68%	89%
$\beta_3(= -2.46)$	FO	-2.56	4.10%	0.015	2.60%	96%	68%
	GQ	-2.55	3.68%	0.012	2.56%	68%	88%
	LP	-2.54	3.14%	0.013	3.42%	68%	85%

**Table 3.9:** Correlated Random Effects: Model 4 ( $\rho = 0.75$ ).

Estimates from converging models among 1000 cases.

While there was no consistent relationship between the magnitude of the inter-subject variation (measured by  $\rho$  in this case) and the fixed-effect parameters estimates and their precisions, we observed that for two parameters with larger magnitude,  $\beta_1$  and  $\beta_3$ , the relative bias increased with the increase of inter-subject variation. For all three approximation techniques, the relative bias increased from 1.3%

to 1.56% for the GQ, from 1.58% to 1.76% for the FO, and from 1.09% to 1.06% for the LP when  $\rho$  varied from 0.0 to 0.75. The mean square error also followed the same trend, however the coverage probability rates and convergence rates did not seem to follow specific direction with respect to  $\rho$ . The convergence was also much easier to achieve with the FO method with a rate of 96% or higher. The GQ and LP were very sensitive to the complexity of the random effect parameters structure and the achieved convergence rates were lower than 70%. The power (coverage probability) is almost similar for both the GQ and LP method, whereas the FO approximation technique has the least power. Tables 3.10, 3.11, 3.12 list the results for the model simulation for  $\rho = 0.0, 0.25, 0.50$ , and 0.75 for each of  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  respectively.



Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	-3.27	-3.420	-3.11	1.30%	0.37%	1.36%	62%	97%
	0.25	-3.27	-3.433	-3.10	1.35%	0.38%	1.33%	64%	98%
	0.50	-3.27	-3.428	-3.14	1.46%	0.40%	1.28%	68%	95%
	0.75	-3.28	-3.422	-3.13	1.56%	0.43%	1.29%	68%	96%
FO	0.00	-3.28	-3.496	-3.11	1.58%	0.50%	1.48%	96%	76%
	0.25	-3.28	-3.501	-2.97	1.64%	0.51%	1.47%	98%	77%
	0.50	-3.28	-3.430	-3.02	1.66%	0.52%	1.41%	96%	77%
	0.75	-3.28	-3.571	-3.15	1.76%	0.52%	1.36%	96%	74%
LP	0.00	-3.26	-3.444	-3.00	1.09%	0.34%	1.43%	68%	96%
	0.25	-3.26	-3.449	-3.00	1.17%	0.35%	1.41%	66%	97%
	0.50	-3.26	-3.412	-2.99	1.16%	0.43%	1.64%	67%	94%
	0.75	-3.27	-3.388	-2.98	1.26%	0.47%	1.69%	68%	94%

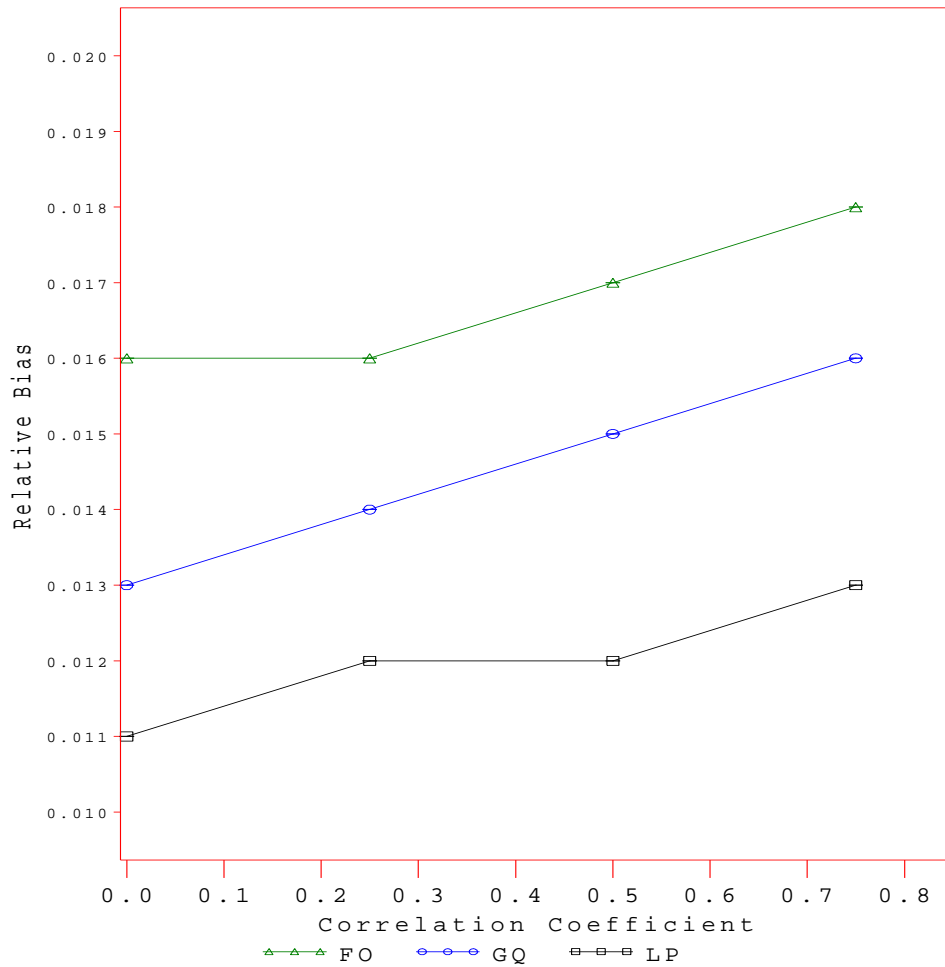
**Table 3.10:** Correlated Model Estimates for  $\beta_1$  (True  $\beta_1=-3.23$ ).  
Estimates from converging models among 1000 cases.

Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	0.56	0.17	1.29	17.57%	3.81%	31.32%	62%	91%
	0.25	0.53	0.12	1.12	10.69%	2.57%	28.70%	64%	93%
	0.50	0.51	0.15	0.99	5.96%	1.82%	25.99%	68%	92%
	0.75	0.49	0.17	0.89	1.53%	1.25%	22.99%	68%	93%
FO	0.00	0.57	-0.04	1.55	18.67%	4.65%	34.59%	96%	69%
	0.25	0.54	-0.10	1.21	12.80%	3.01%	30.09%	98%	72%
	0.50	0.52	0.16	1.04	8.16%	2.01%	26.35%	96%	74%
	0.75	0.49	0.17	0.98	3.41%	1.39%	23.62%	96%	72%
LP	0.00	0.56	0.11	1.41	17.11%	3.74%	31.30%	68%	84%
	0.25	0.53	0.15	1.12	11.44%	2.56%	28.24%	66%	87%
	0.50	0.51	0.15	0.92	7.15%	1.87%	25.84%	67%	89%
	0.75	0.48	0.17	0.86	1.19%	1.27%	23.23%	68%	89%

**Table 3.11:** Correlated Model Estimates for  $\beta_2$  (True  $\beta_2=0.48$ ).  
Estimates from converging models among 1000 cases.

Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	-2.53	-2.80	-2.27	2.72%	1.09%	3.18%	62%	96%
	0.25	-2.53	-2.79	-2.32	2.96%	1.10%	2.98%	64%	95%
	0.50	-2.54	-2.77	-2.33	3.32%	1.15%	2.73%	68%	92%
	0.75	-2.55	-2.75	-2.37	3.68%	1.25%	2.56%	68%	88%
FO	0.00	-2.55	-2.82	-2.31	3.68%	1.55%	3.34%	96%	73%
	0.25	-2.55	-2.84	-2.34	3.83%	1.53%	3.13%	98%	73%
	0.50	-2.55	-2.79	-2.35	3.87%	1.44%	2.85%	96%	73%
	0.75	-2.56	-2.86	-2.33	4.10%	1.46%	2.60%	96%	68%
LP	0.00	-2.52	-2.80	-2.03	2.48%	1.06%	3.29%	68%	94%
	0.25	-2.53	-2.72	-2.03	2.78%	1.09%	3.12%	66%	94%
	0.50	-2.53	-2.75	-2.03	2.88%	1.31%	3.55%	67%	92%
	0.75	-2.54	-2.74	-2.03	3.14%	1.35%	3.42%	68%	85%

**Table 3.12:** Correlated Model Estimates for  $\beta_3$  (True  $\beta_3=-2.46$ ).  
Estimates from converging models among 1000 cases.



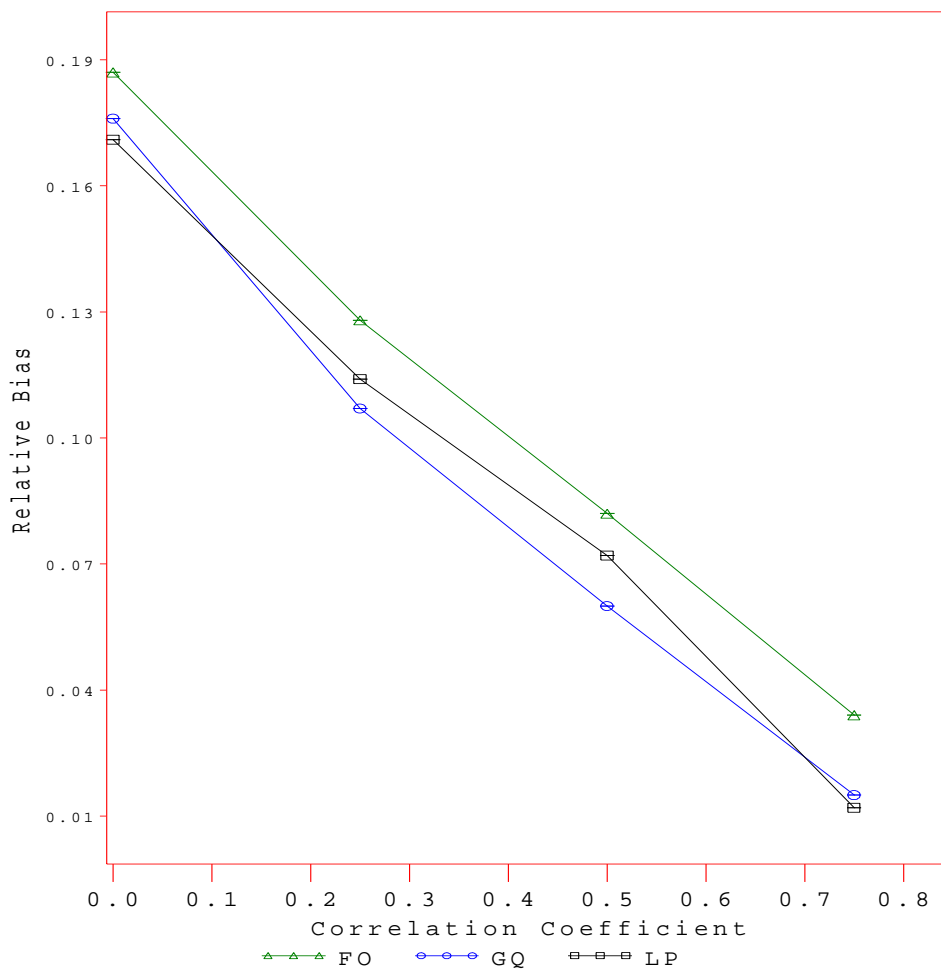
**Figure 3.4:** The Relative Bias Plot for  $\beta_1(\rho)$

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

Overall, the relative bias of  $\beta_1$  estimates resulting from the FO, GQ, and LP approximation methods tends to increase with the increase of the correlation between the random effects.



**Figure 3.5:** The Relative Bias Plot for  $\beta_2(\rho)$

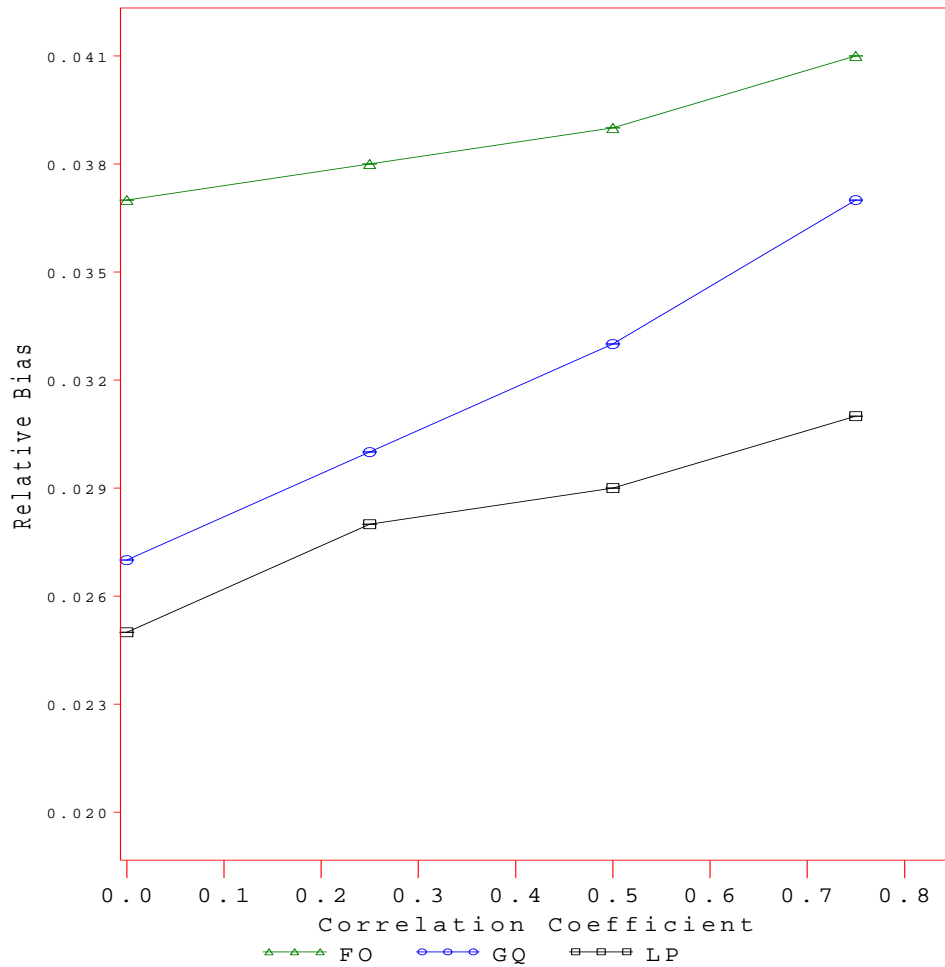
Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

Overall, the relative bias of  $\beta_2$  estimates resulting from the FO, GQ, and LP approximation methods strongly decreases with the increase of the random effects' correlation. It should be noted that the

scale magnitude of  $\beta_2$  is the smallest among the fixed effects.



**Figure 3.6:** The Relative Bias Plot for  $\beta_3(\rho)$

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

The relative bias of  $\beta_3$  estimates resulting from the FO, GQ, and LP approximation methods tends to increase with the increase of the random effects' correlation.

### 3.4 NonCorrelated Random Parameters Model

In this section we assumed a non-correlated random effects model and used the same datasets replicates from the previous section: that is, the datasets were simulated from a NLMM with correlated random effects with the coefficient of variation taking the values:  $\rho = 0.0, 0.25, 0.50$  and  $0.75$  respectively. This assumption was made in order to evaluate the performances of the approximation techniques under the false model, as in the particular case of model 3.2.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.59%	0.005	1.48%	97%	88%
	GQ	-3.28	1.61%	0.008	2.21%	95%	93%
	LP	-3.26	0.88%	0.008	2.52%	94%	92%
$\beta_2(= 0.48)$	FO	0.56	16.77%	0.041	33.18%	97%	81%
	GQ	0.55	15.64%	0.045	35.97%	95%	91%
	LP	0.55	14.20%	0.042	35.49%	94%	90%
$\beta_3(= -2.46)$	FO	-2.55	3.61%	0.015	3.36%	97%	85%
	GQ	-2.53	3.03%	0.018	4.39%	95%	92%
	LP	-2.52	2.27%	0.019	4.93%	94%	92%

**Table 3.13:** NonCorrelated Random Effects: Model 1 ( $\rho = 0.00$ ).  
Estimates from converging models among 1000 cases.



Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.62%	0.005	1.42%	98%	89%
	GQ	-3.28	1.69%	0.007	1.96%	95%	93%
	LP	-3.26	1.00%	0.007	2.28%	94%	92%
$\beta_2(= 0.48)$	FO	0.53	11.65%	0.029	29.97%	98%	84%
	GQ	0.53	10.29%	0.032	32.48%	95%	92%
	LP	0.52	8.49%	0.031	33.02%	94%	89%
$\beta_3(= -2.46)$	FO	-2.55	3.72%	0.015	3.09%	98%	84%
	GQ	-2.54	3.28%	0.016	3.75%	95%	91%
	LP	-2.52	2.47%	0.017	4.57%	94%	92%

**Table 3.14:** NonCorrelated Random Effects: Model 2 ( $\rho = 0.25$ ).  
Estimates from converging models among 1000 cases.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.69%	0.005	1.37%	98%	89%
	GQ	-3.28	1.64%	0.007	1.94%	94%	91%
	LP	-3.26	1.08%	0.008	2.49%	94%	92%
$\beta_2(= 0.48)$	FO	0.51	7.03%	0.020	26.69%	98%	85%
	GQ	0.51	5.79%	0.020	27.34%	94%	93%
	LP	0.49	3.47%	0.021	29.06%	94%	92%
$\beta_3(= -2.46)$	FO	-2.55	3.87%	0.014	2.83%	98%	84%
	GQ	-2.55	3.55%	0.016	3.56%	94%	90%
	LP	-2.52	2.56%	0.017	4.53%	94%	91%

**Table 3.15:** NonCorrelated Random Effects: Model 3 ( $\rho = 0.50$ ).  
Estimates from converging models among 1000 cases.

Parameter	Method	Estimate	Bias	MSE	CV	CR	CP
$\beta_1(= -3.23)$	FO	-3.28	1.71%	0.005	1.31%	98%	84%
	GQ	-3.28	1.57%	0.006	1.76%	93%	92%
	LP	-3.27	1.34%	0.010	2.71%	93%	88%
$\beta_2(= 0.48)$	FO	0.49	2.49%	0.013	23.57%	98%	83%
	GQ	0.49	1.99%	0.011	21.69%	93%	95%
	LP	0.47	-1.02%	0.014	25.35%	93%	92%
$\beta_3(= -2.46)$	FO	-2.56	3.96%	0.014	2.62%	98%	78%
	GQ	-2.55	3.79%	0.014	2.92%	93%	87%
	LP	-2.53	2.92%	0.020	4.84%	93%	83%

**Table 3.16:** NonCorrelated Random Effects: Model 4 ( $\rho = 0.75$ ).  
Estimates from converging models among 1000 cases.

Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	-3.28	-3.643	-3.02	1.61%	0.79%	2.21%	95%	93%
	0.25	-3.28	-3.523	-3.08	1.69%	0.71%	1.96%	95%	93%
	0.50	-3.28	-3.549	-3.10	1.64%	0.68%	1.94%	94%	91%
	0.75	-3.28	-3.522	-3.15	1.57%	0.59%	1.76%	93%	92%
FO	0.00	-3.28	-3.493	-3.05	1.59%	0.50%	1.48%	97%	88%
	0.25	-3.28	-3.524	-3.08	1.62%	0.49%	1.42%	98%	89%
	0.50	-3.28	-3.506	-3.09	1.69%	0.50%	1.37%	98%	89%
	0.75	-3.28	-3.441	-3.14	1.71%	0.49%	1.31%	98%	84%
LP	0.00	-3.26	-3.581	-2.98	0.88%	0.75%	2.52%	94%	92%
	0.25	-3.26	-3.639	-3.00	1.00%	0.65%	2.28%	94%	92%
	0.50	-3.26	-3.888	-2.77	1.08%	0.78%	2.49%	94%	92%
	0.75	-3.27	-3.900	-2.98	1.34%	0.97%	2.71%	93%	88%

**Table 3.17:** NonCorrelated Model Estimates for  $\beta_1$  (True  $\beta_1=-3.23$ ).

Estimates from converging models among 1000 cases.

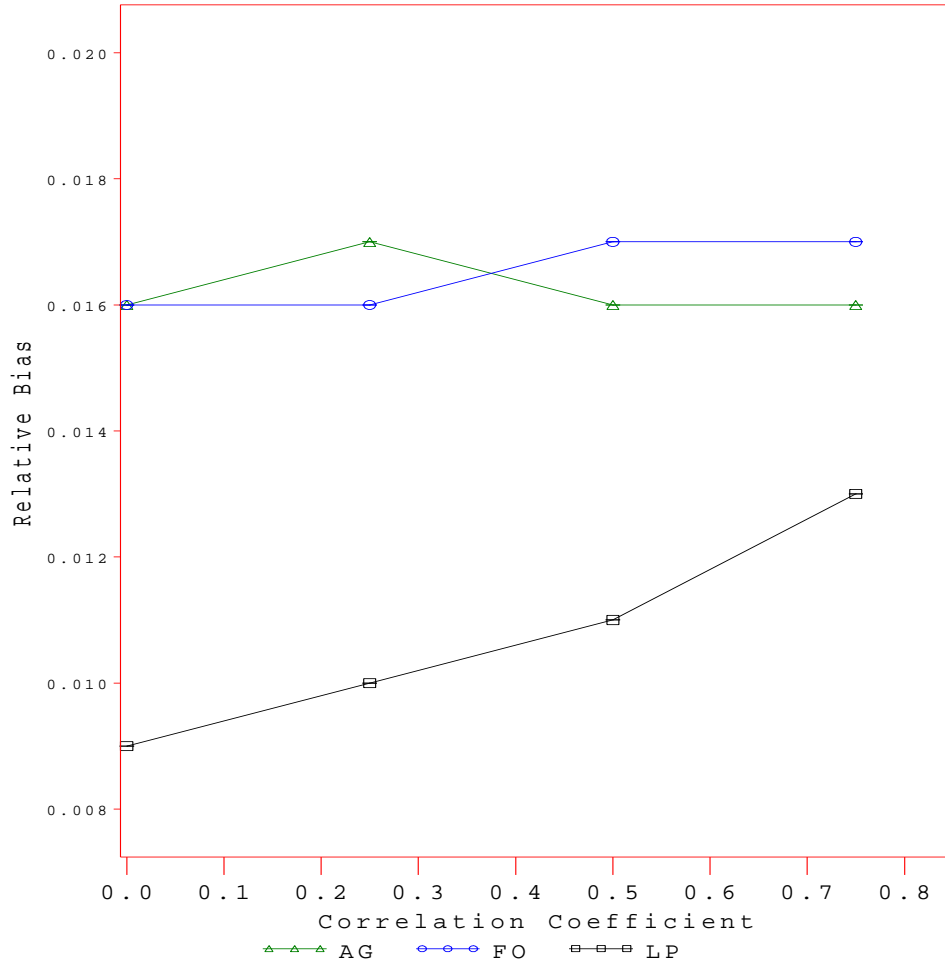
Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	0.55	-0.07	1.34	15.64%	4.52%	35.97%	95%	91%
	0.25	0.53	-0.01	1.40	10.29%	3.18%	32.48%	95%	92%
	0.50	0.51	0.08	1.13	5.79%	1.99%	27.34%	94%	93%
	0.75	0.49	0.17	0.93	1.99%	1.13%	21.69%	93%	95%
FO	0.00	0.56	-0.01	1.36	16.77%	4.08%	33.18%	97%	81%
	0.25	0.53	0.17	1.14	11.65%	2.87%	29.97%	98%	84%
	0.50	0.51	0.10	1.04	7.03%	1.98%	26.69%	98%	85%
	0.75	0.49	0.17	1.19	2.49%	1.35%	23.57%	98%	83%
LP	0.00	0.55	-0.01	1.48	14.20%	4.22%	35.49%	94%	90%
	0.25	0.52	-0.04	1.28	8.49%	3.10%	33.02%	94%	89%
	0.50	0.49	0.01	1.09	3.47%	2.10%	29.06%	94%	92%
	0.75	0.47	-0.02	0.95	-1.02%	1.44%	25.35%	93%	92%

**Table 3.18:** NonCorrelated Model Estimates for  $\beta_2$  (True  $\beta_2=0.48$ ).  
Estimates from converging models among 1000 cases.

Method	$\rho$	Estimate	Min	Max	Bias	MSE	CV	CR	CP
GQ	0.00	-2.53	-3.01	-2.15	3.03%	1.79%	4.39%	95%	92%
	0.25	-2.54	-2.85	-2.24	3.28%	1.56%	3.75%	95%	91%
	0.50	-2.55	-2.87	-2.25	3.55%	1.58%	3.56%	94%	90%
	0.75	-2.55	-2.86	-2.29	3.79%	1.42%	2.92%	93%	87%
FO	0.00	-2.55	-2.91	-2.17	3.61%	1.52%	3.36%	97%	85%
	0.25	-2.55	-2.86	-2.34	3.72%	1.46%	3.09%	98%	84%
	0.50	-2.55	-2.83	-2.32	3.87%	1.43%	2.83%	98%	84%
	0.75	-2.56	-2.81	-2.33	3.96%	1.40%	2.62%	98%	78%
LP	0.00	-2.52	-2.92	-2.03	2.27%	1.85%	4.93%	94%	92%
	0.25	-2.52	-2.96	-2.03	2.47%	1.69%	4.57%	94%	92%
	0.50	-2.52	-2.80	-2.03	2.56%	1.70%	4.53%	94%	91%
	0.75	-2.53	-2.81	-2.03	2.92%	2.02%	4.84%	93%	83%

**Table 3.19:** NonCorrelated Model Estimates for  $\beta_3$  (True  $\beta_3=-2.46$ ).

Estimates from converging models among 1000 cases.



**Figure 3.7:** The Relative Bias Plot for  $\beta_1(\rho)$

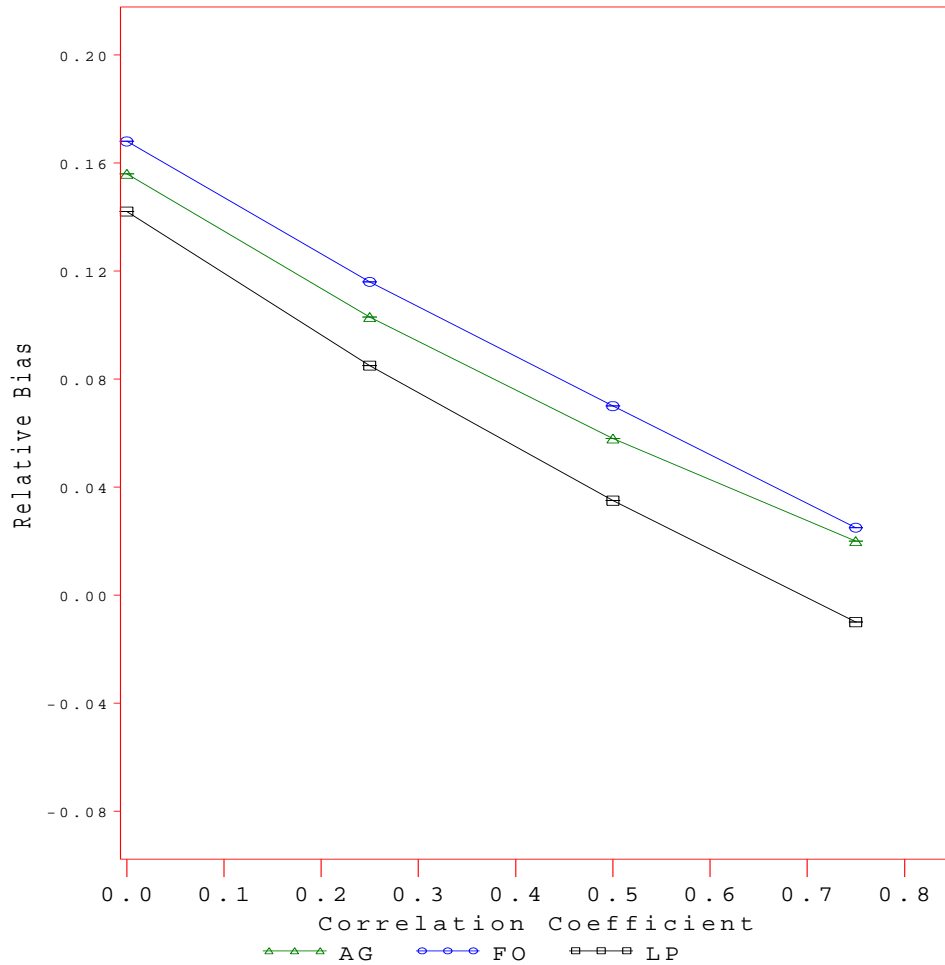
Non-Correlated Random Effect Parameters Model

GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

The relative bias of  $\beta_1$  estimates resulting from the FO and LP approximation methods increases with the increase of the random effects' correlation. However, no similar trend could be observed for the

GQ approximation method.



**Figure 3.8:** The Relative Bias Plot for  $\beta_2(\rho)$

Non-Correlated Random Effect Parameters Model.

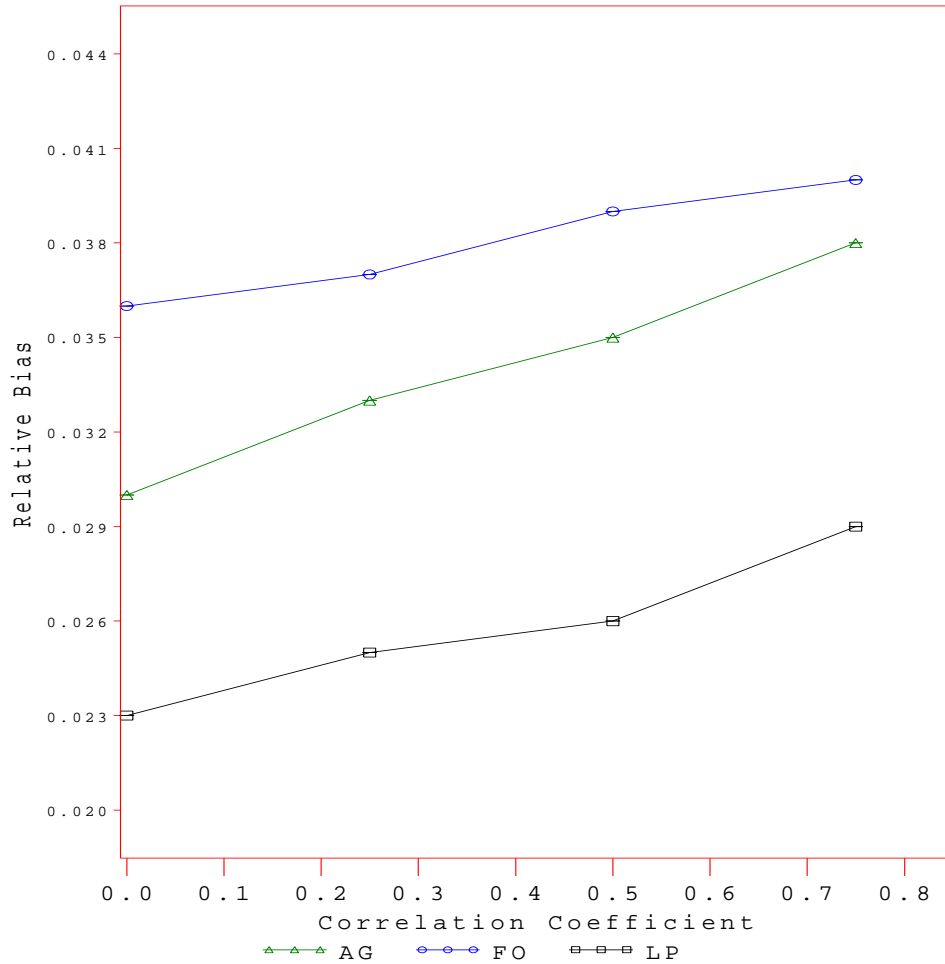
GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

Overall, the relative bias of  $\beta_2$  estimates resulting from the FO, GQ, and LP approximation methods strongly decreases with the increase of the random effects' correlation. It should be noted that the



scale magnitude of  $\beta_2$  is the smallest among the fixed effect parameters.



**Figure 3.9:** The Relative Bias Plot for  $\beta_3(\rho)$

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method,

FO = First Order Method

Overall, the relative bias of  $\beta_3$  estimates resulting from the FO, GQ, and LP approximation methods increases with the increase of the

random effects' correlation.

### 3.5 Conclusion

In this chapter, we have reported the results of simulations studies that we undertook in order to better gain insight into the NLMM and the approximation techniques used to get the estimates of the fixed parameters effects. We focussed on the three most popular approximations: the GQ, LP, and FO methods. It is not appropriate to draw general conclusions from a particular simulation; however, due to the complexity of these models, it has become the standard in the statistical literature that simulation is the only way to investigate the robustness and sensitivity of these techniques, Hartford and Davidian [18] and Pinheiro and Bates [27].

From our simulation, we showed that the LP and GQ methods are very similar and provided the most accurate estimates. Even though the LP is fairly robust to mild deviations, the LP estimates can be extremely biased due to the difficulty of achieving convergence. The LP method is sensitive to misspecification of the inter-individual re-

pones correlation in the model. One also can argue that the GQ, even when it converges it can be inefficient. For example, the maximum estimate is positive when it is supposed to be negative. The lack of coverage probability issue can be explained in part by the number random parameters in the model, but also by the correlation of the random effects, and the need for more observations to compensate for such dependency.

## Chapter 4

### Goodness of Fit

#### 4.1 Motivation

In previous chapters, we tried to assess the performance of three approximation techniques of the likelihood function of model 1.2, however none method could be chosen for all situations, therefore the need for a general goodness of fit is needed. In this section we will evaluate a goodness of fit measure for a particular class of NLMM: The Mixed Logistic Model. First we motivate the use of mixed logistic model through an example of a review from the a drug application. In a clinical trial to evaluate the cumulative irritation properties of the Test product (new product) under maximization conditions in healthy subjects and to compare it to the innovator's (Reference product: ex-

isting product). The mean differences in cumulative irritation scores were compared between the Test and Reference products to determine whether the outcome was similar in both treatment groups. If the upper bound of the 95% confidence interval for the difference between the Test and 1.25 times the Reference's mean was less than or equal to zero, the Test product was deemed to be non-inferior to the Reference product with regard to skin irritation.

Two active patches were applied simultaneously to each subject every 7 days for a total of 3 applications to the same site. The assessment for cumulative skin irritation was combined for 21 days. Let the pair  $(y_{Rij}, y_{Tij})$  be the irritation scores for the Test(T) and the Reference (R) products of a subject  $i, i = 1...120$  at the  $j^{th}$  occasion,  $j = 1...21$  for treatment. The Skin Irritation Scoring Scale is defined as follows:

0 : No evidence of irritation
1 : Minimal irritation
2 : Moderate irritation
3 : Strong irritation

**Table 4.1:** Skin Irritation Scoring Scale

Define  $\mu_T$  to be the mean values for the Test product, and  $\mu_R$  the mean values for the Reference product.

Assuming that  $\mu_R > 0$ , then a comparison of potential skin irritation between the Test and Reference products can be conducted using an ANOVA model and the following hypothesis testing:

$$H_0 : \mu_T - 1.25\mu_R > 0 \text{ vs. } H_1 : \mu_T - 1.25\mu_R < 0$$

The upper bound of a one-sided 95% confidence interval on  $\mu_T - 1.25\mu_R$  was used to establish the non-inferiority of the Test product compared to the Reference product. If the upper bound of the 95% confidence interval for the difference between the Test and 1.25 times

the Reference mean was less than or equal to zero, the Test product was deemed to be non-inferior (not worse than that of an active control [reference] by more than a specified margin 25%) to the reference product in regard to skin irritation.

ANOVA analysis assumes that the data is from a normal distribution. Is this assumption correct? A normal test reveal that both the mean irritation scores and the residual mean score do not come from normal distribution.

Parameter		P-value			
Mean Score	Mean	Shapiro-Wilk	Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Test	1.83	0.000	0.010	0.005	0.005
Reference	1.67	0.000	0.010	0.005	0.005
Residual Mean <sup>1</sup>					
Test	0.00	0.000	0.010	0.005	0.005
Reference	0.00	0.000	0.010	0.005	0.005

**Table 4.2:** Testing for Normality

<sup>1</sup> Difference between the Mean Irritation score and the overall irritation score of the test or reference

All statistical tests provided a significant p-value of less than 5% chance of having normal distribution when both Test and Reference are combined

In addition with a closer look to the data's distribution one can observe that the data can be summarized into few data point values and may not be continuous. A summary of the data can be done as follows:

Mean	Test's frequency	Reference's frequency	Proportion (%)
0.67	1	4	2.4%
1.00	5	9	6.8%
1.33	14	26	19.4%
1.67	26	33	28.6%
2.00	53	27	38.8%
2.33	2	0	1.0%
2.67	1	2	1.5%
4.00	0	1	0.5%
7.00	1	1	1.0%

**Table 4.3:** Range of Dataset: Frequencies of the Mean Irritation Scores

One can argue that the dataset is not normal, or may not even be a continuous response, since 94% of the responses take four values: 1.00, 1.33, 1.67, and 2.00.

An alternative method of analyzing the data is to dichotomize the response into binary response  $X^A_{ij}$ . Where  $A = T, R$  denotes the



particular response in the treatment  $A$  and  $B$  respectively, and where  $X^A_{ij}$  is defined as follows:

$$X^A_{ij} = \begin{cases} 0 & \text{if } y_{Tij} < 1 \\ 1 & \text{if } y_{Tij} \geq 1 \end{cases}$$

Using this new binary response, we analyze the dataset with a mixed logistic model to estimate the odd ratio of the Test and Reference products and its 90% confidence interval.

## 4.2 Mixed Logistic Model

Binary outcomes are very common in medical studies and the mixed logistic model (MLOM) has recently gained a wide range of applications in the regulatory and research communities. While methods to fit mixed logistic regression models are common, methods to evaluate model adequacy and robustness are still an open area of research. Goodness of fit tests for standard logistic regression (SLM) models have been proposed Hosmer et al. [20].

For the MLOM model, Slud and Kedem [42] and Kedem and

Fokianos [23] developed a Goodness of Fit measure for Time-series logistic model with continuous covariates using a classification method. Sturdivant and Hosmer [43] proposed a smoothed residual goodness of fit for a hierarchical logistic regression model. In this section, we evaluate a goodness of fit based on the residual statistics and show that such statistics is asymptotically normal distribution. In addition, we assess the robustness of such statistics with different covariate structure.

In this chapter, we first define a standard logistic regression model with a binary outcome variable, then we introduce a mixed logistic model with a random effect. Then we show how the standardized Sum of Squares Residual (SSR) Statistics of the conditional response on the random effect is asymptotically comparable to a standard normal distribution under the hypothesis that the underline model is correct. We also performed extensive simulations to evaluate SSR Statistics under different model scenarios:

1. under continuous and discrete covariates;
2. under different sample size (number of clusters); and

3. under different sample size (number of observation per cluster).

#### 4.2.1 Standard Logistic Model

Let  $Y_{ij}$  be a binary outcome variable, and let  $\pi_{ij} = P(Y_{ij} = 1)$  and  $x_{ij}$  is a known covariate variable. Such response can be linked to the probability as a simple regression model:  $Y_{ij} = \pi_{ij} + \epsilon_{ij}$ .

A standard logistic regression model is given as follows:

$$\text{logit}(\pi_{ij}) = \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + \beta_1 \times x_{ij} \text{ or,} \quad (4.1)$$

$$\pi_{ij} = \frac{\exp(\beta_0 + \beta_1 \times x_{ij})}{1 + \exp(\beta_0 + \beta_1 \times x_{ij})}$$

$i = 1 \dots N$  is the  $i^{\text{th}}$  subject (cluster) ;  $j = 1 \dots n_i$  is the  $j^{\text{th}}$  observation in the  $i^{\text{th}}$  subject.  $\epsilon_{ij}$  are the measurement error and are assumed to be independent and  $E(\epsilon_{ij}) = 0$ ,  $Var(\epsilon_{ij}) = \sigma_e^2 = \pi_{ij}(1 - \pi_{ij})$ .

#### 4.2.2 Mixed Logistic Model

A mixed logistic regression model with a random effect is given as follows:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \times x_{ij} + b_i \quad (4.2)$$

where  $b_i$  is the random effect parameter, representing the deviation of the observation from the average response within subject  $i$  and is assumed normal distribution  $\sim N(0, \sigma^2_b)$ .

In the following section and for the rest of the chapter we will use matrix notation to rewrite the mixed logistic model as follows:

$$Y = \frac{\exp(X\beta + Zb)}{1 + \exp(X\beta + Zb)} + \epsilon = \pi(\eta) + \epsilon, \text{ and } \eta = X\beta + Zb \quad (4.3)$$

Where:

- $Y$  : is  $N \times 1$  vector of binary responses
- $\beta$  : is  $p \times 1$  vector of unknown fixed effect parameter
- $X$  : is  $N \times P$  design matrix
- $b$  : is  $q \times 1$  vector of random effects and represents the difference between the mixed model and the standard logistic regression model

- $Z$  : is  $N \times q$  design matrix
- $\text{Var}(\epsilon) = \text{diag}[\pi(1 - \pi)]$

The vector and/or matrix notations are used in operations componentwise and not as a whole, for example:

$$Y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} \frac{\exp(X_1\beta_1 + Z_1b_1)}{1 + \exp(X_1\beta_1 + Z_1b_1)} + \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \frac{\exp(X_N\beta_N + Z_Nb_N)}{1 + \exp(X_N\beta_N + Z_Nb_N)} + \epsilon_N \end{bmatrix}$$

it can be shown that, if  $f(x) = \frac{\exp(x)}{1 + \exp(x)}$ , then  $f'(x) = f(x)(1 - f(x))$ , applying this formula to  $\pi(\eta)$ , we have:

$$\frac{d\pi}{d\eta} = \pi(\eta)(1 - \pi(\eta)) = D \tag{4.4}$$

Where  $D$  is an  $N \times N$  diagonal matrix.

Using a first order Taylor approximations as in chapter 1, we expand  $\pi(\eta)$  around  $\hat{\eta} = X\hat{\beta} + Z\hat{b}$  to get the first order approximation of the conditional marginal response. Here  $\hat{b}$  are known parameters.

$$\pi(\eta) \approx \pi(\hat{\eta}) + \pi'(\hat{\eta})(\eta - \hat{\eta}) = \pi(\hat{\eta}) + \pi'(\hat{\eta})(\beta - \hat{\beta})X + \pi'(\hat{\eta})(b - \hat{b})Z$$

Replacing this approximation in equation 4.4, we have:

$$Y = \pi(\hat{\eta}) + \pi'(\hat{\eta})(\beta - \hat{\beta})X + \pi'(\hat{\eta})(b - \hat{b})Z + \epsilon$$

$$\text{Setting, } X_1 = \hat{D}X, Z_1 = \hat{D}Z \text{ and } Y_1 = Y - \pi(\hat{\eta}) + X_1\hat{\beta} + Z_1\hat{b}$$

where

$$\hat{D} = \text{Diag}[\pi(\hat{\eta})(1 - \pi(\hat{\eta}))] = [\pi(\hat{\eta})(1 - \pi(\hat{\eta}))]I_N, \text{ where } I_N \text{ is the}$$

$N \times N$  diagonal matrix, we have a general linear form as follows:

$$Y_1 = X_1\beta + Z_1b + \epsilon \tag{4.5}$$

This approximation is the basis for the FO approximation techniques described in Chapter 1.

### 4.3 Sum of Squares Residual Statistics (SSR)

Let  $\hat{e} = Y - \hat{\pi}$ , where,  $y_i$  is the observed response for the  $i^{th}$  subject (Cluster),  $\hat{\pi}$  is the estimated response for the  $i^{th}$  subject (Cluster). We define the residual sum of squares  $\hat{R}$  as follows:

$$\hat{R} = \hat{e}^T \hat{e} = \sum_{i=1}^N (y_i - \hat{\pi}_i)^2$$

Hosmer et. Al. [20] calculated approximation of the asymptotic moments of  $\hat{R}$  for the standard logistic regression model as follows:

$$E(\hat{R}) \approx \text{Trace}(D), \quad (4.6)$$

$$\text{Var}(\hat{R} - \text{Trace}(D)) \approx P^T (I - V) D P$$

$P$  is a vector with components  $P_i = 1 - 2\hat{\pi}_i$ ,  $D$  is the covariance matrix  $D = \hat{\pi}_i(1 - \hat{\pi}_i)I_N$ , and  $V = DX(X^TDX)^{-1}X^T$

Using the results of Hosmer in 4.7 the following standardized statistics can be compared to a standard normal distribution:

$$\hat{H}_c = \frac{\hat{R} - \text{Trace}(\hat{D})}{\sqrt{\hat{\text{Var}}[\hat{R} - \text{Trace}(\hat{D})]}} \quad (4.7)$$

Now we will derive similar approximation for the mixed logistic

regression model. Using 4.5 and by multiplying both sides of the equation,  $Y_1 = X_1\beta + Z_1b + \epsilon$  by  $\hat{D}^{-1}$ , and replacing  $X_1 = \hat{D}X$ ,  $Z_1 = \hat{D}Z$  and  $Y_1 = Y - \pi(\hat{\eta}) + X_1\hat{\beta} + Z_1\hat{b}$  in the same equation we have:

$$\hat{D}^{-1}Y_1 = \hat{D}^{-1}(X_1\beta + Z_1b + \epsilon),$$

$$\hat{D}^{-1}(Y - \hat{\pi} + X_1\hat{\beta} + Z_1\hat{b}) = \hat{D}^{-1}(X_1\beta + Z_1b + \epsilon),$$

$$\hat{D}^{-1}(Y - \hat{\pi}) + \hat{D}^{-1}(X_1\hat{\beta} + Z_1\hat{b}) = X\beta + Zb + \epsilon_1,$$

$$\hat{D}^{-1}(Y - \hat{\pi}) + X\hat{\beta} + Z\hat{b} = X\beta + Zb + \epsilon_1,$$

Let  $Y_2 = \hat{D}^{-1}(Y - \hat{\pi}) + X\hat{\beta} + Z\hat{b}$ , then we have an approximate linear version of the mixed logistic model that can be written as follows:

$$Y_2 = X\beta + Zb + \epsilon \tag{4.8}$$

Harville [17] showed that for the model in 4.8, if  $\hat{\beta}$  and  $\hat{b}$  are components of any solution then they must satisfy the following (penalized quasi-likelihood equation) linear system:



$$\begin{bmatrix} X^T \hat{D}X & X^T \hat{D}Z \\ Z^T \hat{D}X & \hat{\Sigma}^{-1} + Z^T \hat{D}Z \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X^T \hat{D}Y_2 \\ Z^T \hat{D}Y_2 \end{bmatrix} \quad (4.9)$$

In addition,  $\hat{\beta}$  is also a solution to following:

$$(X^T V^{-1} X) \hat{\beta} = X^T V^{-1} Y$$

Let the parameters vector  $\hat{\gamma} = \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix}$ , and

$H = \begin{bmatrix} X & Z \end{bmatrix}$  then 4.10 can be written as follows:

$$\begin{bmatrix} H \hat{D} H + \begin{bmatrix} 0 & 0 \\ 0 & \hat{\Sigma}^{-1} \end{bmatrix} \end{bmatrix} \hat{\gamma} = H \hat{D} Y_2$$

We set  $O = \begin{bmatrix} 0 & 0 \\ 0 & \hat{\Sigma}^{-1} \end{bmatrix}$ , then we have:

$[H \hat{D} H + O] \hat{\gamma} = H \hat{D} Y_2$ , define  $q(\hat{\gamma})$  as follows;

$q(\hat{\gamma}) = [H \hat{D} H + O] \hat{\gamma} - H \hat{D} Y_2 = 0$  substituting  $Y_2 = \hat{D}^{-1}(Y - \hat{\pi}) + X \hat{\beta} + Z \hat{b}$  in  $q$  we have:

$q(\hat{\gamma}) = [H \hat{D} H + O] \hat{\gamma} - H \hat{D} (\hat{D}^{-1}(Y - \hat{\pi}) + X \hat{\beta} + Z \hat{b})$  and adding

the terms in  $\hat{\gamma}$ , we have:

$$q(\hat{\gamma}) = O\hat{\gamma} - H(Y - \hat{\pi})$$

As previously stated with regard to the logistic function derivative, it can be shown that if

$$f(x) = \frac{\exp(kx)}{1+\exp(kx)}, \text{ then } f'(x) = kf(x)(1 - f(x)), \text{ therefore if}$$

$$\hat{\pi} = \frac{\exp(H\hat{\gamma})}{1 + \exp(H\hat{\gamma})}, \text{ then } \frac{d\hat{\pi}}{d\hat{\gamma}} = \hat{\pi}(1 - \hat{\pi})H = \hat{D}H$$

$$\frac{dq}{d\hat{\gamma}} = H^T \hat{D}H + O$$

Using a first order Taylor expansion of  $q$  around  $\hat{\gamma} = \gamma$ , gives:

$$q(\hat{\gamma}) \approx q(\gamma) + \frac{dq}{d\hat{\gamma}}(\hat{\gamma} - \gamma) = O\gamma - H^T(Y - \pi) + (H^T \hat{D}H + O)(\hat{\gamma} - \gamma) = 0$$

Solving for  $(\hat{\gamma} - \gamma)$  we have:

$$(\hat{\gamma} - \gamma) = [H^T \hat{D}H + O]^{-1}[H^T(Y - \pi) - O\gamma] \quad (4.10)$$

Using another first order Taylor expansion this time of  $\hat{\pi}$  around  $\hat{\gamma} = \gamma$ , gives:

$$\hat{\pi} \approx \pi + \frac{d\hat{\pi}}{d\hat{\gamma}}(\hat{\gamma} - \gamma) = \pi + DH(\hat{\gamma} - \gamma) \quad (4.11)$$

replacing  $(\hat{\gamma} - \gamma)$  by its value from equation 4.10, gives:

$$\begin{aligned} \hat{\pi} &= \pi + DH[H^T DH + O]^{-1}[H^T(Y - \pi) - O\gamma] \\ &= \pi + DH[H^T DH + O]^{-1}H^T(Y - \pi) - DH[H^T DH + O]^{-1}O\gamma \\ &= \pi + Ve - w \end{aligned} \quad (4.12)$$

where

$$\begin{aligned} V &= DH[H^T DH + O]^{-1}H^T \\ w &= DH[H^T DH + O]^{-1}O\gamma \\ e &= Y - \pi \end{aligned}$$

using this result and the previous ones, we can express the estimated residuals  $\hat{e}$  as a function of the true residuals  $e$  as follows:

$$\begin{aligned}
\hat{e} &= Y - \hat{\pi} \\
&\approx Y - [\pi + V(Y - \pi) - w] \\
&= (Y - \pi) - V(Y - \pi) + w \\
&= (I - V)e + w
\end{aligned} \tag{4.13}$$

Using similar approximation as Hosmer et. Al. [20], we compute the asymptotic moments of the following centered distribution:

$$S\hat{S}R = \frac{\hat{R} - \text{Trace}(\hat{D})}{\sqrt{\hat{\text{Var}}[\hat{R} - \text{Trace}(\hat{D})]}}$$

$$\begin{aligned}
S\hat{S}R - \text{trace}(\hat{D}) &= \hat{e}^T \hat{e} - \hat{\pi}^T (\mathbf{1} - \hat{\pi}) \\
&= (y - \hat{\pi})^T (y - \hat{\pi}) - \hat{\pi}^T \mathbf{1} + \hat{\pi}^T \hat{\pi} \\
&= y^T y - \hat{\pi}^T y - y^T \hat{\pi} + \hat{\pi}^T \hat{\pi} - \hat{\pi}^T \mathbf{1} + \hat{\pi}^T \hat{\pi} \\
&= \mathbf{1}^T y - 2\hat{\pi}^T y - \mathbf{1}^T \hat{\pi} + 2\hat{\pi}^T \hat{\pi} \\
&= (\mathbf{1} - 2\hat{\pi})^T y - (\mathbf{1} - 2\hat{\pi})^T \hat{\pi} \\
&= (\mathbf{1} - 2\hat{\pi})^T (y - \hat{\pi}) \\
&= (\mathbf{1} - 2\hat{\pi})^T \hat{e}
\end{aligned}$$

Therefore, using this and the results from 4.13 and noting that  $E(e) = 0$ , and the results from Seber and Lee [37] for the moments of quadratic form we have:

$$\begin{aligned}
E[S\hat{S}R - \text{trace}(\hat{D})] &= E[\hat{e}^T \hat{e} - \text{trace}(\hat{D})] \\
&= E[(\mathbf{1} - 2\hat{\pi})^T] \\
&\approx E[(\mathbf{1} - 2(\pi - w + Ve)^T)[(I - V)e + w]] \\
&= E[\mathbf{1}^T e - \mathbf{1}^T V e + \mathbf{1}^T w - 2\pi^T e + 2\pi^T V e - \\
&\quad 2\pi^T w + 2w^T e - 2w^T V^T e + \\
&\quad 2w^T w - 2e^T V^T e + 2e^T V^T V e - 2e^T V^T w] \\
&= \mathbf{1}^T w - 2\pi^T w + 2w^T e + 2w^T w - E(2e^T V^T e) + \\
&\quad E(2e^T V^T V e) \\
&= \mathbf{1}^T w - 2\pi^T w + 2w^T w - 2E[e^T V^T (I - V)e]
\end{aligned}$$

$$\begin{aligned}
E[\hat{T} - \text{trace}(\hat{D})] &\approx (\mathbf{1}^T - 2\pi^T + 2w^T)w - \\
&2\text{trace}[V(I - V)D] \tag{4.14}
\end{aligned}$$

Using a similar argument, we can approximate the variance as follows:

$$\begin{aligned}
\text{Var}[\hat{T} - \text{trace}(\hat{D})] &= \text{Var}[(\mathbf{1} - 2\hat{\pi})^T \hat{e}] \\
&\approx \text{Var}(\mathbf{1} - 2\hat{\pi})^T [(I - V)e + w] \\
&= (\mathbf{1} - 2\hat{\pi})^T \text{Var}[(I - V)e](\mathbf{1} - 2\hat{\pi}) \\
&= (\mathbf{1} - 2\hat{\pi})^T (I - V) \text{Var}(e) (I - V)^T (\mathbf{1} - 2\hat{\pi}) \\
&= (\mathbf{1} - 2\hat{\pi})^T (I - V) D (I - V)^T (\mathbf{1} - 2\hat{\pi}) \quad (4.15)
\end{aligned}$$

## 4.4 Simulations Studies

To evaluate the adequacy and robustness of the SSR we simulate data from the following 3 different models:

- Model 1 :  $\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 U_i[-1, 1] + \beta_2 U_{ij}[-2, 2] + b_i$
- Model 2 :  $\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 B_i(0.5) + \beta_2 U_{ij}[-1, 1] + b_i$
- Model 3 :  $\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 B_i(0.5) + \beta_2 B_{ij}(0.25) + b_i$

Where  $U[a, b]$  is the uniform distribution between  $[a, b]$ ,  $B(p)$  is the Bernoulli distribution with success proportion  $p$  and  $b_i$  is a random effect with mean zero and variance  $\sigma_u^2$ . We assume that the simulated datasets each has 20 clusters (20 groups) and each group has a number

of observations  $M = 10$ , and  $M = 20$  respectively.

The true parameter values to be used to generate the datasets replicates are as follows:

$$\beta_0 = 2.50,$$

$$\beta_1 = 0.250, \text{ and}$$

$$\beta_2 = 0.075.$$

The average power of the SSR test statistics was computed based on the significance level  $\alpha$  of 5% and 10% respectively and the 95% confidence intervals associated with such power were constructed. For each case 1000 replicates were generated and the correct model was used to get the parameters estimates, means, variances and the SSR statistics. The hypothesis of detecting the correct model was rejected if the p-value of the SSR statistics was outside the interval  $[0.05, 0.95]$  or  $[0.10, 0.90]$  corresponding to two significant levels  $\alpha = 0.05$ , and  $\alpha = 0.1\%$  respectively.



#### 4.4.1 Results

**Model 1:** For model 1 with two continuous covariates (Uniform) and with a number of groups  $N = 20$  subjects, the power of the test statistics was 95.0% for a group size of  $M = 10$ . The type I error in this case was 5%, and the 95% confidence interval was [93.6%, 96.4%]. Similar results for the power were also similar for  $\alpha$  level of 0.10%. However, the power was low when the group size increased to  $M = 20$  and averaged between 78.2% and 58.2% for  $\alpha$  level of 0.05% and 0.10% respectively. The results of the simulations are presented in the table 4.4. The histogram and the percentiles of the standardized SSR distribution with that of a standard normal are presented in the figures 4.1, 4.2, 4.3, and 4.4.

N	M	$\alpha$	Mean	Lower 95% CI	Upper 95% CI
20	10	0.05	99.6%	99.2%	100.0%
20	20	0.05	97.0%	95.9%	98.1%
20	10	0.10	98.2%	97.4%	99.0%
20	20	0.10	92.5%	90.9%	94.1%

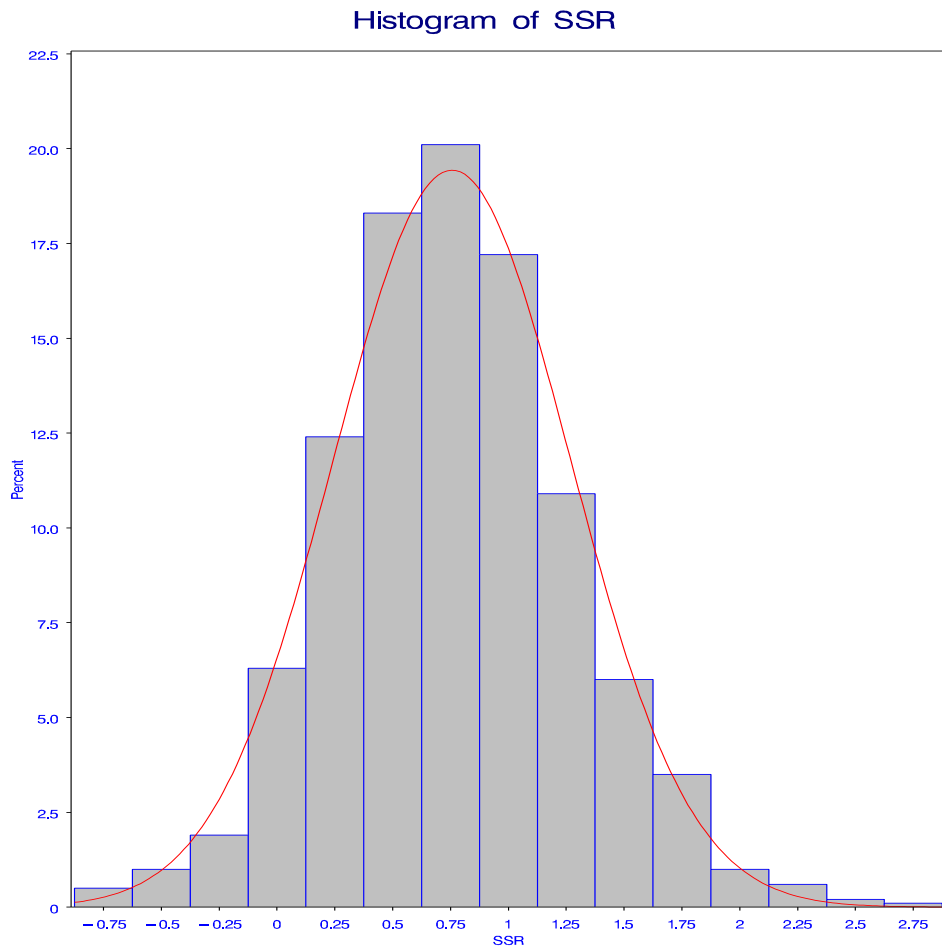
**Table 4.4:** Power Estimate with 95% CI

Model 1:  $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 U_i[-1, 1] + \beta_2 U_{ij}[-2, 2] + b_i$

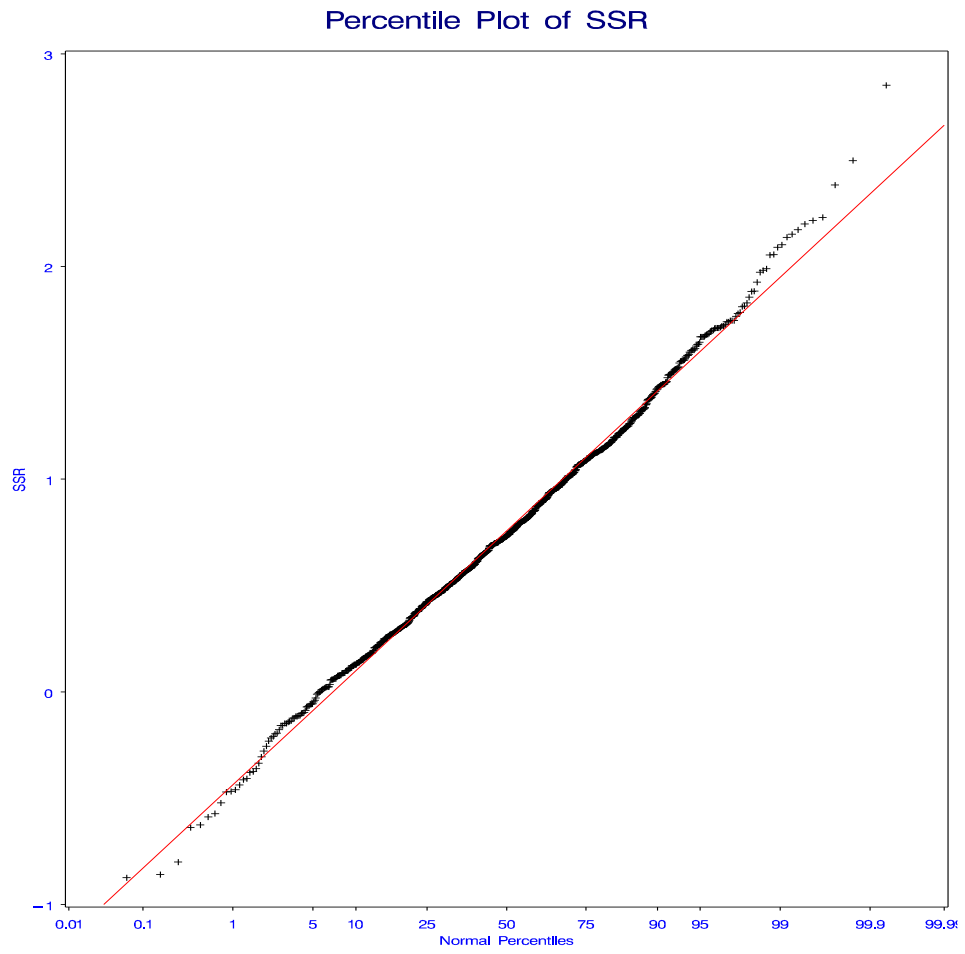
N = The Number of Clusters

M = The Cluster Size

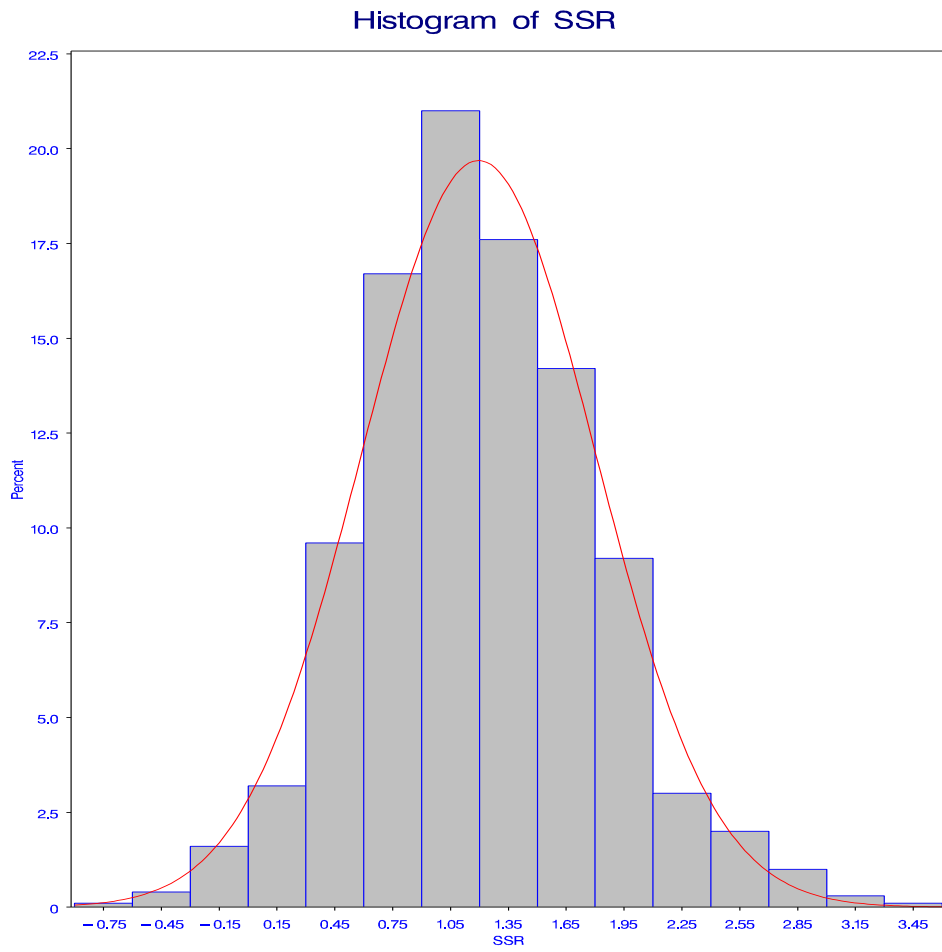
CI: The 95% Confidence Interval



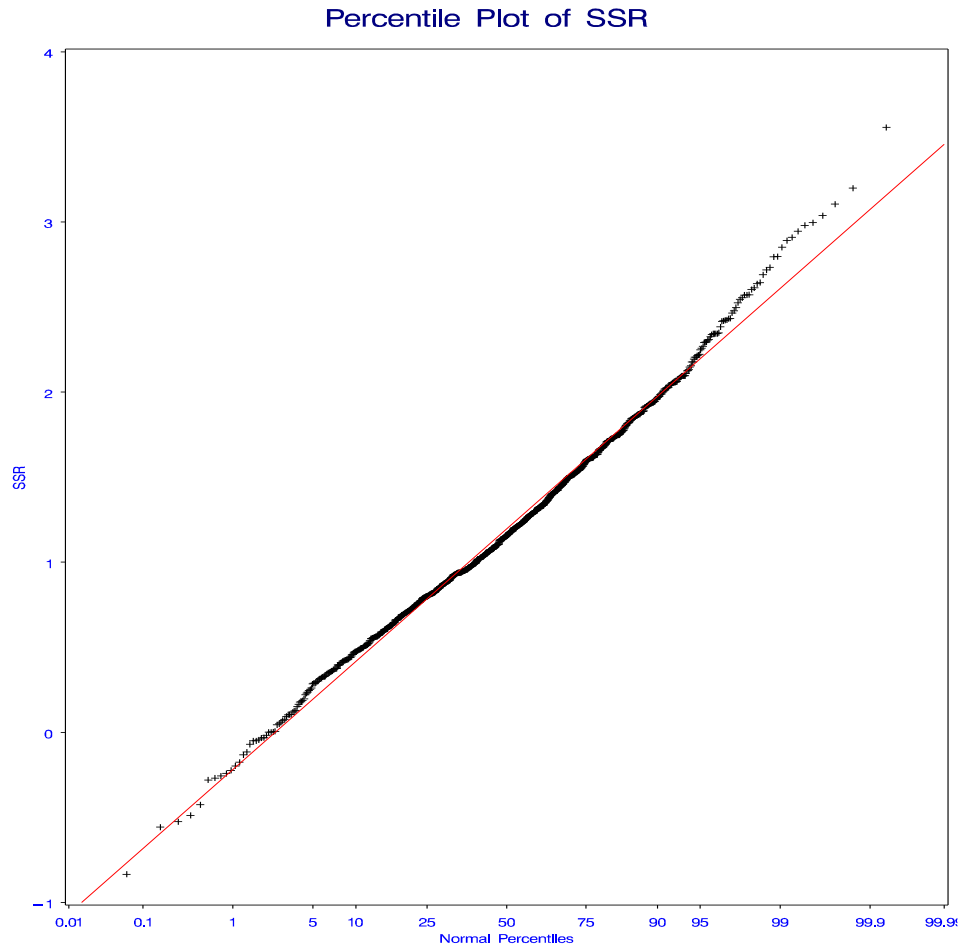
**Figure 4.1:** The SSR Histogram for Model 1 with  $N = 20$  and  $M = 10$



**Figure 4.2:** Percentiles Plot for Model 1 with  $N = 20$  and  $M = 10$



**Figure 4.3:** The SSR Histogram for Model 1 with  $N = 20$  and  $M = 20$



**Figure 4.4:** Percentiles Plot for Model 1 with  $N = 20$  and  $M = 20$

**Model 2:** For this model with a cluster specific binary covariate (Bernoulli) and with a continuous covariate and total number of clusters  $N = 20$  clusters (groups), the average power of the test statistics at the  $\alpha$  level of 0.05% , was 99.6% for a cluster size of  $M = 10$ , and 97% for a cluster of size  $M = 20$ . The acceptance rates seem to be inflated and the 95% confidence intervals were [99.2%, 100%], and

[95.9%, 98.10%] for the cluster size of  $M = 10$ , and  $M = 20$  respectively, with the significance level  $\alpha = 0.05\%$  falling outside the confidence interval, meaning that type I error was much less frequently observed than it should be. This model tends to fail to reject the hypothesis when it should have.

Similar results for the power were similar for the significance level  $\alpha = 0.1$  and again here we observed that, the power decreased with the increase of the cluster size. Here again, the type I error level of 0.05 was outside the 95% confidence intervals for all three sample size scenarios considered. The results of the simulations are presented in the table 4.5. The histogram and the percentiles plots of the standardized SSR distribution and those of standard normal distributions are presented in the figures 4.1, 4.2, 4.3, and 4.4.

N	M	$\alpha$	Mean	Lower 95% CI	Upper 95% CI
20	10	0.05	99.6%	99.2%	100.0%
20	20	0.05	97.0%	95.9%	98.1%
20	10	0.10	98.2%	97.4%	99.0%
20	20	0.10	92.5%	90.9%	94.1%

**Table 4.5:** Power Estimate with 95% CI

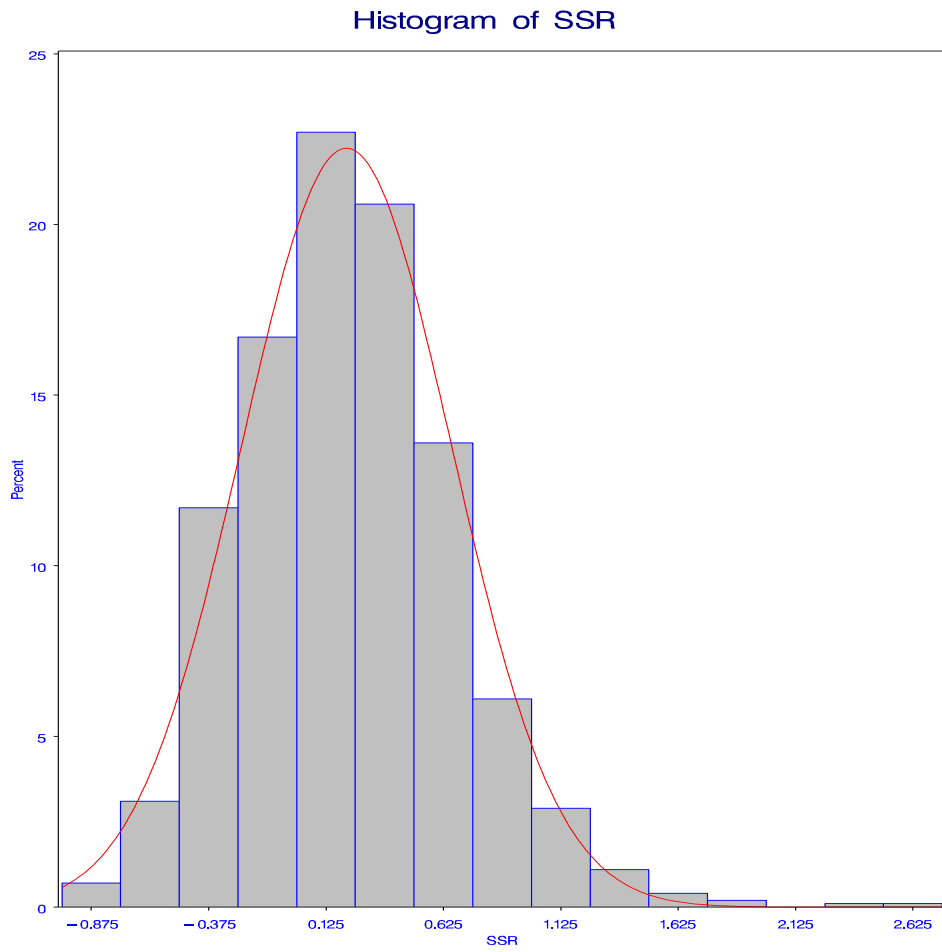
Model 2:  $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 B_i(0.5) + \beta_2 U_{ij}[-1, 1] + b_i$

N = The Number of Clusters

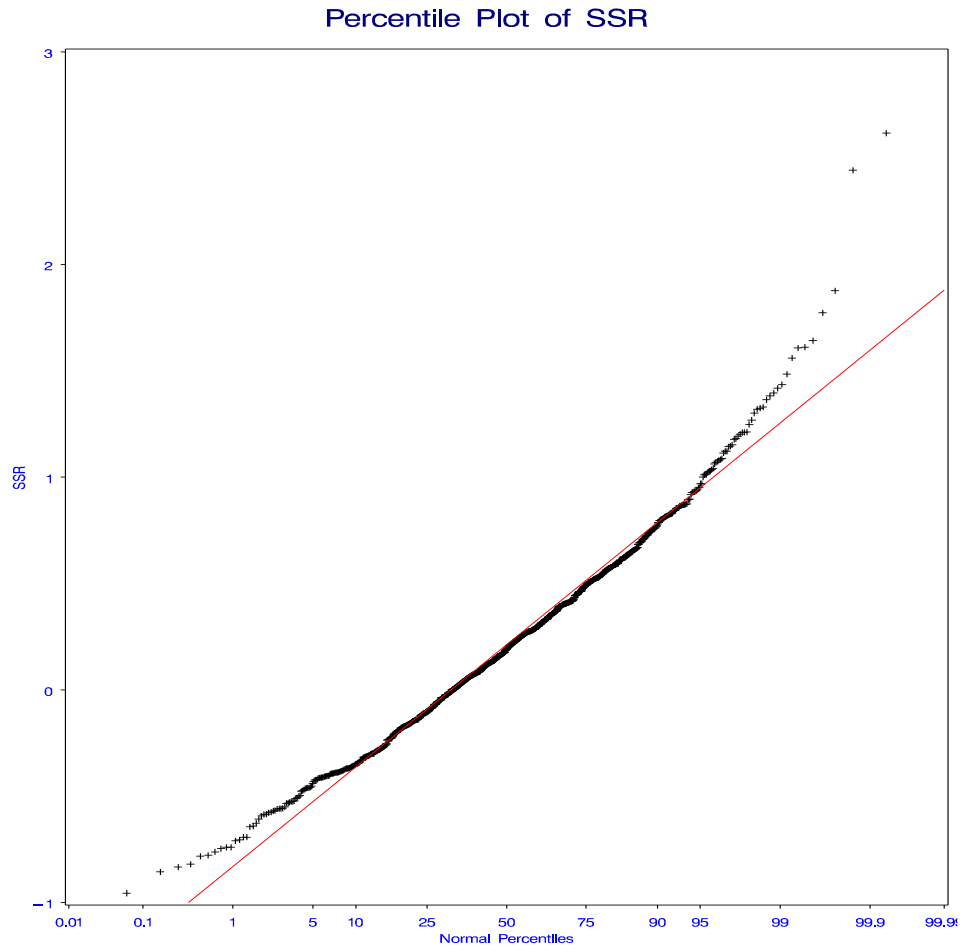
M = The Cluster Size

CI: The 95% Confidence Interval

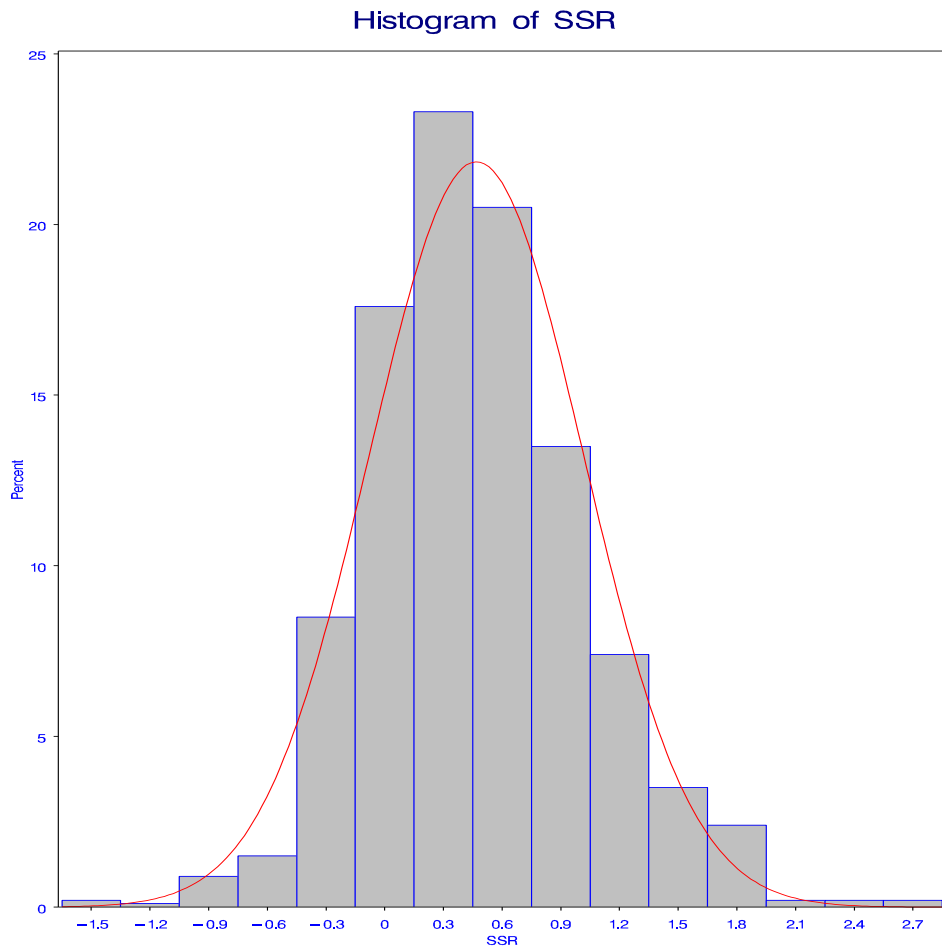




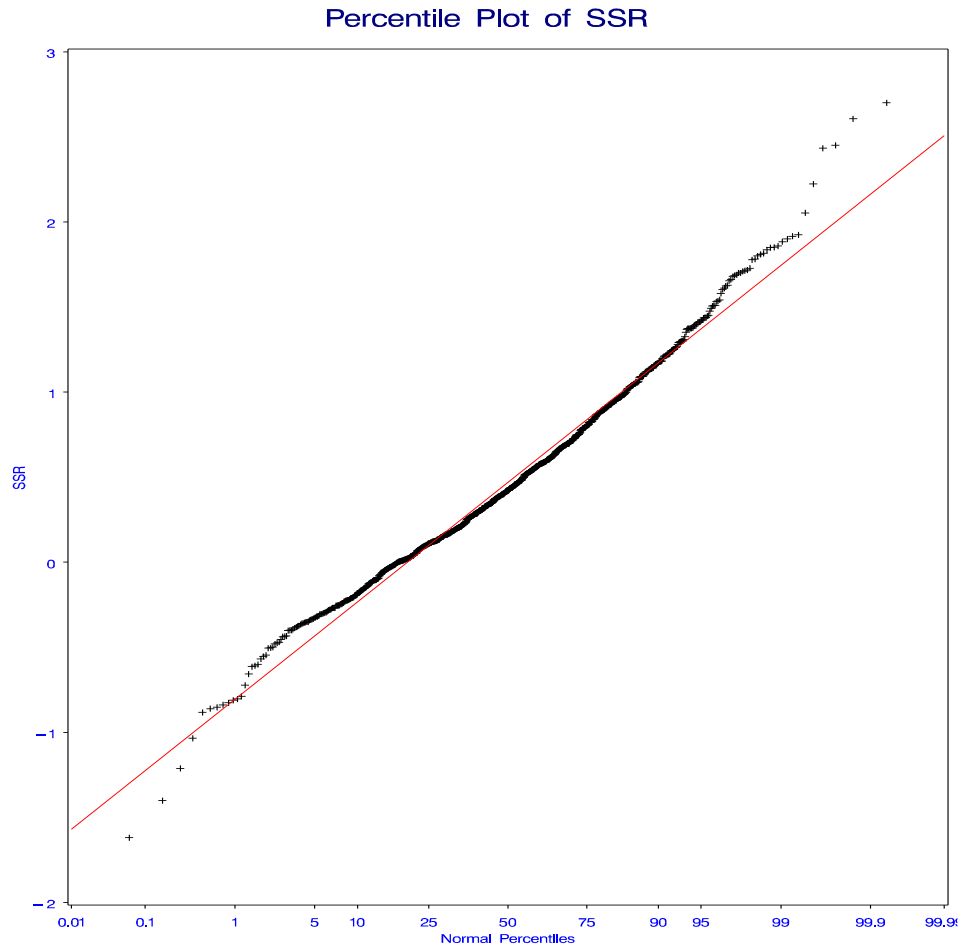
**Figure 4.5:** The SSR Histogram for Model 2 with  $N = 20$  and  $M = 10$



**Figure 4.6:** Percentiles Plot for Model 2 with  $N = 20$  and  $M = 10$



**Figure 4.7:** The SSR Histogram for Model 2 with  $N = 20$  and  $M = 20$



**Figure 4.8:** Percentiles Plot for Model 2 with  $N = 20$  and  $M = 20$

**Model 3:** For this model with two binary covariates (generated from Bernoulli's distribution) and with the total number of clusters  $N = 20$  (groups), the power of the test statistics averages about 84.3% for a cluster of size  $M = 10$ , and 56.9% for a cluster of size  $M = 20$ . In addition the 95% confidence intervals were [82.0%, 86.6%] and [53.8%, 60.0%] for the cluster of size  $M = 10$  and  $M = 20$  respectively.

Similar results for the average power of the test statistics were comparable for a significance level  $\alpha = 0.10\%$ . Here we observed that the power decreased with the increase of the cluster size after reaching an “optimal value” (cluster size  $M = 10$ ). The results of the simulations are presented in table 4.6. The histogram of the standardized SSR statistics as well as the probability percentiles plots are presented in the figures 4.5, 4.6, 4.7, and 4.8.

N	M	$\alpha$	Mean	Lower 95% CI	Upper 95% CI
20	10	0.05	84.3%	82.0%	86.6%
20	20	0.05	56.9%	53.8%	60.0%
20	10	0.10	69.1%	66.2%	72.0%
20	20	0.10	35.1%	32.1%	38.1%

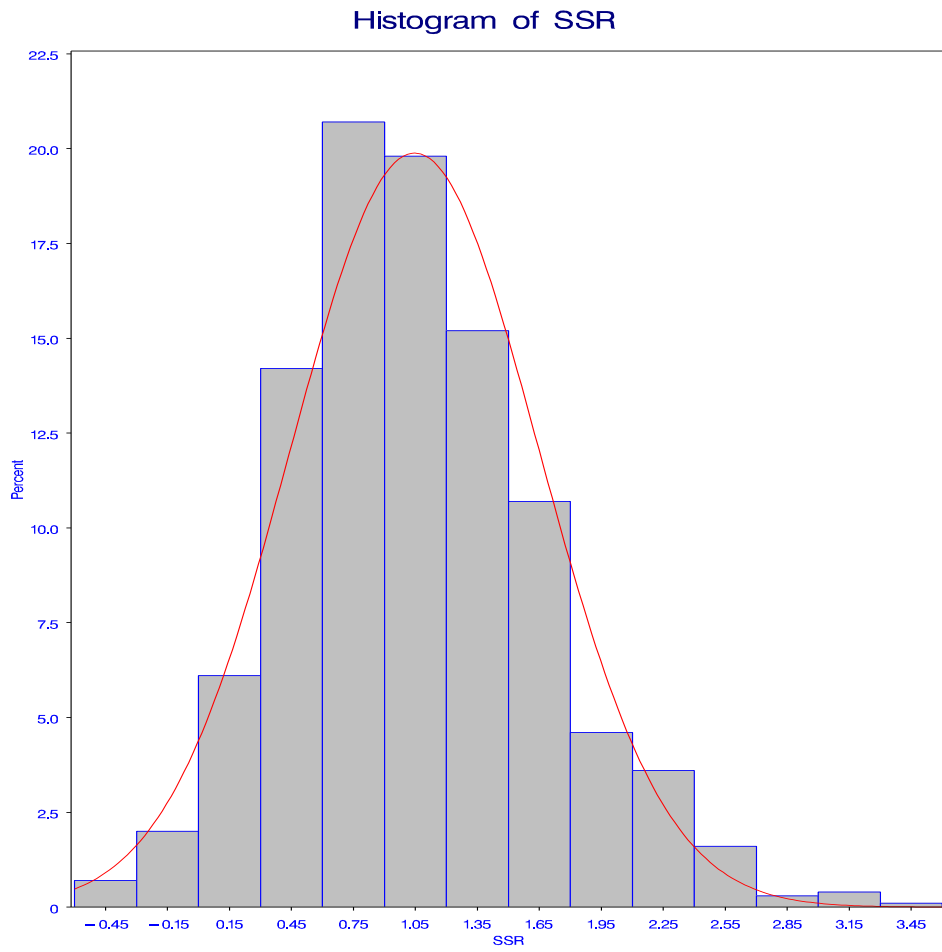
**Table 4.6:** Power Estimate with 95% CI

Model 3:  $\text{logit}(\pi_{it}) = \beta_0 + \beta_1 B_i(0.5) + \beta_2 B_{ij}(0.25) + b_i$

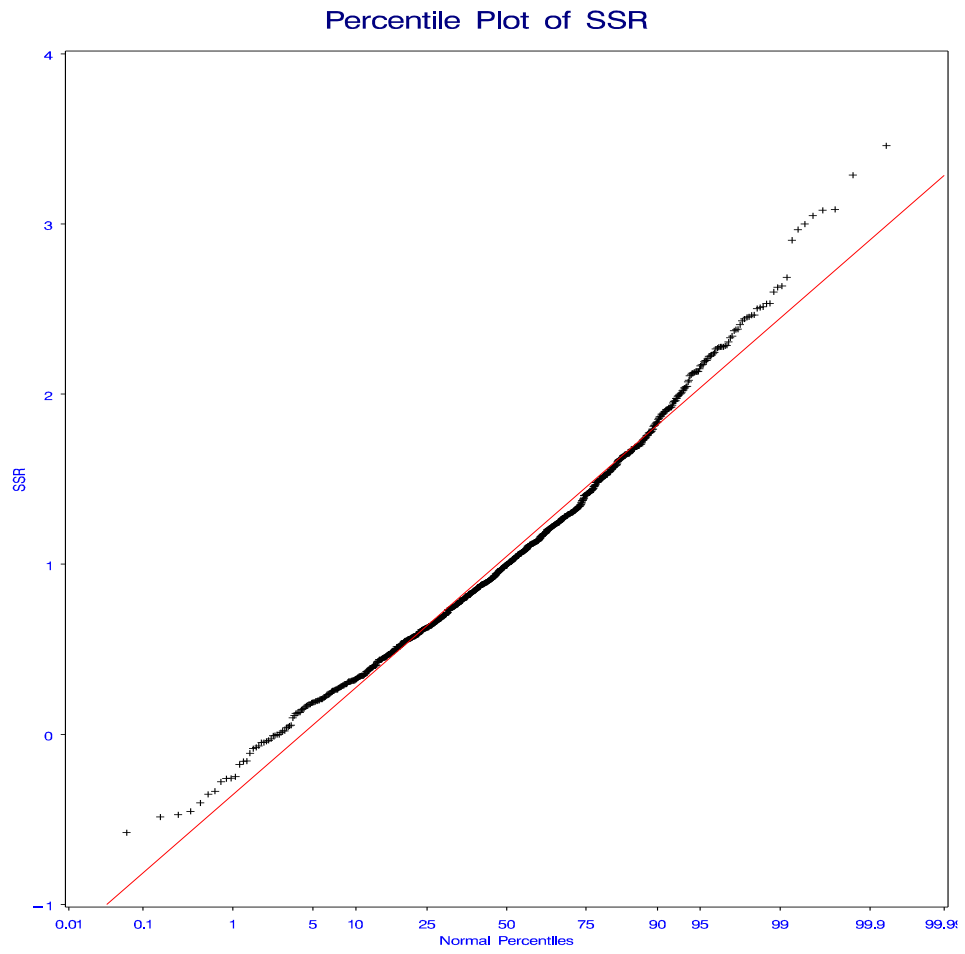
N = the Number of Clusters

M = The Cluster Size

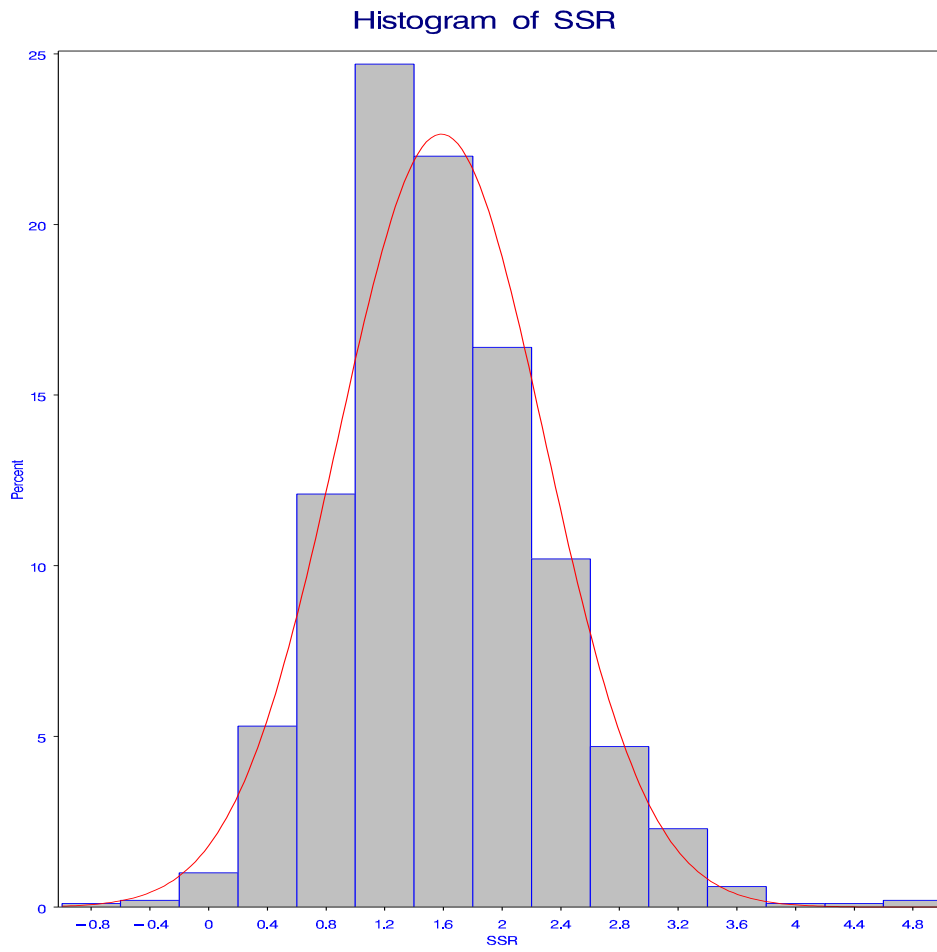
CI: The 95% Confidence Interval



**Figure 4.9:** The SSR Histogram for Model 3 with  $N = 20$  and  $M = 10$

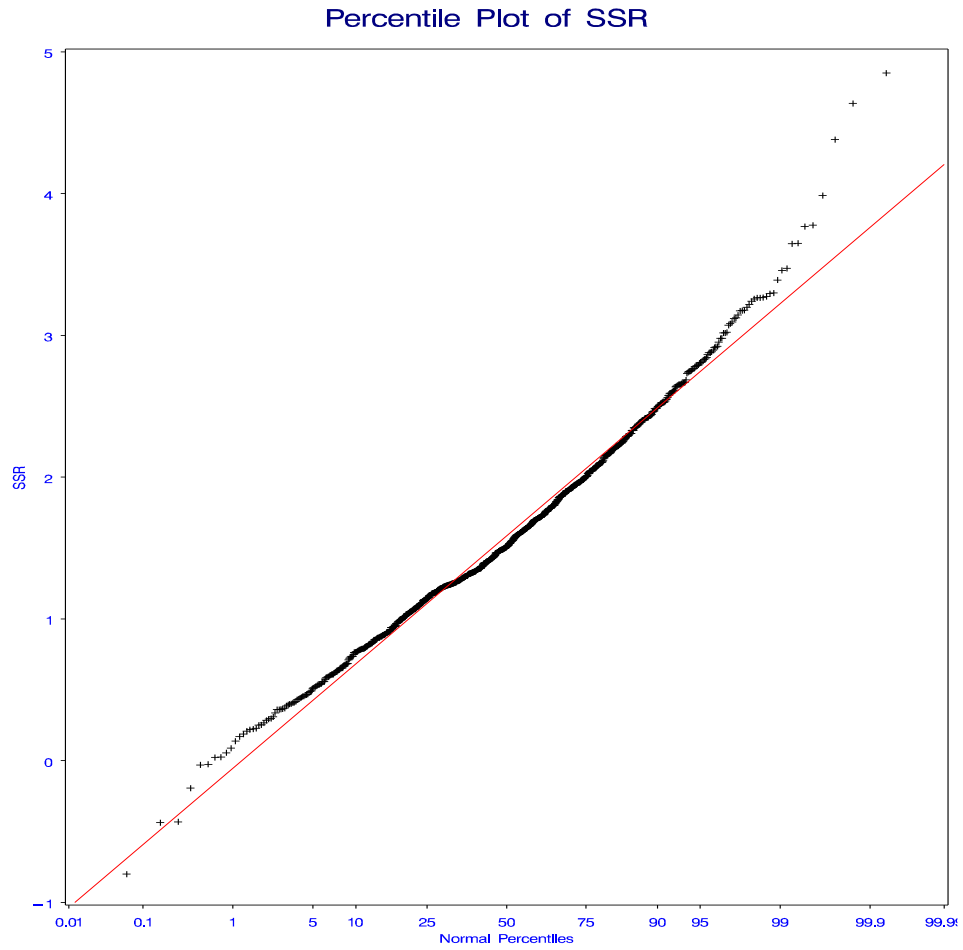


**Figure 4.10:** Percentiles: Model 3 with  $N = 20$  and  $M = 10$



**Figure 4.11:** The SSR Histogram for Model 3 with  $N = 20$  and  $M = 20$





**Figure 4.12:** Percentiles: Model 3 with  $N = 20$  and  $M = 20$

## 4.5 Conclusions

We were able to show that the proposed sum of squares residual statistics of Hosmer et al. [20] and Sturdivant and Hosmer [43] works well for a class of mixed logistic regression model with the presence of a continuous covariate, with a modest to large sample size dataset.

However, the same statistics failed to provide an adequate power to detect the correct model when the covariate are binary variables.

## Chapter 5

### Conclusions and Future work

In this dissertation, we aimed to gain insight into the efficiency of robustness of methods of approximation of the likelihood function of model 1.2. We focused on the three most popular approximations, the First Order (FO) method of Beal and Sheiner [5], the Laplace approximation (LP) of Wolfinger [49] and the Gaussian Quadrature (GQ) of Davidian and Gallant [10]. While no method outperforms the others on a consistent basis, both the GQ and LP showed the most efficiency and robustness, especially when the inter-individual correlation is small. The FO approximation provided the best convergence rate under correlated model as well as a model with two many parameters. We showed that when the random effects are not correlated, the GQ and the LP methods outperformed the FO method.

In the Theophylline dataset's model where the random effects are independent, model 2.1, the GQ and LP provided acceptable bias and confidence interval coverage. However for the Carbo Dioxide dataset's model where the random effects are correlate, model 2.2 and there are covariates in the model and large number of parameters that need to be estimated, the GQ and LP showed high sensitivity to the model's complexity and lack of robust estimates.

Efficiency of the estimates were found to depend on the magnitude of the inter- and intra-subject coefficients of variation. In particular, as the coefficient of variation between and within subject variation increased, the bias from the FO estimates increased. The FO method has been widely used and there are many examples for which it performs well. It has the advantage that it is exact when the model is linear in the random effects. It also has the advantage of being computationally simple and it provides a reasonable convergence rates. However as the variance of  $b_i$  increases (the variance of the random effects is much greater than zero), the accuracy of the the FO approximation decreases.

The effect of sample size was explored and the degree of improvement of the estimates seem not to warrant the costs associated with the moderate sample size increases. The increase of the sample size for the more complex model in the Carbon Dioxide model did not provide the expected nominal value of significance for the coverage probability of 95%.

The effects of reparametrization on the precision of the estimates in the context of Nonlinear Effects models has been significantly covered in the literature, Roe [34], Bates and Watt [3], and Hougaard [19]. From our simulation studies, the Carbon Dioxide model 2.2 is relatively close enough to a linear and therefore the Hougaard's measure of Skewness Ratkowsky [31] for the fixed effects did not change much from the original parametrization. For the Theophylline model 2.1, adding the exponential components into the model by adding an exponential transformation of the parameters to make sure the parameters of interest stay positive, While needed for scientific interpretability, led to poorer estimation properties. The "close-to-linear" measure went from an average of 0.0784, considered "very close-to-linear be-

havior”, to 0.25 where the skewness is apparent. At the same time the correlation between the coefficients stayed the same.

In the last section we showed that the proposed sum of squares residual statistics of Hosmer et al. [20] and Sturdivant and Hosmer [43] works well for a class of mixed logistic regression models with the presence of a continuous covariate, with a modest to large sample size dataset. However the same statistic failed to provide an adequate power to detect the correct model when only binary covariates are present in the model.

## **5.1 Recommendations for the Use of NLMM**

In our simulation studies, we aimed to point out some of the issues faced in practice when analyzing data that have a nonlinear mixed-effects structure, especially in the small to moderate sample sizes settings. NLMM are usually very complicated numerical problems, and as a result convergence issues commonly arise. Following are suggestions for NLMM users such as Pharmacologist and statistician for a better usage of NLMM models:

1. One way to address this is to use the FO approximation, if convergence is met then use the estimates value as starting values for the LP or GQ approximations to get and more accurate estimates.
2. When using these approximation methods to estimate model parameters, it would be helpful first to classify the model, into " a close to linear model" or "too far from linear model" and based on such classification use the FO for the first type and the LP or GQ for the second type.
3. Another useful classification of the model before deciding which approximation method to use, is the number of parameters in the model, for model with few parameters (e.g. less than 4 parameters) use the FO method and for a model with too many parameters (e.g., more than 5 paramors), use LP or GQ approximation method.
4. A simple way of overcoming the problem of nonconvergent is to use a single random effect parameter to find out what the estimate

for the associated fixed effect is, and do the same procedure for each random effect, then use these estimates as starting values to get final estimates.

5. It is also helpful to weight the use of too many random effect benefit and the complexity added to the model by including so many random effects.

## 5.2 Future work

We hope that more focus will be given to the inference involving small sample settings, and one needs to always keep in mind that most of the current computing tools that are implemented in SAS and S-plus for NLMM are based on the large sample theory and that care must be taken when using the NLMM with small to moderate size datasets.

The linearity of the approximation of NLMM by a semilinear model needs to be further explored using Ad HOC analysis such as the Hougaard's measure of Skewness Ratkowsky [31]. The "close-to-linear" nonlinear regression model, first described by Ratkowsky [32],



is a model that produces parameters having properties similar to those produced by a linear regression model. That is, the least squares estimates of the parameters are close to being unbiased, normally distributed, and have minimum variance estimators.

A nonlinear regression model sometimes fails to be close to linear due to the properties of a single parameter. When this occurs, bias in the parameters can render inferences using the reported standard errors and confidence limits invalid. An often suggested fix to the problem is with reparametrization, replacing the offending parameter with one that has better estimation properties.

Let  $H_i$  be the Hougaard's skewness measure for the  $i^{th}$  parameter  $\theta_i$ ,  $H_i$  is defined as follows:

$$H_i = E[\hat{\theta}_i - E(\hat{\theta}_i)]^3 = -(MSE)^2 \sum_{jkl}^{np} L^{ij} L^{ik} L^{il} (W_{jkl} + W_{kjl} W_{ljk})$$

Where  $L = (X^T X)^{-1}$  and  $W_{jkl} = \sum_{m=1}^n J_m^j H_m^{kl}$ ,  $J_m$  is the Jacobian vector and  $H_m$  is the Hessian matrix evaluated at observation  $m$ . This third moment is normalized using the standard error as:

$$H_i = \frac{E[\hat{\theta}_i - E(\hat{\theta}_i)]^3}{(MSE \times L^{ii})^{\frac{3}{2}}}$$

According to Ratkowsky [32], if  $|H_i| < 0.1$ , the estimator of pa-

parameter is very close-to-linear in behavior and, if  $0.1 < |H_i| < 0.25$ , the estimator is reasonably close-to-linear. If  $|H_i| > 0.25$ , the skewness is very apparent. For  $|H_i| > 0.25$ , the nonlinear behavior is considerable.

Another issue that deserves further investigation is the effects of relative magnitude of the scales of the parameters in the models, on the estimations. For example, the smallest parameter in the model seems to have the worse precision. Pinheiro and Bates [27] suggested using a normalized, scale invariant version of the variance-covariance matrix. There are different ways of normalizing  $\Delta$  in model 1.2. they proposed the use of the coefficient of variation (CV) matrix  $D_{CV}$  whose elements are defined as follows:

$$[D_{CV}]_{ij} = \frac{[D]_{ij}}{|\beta_{k(i)}\beta_{k(j)}|},$$

where  $\beta_k$  represents the  $k^{th}$  fixed effect and  $k(i), k(j)$  represent the indices of the fixed effects associated with the  $i^{th}$  and  $j^{th}$  random effects.

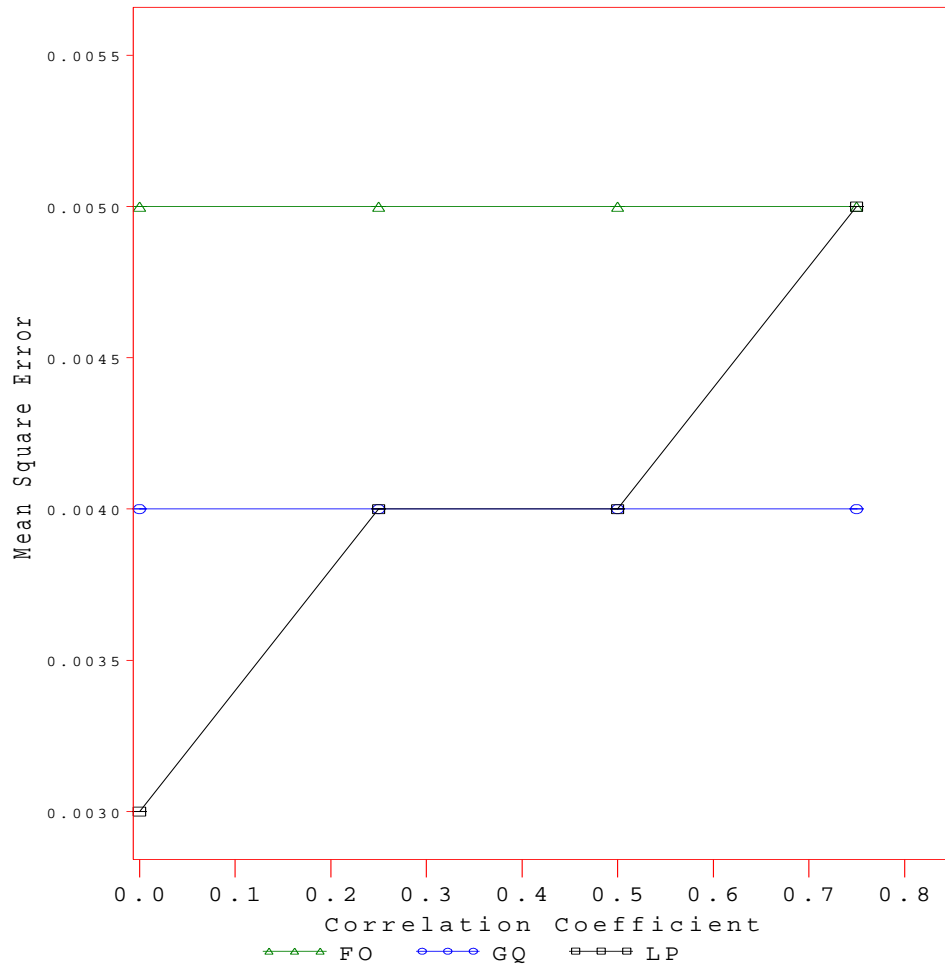
NLMM models provide a tool for analyzing repeated measurements data in which the relationship between the explanatory and response variables can be modeled as a single function, allowing the parameters to differ between individuals. In addition, these techniques

recognize that the variability associated with the response variable for a given individual may depend on the response value in a way that is similar for all individuals. This could be due, for example, to properties associated with measurement error and it would be helpful to explore modeling the measurement error as time dependent, such as Times Series type correlation and explore the improvement and complexity such approach may add to the model and to the approximation methods and their estimation. For example measurement error in model 1.2,  $\epsilon_{ij} \sim N(0, \sigma^2)$ ,  $\sigma^2$  can be modeled as a time dependent function as follows:

$$\sigma^2 = \exp[(1 - t_{ij}) \log(\rho)],$$
 where  $\rho$  is an unknown parameter that need to be estimated.

# Appendix A

## Figures of Precision Parameters

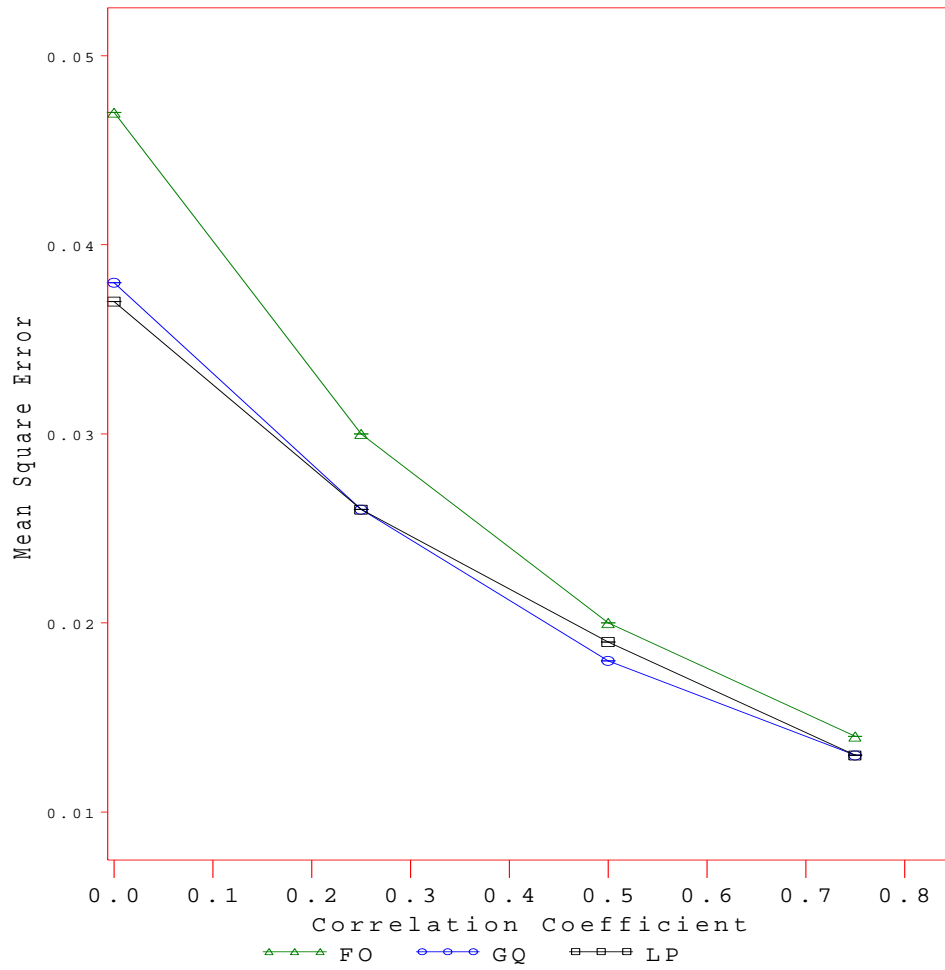


**Figure A-1:** The Mean Square Error Plot for  $\beta_1$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

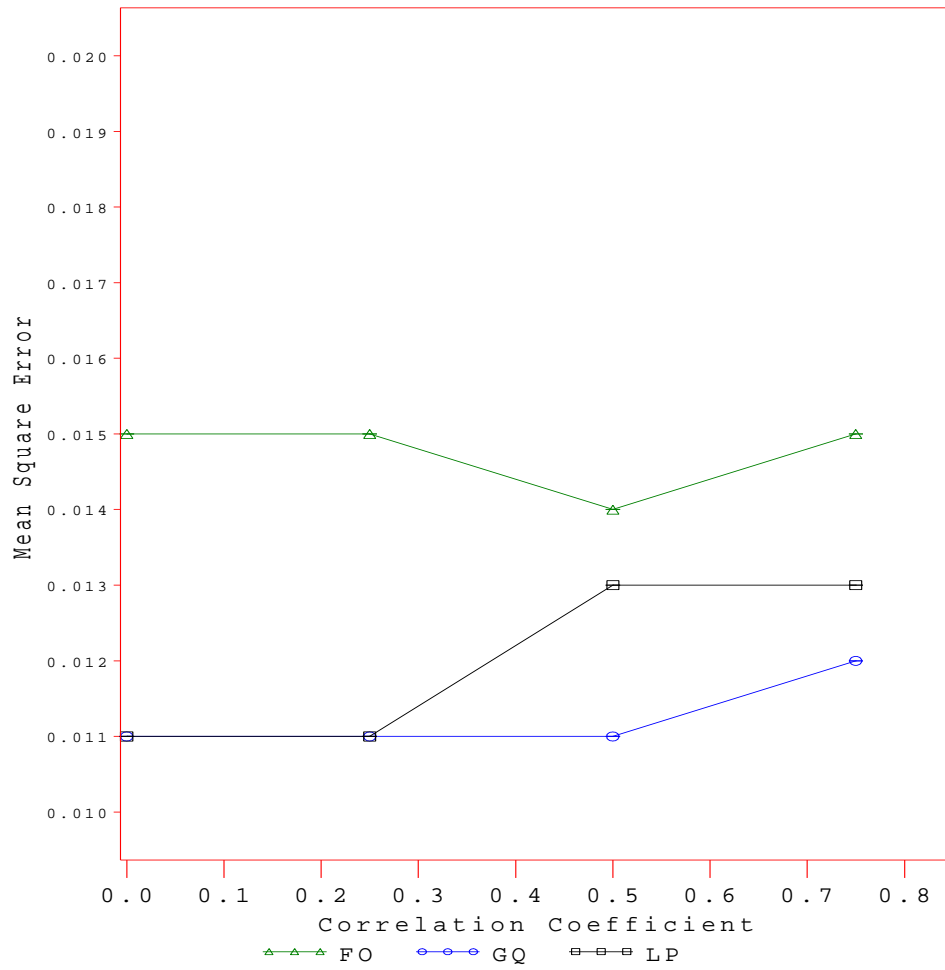


**Figure A-2:** The Mean Square Error Plot for  $\beta_2$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

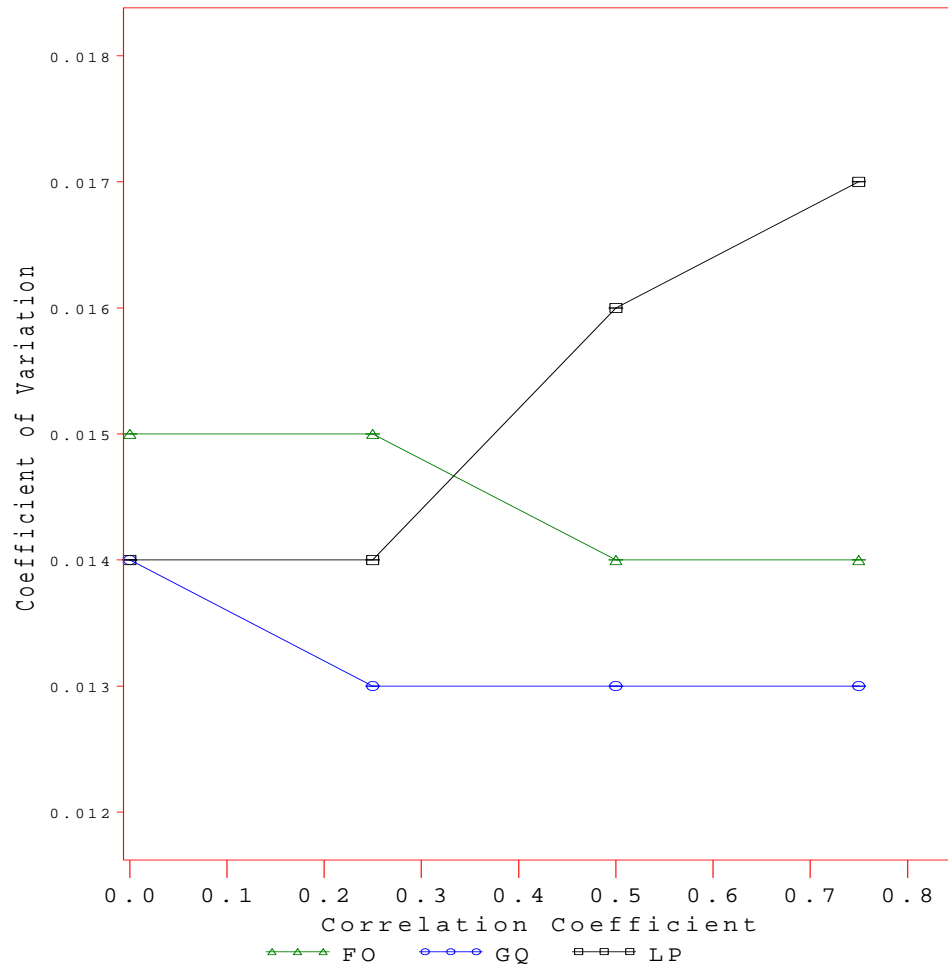


**Figure A-3:** The Mean Square Error Plot for  $\beta_3$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

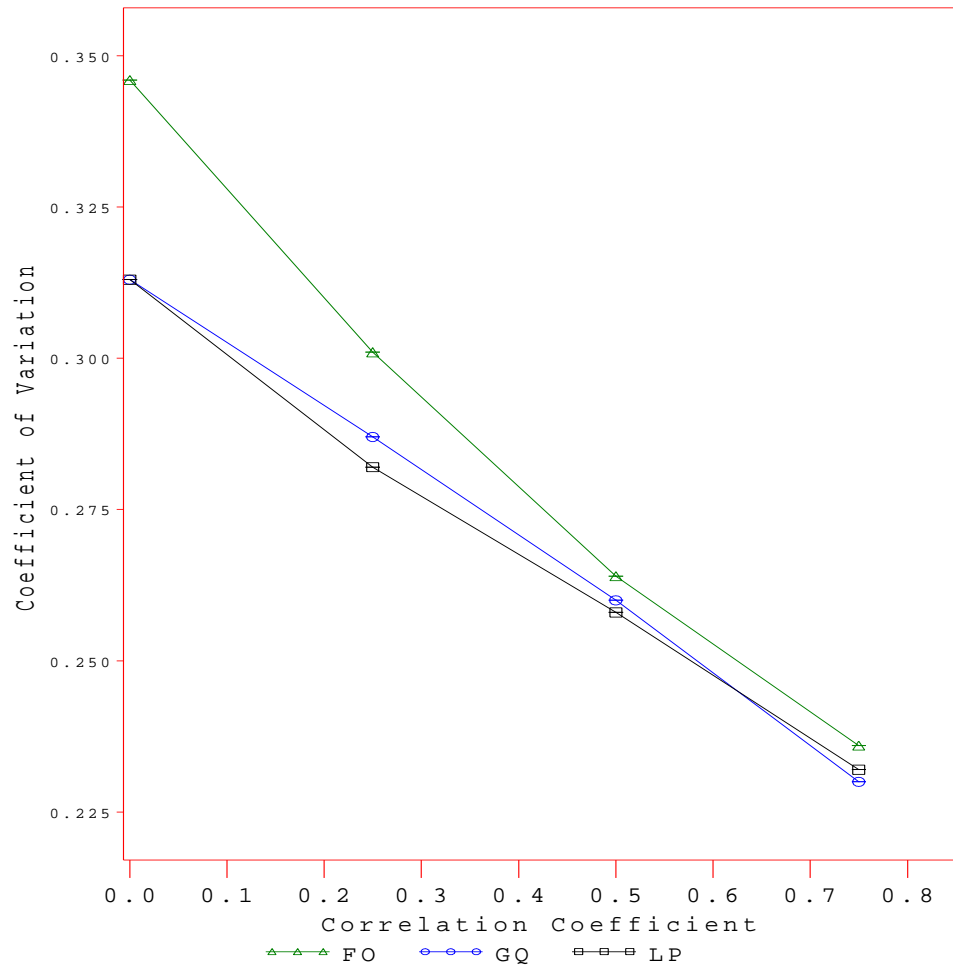


**Figure A-4:** The Coefficient of Variation Plot for  $\beta_1$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method



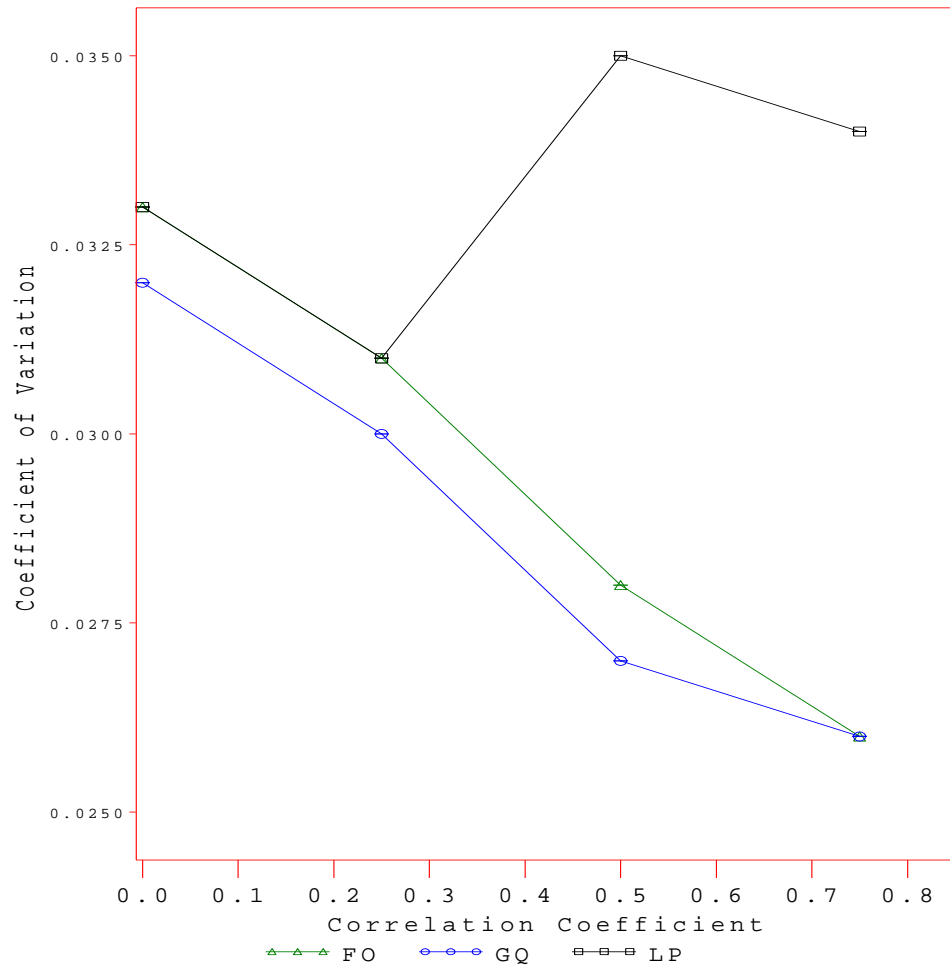
**Figure A-5:** The Coefficient of Variation Plot for  $\beta_2$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method



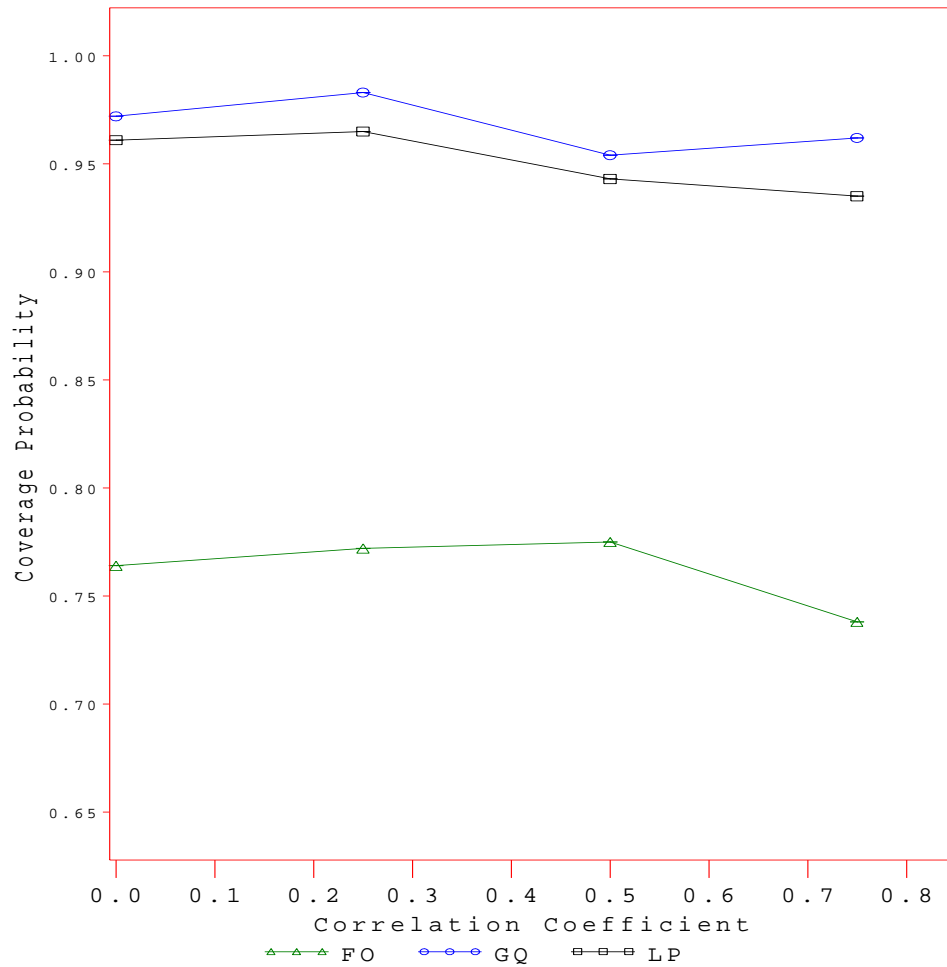


**Figure A-6:** The Coefficient of Variation Plot for  $\beta_3$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

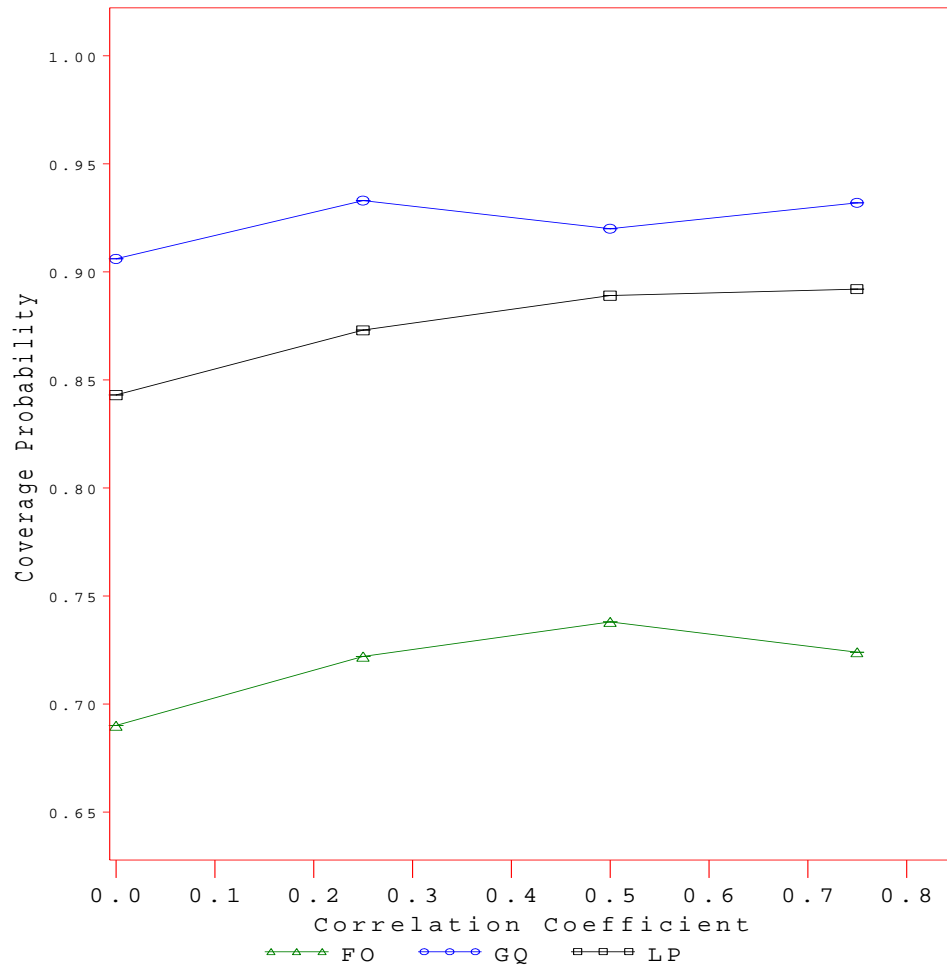


**Figure A-7:** The Coverage Probability Plot for  $\beta_1$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

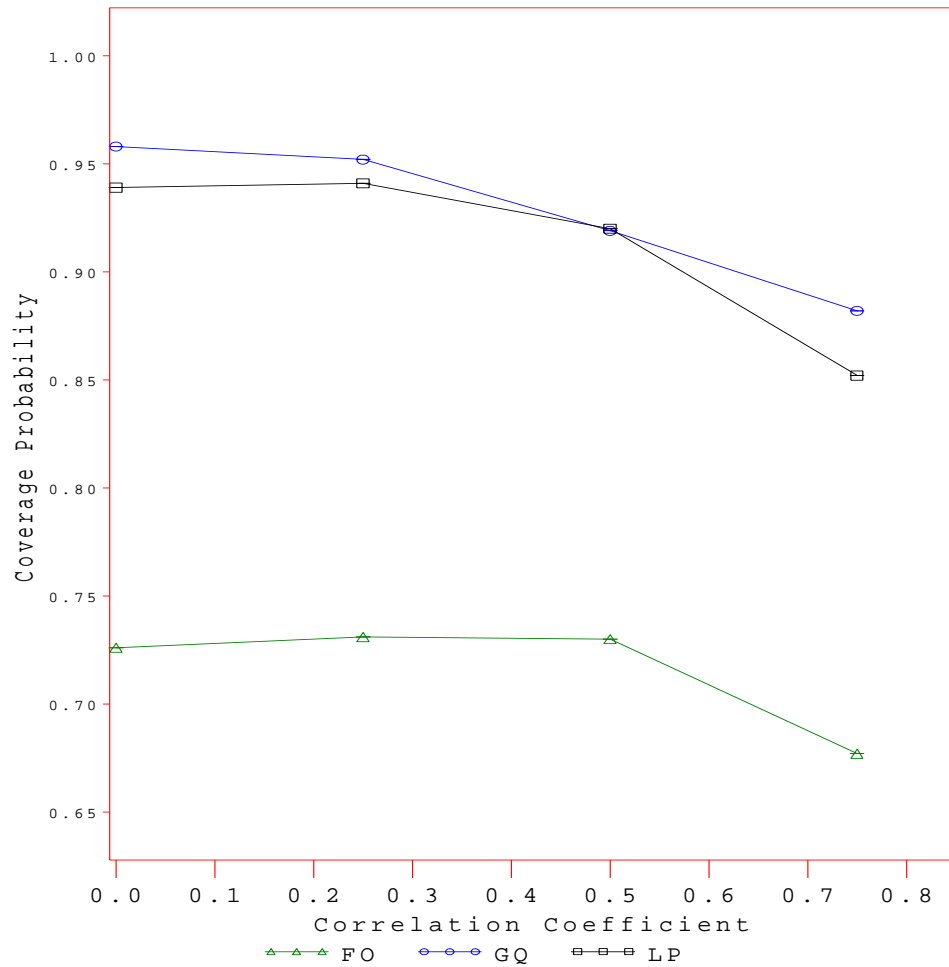


**Figure A-8:** The Coverage Probability Plot for  $\beta_2$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

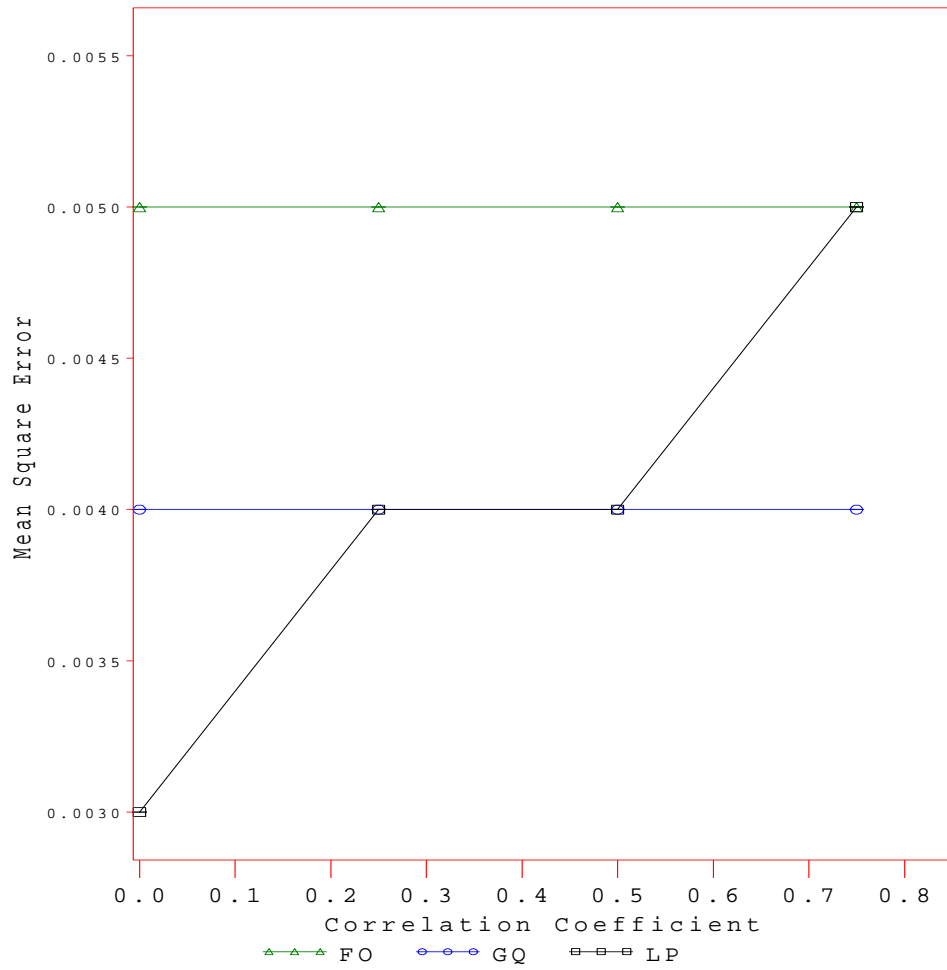


**Figure A-9:** The Coverage Probability Plot for  $\beta_3$  against  $\rho$  values.

Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

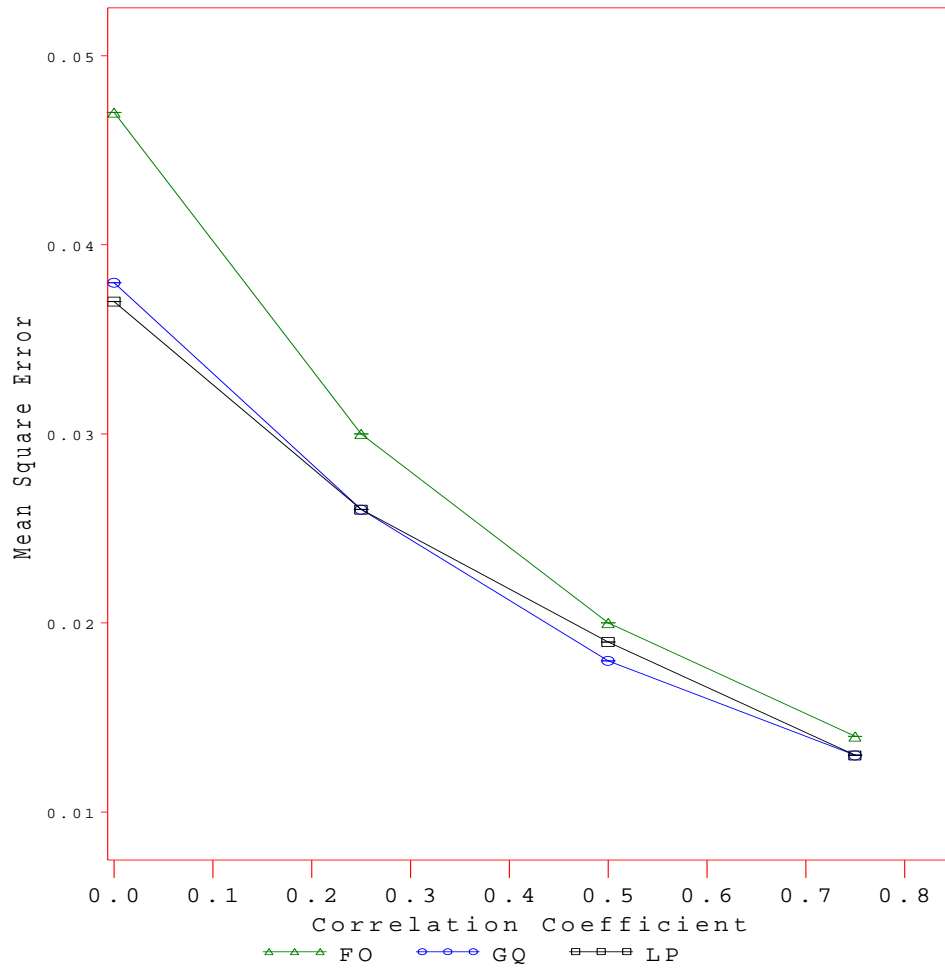


**Figure A-10:** The Mean Square Error Plot for  $\beta_1$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

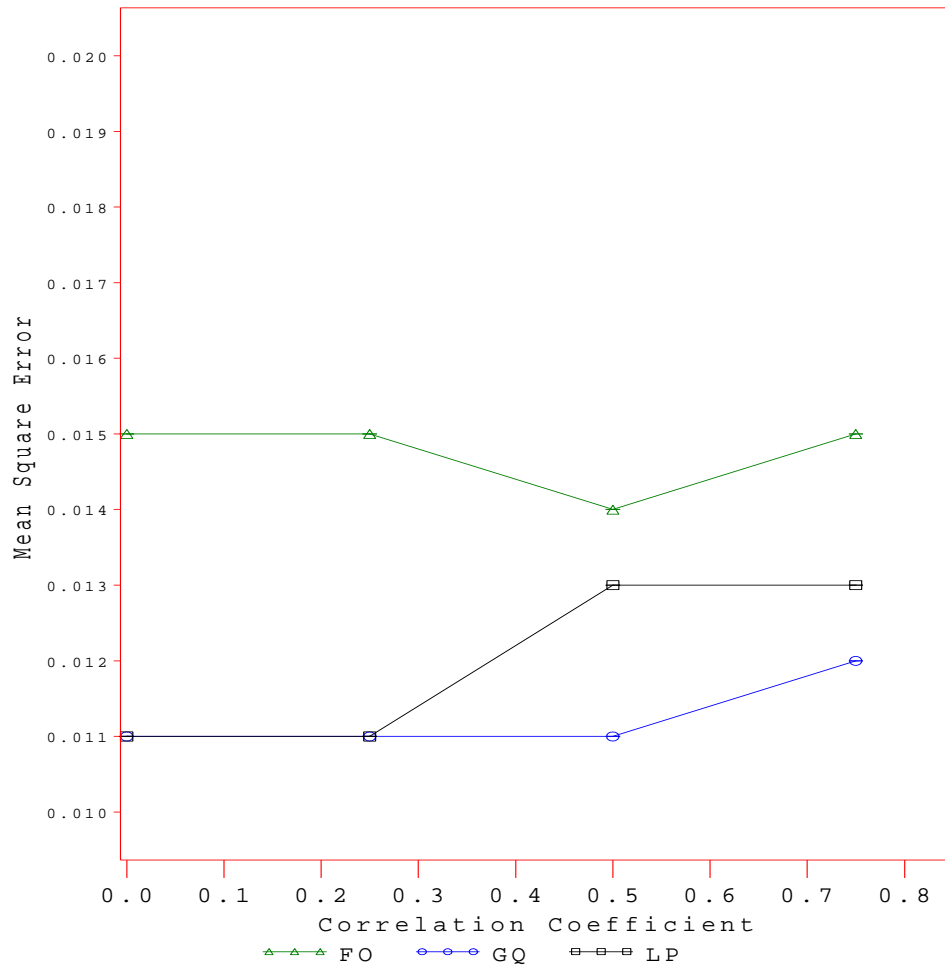


**Figure A-11:** The Mean Square Error Plot for  $\beta_2$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

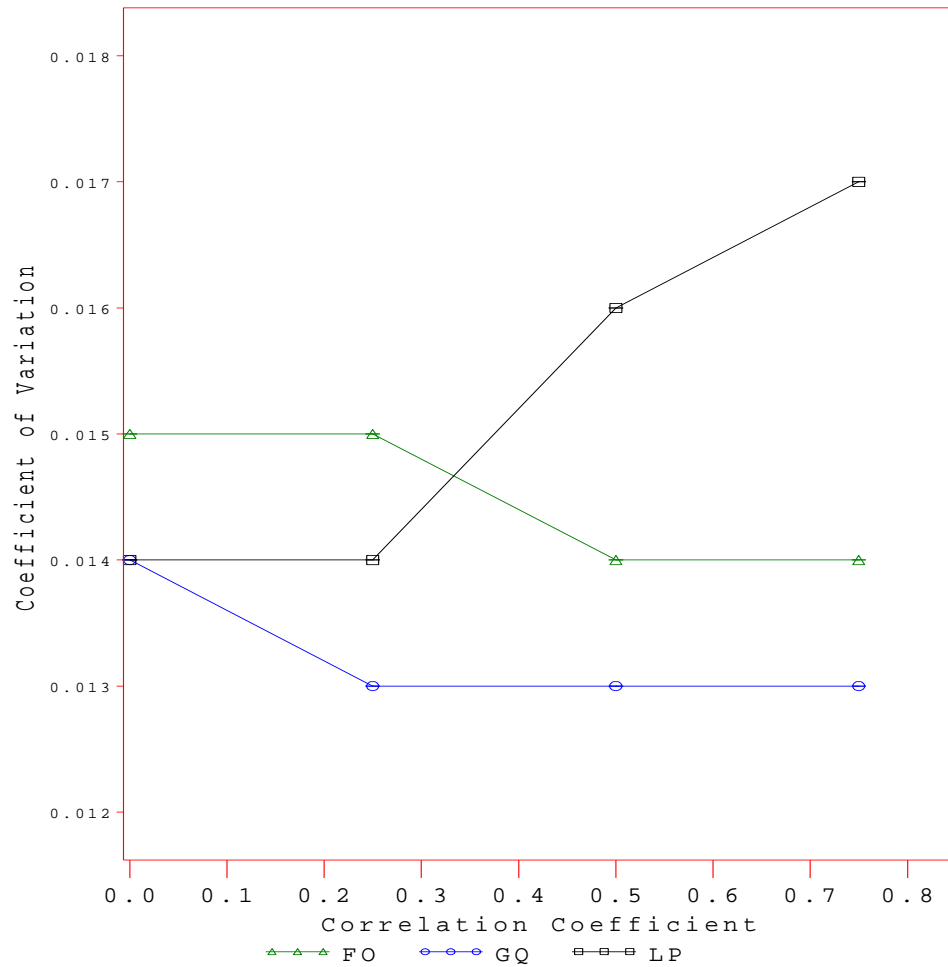


**Figure A-12:** The Mean Square Error Plot for  $\beta_3$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method



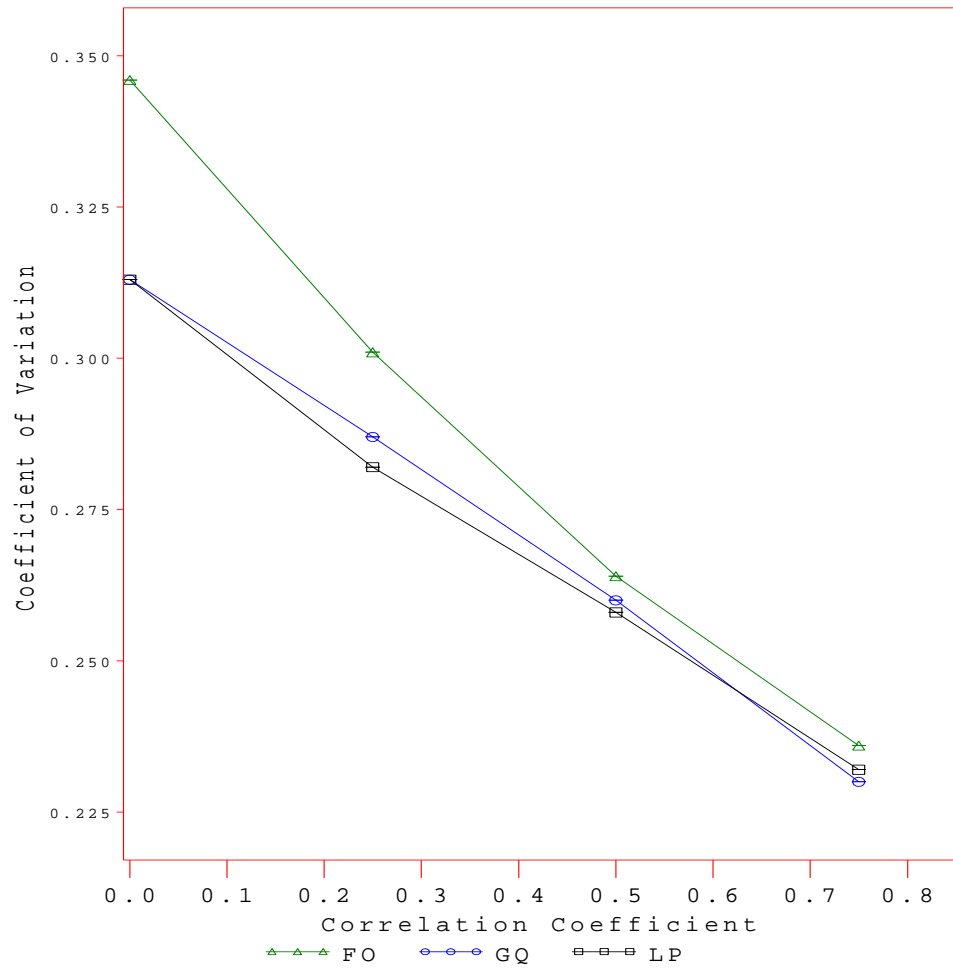
**Figure A-13:** The Coefficient of Variation Plot for  $\beta_1$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method



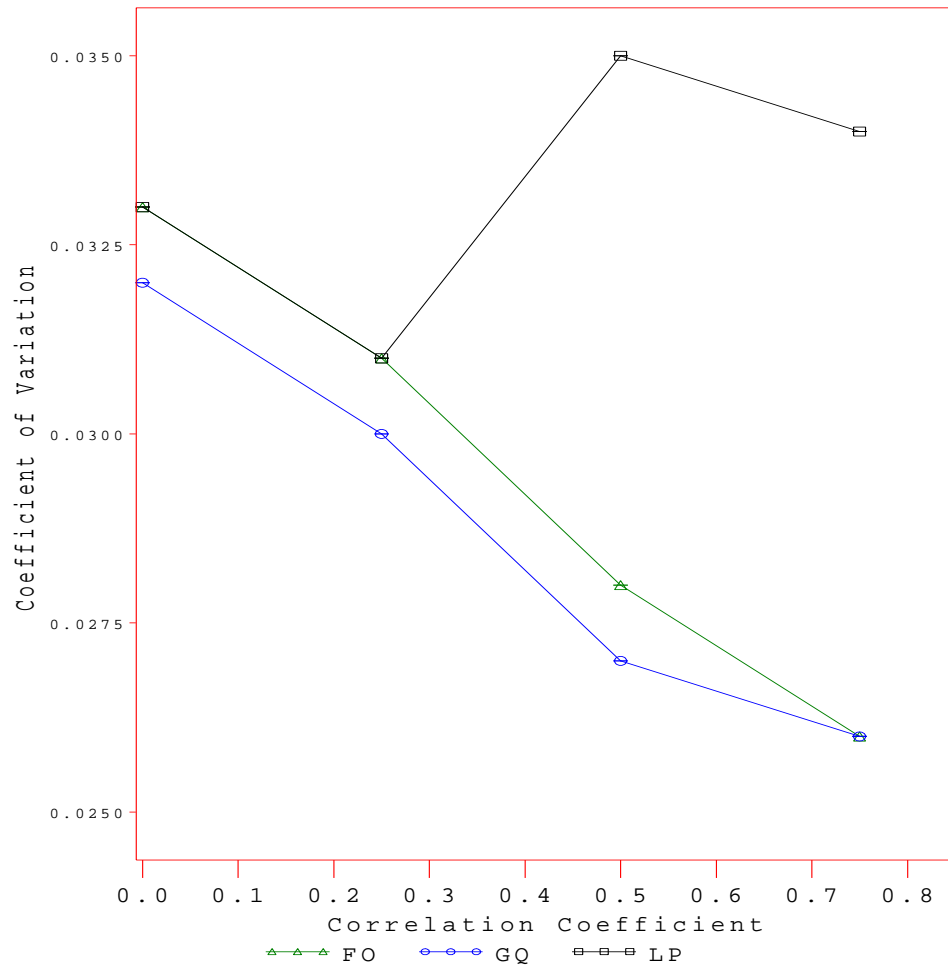


**Figure A-14:** The Coefficient of Variation Plot for  $\beta_2$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

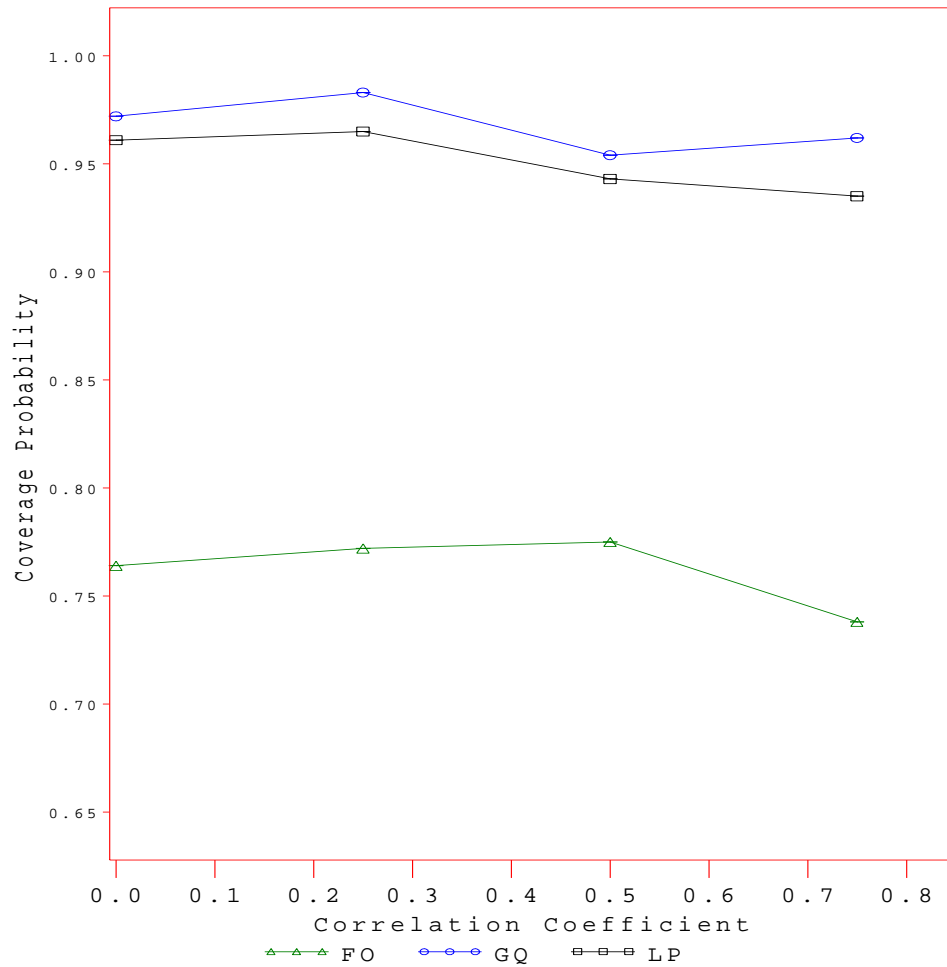


**Figure A-15:** The Coefficient of Variation Plot for  $\beta_3$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

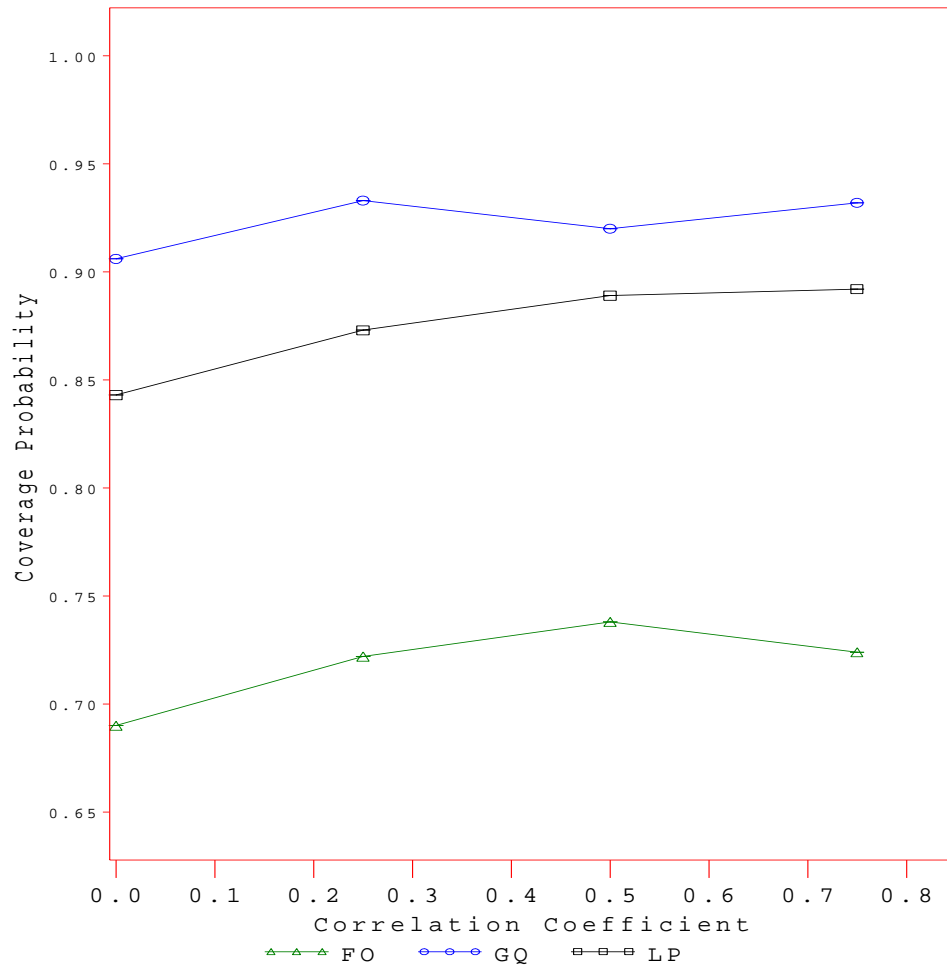


**Figure A-16:** The Coverage Probability Plot for  $\beta_1$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

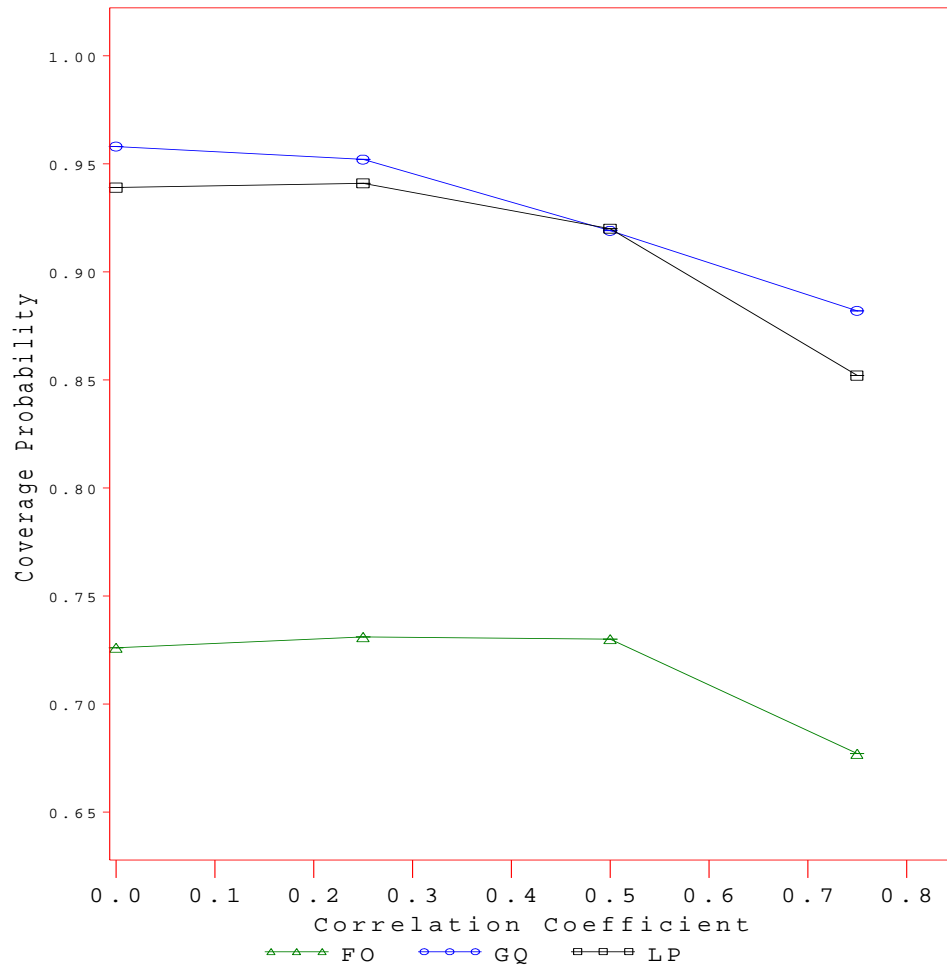


**Figure A-17:** The Coverage Probability Plot for  $\beta_2$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method



**Figure A-18:** The Coverage Probability Plot for  $\beta_3$  against  $\rho$  values.

Non-Correlated Random Effect Parameters Model.

GQ = Gaussian Quadrature Method, LP = Laplacian Method, and

FO = First Order Method

## BIBLIOGRAPHY

- [1] Abramowitz M. and Stegun, I. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- [2] Akaike Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716723.
- [3] Bates, D. and Watt, D. (1988). *Nonlinear regression analysis and its applications*. Wiley, New York.
- [4] Beal, S., and Sheiner, L., (1980). The nonmem system. *American Statistician*, **34**, 118-119.
- [5] Beal, S. and Sheiner, L. (1982). Estimating population kinetics. *Critical Reviews in Biomedical Engineering*, **8(3)**, 195-222.
- [6] Beitler, P.J., Landis, J.R. (1985). A mixed effects model for categorical data. *Biometrics*, **41**, 9911000.
- [7] Brown, H. and Kempton, R. (1994). The application of REML in clinical trials. *Statistics in Medicine*, **13**, 1601-1618.
- [8] Carroll, R. J. and Ruppert, D. (1982). A comparison between maximum like-

- likelihood and generalized least squares in a heteroscedastic linear model. *Journal of the American Statistical Association*, **77**, 878-882.
- [9] Davis, P. and Rabinowitz, P. (1984). *Methods of numerical integration*. Academic Press, New York.
- [10] Davidian, M. and Gallant, R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, **80**, 475-488.
- [11] Davidian, M. and Giltinan, D. (1993). Some general estimation methods for nonlinear mixed effects models. *Journal of Biopharmaceutical Statistics*, **3**, 23-55.
- [12] Davidian, M. and Giltinan, D. (1995). *Nonlinear models for repeated measurement data*. Chapman and Hall, London.
- [13] Food and Drug Administration (2003). *Bioavailability and bioequivalence studies for orally administered drug products, general considerations*. Rockville (MD): FDA.
- [14] Gallant, A. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics*, **3**, 35-50.
- [15] Gibaldi, M. and Perrier, D. (1982). *Pharmacokinetics*. Marcel Dekker, New York.
- [16] Hand, D. and Crowder, M. (1996). *Practical longitudinal data analysis*. Chapman and Hall. London.
- [17] Harville, D. (1977). Maximum likelihood approaches to variance components

- estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320-340.
- [18] Hartford, A., and Davidian, M. (2000). Consequences of misspecifying distributional assumptions in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, **34**, 139-164.
- [19] Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *Journal of the Royal Statistical Society Series B*, **47**, 103-114.
- [20] Hosmer, D.W. et al. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, **16**, 965-980.
- [21] Jeske, D. and Harville, D. (1988). Prediction-interval procedures and fixed effects confidence interval procedures for mixed linear models. *Communication in Statistics Theory and Methods*, **17**, 1053-1087.
- [22] Kacker, A. and Harville, D. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, **79**, 853-862.
- [23] Kedem, B. and Fokianos, K. (2002). *Regression models for time series analysis*, Wiley, New York.
- [24] Kenward, M. and Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983-997.
- [25] Lindstrom, M. and Bates, M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673-687.



- [26] Pinheiro, J. and Bates, D. (1995). Approximations to log-likelihood function in the nonlinear mixed effects model. *Computational and Graphical Statistics*, **4(1)**, 12-35.
- [27] Pinheiro, J. and Bates, D. (2001). *Mixed effects models in S and S-Plus*. Springer, New York.
- [28] Press, William H., Teukolsky, Saul A. , Vetterlink, William T., and Flannery, Brian P. (1992). *Numerical recipes in C: The art of scientific computing (second edition)*. Cambridge University Press., <http://www.nr.com>.
- [29] Price, C. J., Kimmel, C. A., Tyl, R. W., and Marr, M. C. (1985). The developmental toxicity of ethylene glycol in rats and mice. *Toxicological Applications in Pharmacology*, **81**, 825-839.
- [30] Racine-Poon, A. (1985). Bayesian approach to nonlinear random effects models. *Biometrics*, **41**, 1015-1023.
- [31] Ratkowsky, D. (1983). *Nonlinear regression modeling*. Marcel Dekker, New York.
- [32] Ratkowsky, D. (1990). *Handbook of nonlinear regression models*. Marcel Dekker, New York.
- [33] Ross, G. J. S. (1970). The efficient use of the function minimization in nonlinear maximum likelihood estimation. *Applied Statistics*, **20**, 7-14.
- [34] Roe, D. J. (1997). Comparison of population pharmacokinetic modeling methods using simulated data, Results from the population modeling work group. *Statistics in Medicine*, **16**, 1241-1262.

- [35] SAS Institute Inc. (2000). *The nlmixed procedure, SAS/STAT User's guide, Version 8*. Cary, NC.
- [36] Satterthwaite F.E. (1941). Synthesis of variance. *Psychometrika*, **6**, 309-316.
- [37] Seber, G. and Alan A. (2003). *Linear regression analysis*. John Wiley and Sons, New York.
- [38] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6(2)**, 461-464.
- [39] Searle, S.R. (1971). *Linear models*. Wiley, New York.
- [40] Sheiner, L. and Beal, S. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten Model: routing clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, **8**, 553-571.
- [41] Sheiner L., Rosenberg, B., and Melmon, K. (1972). Modeling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research*, **5**, 441-459.
- [42] Slud, E. and Kedem, B. (1996). Partial likelihood analysis of logistic regression and autoregression. *Statistica Sinica*, **4**, 89-106.
- [43] Sturdivant R., Hosmer, D. (2007). A smoothed residual based goodness of fit statistics for logistic hierarchical regression models. *Computational Statistics and Data Analysis*, **51**, 3898-3912.
- [44] Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New York.

- [45] Vonesh, E. and Chinchilli, V. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, Inc..
- [46] Vonesh, E. (1996). A note on the use of laplace's approximation for nonlinear mixed effects models. *Biometrika* **83**, 447-452.
- [47] Vonesh, E. (1992). Nonlinear models for the analysis of longitudinal data. *Statistics in Medicine*, **11**, 1929-1954.
- [48] Walker, S. (1996). An EM algorithm for nonlinear random effects models. *Biometrics*, **52**, 934-944.
- [49] Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, **80**, 791-795.

## VITA

Born in December 1970, Nouakchott, Mauritania.

B.S. Mathematics, 1995, Universit Henri Poincar Nancy 1, Nancy, France.

D.E.A (Master) Mathematical Economics, 1996, Universit Pierre et Marie Curie, Paris, France.

M.S. Applied Mathematics, 2003, University of Maryland College Park, Maryland.

May 2000 to March 2003, Statistical Analyst, American Petroleum Institute (API), Washington, DC.

March 2003- Present, Biostatistician, Food and Drug Administration (FDA), Silver Spring, Maryland.