Stony Brook University

Power Analysis of the Likelihood Ratio Test for Logistic Regression Mixtures

A Dissertation Presented

by

Minyoung Lee

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

May 2011

Stony Brook University

The Graduate School

Minyoung Lee

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Nancy R. Mendell – Dissertation Advisor
Professor, Department of Applied Mathematics and Statistics

Stephen J. Finch – Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Hongshik Ahn
Professor, Department of Applied Mathematics and Statistics

Barbara Nemesure
Associate Professor, Department of Preventive Medicine

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Power Analysis of the Likelihood Ratio Test for Logistic Regression Mixtures

by

Minyoung Lee

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2011

Finite mixture models emerge in many applications, particularly in biology, psychology and genetics. This dissertation focused on detecting associations between a quantitative explanatory variable and a dichotomous response variable in a situation where the population consists of a mixture. That is, there is a fraction of the population for whom there is an association between the quantitative predictor and the response and there is a fraction of individuals for whom there is no association between the quantitative predictor and the response.

We developed the Likelihood Ratio Test (LRT) in the context of ordinary logistic regression models and logistic regression mixture models. However, the classical theorem for the null distribution of the LRT statistics can not be applied to finite mixture alternatives. Thus, we

conjectured that the asymptotic null distribution of the LRT statistics held. We investigated how the empirical and fitted null distribution of the LRT statistics compared with our conjecture. We found that the null distribution appears to be well approximated by a 50:50 mixture of chi-squared distributions, i.e., $0.5\chi_1^2 + 0.5\chi_2^2$ with respect to the critical values. Based on this null distribution, simulation studies were conducted to compare the power of the ordinary logistic regression models to the logistic regression mixture models. The logistic regression mixture models resulted in the improvement in power to detect the association between the two variables, compared with the ordinary logistic regression models. We found the significant factors in the *improvement* of the power by modeling the odds ratio in the improvement (logistic mixture model vs. ordinary logistic regression model). Essentially, the only factors that affected *improvement* in power were slope and mixing proportion. In addition, we compared the precision of these two approaches. This mixture model can be widely applied in large sample surveys with non-response and in missing data problems.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank all people who have helped and inspired me during my doctoral study. I would never have been able to finish my dissertation without the guidance of my advisor, the committee members, help from friends, and support from my family.

Above all, I would like to express my deepest gratitude to my advisor, Dr. Nancy R. Mendell, for her excellent guidance, caring, patience, and supporting me throughout the year. It has been an honor to be her student. She is always my good genius.

I would also like to thank my dissertation committee, Dr. Stephen J. Finch, Dr. Hongshik Ahn, and Dr. Barbara Nemesure for their direction and invaluable advice.

Last but not least, I am truly thankful to my family and friends for all their love and encouragement. Their unconditional and generous support inspired me with confidence.

Chapter 1

Introduction and Literature Review

1.1 Introduction

The switching regression model was originally proposed by Quandt (1972) and Ramsey (1975). This model has two or more components of a probability density function that is the mixture of normal densities. This dissertation considers the logistic switching regression model. That is, our attention is focused on a switching model which has two components in the context of the logistic regression. The logistic switching regression model can be said to be the equivalent to the finite mixture model (Pearson, 1894) for the logistic regression relationship. In that sense, we refer to this model as a logistic regression mixture model. Particularly, it focuses on the case that two logistic regression equations differ only in their slopes and one of the slopes is assumed to be zero. This model is motivated by Fienberg et al. (1985) and it is written as follows

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \quad \text{with probability } \pi \tag{1.1.1}$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 \quad \text{with probability } 1-\pi\,. \tag{1.1.2}$$

Here, p is the probability that the dichotomous response variable Y equals 1 and x is a quantitative explanatory variable. However, there are no explicit expressions for obtaining the maximum likelihood estimates (MLE) of the parameters of interest under this model. The Expectation Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) has been widely used to obtain the MLEs in this kind of mixture model.

This study involves a situation where one is conducting the likelihood ratio test (LRT) for an association between a quantitative explanatory variable and a dichotomous response variable. The purpose is to compare the power of the LRT in the case of fitting the single ordinary logistic regression model to the power of the LRT upon fitting the logistic regression mixture model defined in Equation (1.1.1) and (1.1.2). The power analysis will be conducted using simulation and the power will be assessed and compared in several different scenarios by sample size, effect size, intercept, and mixing proportion of observed mixture populations.

A comprehensive review of the literature is presented in the following section. It includes the theoretical background and the outline of the problem in this dissertation. Chapter 2 details the EM algorithm that is used for finding MLEs in logistic regression mixture models, including simulation results for the estimates of the parameters obtained.. We investigate the null distribution of LRT statistics for power studies in Chapter 3. Power analyses are discussed in Chapter 4, based on two approaches: ordinary logistic regression models and logistic regression mixture models. In addition, we compare the precision of estimates in the context of these two approaches. Chapter 5 contains the conclusions and the directions for future study.

1.2 Literature Review

1.2.1 Finite Mixture Models

The finite mixture model was proposed to analyze heterogeneous data. The model allows for combination of the samples from different populations in a single sample. One of the first major analyses involving the use of mixture models was Pearson's study (1894). The study was about the frequency distribution of measurements of the carapace of 2,000 female shore crabs, provided by Weldon (1893). Half of the samples were obtained from crabs at Plymouth Sound, and the remaining samples were from the Bay of Naples. Weldon observed that the measurements of the frontal breadth of the shore crabs at the Bay of Naples were generated from an asymmetric frequency distribution. Pearson demonstrated that a two component Gaussian mixture density fit the data. After his study, the mixture model-based approach has been widely used in many fields in the biological, physical, and social sciences because of the flexibility of the mixture model.

In general, the observations $y_1, \Lambda\ , y_n$ are said to arise from a finite mixture distribution, if the probability density function $f(y)$ of this distribution has the following form:

$$f(y \mid \Theta) = \pi_1 f_1(y \mid \theta_1) + \pi_2 f_2(y \mid \theta_2) + \Lambda\ + \pi_m f_m(y \mid \theta_m) \qquad (1.2.1)$$

Here, $y = (y_1, \Lambda\ , y_n)'$ denotes the random vector and $\Theta = (\theta_1, \Lambda\ , \theta_m; \pi_1, \Lambda\ , \pi_m)$ denotes the vector of all parameters. The π_j denotes the relative proportion of j^{th} component density function $f_j(y \mid \theta_j)$: that is, $0 \le \pi_j \le 1$ and $\sum_{j=1}^{m} \pi_j = 1$.

Since these finite mixture models have multiple maxima in the mixture likelihood function it is usually difficult to obtain the estimates of the parameters using the maximum likelihood method. A number of papers have dealt with the problem of estimating the parameters of the finite mixture models using the maximum likelihood method. However, it was the paper by Dempster, Laird, and Rubin (1977) that illustrated the application of the EM (Expectation-Maximization) algorithm to the finite mixture models. The EM algorithm can be applied to obtain the MLEs of the parameters of the finite mixture model. This algorithm is an efficient iterative procedure to compute the MLE in the presence of missing or hidden data. Section 1.2.4 discusses the EM algorithm in detail.

1.2.2 Switching Regression Models

The switching regression model is an exogenous switching model proposed by Quandt (1972). This model generalized a problem of mixture distributions (Day, 1969). If a *priori* information on how the sample is partitioned into the corresponding regime is provided, it is called a switching regression model with known sample separation. Otherwise, it is a switching regression model with unknown sample separation.

The simplest formulation of the switching regression model consists of two regression equations as follows:

$$y_i = x_i^T \beta_1 + \varepsilon_{1i}, \quad \text{with probability } \pi \tag{1.2.2}$$

$$\text{and} \qquad y_i = x_i^T \beta_2 + \varepsilon_{2i}, \quad \text{with probability } 1-\pi, \tag{1.2.3}$$

where $\varepsilon_{1i} \sim N(0, \sigma_1^2)$, $\varepsilon_{2i} \sim N(0, \sigma_2^2)$, and $x^T = (x_1, \Lambda, x_p)$ is a vector of p independent variables. The sample in this model is generated from distinct regression equations. In other words, the i^{th} observed dependent variable y_i is generated either from Equation (1.2.2) or from Equation (1.2.3), but never both.

The traditional interest in the switching regression model involves the following issues: (1) testing the null hypothesis that no switch in regimes exists against the alternative that the observations were generated by two or more distinct regression equations, (2) estimating the corresponding regression equations for each regime, and (3) classifying the observations into underlying regimes. Various special cases of these problems have been treated in the literature. When, under the alternative hypothesis, the information on sample separation is given, the problem of testing the null hypothesis was solved exactly by a test (Chow, 1960). Each of the equations can be estimated by standard methods such as ordinary least squares. When sample separation is unknown, Quandt (1958) derived the relevant likelihood ratio test statistic λ to test the null hypothesis that no switch occurred. The results of the sampling experiment performed by Quandt(1960) led to the rejection of the hypothesis that $-2\log\lambda$ has the χ^2 distribution. The maximum likelihood estimation methods were suggested by Goldfeld and Quandt (1972), Hartley (1978), and Kiefer (1980). Other estimation methods based on moment generating functions were investigated in Quandt and Ramsey (1978). In these cases, the estimation of the switching regression is equivalent to the estimation of the parameters of mixtures of normal distribution since there is the assumption that the observations were generated from a mixture of two normal densities. In other words, the switching regression models are equivalent to the finite mixture models of regression relationship.

In this dissertation, we consider the finite mixture models of logistic regression relationships. This mixture model can be considered the switching regression models in the context of logistic regression equations. Particularly, we deal with the situation where some subjects are unaffected by treatment. That is, our generated data sets include the variable X of zero. This situation can be related to the problem of detecting a treatment effect when the treatment group contains non-responders which was considered by Good (1979). Good used a mixture to describe the distribution of the responses in treatment group and he represented the distribution of the affected group with a shift in the mean of the distribution of the unaffected group and used a mixture of the two components for the distribution of the treatment group.

1.2.3 The Ordinary Logistic Regression and Parameter Estimation

Logistic Regression Model

This dissertation considers a two component mixture model for binary variables in the context of logistic regression. Logistic regression is a method that can be used for assessing association between a categorical response variable and quantitative explanatory variables. The fitted logistic regression model can also be used for predicting the probability of occurrence of an event. The general logistic regression model is

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_i \beta_i x_i \ . \qquad (1.2.4)$$

Here, p denotes the probability of a particular outcome of a dichotomous or polytomous response variable, and the $\{x_i\}$ are observed values corresponding to a set of explanatory variables. In the case of two or more explanatory variables, these explanatory variables can be quantitative or

qualitative or both. In the case of a single explanatory variable, logistic regression is used only for a quantitative explanatory variable.

Parameter Estimation

Throughout this research parameter estimation is based on the method of maximum likelihood, introduced by Fisher (1921). It is the most popular technique for obtaining estimators because of the desirable asymptotic properties of MLEs; consistency, invariance, normality and efficiency. In general, if $x = (x_1, \Lambda\ x_n)$ are a set of independent and identically distributed (i.i.d.) values in a random sample of size *n* from a population with parameter θ and probability density function $f(x \mid \theta)$, the likelihood function is defined by

$$L(\theta \mid x) = \prod_{i=1}^{n} f(x_i \mid \theta). \tag{1.2.5}$$

Also, the log likelihood function is represented by $\lambda(\theta \mid x) = \log L(\theta \mid x)$. The MLE $\hat{\theta}$ of θ can be obtained by maximizing the likelihood, which is equivalent to maximizing log likelihood since the logarithm is a continuous strictly increasing function over the range of the likelihood. Therefore the MLE is,

$$\hat{\theta} = \arg\max_{\theta} \lambda(\theta \mid x). \tag{1.2.6}$$

For purposes of simplicity, suppose that the logistic model has only a single quantitative explanatory variable; i.e., the single logistic regression model in this research is then

$$\log(\frac{p_i}{1-p_i}) = \alpha + \beta x_i = \mathbf{x_i}^T \theta, \tag{1.2.7}$$

where $\mathbf{x_i} = (1, x_i)^T$ for $i = 1, \ldots, n$ and $\theta = (\alpha, \beta)^T$. Since the response variable, $y = (y_1, \Lambda, y_n)$ has a Bernoulli distribution with probability p_i, the likelihood is

$$L(\theta \mid y) = \prod_{i=1}^{n} p_i^{\,y_i} (1-p_i)^{1-y_i} \ . \tag{1.2.8}$$

From Equation (1.2.7) p_i is as follows:

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} = \frac{e^{\mathbf{x_i}^T \theta}}{1+e^{\mathbf{x_i}^T \theta}}. \tag{1.2.9}$$

Therefore, the log likelihood can be written as

$$\lambda(\theta \mid x, y) = \sum_{i=1}^{n} \left[y_i \log p_i + (1-y_i) \log(1-p_i) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \log\left(\frac{p_i}{1-p_i} \right) + \log(1-p_i) \right]$$

$$= \sum_{i=1}^{n} \left[y_i (\alpha + \beta x_i) + \log\left(\frac{1}{1+e^{\alpha+\beta x_i}} \right) \right]. \tag{1.2.10}$$

To obtain a MLE $\hat{\theta} = (\hat{\alpha}, \hat{\beta})^T$ of $\theta = (\alpha, \beta)^T$, the following equation should be solved.

$$\frac{\partial \lambda(\theta \mid x, y)}{\partial \theta} = \begin{pmatrix} \frac{\partial \lambda(\theta \mid x, y)}{\partial \alpha} \\ \frac{\partial \lambda(\theta \mid x, y)}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \left(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \right) \\ \sum_{i=1}^{n} \left[x_i \left(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \right) \right] \end{pmatrix} = \mathbf{0}, \tag{1.2.11}$$

$$\text{i.e., } \sum_{i=1}^{n}\left(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \right) = 0 \text{ and } \sum_{i=1}^{n}\left[x_i\left(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \right) \right] = 0 . \quad (1.2.12)$$

However, since Equation (1.2.12) cannot be solved explicitly, it is solved numerically using the Newton-Raphson method.

Newton-Raphson Method

The Newton-Raphson method is a general technique for finding roots of the equation $f(x) = 0$ in an iterative manner (McCulloch & Searle, 2001). This algorithm was described by Jennrich and Schluchter (1986), Lindstorm and Bates (1988), and Press et al. (1996) in detail.

Given some initial point x_0, an updated value is obtained by solving

$$f(x) = 0 \cong f(x_0) + (x - x_0) f'(x_0)$$

for x; or

$$x \cong x_0 - \frac{f(x_0)}{f'(x_0)} . \quad (1.2.13)$$

This method suggests one approach for obtaining an iterative solution of the MLE. Applying Equation (1.2.13) to solve Equation (1.2.12), $f = \partial\lambda(\theta \mid x, y)/\partial\theta$ and $f' = \partial\lambda^2(\theta \mid x, y)/\partial^2\theta$. The matrix of second derivative of the log likelihood, the so-called Hessian matrix, $\mathbf{H}$, is

$$\mathbf{H} = \frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \theta} = \frac{\partial \lambda^2(\theta \mid x, y)}{\partial \alpha \partial \beta} = \begin{pmatrix} \frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \alpha} & \frac{\partial \lambda^2(\theta \mid x, y)}{\partial \alpha \partial \beta} \\ \frac{\partial \lambda^2(\theta \mid x, y)}{\partial \beta \partial \alpha} & \frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \beta} \end{pmatrix}. \qquad (1.2.14)$$

Here, the components of the matrix $\mathbf{H}$ are as follows:

$$\frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \alpha} = -\sum_{i=1}^{n} \frac{e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} p_i(1-p_i),$$

$$\frac{\partial \lambda^2(\theta \mid x, y)}{\partial \alpha \partial \beta} = \frac{\partial \lambda^2(\theta \mid x, y)}{\partial \beta \partial \alpha} = -\sum_{i=1}^{n} x_i \frac{e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} x_i p_i(1-p_i),$$

and $\frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \beta} = -\sum_{i=1}^{n} x_i^{\,2} \frac{e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} x_i^{\,2} p_i(1-p_i)$.

Therefore, the matrix $\mathbf{H}$ can be written as

$$\mathbf{H} = \frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \theta} = \begin{pmatrix} -\sum_{i=1}^{n} p_i(1-p_i) & -\sum_{i=1}^{n} x_i p_i(1-p_i) \\ -\sum_{i=1}^{n} x_i p_i(1-p_i) & -\sum_{i=1}^{n} x_i^{\,2} p_i(1-p_i) \end{pmatrix} = -\mathbf{X^T V X}, \qquad (1.2.15)$$

where $\mathbf{X}$ is the model matrix with $\mathbf{x_i}^T$ as its i^{th} row and $\mathbf{V}$ is a diagonal matrix with diagonal entries $p_i(1-p_i)$. Similarly, Equation (1.2.11) can be rewritten by

$$\frac{\partial \lambda(\theta \mid x, y)}{\partial \theta} = \mathbf{X^T}(\mathbf{y} - \mathbf{p}), \qquad (1.2.16)$$

where $\mathbf{y} = (y_1, \Lambda\ , y_n)^T$ and $\mathbf{p} = (p_1, \Lambda\ , p_n)^T$.

Using the Newton-Raphson method, given a current estimate $\hat{\theta}^{(k)}$ an updated estimate $\hat{\theta}^{(k+1)}$ can be obtained as follows:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \left(\frac{\partial \lambda^2(\theta \mid x, y)}{\partial^2 \theta}\bigg|_{\theta=\hat{\theta}^{(k)}} \right)^{-1} \left(\frac{\partial \lambda(\theta \mid x, y)}{\partial \theta}\bigg|_{\theta=\hat{\theta}^{(k)}} \right)$$

$$= \hat{\theta}^{(k)} + (\mathbf{X^T V}^{(k)}\mathbf{X})^{-1}\mathbf{X^T}(\mathbf{y} - \mathbf{p}^{(k)}). \qquad (1.2.17)$$

In the above Equation (1.2.17), $\mathbf{p}^{(k)}$ is the vector of fitted probabilities from the k^{th} iteration with the i^{th} entry of which is

$$p_{i,k} = \frac{e^{\mathbf{x_i}^T \hat{\theta}^{(k)}}}{1 + e^{\mathbf{x_i}^T \hat{\theta}^{(k)}}},$$

and $\mathbf{V}^{(k)}$ is a diagonal matrix with diagonal entries $p_{i,k}(1 - p_{i,k})$.

The iteration continues until $|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|$ is sufficiently small to indicate convergence. Under reasonable assumptions concerning the likelihood function and a sufficiently accurate starting value $\hat{\theta}^{(0)}$, the sequence of iterated estimates $\{\hat{\theta}^{(k)}\}$ produced by the Newton-Raphson method result in quadratic convergence to a solution of Equation (1.2.11). If the log likelihood function is concave and unimodal, then the sequence of values $\{\hat{\theta}^{(k)}\}$ converge to the MLE of θ.

1.2.4 The Expectation Maximization (EM) Algorithm

The Expectation Maximization (EM) algorithm is a widely applicable approach for the iterative computation of maximum likelihood estimates when the calculations via the Newton-

Raphson method do not converge to a global maximum. The formulation of the EM algorithm in its present generality was given by Dempster, Laird, and Rubin in 1977. In particular, this algorithm has become a popular tool in statistical estimation involving incomplete data or for problems which can be posed as a similar form, such as mixture models, since in these cases the likelihood functions are generally intractable. If some latent variables or hidden variables are included, the data is regarded as being incomplete since the values of the hidden variables are unknown.

The main idea of the EM algorithm is to consider the original data as being incomplete and to add some latent or hidden variable since the complete data has a much simpler likelihood function for the purpose of finding a maximum. Then we can maximize the likelihood for the incomplete data through maximizing the expected log likelihood for the complete data. The expectation is taken over all possible values of the latent or hidden variable.

Each iteration of the EM algorithm consists of two steps – the E (expectation) step and the M (maximization) step. First, one initializes all the parameters randomly or heuristically according to any *prior* knowledge about the optimal parameter value. Then, the updated estimates are iteratively obtained by repeating the E step and the M step. In the E step, one computes the expected log likelihood for the complete data. The expectation is taken with respect to the computed conditional distribution of the latent or hidden variables given the current settings of the parameters and the observed data. In the M step, all the parameters are re-estimated by maximizing the expectation of the complete log likelihood. Once a set of parameter values is generated from the starting values of the parameters, the algorithm repeats the E step and M step to obtain the next updated estimates of the parameters. This process continues until

the value of the likelihood converges, i.e., reaching a global maximum. The derivation and the application of this EM algorithm in the presence of mixture are discussed in Chapter 2 in detail.

The Convergence of the EM algorithm

Dempster, Laird, and Rubin (1977) established fundamental properties of the EM algorithm. In particular these properties imply that typically in practice the sequence of EM estimates will converge to a local maximum of the log likelihood function. In general, if the log likelihood has several maxima, the convergence depends on the choice of starting point.

Wu (1983) demonstrated the properties of the convergence of the EM algorithm in detail. Wu mentioned the problem that the convergence of the likelihood does not automatically imply the convergence of the updated parameter. On this same concern, Boyles (1983) gives an example of a generalized EM algorithm that converges to the circle of the unit radius and not to a single point. Lansky, Casella, McCulloch, and Lansky (1992) establish some invariance, convergence, and rates of convergence results. The convergence properties of the EM algorithm are discussed in detail by McLachlan and Krishnan (1996).

1.2.5 The Likelihood Ratio Test and Bootstrap methods

Likelihood Ratio Test

The likelihood ratio test (LRT) proposed by Neyman and Pearson (1928) is a general statistical method for making a decision between two hypotheses. To construct the LRT, recall the likelihood function (1.2.5) as follows:

$$L(\theta \mid x) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

The LRT statistic for testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_0{}^c (= \Theta_1)$ is,

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta \mid x)}{\sup_{\Theta_1} L(\theta \mid x)}. \tag{1.2.18}$$

This statistic is related to the MLE. The numerator and denominator of the above Equation (1.2.18) can be calculated by finding the MLE of θ under the null and the alternative hypothesis, respectively, and then by substituting the MLE back into the corresponding likelihood functions.

The LRT is any test that has a rejection region of the form $\{x : \lambda(x) \le k\}$, where k is any number satisfying $0 \le k \le 1$. That is, the LRT compares the plausibility of the θ values in the null hypothesis with that in the alternative. Small values of the LRT statistics are interpreted as being evidence against the null hypothesis. Hence, it leads to reject the null hypothesis. To define a level α test, the constant k must be chosen so that

$$\sup_{\theta \in \Theta_0} P_\theta(\lambda(x) \le k) \le \alpha. \tag{1.2.19}$$

The Neyman-Pearson lemma (Neyman & Pearson, 1933) demonstrates that the LRT is most efficient in the sense that it minimizes the probability of type II error rate among all level α tests that have the same significance level α.

If the distribution of the LRT statistics corresponding to the null and alternative hypothesis can be explicitly determined, then decision regions can be directly obtained from the distribution. In most cases, however, since the exact distribution is unknown, it is difficult to

determine decision regions exactly. Hence, to obtain the asymptotic decision regions, the following asymptotic distribution of the LRT (Cox and Hinkley, 1974) can be used. They state the following:

Under the regularity conditions, if $\theta \in \Theta_0$, *the distribution of the statistic* $-2\log\lambda(x)$ *converges to a chi squared distribution as the sample size* $n \to \infty$, *i.e.,*

$$-2\log\lambda(x) \xrightarrow{d} \chi^2_{df} \quad as \; n \to \infty. \tag{1.2.20}$$

Here, the degrees of freedom of the chi squared distribution, df equal to the difference between the number of free parameters in Θ_0 and the number of free parameters in Θ_1. Rejection of the null hypothesis for small values of $\lambda(x)$ is equivalent to rejection for large values of $-2\log\lambda(x)$. Therefore, H_0 can be rejected if $-2\log\lambda(x) \geq \chi^2_{df,\alpha}$.

However, these conditions do not hold in the case where we test against mixture alternatives. Since there is generally a relationship between parameters under mixture alternative hypothesis, the asymptotic chi-square distribution cannot be directly used. It has been proposed that under these nonstandard conditions the null distribution of LRT statistics is a mixture of central chi-squared distributions. Important contributions to the understanding of this asymptotic behavior of the LRT statistics in this situation have been made by, for example, Self and Liang (1987) and Stram and Lee (1994). Self and Liang (1987) found the asymptotic distribution of LRT statistics using a projection of a normal variable onto a tangent cone of the parameter space. They considered the special case where one parameter is specified under the null hypothesis and it falls on the boundary. No other parameters are on the boundary. For this case they derived the

null distribution to be a 50:50 mixture of χ_0^2 and χ_1^2 distribution (Case 5 in Self and Liang, 1987).

Stram and Lee (1994) showed that the asymptotic null distribution of LRT statistics for testing when two parameters are on the boundary is a 50:50 mixture of a χ_1^2 and a χ_2^2 (Case 2 in Stram and Lee, 1994). By extension, in their Case 3 they proposed the asymptotic distribution of the statistics is a 50:50 mixture of χ_{d-1}^2 and χ_d^2 with *d* the number of parameters added by their mixture alternative. Our case is more like Stram and Lee's case (1994) rather than Self and Liang's (1987).

Bootstrap Methods

The bootstrap is a re-sampling technique for estimating the precision of a parameter estimate. This method was invented by Bradley Efron (1979) and further developed by Efron and Tibshirani (1993). The basic idea of this method is that the original sample represents the population from which it was drawn, so the replicated samples redrawn from this original sample represent what would get if we took many samples from the population. These replicated samples are generally called bootstrap samples in this procedure. For each bootstrap sample a statistic of interest is generated. The distribution of the statistic, based on many bootstrap samples, represents the sampling distribution of the statistic, based on many samples from the population. Particularly, it provides an alternative to large sample techniques when asymptotic properties are not met or when the standard error of the estimate has complicated mathematical characteristics. The power of the bootstrap lies in the fact that the method applies to almost any estimator, no matter how complicated. Also, in practice, it is a computer-intensive method for approximating the sampling distribution of any statistic derived from a random sample.

Therefore, the only requirement is a computer program to calculate the estimator from a sample and a method to redraw samples.

In this dissertation, we used this bootstrap procedure to investigate the empirical null distribution of LRT statistics in Chapter 3. On the basis of many bootstrap samples from this procedure the LRT statistics were generated and the empirical distribution of the statistics was constructed. This method will be described in more detail in Section 3.2.

1.3 Outline of the Dissertation

1.3.1 The Problem

This dissertation considers a situation where we are testing for an association between a quantitative explanatory variable X and a dichotomous response variable Y. Our concern is to compare the power of the test based on an ordinary logistic regression model with the test based on a logistic regression mixture model. Data are drawn from a two component logistic regression mixture model which has equal intercepts and unequal slopes. Specifically, we consider the case where one of the slopes equals zero. That is, the population consists of a fraction π for which the response variable Y depends on X and a fraction $1 - \pi$ where it is independent. The logistic regression mixture model is as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \qquad \text{with probability } \pi \tag{1.3.1}$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 \qquad \text{with probability } 1 - \pi\,. \tag{1.3.2}$$

Here, *p* indicates the conditional probability that the binary response variable Y has the value 1 in a sample with X = x and π denotes a mixing proportion with a value lying between 0 and 1.

We are interested in fitting the logistic regression mixture model to generated data sets and two main goals of this research are as follows:

(1) To determine the power of detecting the relationship between Y and X upon estimating all of the parameters in this mixture model (Equation (1.3.1) and (1.3.2)).

(2) To compare this power to the power one obtains using ordinary logistic regression (which implicitly assumes $\pi = 1$).

Two different alternatives are considered according to the corresponding fitted models. The first is based on the assumption that the data fits the ordinary logistic regression model (H_a^h) and the second is based on the logistic regression mixture model (H_a^m) as follows:

Alternative Ⅰ – The ordinary logistic regression model.

$$H_a^h : \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \tag{1.3.3}$$

Alternative Ⅱ – The logistic regression mixture model.

$$H_a^m : \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \quad \text{with probability } \pi \tag{1.3.4}$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 \quad \text{with probability } 1-\pi \tag{1.3.5}$$

The common null hypothesis considered then is,

$$H_0 : \log\left(\frac{p}{1-p}\right) = \beta_0. \tag{1.3.6}$$

We used the above notation for the alternatives to indicate each alternative is based on the homogeneous (*h*) and mixture (*m*) population, respectively.

Additionally, we will consider the relative precision of these two methods by comparing the bias of the estimates of the regression parameters and the mean squared error.

1.3.2 Data Generation

To generate the mixture data sets used in this dissertation we consider the case where we have data on the four different values of the quantitative explanatory variable X: 0, 1, 2, and 3, in equal proportion (25%). For example, when the total number of observations in the sample is one hundred (n = 100) the number of observations per X value is twenty five ($n_x = 25$). That is, for the sample size of n, the number of observations per X value is $n / 4$.

Based on this data structure and given the parameters setting, first we generated random numbers of size n from uniform distribution $U(0, 1)$. According to the value of the random number compared with the given mixing proportion, each observation was assigned to the first component (Equation 1.3.1) or the second component (Equation 1.3.2) in the logistic regression mixture model defined in Section 2.1. Then, the probability p_1 and p_2 were calculated for each component. The probability $Pr(Y = 1)$ for each combination of the parameter β_0 and β_1 simulated can be found in Appendix A.

Finally, another set of random numbers of size n was generated to assign the value of the binary response variable Y (0 or 1) for each observation. This data generation was done using a C++ program with some functions in the GNU Scientific Library (GSL) 1.14. The MT19937

generator of Makoto Matsumoto and Takuji Nishimura (described in the GSL reference manual) was used as a random number generator.

Chapter 2

Likelihood Ratio Test in the Presence of Mixture

2.1 The Likelihood Ratio Test (LRT)

To test the hypotheses described in Chapter 1 we usually conduct the Likelihood Ratio Test (LRT). Based on our models, the likelihood function L_h under the alternative hypothesis H_a^h for the ordinary logistic regression model is

$$L_h = \prod_{i=1}^{n} p(y_i, x_i; \beta_0, \beta_1)$$

$$= \prod_{i=1}^{n} \frac{\{\exp(\beta_0 + \beta_1 x_i)\}^{y_i}}{1 + \exp(\beta_0 + \beta_1 x_i)}, \qquad (2.1.1)$$

and the likelihood function L_m under the alternative hypothesis H_a^m for the logistic regression mixture model is as follows:

$$L_m = \prod_{i=1}^{n} \left[\pi \cdot p(y_i, x_i; \beta_0, \beta_1) + (1 - \pi) \cdot p(y_i, x_i; \beta_0)\right]$$

$$= \prod_{i=1}^{n} \left[\pi \frac{\{\exp(\beta_0 + \beta_1 x_i)\}^{y_i}}{1 + \exp(\beta_0 + \beta_1 x_i)} + (1 - \pi) \frac{\{\exp(\beta_0)\}^{y_i}}{1 + \exp(\beta_0)}\right]. \qquad (2.1.2)$$

Also, the likelihood function L_0 under the common null hypothesis H_0 is:

$$L_0 = \prod_{i=1}^{n} p(y_i, x_i; \beta_0)$$

$$= \prod_{i=1}^{n} \frac{\{\exp(\beta_0)\}^{y_i}}{1 + \exp(\beta_0)}. \qquad (2.1.3)$$

Through these, the test statistic, $G^2 = -2 \ln \Lambda$ is then calculated. Here, Λ is the ratio of the maximum value of the likelihood function under the null hypothesis (Equation (2.1.3)) and the maximum value of the likelihood function under the alternative hypothesis (Equation (2.1.1) or Equation (2.1.2)) being considered. That is, $\Lambda = \frac{L_{0,\max}}{L_{h,\max}}$ for the ordinary logistic regression model, and $\Lambda = \frac{L_{0,\max}}{L_{m,\max}}$ for the logistic regression mixture model. Therefore, the LRT statistics are computed by using the maximum likelihood estimates to obtain the maximum value of the corresponding likelihood function. In the presence of mixture the maximum likelihood estimates obtained using the Expectation Maximization (EM) algorithm are used.

2.2 Maximum Likelihood Estimation

2.2.1 The Expectation Maximization (EM) Algorithm

Suppose that $\mathbf{X} = \{x_1, \Lambda, x_n\}$ is a sample data set consisting of n independent observations and $p(x_i; \theta)$ is the probability density function. To obtain the Maximum

Likelihood Estimates (MLE) $\hat{\theta}$ of the parameter values, θ, maximizing the log likelihood is needed. However, if some latent variables exist, explicitly finding the MLE is not easy. Let $\mathbf{Z} = \{z_1, \Lambda\ , z_n\}$ denote the latent variable that indicates which component generates the corresponding observations and let Q be some distribution over the latent variable $\mathbf{Z}$. Then the log likelihood for the original incomplete data is rewritten as follows:

$$\lambda(\theta) = \log p(\mathbf{X};\theta) = \log \int p(\mathbf{X}, \mathbf{Z};\theta) d\mathbf{Z}$$

$$= \log \int Q(\mathbf{Z}) \frac{p(\mathbf{X}, \mathbf{Z};\theta)}{\mathbf{Q}(\mathbf{Z})} d\mathbf{Z}$$

$$\geq \int Q(\mathbf{Z}) \log \frac{p(\mathbf{X}, \mathbf{Z};\theta)}{\mathbf{Q}(\mathbf{Z})} d\mathbf{Z} \qquad (2.2.1)$$

$$= \int Q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z};\theta) d\mathbf{Z} - \int Q(\mathbf{Z}) \log Q(\mathbf{Z}) d\mathbf{Z}\ . \qquad (2.2.2)$$

Here, Equation (2.2.1) can be obtained by using Jensen's inequality. From the above derivation, Equation (2.2.2) is the lower-bound of the value of the log likelihood for the complete data. Since $Q(\mathbf{Z})$ is an arbitrary distribution, it is independent of θ. Therefore, in order to maximize the lower-bound with respect to θ, it suffices to simply maximize the first term of Equation (2.2.2), i.e.,

$$\int Q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z};\theta) d\mathbf{Z} = E_Q[\log p(\mathbf{X}, \mathbf{Z};\theta)]\ . \qquad (2.2.3)$$

This expected log likelihood for the complete data is computed in the E step. Then we need to maximize the expected log likelihood for the complete data, where the expectation is taken with respect to $Q(\mathbf{Z})$. This is the M step of the EM algorithm.

On the other hand, for the computation of the expected complete log likelihood, $Q(\mathbf{Z})$ should be chosen. If we set $Q(\mathbf{Z}) = p(\mathbf{Z} \mid \mathbf{X}; \theta)$ in Equation (2.2.1) to compute the expected log likelihood for the complete data, then the value of the lower-bound becomes the log likelihood for the incomplete data as follows:

$$\int Q(\mathbf{Z}) \log \frac{p(\mathbf{X}, \mathbf{Z}; \theta)}{\mathbf{Q}(\mathbf{Z})} d\mathbf{Z} = \int p(\mathbf{Z} \mid \mathbf{X}; \theta) \log \frac{p(\mathbf{X}, \mathbf{Z}; \theta)}{p(\mathbf{Z} \mid \mathbf{X}; \theta)} d\mathbf{Z}$$

$$= \int p(\mathbf{Z} \mid \mathbf{X}; \theta) \log \frac{p(\mathbf{Z} \mid \mathbf{X}; \theta) p(\mathbf{X}; \theta)}{p(\mathbf{Z} \mid \mathbf{X}; \theta)} d\mathbf{Z}$$

$$= \int p(\mathbf{Z} \mid \mathbf{X}; \theta) \log p(\mathbf{X}; \theta) d\mathbf{Z}$$

$$= \log p(\mathbf{X}; \theta) \int p(\mathbf{Z} \mid \mathbf{X}; \theta) d\mathbf{Z}$$

$$= \log p(\mathbf{X}; \theta) .$$

Hence, when computing the expected complete log likelihood (2.2.3), the expectation should be taken with respect to the conditional distribution of the latent variable $\mathbf{Z}$ given the observed data $\mathbf{X}$, i.e.,

$$Q(\mathbf{Z}) = p(\mathbf{Z} \mid \mathbf{X}; \theta) . \qquad (2.2.4)$$

2.2.2 The MLE based on the EM Algorithm

The maximum likelihood estimates were calculated under the alternative hypothesis in the logistic regression mixture model defined earlier in Section 1.3. These maximum likelihood estimates were in turn used to obtain the corresponding LRT statistic. The procedure used for the

customized EM algorithm is given in this section in the context of the logistic regression mixture model.

- **E step**

Suppose that the latent variable $\mathbf{Z}$ in the logistic regression mixture model consists of n two-dimensional vectors, i.e., $z_i = (z_{i1}, z_{i2})$ for $i = 1, \ldots, n$. Here, each element of the vectors is 1 or 0 to indicate the corresponding component that the i^{th} observation comes from. That is, $z_i = (1, 0)$ means the i^{th} observation y_i is from the first component. In the same manner, $z_i = (0, 1)$ indicates that y_i is from the second component. Also, let $\hat{\theta}^{(0)} = (\hat{\beta}_0^{(0)}, \hat{\beta}_1^{(0)}, \hat{\pi}^{(0)})$ be the starting values of the parameters of the logistic regression mixture model. Then, with these starting values the log likelihood for the complete data, including the latent variable $\mathbf{Z}$, can be computed as follows:

$$\lambda_c(\hat{\theta}^{(0)}) = \log p(\mathbf{X}, \mathbf{Y}, \mathbf{Z}; \hat{\theta}^{(0)}) = \log \prod_{i=1}^{n} p(x_i, y_i, z_i; \hat{\theta}^{(0)})$$

$$= \log \prod_{i=1}^{n} \prod_{m=1}^{2} [p(y_i \mid x_i, z_{im} = 1; \hat{\theta}^{(0)}) p(z_{im} = 1)]^{z_{im}}$$

$$= \sum_{i=1}^{n} \sum_{m=1}^{2} [z_{im} \log p(y_i \mid x_i, z_{im} = 1; \hat{\theta}^{(0)}) + z_{im} \log \hat{\pi}_m^{(0)}]. \qquad (2.2.5)$$

Thus, the expected complete log likelihood is given by

$$E_Q[\lambda_c(\hat{\theta}^{(0)})] = \sum_{i=1}^{n} \sum_{m=1}^{2} [E_Q(z_{im})^{(0)} \log p(y_i \mid x_i, z_{im} = 1; \hat{\theta}^{(0)}) + E_Q(z_{im})^{(0)} \log \hat{\pi}_m^{(0)}] \qquad (2.2.6)$$

where $Q(\mathbf{Z}) = p(\mathbf{Z} \mid \mathbf{X}, \mathbf{Y}; \theta)$ as derived in Equation (2.2.4).

In Equation (2.2.6) the starting value of the expectation $E_Q(z_{im})^{(0)}$ in the first iteration is computed in the following

$$
\begin{aligned}
E_Q(z_{im})^{(0)} &= p(z_{im}=1 \mid x_i, y_i; \hat{\theta}^{(0)}) \\
&= \frac{p(y_i \mid x_i, z_{im}=1; \hat{\theta}^{(0)})\, p(z_{im}=1)}{\sum_{k=1}^{2} p(y_i \mid x_i, z_{ik}=1; \hat{\theta}^{(0)})\, p(z_{ik}=1)} \\
&= \frac{p(y_i \mid x_i, z_{im}=1; \hat{\theta}^{(0)})\, \hat{\pi}_m{}^{(0)}}{\sum_{k=1}^{2} p(y_i \mid x_i, z_{ik}=1; \hat{\theta}^{(0)})\, \hat{\pi}_k{}^{(0)}}
\end{aligned}
\tag{2.2.7}
$$

for each observation and component ($i = 1, \ldots, n$ and $m = 1, 2$). This is the E step in the EM algorithm.

- **M step**

The M step maximizes the expected complete log likelihood which was defined in Equation (2.2.6) with respect to the parameters that are to be estimated, i.e., $\theta = (\beta_0, \beta_1, \pi)$ to obtain the updated estimates of the parameters. The updated estimate of β_1 can be computed by solving the equation, $\frac{\partial}{\partial \beta_1} E_Q[\lambda_c(\hat{\theta}^{(0)})] = 0$. This equation can be rewritten as

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} E_Q[\lambda_c(\hat{\theta}^{(0)})] &= \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} \sum_{m=1}^{2} \left[E_Q(z_{im})^{(0)} \log\, p(y_i \mid x_i, z_{im}=1; \hat{\theta}^{(0)}) + E_Q(z_{im})^{(0)} \log\, \hat{\pi}_m{}^{(0)} \right] \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} \sum_{m=1}^{2} \left[E_Q(z_{im})^{(0)} \log\, p(y_i \mid x_i, z_{im}=1; \hat{\theta}^{(0)}) + E_Q(z_{im})^{(0)} \log\, \hat{\pi}_m{}^{(0)} \right]
\end{aligned}
$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} \Big[E_Q(z_{i1})^{(0)} \log\, p(y_i \mid x_i, z_{i1} = 1; \hat{\theta}^{(0)}) + E_Q(z_{i1})^{(0)} \log\, \hat{\pi}_1{}^{(0)}$$

$$+ E_Q(z_{i2})^{(0)} \log\, p(y_i \mid x_i, z_{i2} = 1; \hat{\theta}^{(0)}) + E_Q(z_{i2})^{(0)} \log\, \hat{\pi}_2{}^{(0)} \Big]$$

$$= \sum_{i=1}^{n} \left[E_Q(z_{i1})^{(0)} \frac{\partial}{\partial \beta_1} \log\, p(y_i \mid x_i, z_{i1} = 1; \hat{\theta}^{(0)}) \right] \tag{2.2.8}$$

$$= \sum_{i=1}^{n} \left[E_Q(z_{i1})^{(0)} x_i \left(y_i - \frac{e^{\hat{\beta}_0{}^{(0)} + \hat{\beta}_1{}^{(0)} x_i}}{1 + e^{\hat{\beta}_0{}^{(0)} + \hat{\beta}_1{}^{(0)} x_i}} \right) \right] = 0 . \tag{2.2.9}$$

In Equation (2.2.8), $\log\, p(y_i \mid x_i, z_{i1} = 1)$ can be considered as in the case of the ordinary logistic regression, since the information on the component $z_{i1} = 1$ is already given. That is,

$$\log\, p(y_i \mid x_i, z_{i1} = 1) = \frac{(e^{\beta_0 + \beta_1 x_i})^{y_i}}{1 + e^{\beta_0 + \beta_1 x_i}}. \tag{2.2.10}$$

However, since the above Equation (2.2.9) does not have an explicit solution, it should be solved iteratively using the Newton-Raphson method. In other words, to obtain the MLE of β_1 the Newton-Raphson method is used within each M step in the EM algorithm. The updated estimate is computed as:

$$\hat{\beta}_1^{(1)} = \hat{\beta}_1^{(0)(k+1)}$$

$$= \hat{\beta}_1^{(0)(k)} - \left(\frac{\partial^2 E_Q[\lambda_c(\hat{\theta}^{(0)})]}{\partial \beta_1{}^2} \Big|_{\hat{\beta}_1^{(0)} = \hat{\beta}_1^{(0)(k)}} \right)^{-1} \left(\frac{\partial E_Q[\lambda_c(\hat{\theta}^{(0)})]}{\partial \beta_1} \Big|_{\hat{\beta}_1^{(0)} = \hat{\beta}_1^{(0)(k)}} \right) \tag{2.2.11}$$

In addition, the updated estimate of β_0 can be computed by solving the equation $\frac{\partial}{\partial \beta_0} E_Q[\lambda_c(\hat{\theta}^{(0)})] = 0$. In the same way the equation to be solved is given by

$$\frac{\partial}{\partial \beta_0} E_Q[\lambda_c(\hat{\theta}^{(0)})] = \sum_{i=1}^{n} \left[E_Q(z_{im})^{(0)} \frac{\partial}{\partial \beta_0} \log\, p(y_i \mid x_i, z_{im} = 1; \hat{\theta}^{(0)}) \right]$$

$$= \sum_{i=1}^{n} \left[E_Q(z_{im})^{(0)} \left(y_i - \frac{e^{\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_i}}{1 + e^{\hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} x_i}} \right) \right] = 0. \qquad (2.2.12)$$

Similarly, the Newton-Raphson method can be used within each M step since Equation (2.2.12) also cannot be solved explicitly. That is, the updated estimate of β_0 can be calculated as:

$$\hat{\beta}_0^{(1)} = \hat{\beta}_0^{(0)(k+1)}$$

$$= \hat{\beta}_0^{(0)(k)} - \left(\frac{\partial^2 E_Q[\lambda_c(\hat{\theta}^{(0)})]}{\partial \beta_0^{\,2}} \bigg|_{\hat{\beta}_0^{(0)} = \hat{\beta}_0^{(0)(k)}} \right)^{-1} \left(\frac{\partial E_Q[\lambda_c(\hat{\theta}^{(0)})]}{\partial \beta_0} \bigg|_{\hat{\beta}_0^{(0)} = \hat{\beta}_0^{(0)(k)}} \right). \qquad (2.2.13)$$

Finally, to maximize the expected complete log likelihood with respect to π_m, the Lagrange multiplier λ can be used since there is the constraint that $\sum_{m=1}^{2} \pi_m = 1$. Therefore, the following equation can be considered:

$$G(\hat{\theta}^{(0)}) = E_Q[\lambda_c(\hat{\theta}^{(0)})] - \lambda(\sum_{m=1}^{2} \pi_m - 1). \qquad (2.2.14)$$

Then, by differentiating the above Equation (2.2.14), the equation to be solved is given by

$$\frac{\partial}{\partial \pi_m} G(\hat{\theta}^{(0)}) = \frac{\partial}{\partial \pi_m} E_Q[\lambda_c(\hat{\theta}^{(0)})] - \lambda = \sum_{i=1}^{n} \frac{E_Q(z_{im})^{(0)}}{\pi_m} - \lambda = 0. \quad (2.2.15)$$

The above Equation (2.2.15) is equivalent to $\sum_{i=1}^{n} E_Q(z_{im})^{(0)} - \lambda \pi_m = 0$. Summing this equation over all *m*, the following result is obtained:

$$\sum_{m=1}^{2} \sum_{i=1}^{n} E_Q(z_{im})^{(0)} - \lambda \sum_{m=1}^{2} \pi_m = n - \lambda = 0 \quad (2.2.16)$$

Hence, from Equation (2.2.15) and (2.2.16) the updated estimate of π_m is

$$\hat{\pi}_m{}^{(1)} = \frac{\sum_{i=1}^{n} E_Q(z_{im})^{(0)}}{n}. \quad (2.2.17)$$

The stopping criteria that we choose for the Newton-Raphson method within the M step of the EM algorithm is based on the relative change of the parameter values in consecutive iterations, $| \hat{\beta}_0{}^{(t)} - \hat{\beta}_0{}^{(t-1)} | < 10^{-5}$ and $| \hat{\beta}_1{}^{(t)} - \hat{\beta}_1{}^{(t-1)} | < 10^{-5}$. As for the EM algorithm, once the updated estimates of the parameters $\hat{\theta}^{(1)}$ are computed starting with the starting values of the parameters $\hat{\theta}^{(0)}$, the new updated estimates $\hat{\theta}^{(2)}$ can be obtained by the same procedure with the previous updated estimates $\hat{\theta}^{(1)}$. This process continues until the log likelihood converges. However, since there is no guarantee of a unique stationary pair $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\pi})$ to maximize the log likelihood, several random starting points are needed to find a global maximum value of the likelihood function. The method to select the starting values and the number of starting points for the EM algorithm are described in the following sections.

2.2.3 Selection of Starting Values for the EM Algorithm

The main drawbacks of the EM algorithm are its slow convergence and the dependence on the choice of starting values for the unknown parameters used. The choice of starting values is important in the EM algorithm since it affects the speed of convergence and the ability to find the global maximum. The method of selecting the starting values has been dealt with in various studies. We are interested in choosing starting values ($\hat{\pi}^{(0)}$, $\hat{\beta}_0^{(0)}$, $\hat{\beta}_1^{(0)}$) under the alternative hypothesis H_a^m in the context of the logistic regression mixture model. Recall the alternative hypothesis is as follows:

Alternative Ⅱ - The logistic regression mixture model.

$$H_a^m : \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \qquad \text{with probability } \pi$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 \qquad \text{with probability } 1-\pi$$

Two different methods were considered to select random starting values of the parameters. The two ways are both based on fitting the ordinary logistic regression model. Suppose that the fitted estimates in the context of the ordinary logistic regression model are $\tilde{\beta}_0$ and $\tilde{\beta}_1$. Method (1) involves using only the fitted estimate $\tilde{\beta}_1$ for a starting value of β_1. Then, a starting value of the common intercept β_0 is selected using the observed conditional probability that the value of Y equals one given X = 0 in a sample. In Method (2), the fitted estimates $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are used for selecting starting values of β_0 and β_1, respectively. The procedure of the first method is described as follows:

Method (1)

· Step 1. Choose a starting value for the mixing proportion, $\hat{\pi}^{(0)}$:

To find an optimal starting value for the mixing proportion $\hat{\pi}^{(0)}$, a uniform (0,1) random number is generated.

· Step 2. Calculate a starting value for the common intercept, $\hat{\beta}_0^{(0)}$:

A starting value of the common intercept is selected using the conditional probability that the value of Y equals 1 given X=0 in the observed sample since the following equation holds when the value of X is 0:

$$\log\left(\frac{p_{Y=1|X=0}}{1-p_{Y=1|X=0}}\right) = \hat{\beta}_0 + \hat{\beta}_1 \cdot 0 = \hat{\beta}_0^{(0)} \qquad (2.2.18)$$

· Step 3. Select a starting value for the nonzero slope, $\hat{\beta}_1^{(0)}$:

We fit an ordinary logistic regression model to the overall data set $\{(x_1, y_1), \Lambda\ , (x_n, y_n)\}$. Let $\tilde{\beta}_1$ be the estimate of the slope in the ordinary logistic regression model. According to the definition of the logistic regression mixture model, the starting estimate $\hat{\beta}_1^{(0)}$ for the nonzero slope β_1 can be obtained by using the starting value for the mixing proportion $\hat{\pi}^{(0)}$. The value of the nonzero slope β_1 is affected by the corresponding mixing proportion starting from the estimate of the slope $\tilde{\beta}_1$ in the ordinary logistic regression model. The estimate of the slope $\tilde{\beta}_1$ is represented in the following way:

$$\tilde{\beta}_1 = \pi \cdot \beta_1 + (1 - \pi) \cdot 0 = \pi\beta_1 . \quad (2.2.19)$$

From the above Equation (2.2.19), we expect that the value of the nonzero slope is calculated as

$$\beta_1 = \frac{\tilde{\beta}_1}{\hat{\pi}^{(0)}} . \quad (2.2.20)$$

However, the above value obtained by Equation (2.2.20) can be unrealistically large whenever the starting value of the mixing proportion $\hat{\pi}^{(0)}$ is close to zero. In order to avoid this we adjusted the value by taking minimum with a certain number. In this dissertation, we take the starting value for the nonzero slope as follows:

$$\hat{\beta}_1^{(0)} = \min(10, \frac{\tilde{\beta}_1}{\hat{\pi}^{(0)}}) . \quad (2.2.21)$$

Method (2)

· Step 1. Choose a starting value for the mixing proportion, $\hat{\pi}^{(0)}$ in the same way as Method (1).

· Step 2. Fit an ordinary logistic regression model to the overall data set, and let $\tilde{\beta}_0$ and $\tilde{\beta}_1$ be the fitted estimates in the ordinary logistic regression model.

· Step 3. Select starting values for the common intercept β_0 and the nonzero slope β_1:

$$\hat{\beta}_0^{(0)} = \tilde{\beta}_0 \text{ and } \hat{\beta}_1^{(0)} = \min(10, \frac{\tilde{\beta}_1}{\hat{\pi}^{(0)}}) .$$

When we used Method (1) to select the starting values for the EM algorithm, the algorithm did not reach a global maximum of the value of the likelihood, or it was confronted by a division-by-zero error within the Newton-Raphson method. Meanwhile, the EM algorithm found the global maximum using Method (2) as the way to select the starting values. Therefore, Method (2) is used for selecting the starting values throughout this study.

2.3 Simulated Results of the Estimation

2.3.1 The Number of Random Starting Points for the EM Algorithm

Although iterations of the EM algorithm always lead to non-decreasing values of the likelihood, there is no proof of the uniqueness of a maximum likelihood value. The number of random starting points is important to assure that the observed maximum likelihood is a global one. To specify the number of random starting points required to get a global maximum, the maxima of the LRT statistics are compared at specified numbers of random starting points used in the EM algorithm. Since an observed negative maximum value of the LRT statistics indicates a local maximum we can investigate this condition as well as the convergence by comparing the values of the LRT statistics obtained using different starting values.

We obtain the maximum log likelihood and the corresponding maximum likelihood estimates for each set of initial starting points in this simulation. Then the LRT statistic is computed, and I choose the largest value of the LRT statistics comparing all the values obtained from each set of starting points. If these largest values converge to a certain value, we conclude

that the EM algorithm with the specified number of random starting points has reached the global maximum.

For simplicity, the intercept in the regression model is fixed to zero ($\beta_0 = 0$) and the slope is set to one ($\beta_1 = 1$). Under the alternative, the five mixing proportions π = 0.1, 0.3, 0.5, 0.7, and 0.9 are considered for each sample size n = 100, 200, and 400. Since the stopping criteria in the EM algorithm that I choose to use is based on the relative change of the log likelihood function in consecutive iterations $| \lambda^{(t)} - \lambda^{(t-1)} | < 10^{-5}$, the number of random starting points needed to find the global maximum can be chosen at a certain point that the difference in consecutive maxima of LRT statistics becomes small enough in the same context.

Table 2.1 reports a relationship between the relative change in the values of LRT statistics and the number of random starting points for the case where the data is generated under the null hypothesis (H_0) described in Section 1.3. The relative change in the values of LRT statistics with 45 or more starting points is less than 10^{-4}. Under the mixture alternative hypothesis (H_a^m), this value is less than 10^{-3} after 45 or more starting points. Therefore, 45 random starting points are used in each sample for the EM algorithm with the tolerance 10^{-5}, and we choose the maximum of 45 maxima to obtain the global maximum of the log likelihood functions in this power study.

Table 2.1 – Maximum LRT statistics for selected numbers of Random Starting Points (RSPs) under the null alternative hypothesis in logistic regression mixture models with the difference between maximum LRT statistics (Δ)

Sample Size (n)[1]	Number of RSPs	Under H_0		Under H_a^m									
		mixing prop. = 0.0		mixing prop. = 0.1		0.3		0.5		0.7		0.9	
			Δ		Δ		Δ		Δ		Δ		Δ
100	1	0.1505		0.1326		1.3601		7.2318		10.8034		18.8370	
	10	0.1505	0.0000	0.1326	0.0000	1.3804	0.0203	7.2318	0.0000	10.8034	0.0000	18.8370	0.0000
	20	0.1505	0.0000	0.1331	0.0004	1.3804	0.0000	7.2318	0.0000	10.8034	0.0000	18.8370	0.0000
	30	0.1505	0.0000	0.1331	0.0001	1.3804	0.0000	7.2318	0.0000	10.8034	0.0000	18.8370	0.0000
	45	**0.1511**	**0.0005**	**0.1332**	**0.0001**	**1.3839**	**0.0034**	**7.2318**	**0.0000**	**10.8034**	**0.0000**	**18.8381**	**0.0011**
	60	0.1511	0.0000	0.1333	0.0000	1.3839	0.0000	7.2318	0.0000	10.8034	0.0000	18.8381	0.0000
	100	0.1511	0.0000	0.1333	0.0001	1.3839	0.0000	7.2329	0.0010	10.8034	0.0000	18.8381	0.0000
	150	0.1511	0.0000	0.1333	0.0000	1.3840	0.0001	7.2330	0.0001	10.8039	0.0005	18.8385	0.0004
	200	0.1511	0.0000	0.1333	0.0000	1.3840	0.0000	7.2330	0.0000	10.8039	0.0000	18.8385	0.0000
200	1	0.0040		1.7395		4.1711		13.493		18.5709		37.0914	
	10	0.0040	0.0000	1.7725	0.0330	4.2023	0.0312	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	20	0.0040	0.0000	1.7725	0.0000	4.2023	0.0000	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	30	0.0040	0.0000	1.7725	0.0000	4.2023	0.0000	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	45	**0.0040**	**0.0000**	**1.7725**	**0.0000**	**4.2023**	**0.0000**	**13.493**	**0.0000**	**18.5709**	**0.0000**	**37.0914**	**0.0000**
	60	0.0040	0.0000	1.7725	0.0000	4.2023	0.0000	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	100	0.0040	0.0000	1.7725	0.0000	4.2023	0.0000	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	150	0.0040	0.0000	1.7725	0.0000	4.2023	0.0000	13.493	0.0000	18.5709	0.0000	37.0914	0.0000
	200	0.0040	0.0000	1.7725	0.0001	4.2024	0.0001	13.493	0.0000	18.5721	0.0012	37.0914	0.0000

Note 1. The number of observations per X values $n_x = n / 4$ for each sample size n. 2. The cases with larger sample size ($n > 200$) are not shown because the results are consistent with the above results (45 starting points are needed).

2.3.2 MLE in the Logistic Regression Mixture

In order to verify that the customized EM algorithm works well in the logistic regression mixture model defined in this dissertation, I investigated the means and standard errors of the MLEs obtained by the EM algorithm with 45 random starting points, which are based on 1,000 replicates. We considered the logistic regression mixture model where it has the value of zero as the common intercept ($\beta_0 = 0$) and the value of one as the nonzero slope ($\beta_1 = 1$). Also, four sample sizes n = 100, 200, 400, 1,000, and 2,000 were considered; i.e., the number of observations per X value is n_x = 25, 50, 100, 250, and 500, respectively.

The parameters setting for this simulation study are shown in Table 2.2, and it contains the mean and standard error of the estimated values of the parameters. As one can see, the large sample theorems for the expected values of the MLE only hold for extremely large samples in the case of this likelihood. In a rough way, the EM algorithm seems to work well in estimating the value of β_0 regardless of sample size and true values. Meanwhile, as sample size increases the expected values of the estimates approach true values and the standard error of the estimates decreases, particularly in the estimated results of β_1. From the viewpoint of mixing proportions π, as the true value of π increases, the bias of the estimated value decreases. For the estimate $\hat{\beta}_1$, we can roughly compare the precision based on the expected bias of $\hat{\beta}_1$ in ordinary logistic regression models. We can expect the bias ($\hat{\beta}_1$) in the context of ordinary logistic regression to be $E(\hat{\beta}_1 - \beta) = E(\hat{\beta}_1) - \beta_1 = \pi\beta_1 + (1 - \pi)0 - \beta_1 = (\pi - 1)\beta_1$. The bias of $\hat{\beta}_1$ appears to be consistently smaller than the expected bias obtained by ordinary logistic regression for large samples ($n > 400$).

Table 2.2 – Simulated mean MLEs with the standard error (in parentheses) under the logistic regression mixture population based on 1,000 replicates in each case: four sample sizes n =100, 200, 400, 1,000, and 2,000 are considered and 45 random starting points are used

Sample Size (n)[1]	π	β_0	β_1	***Expected* Bias($\hat{\beta}_1$)[2]**	Mean MLEs (SE)		
					$\hat{\pi}$	$\hat{\beta}_0$	$\hat{\beta}_1$
100	0.1	0	1	**-0.9**	0.54 (0.01)	-0.01 (0.01)	**0.40* (0.06)**
	0.3	0	1	**-0.7**	0.58 (0.01)	-0.03 (0.01)	**0.87 (0.06)**
	0.5	0	1	**-0.5**	0.66 (0.01)	-0.07 (0.01)	**1.26* (0.05)**
	0.7	0	1	**-0.3**	0.76 (0.01)	-0.06 (0.01)	**1.36 (0.04)**
	0.9	0	1	**-0.1**	0.89 (0.00)	-0.06 (0.01)	**1.34 (0.03)**
200	0.1	0	1	**-0.9**	0.54 (0.01)	0.00 (0.01)	**0.40* (0.05)**
	0.3	0	1	**-0.7**	0.52 (0.01)	0.00 (0.01)	**1.12* (0.06)**
	0.5	0	1	**-0.5**	0.61 (0.01)	-0.04 (0.01)	**1.35* (0.05)**
	0.7	0	1	**-0.3**	0.74 (0.01)	-0.05 (0.01)	**1.40 (0.04)**
	0.9	0	1	**-0.1**	0.89 (0.00)	-0.05 (0.01)	**1.29 (0.02)**
400	0.1	0	1	**-0.9**	0.49 (0.01)	-0.01 (0.01)	**0.47* (0.05)**
	0.3	0	1	**-0.7**	0.50 (0.01)	-0.02 (0.01)	**1.14* (0.05)**
	0.5	0	1	**-0.5**	0.60 (0.01)	-0.02 (0.01)	**1.27* (0.04)**
	0.7	0	1	**-0.3**	0.75 (0.00)	-0.02 (0.01)	**1.14* (0.02)**
	0.9	0	1	**-0.1**	0.90 (0.00)	-0.03 (0.01)	**1.12* (0.01)**
1,000	0.1	0	1	**-0.9**	0.47 (0.01)	0.00 (0.00)	**0.43* (0.04)**
	0.3	0	1	**-0.7**	0.43 (0.01)	-0.01 (0.00)	**1.23* (0.05)**
	0.5	0	1	**-0.5**	0.56 (0.00)	-0.01 (0.00)	**1.10* (0.02)**
	0.7	0	1	**-0.3**	0.73 (0.00)	-0.01 (0.00)	**1.06* (0.01)**
	0.9	0	1	**-0.1**	0.90 (0.00)	-0.01 (0.00)	**1.05* (0.01)**

Note 1. The number of observations per X values $n_x = n / 4$ for each sample size n

2. Bias($\hat{\beta}_1$) is the *expected* bias of the estimates $\hat{\beta}_1$ in ordinary logistic regression:
Bias($\hat{\beta}_1$) = E($\hat{\beta}_1 - \beta$) = E($\hat{\beta}_1$) – $\beta_1 = \pi\beta_1 + (1 - \pi)0 - \beta_1 = (\pi - 1)\beta_1$.

* : Smaller bias compared with the expected bias of $\hat{\beta}_1$ in ordinary logistic regression

Table 2.2 (Continued) – Simulated mean MLEs with the standard error (in parentheses) under the logistic regression mixture population based on 1,000 replicates in each case: four sample sizes n =100, 200, 400, and 1,000 are considered and 45 random starting points are used

Sample Size (n)[1]	π	β_0	β_1	***Expected* Bias($\hat{\beta}_1$)[2]**	Mean MLEs (SE)		
					$\hat{\pi}$	$\hat{\beta}_0$	$\hat{\beta}_1$
2,000	0.1	0	1	**-0.9**	0.47 (0.01)	0.01 (0.00)	**0.38* (0.04)**
	0.3	0	1	**-0.7**	0.41 (0.01)	-0.00 (0.00)	**1.12* (0.04)**
	0.5	0	1	**-0.5**	0.54 (0.00)	-0.00 (0.00)	**1.02* (0.01)**
	0.7	0	1	**-0.3**	0.72 (0.00)	-0.00 (0.00)	**1.01* (0.01)**
	0.9	0	1	**-0.1**	0.90 (0.00)	-0.00 (0.00)	**1.01* (0.01)**

Note 1. The number of observations per X values $n_x = n / 4$ for each sample size n

2. Bias($\hat{\beta}_1$) is the *expected* bias of the estimates $\hat{\beta}_1$ in ordinary logistic regression:
Bias($\hat{\beta}_1$) = E($\hat{\beta}_1 - \beta$) = E($\hat{\beta}_1$) – $\beta_1 = \pi\beta_1 + (1 - \pi)0 - \beta_1 = (\pi - 1)\beta_1$.
* : Smaller bias compared with the expected bias of $\hat{\beta}_1$ in ordinary logistic regression

We also investigated the MLEs in the similar way under the null hypothesis that there is no association between the quantitative explanatory variable and the response variable ($\pi = 0$). The simulation results are shown in Table 2.3. The means and standard errors of MLEs were obtained based on 1,000 replicates. In this case we also set β_0 equal zero and considered four sample sizes of n = 100, 200, 400, 1,000, and 2,000 were considered (n_x = 25, 50, 100, 250, and 500). Thus, the estimate of the intercept should equal zero (on average) and the estimated value of slope should also equal zero (on average) if the EM algorithm still works under the null hypothesis as well. Moreover, the estimates of the mixing proportion are expected to follow a uniform distribution with mean of 0.5.

Table 2.3 shows the mean and standard error of the estimates under the null hypothesis for each sample size. The mean estimates of the mixing proportion are 0.5 and the estimated intercept is always within standard error of 0.01 with any sample size. Meanwhile, the standard error of the slope estimate decreases as sample size increases. The distribution of the estimates of $\hat{\pi}$ for sample size of 2,000 can be found in Appendix B.

Table 2.3 – Simulated mean MLEs with the standard error (in parentheses) under the null hypothesis of no association ($\pi = 0$, $\beta_0 = 0$) based on 1,000 replicates in each case: four sample sizes n =100, 200, 400, 1,000, and 2,000 are considered and 45 random starting points are used

			Mean MLEs (SE)		
Sample Size (n)[1]	π	β_0	$\hat{\pi}$	$\hat{\beta}_0$	$\hat{\beta}_1$
100	0	0	0.52 (0.01)	0.01 (0.01)	0.11 (0.10)
200	0	0	0.51 (0.01)	-0.02 (0.01)	0.03 (0.09)
400	0	0	0.50 (0.01)	0.01 (0.01)	-0.06 (0.06)
1,000	0	0	0.50 (0.01)	0.00 (0.00)	-0.01 (0.03)
2,000	0	0	0.50 (0.01)	-0.00 (0.00)	0.01 (0.00)

Note 1. The number of observations per X values $n_x = n / 4$ for each sample size n

Based on the results, we conclude that the EM algorithm works well for observed samples under both the alternative hypothesis and the null hypothesis considered in this dissertation. However, as sample size increases, the estimates obtained by the EM algorithm approach the true values of the corresponding parameters. Therefore, a sample size at least 400 is needed to obtain accurate estimates using this EM algorithm in the logistic regression mixture model. We will investigate the precision of the estimates obtained from this mixture model by comparing the bias and the mean squared error with the estimates in ordinary logistic regression in Chapter 4.

Chapter 3

The Null Distribution of the Likelihood Ratio Test Statistics

3.1 Asymptotic Null Distribution of LRT Statistics

The usual test of the null hypothesis (H_0) against the alternative (H_a^h and H_a^m) described earlier is the likelihood ratio test (LRT). Based on the classical asymptotic theorem for the null distribution of LRT statistics, we would expect the asymptotic distribution of the LRT statistics to be chi-square distribution with 1 degree of freedom and with 2 degrees of freedom for our homogeneous alternative (H_a^h) and mixture alternative (H_a^m), respectively. However, this classical asymptotic null distribution does not hold in the case where we test the common null hypothesis (H_0) against our mixture alternative (H_a^m) as we discussed in Chapter 1.

Instead, based on the asymptotic results on the boundary of the parameter space, we conjectured that the *asymptotic* null distribution of the LRT statistics may be $0.5\chi_1^2 + 0.5\chi_2^2$ to test the mixture alternative hypothesis in our power study. We will verify this conjecture by using the empirical null distribution and the fitted null distribution in Section 3.2 and 3.3. Concurrently, we will see if the null distribution is invariant to generating models with the value of the parameter β_0 under the null hypothesis.

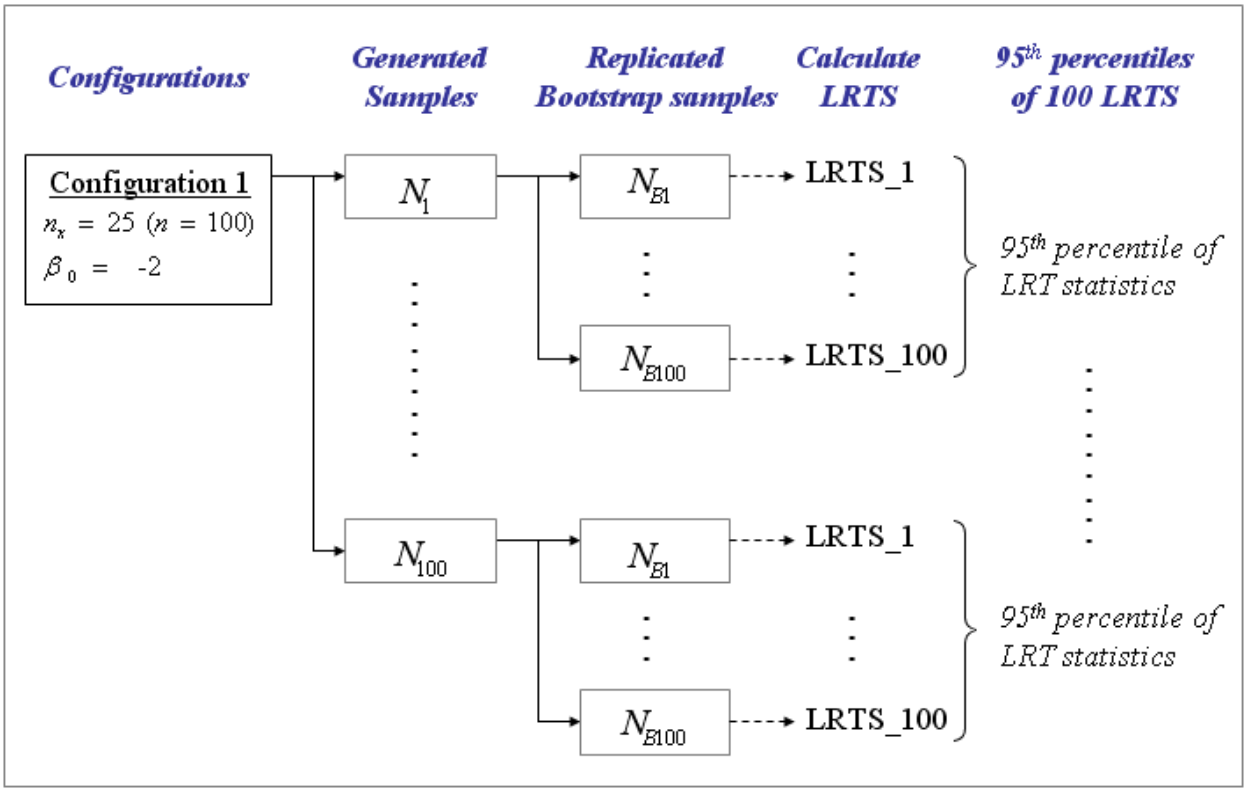
3.2 Empirical Null Distribution of the LRT Statistics

We want to verify that an asymptotic chi-square distribution also holds true for the LRT test in the situation we are considering. Thus, the empirical null distribution of the LRT statistics was obtained through simulation. We compared the theoretical asymptotic distribution with the empirical null distribution found from the simulation. The 95th percentile of the empirical null distribution of the LRT statistics and corresponding 95% confidence intervals are computed.

3.2.1 Data Simulation

Five different values of the parameter β_0 were considered to generate the null distributions that there is no association between an explanatory variable and a response variable, that is, β_0 = -2, -1, 0, 1, and 2. Also, we considered three different sample sizes per configuration to model both small and large samples, that is, three different sizes for each value of the explanatory variable x = 0, 1, 2, and 3 were considered; n_x = 25, 50, and 100 (i.e., n = 100, 200, and 400). The values of the parameter β_1 and π were fixed (β_1 = 1 and π = 0.5) because the null distribution of LRT statistics is not affected by these values. For each configuration one hundred samples were generated and for each sample one hundred bootstrap samples were replicated under the null hypothesis. LRT statistics were calculated for each bootstrap sample against the corresponding generated sample for each configuration, and the 95th percentile values were obtained among the one hundred LRT statistics based on each generated sample. Figure 3.1 shows this bootstrap procedure for the configuration with n_x = 25 and β_0 = -2.

Figure 3.1 – The bootstrap procedure to construct the empirical null distribution of the LRT statistics for the configuration with n_x = 25 and β_0 = -2 : 100 generated samples are considered and 100 bootstrap samples are replicated under the null hypothesis for each sample.



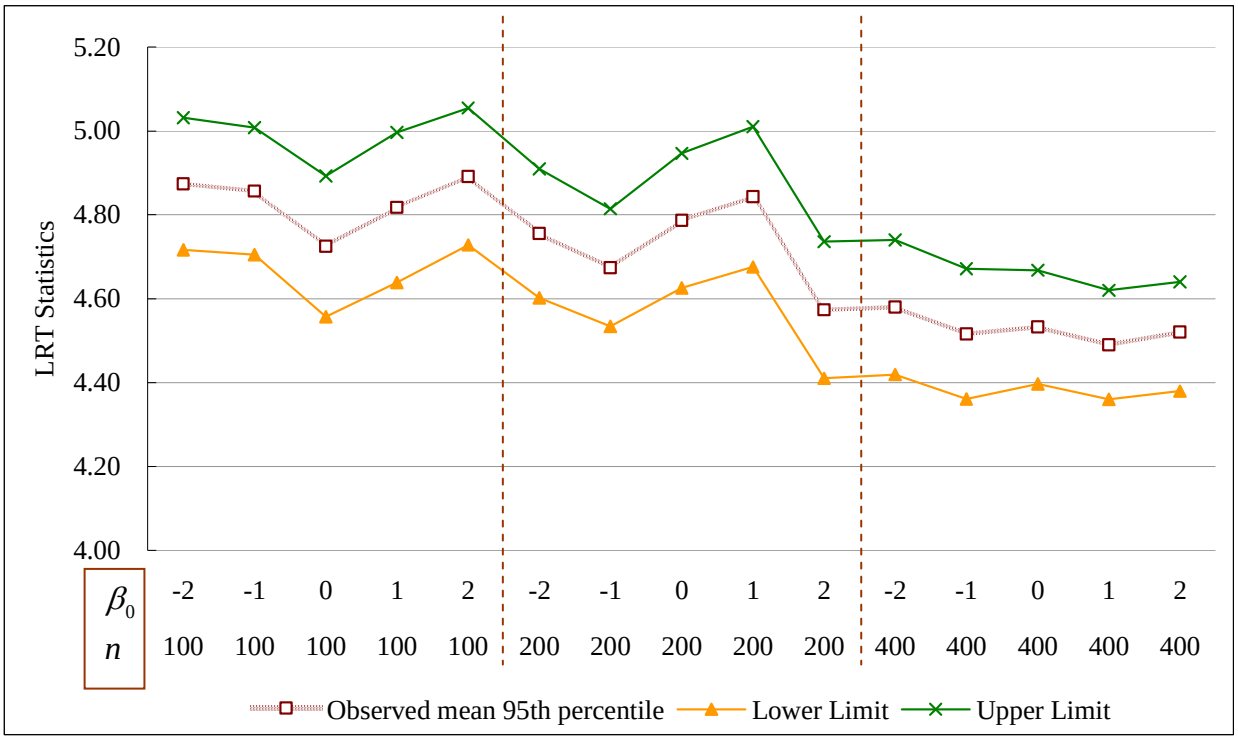
Note. This procedure is also applied for the other fourteen combinations of n_x and β_0 (n_x = 25, 50, and 100; β_0 = -2, -1, 0, 1, and 2)

3.2.2 Simulation Results of the Empirical Null Distribution

For each configuration (n_x = 25, 50, and 100; β_0 = -2, -1, 0, 1, and 2), we computed the mean and variance of the 95% percentile of the LRT statistics for each model under the null hypothesis based on 100 simulations per n_x and β_0 combination. In addition, the 95% confidence intervals for the empirical 95th percentile of the LRT statistics were constructed. The 95% confidence intervals for these five different models are displayed in Figures 3.2 for each sample size.

As one can see by looking at Figure 3.2, it appears that the majority of the 95% confidence intervals of the 95th percentile overlapped. That is, we observed apparent invariance to the generating model under the null hypothesis for the LRT statistics. Hence, we combined the 50,000 LRT statistics obtained from the simulations to find a more precise estimate of the 95th percentile of the empirical null distribution for each sample size.

Figure 3.2 – The 95% confidence intervals for the mean 95th percentile of the LRT statistics according to the values of β_0 : β_0 = -2, -1, 0, 1, and 2 are considered and three sample sizes are also considered (n = 100, 200, and 400).



Note. The observed mean of the bootstrap 95th percentile and 95% confidence intervals are based on 100 simulated samples per β_0 value (N_B =100 bootstrap samples under H_0 per simulated sample).

We found the 95th percentile of the LRT statistics under the null hypothesis for sample size of 100 to be 5.1 and the 95% confidence interval for the empirical 95th percentile was [5.0, 5.2]. In the similar manner, the 95th percentile of the LRT statistics for sample size of 200 was 4.9 and the 95% confidence interval for the empirical 95th percentile was [4.9, 5.0]. We also found the 95th percentile of the LRT statistics for sample size of 400 to be 4.6. The 95% confidence interval for the empirical 95th percentile was [4.6, 4.7].

Table 3.1 shows these empirical 95th percentile values of the LRT statistics and corresponding 95% confidence intervals obtained from the combined 50,000 LRT statistics for each sample size. These confidence limits were computed from non-parametric methods (Snedecor and Cochran, 1967). Table 3.2 contains the mean, variance, and selected percentiles from the simulated null distribution of the LRT statistics for sample size 100, 200, and 400. The means, variances and percentiles monotonically decrease. The table also reports the values for selected chi-squared distributions. The percentiles of the empirical null distributions lie between the values for the distribution of $0.5\chi_1^2 + 0.5\chi_2^2$ and χ_1^2. Thus, we expect that the simulated LRT statistics follow the mixture of chi-squared distributions with 1 degree of freedom and 2 degrees of freedom, where the fraction of the chi-squared distribution with 1 degree of freedom would be between 0.5 and 1.

Table 3.1 – The empirical 95th percentile of the LRT statistics and corresponding 95% confidence interval based on the combined 50,000 LRT statistics

Sample Size (n)	Empirical 95th percentile of LRT statistics	95% Confidence Interval	
		Lower Limit	Upper Limit
100	5.1	5.0	5.2
200	4.9	4.9	5.0
400	4.6	4.6	4.7

Note 1. The number of observations per x values $n_x = n / 4$ for each sample size n

2. The results are based on combined LRT statistics with N_B = 100 bootstrap samples, N = 100 replications, and five configurations (β_0 = -2, -1, 0, 1, and 2)

Table 3.2 – Summary of empirical null distribution of LRT statistics of the null hypothesis of ordinary logistic regression models versus the alternative of logistic regression mixture models

Null distribution	Sample Size (n)	Mean	Variance	Percentiles of LRT statistics		
				90%	95%	99%
χ^2_2		2	4	4.6	6.0	9.2
$0.5\chi^2_1 + 0.5\chi^2_2$		1.5	3.25	3.8	5.1	8.3
χ^2_1		1	2	2.7	3.8	6.6
Empirical distribution	100	1.35	2.94	3.7	5.1	9.0
	200	1.30	2.83	3.6	4.9	8.0
	400	1.22	2.63	3.3	4.6	7.9

Note 1. The number of observations per x values $n_x = n / 4$ for each sample size n

2. Empirical null distribution is based on the 50,000 combined LRT statistics per line

3.3 Fitted Null Distribution of the LRT Statistics

Wilson and Hilferty (1931) showed that the cube root of a chi-square distribution is approximately normal. Consequently, if the null distribution of a statistic were of the form $p\chi_0^2 + (1-p)\chi_v^2$, then the distribution of the cube root of the statistic would be a mixture of a fraction of zero with a proportion of *p* and an approximately normal distribution. In the context of this fact and our conjecture on the asymptotic null distribution of the LRT statistics in Section 3.1, we consider the form $p\chi_{v-1}^2 + (1-p)\chi_v^2$ as the null distribution of the LRT statistics and evaluate the fit of the distribution of the LRT statistics under the null hypothesis to this form. This approach also has a thread of connection with the expectation of empirical null distribution in Section 3.2.2.

Based on the combined LRT statistics, let the null distribution of the statistics be the form of $p\chi_{v-1}^2 + (1-p)\chi_v^2$. Then we can obtain the estimate $\hat{p}$ and $\hat{v}$ of the parameters *p* and *v* using the mean and variance of the LRT statistics for each sample size as follows.

$$\begin{aligned} E(LRTS) &= \bar{L} \\ &= p(v-1) + (1-p)v \\ &= v - p \end{aligned} \tag{3.3.1}$$

$$\begin{aligned} Var(LRTS) &= S^2 \\ &= E(LRTS^2) - [E(LRTS)]^2 \\ &= [p\{2(v-1) + (v-1)^2\} + (1-p)(2v + v^2)] - (v-p)^2 \\ &= 2v - p - p^2 \end{aligned} \tag{3.3.2}$$

From Equation (3.3.1) and (3.3.2) the estimates of the mixing proportion *p* and the degrees of freedom *v* are computed in the form of $p\chi^2_{v-1} + (1-p)\chi^2_v$. That is,

$$\hat{p} = \frac{1 \pm \sqrt{1 + 4(2\bar{L} - S^2)}}{2} \tag{3.3.3}$$

$$\hat{v} = \bar{L} + \hat{p} \tag{3.3.4}$$

As a result, we estimate the empirical null distribution of the LRT statistics from the 50,000 combined LRT statistics described above. The fitted null distributions of LRT statistics are $0.41\chi^2_{0.76} + 0.59\chi^2_{1.76}$, $0.61\chi^2_{0.91} + 0.39\chi^2_{1.91}$, and $0.74\chi^2_{0.95} + 0.26\chi^2_{1.95}$ for sample size 100, 200, and 400, respectively. The fraction of chi-square distribution with smaller value of the degrees of freedom increases and the values of the degrees of freedom also increase as sample size increases. That is, while we observed apparent invariance to the generating models under the null hypothesis, there is some dependence on sample size.

In Table 3.3, we report the 90th, 95th, and 99th percentiles, and other relevant summary statistics for sample size. Q-Q plots are used to compare the observed null distribution with the fitted null distribution of the LRT statistics. Although the fitted null distribution does not work well in the upper quartiles when sample size is small, there mostly appears to be no difference in the null distribution when sample size is large (Figure 3.3).

Table 3.3 – The mean and variance of the LRT statistics for each sample size and corresponding estimated values of parameters p and v in the form of $p\chi^2_{v-1} + (1-p)\chi^2_v$

			Fitted parameters of null distribution		Percentiles		
Sample Size (n)	Mean	Variance	$\hat{p}$	$\hat{v}$	90%	95%	99%
100	1.35	2.94	0.41	1.76	3.5	4.8	8.0
200	1.30	2.83	0.61	1.91	3.4	4.7	7.8
400	1.22	2.63	0.74	1.95	3.2	4.5	7.5

Note: The results are based on the 50,000 simulated LRT statistics per line

Figure 3.3 – Q-Q plots of the null distribution of LRT statistics for sample size: comparing observed null distribution with fitted null distribution of LRT statistics

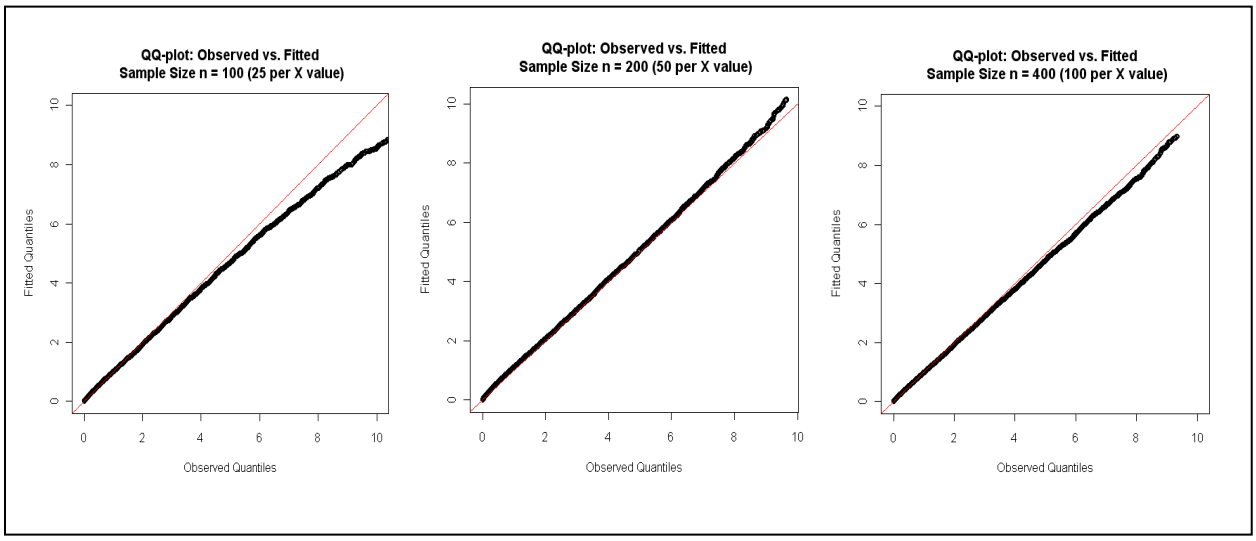


Table 3.4 summarizes the 95th percentile values of LRT statistics in the following cases of the null distribution; (1) the asymptotic null distribution, $0.5\chi^2_1 + 0.5\chi^2_2$, (2) the empirical null distribution, and (3) the fitted null distribution with the form of $\hat{p}\chi^2_{\hat{v}-1} + (1-\hat{p})\chi^2_{\hat{v}}$ for specified sample size.

Table 3.4 – Summary of the 95th percentile selected for sample size.

Sample Size (n)	The 95th percentile		
	Asymptotic[1] null distribution	Empirical [2] null distribution	Fitted [3] null distribution
100	5.1	5.1	4.8
200	5.1	4.9	4.7
400	5.1	4.7	4.5

Note 1. For the asymptotic null distribution, $0.5\chi_1^2 + 0.5\chi_2^2$ is conjectured.

2. Empirical null distribution is based on the 50,000 combined LRT statistics per line: given in Table 3.2
3. Fitted null distribution is estimated as form of $\hat{p}\chi^2_{\hat{\nu}-1} + (1 - \hat{p})\chi^2_{\hat{\nu}}$ based on the 50,000 combined LRT statistics per line: given in Table 3.3

Table 3.5 shows the Type I error rates of the LRT under each generating null model for each sample size. For each configuration one thousand replications were used to obtain the Type I error rates. The Type I error rates were estimated using three different null distributions of the LRT statistics described in this Chapter 3. The corresponding 95th percentile values of LRT statistics for each case are shown in Table 3.4.

Table 3.5 – Type I Error rates of the LRT under the generating null hypothesis using asymptotic, empirical, and fitted null distribution of LRT statistics (α = 0.05)

Sample Size (n)	Generating Model (β_0)	Type I Error rates of the LRT [4]		
		Asymptotic[1] Null distribution	Empirical [2] null distribution	Fitted [3] null distribution
100	-2	0.046	0.057	0.062
	-1	0.044	0.052	0.067
	0	0.056	0.056	0.059
	1	0.048	0.045	0.057
	2	0.043	0.052	0.067
	Average	**0.047**	**0.052**	**0.062**
200	-2	0.056	0.060	0.050
	-1	0.051	0.043	0.053
	0	0.044	0.048	0.053
	1	0.043	0.045	0.046
	2	0.049	0.052	0.053
	Average	**0.049**	**0.050**	**0.051**
400	-2	0.040	0.036	
	-1	0.041	0.050	0.063
	0	0.043	0.051	0.052
	1	0.044	0.044	0.046
	2	0.043	0.043	0.045
	Average	**0.042**	**0.045**	**0.049**

Note 1. For the asymptotic null distribution, $0.5\chi_1^2 + 0.5\chi_2^2$ is conjectured.

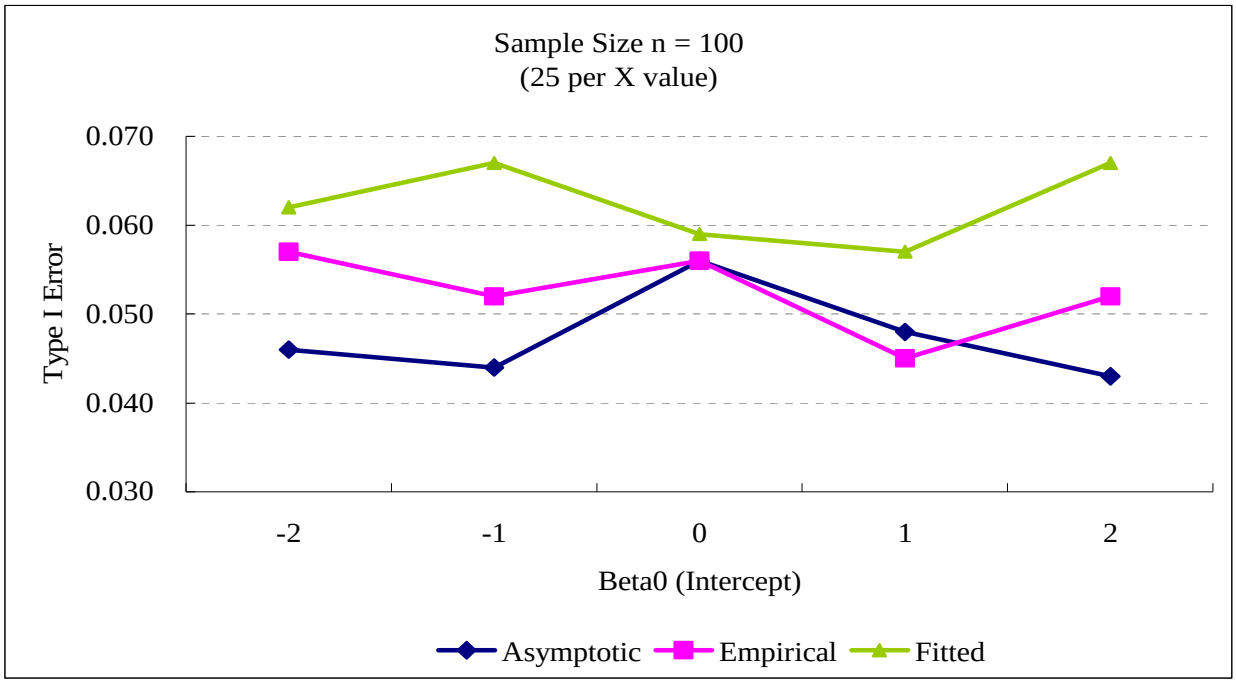
2. Empirical null distribution is based on the 50,000 combined LRT statistics per line: given in Table 3.2

3. Fitted null distribution is estimated as form of $\hat{p}\chi^2_{\hat{\nu}-1} + (1 - \hat{p})\chi^2_{\hat{\nu}}$ based on the 50,000 combined LRT statistics per line: given in Table 3.3

4. Simulation results are based on 1,000 replications per line and the 95% margin of error is ±0.01 for each configuration

As one can see in Table 3.5, the Type I error rates of the LRT seem close to the nominal value of 0.05 for each of 95th percentile value within a 95% margin of error ($\pm$0.01). However, when sample size is small (n = 100) the average Type I error rate (0.062) is slightly larger than the nominal value. It is because the modeled null distribution for n =100 does not fit well in the upper quartiles (Figure 3.3). Therefore, we use the asymptotic null distribution of $0.5\chi_1^2 + 0.5\chi_2^2$ to calculate the power of the test with mixture alternative hypothesis (H_a^m). Figure 3.4 – 3.6 illustrate the Type I error rates of the LRT for the five different generating models for sample size.

Figure 3.4 – Type I error rates of the LRT under the generating null hypothesis using asymptotic, empirical, and fitted null distribution of LRT statistics with sample size n = 100.



Note. Simulation results are based on 1,000 replications and the 95% margin of error is $\pm$0.01 for each configuration

Figure 3.5 – Type I error rates of the LRT under the generating null hypothesis using asymptotic, empirical, and fitted null distribution of LRT statistics with sample size n = 200.

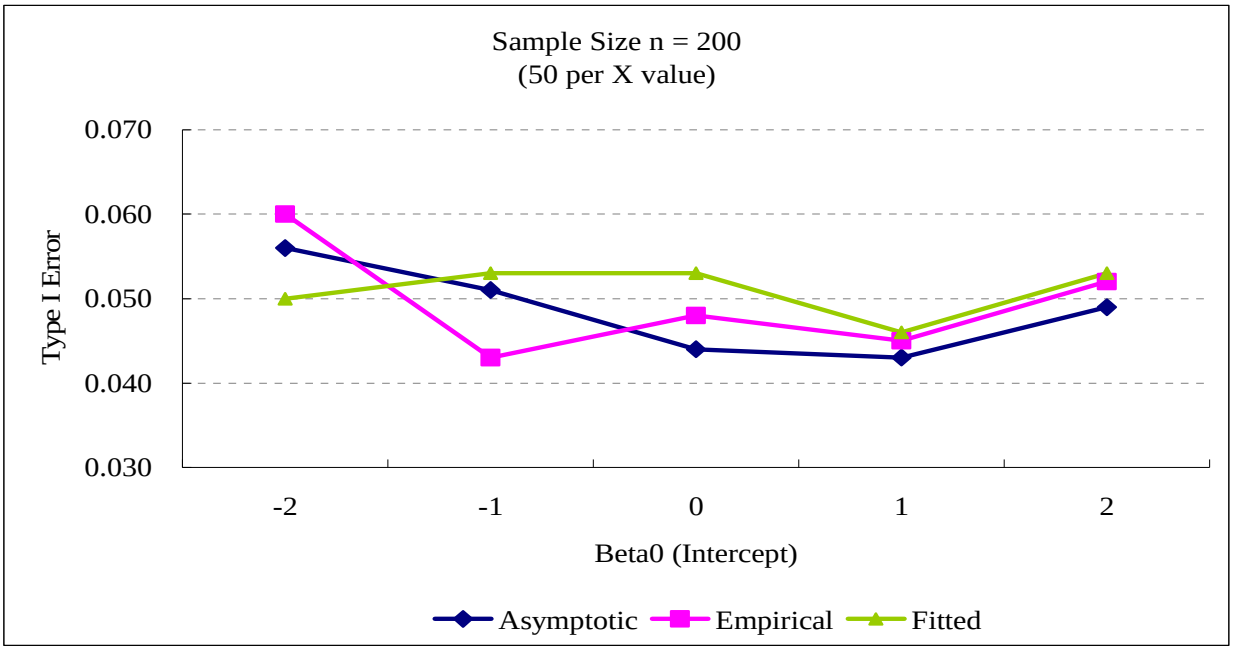
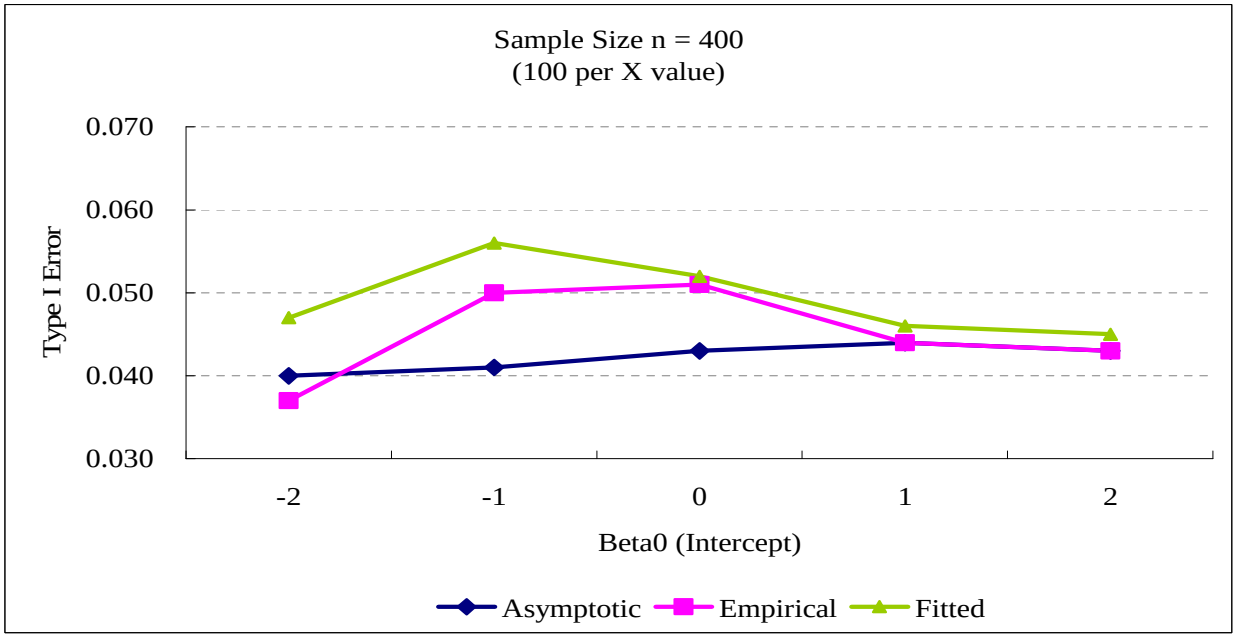


Figure 3.6 – Type I error rates of the LRT under the generating null hypothesis using asymptotic, empirical, and fitted null distribution of LRT statistics with sample size n = 400.



Note. Simulation results are based on 1,000 replications and the 95% margin of error is ±0.01 for each configuration

Chapter 4

Power Study

of the Likelihood Ratio Test

4.1 Data Simulation

We evaluated the performance of the hypothesis tests about the association in the context of ordinary logistic regression and logistic regression mixture models defined in Section 1.3. Based on the simulation results in Chapter 3 we concluded the null distribution of LRT statistics is invariant to parameter setting of the generating models under the null hypothesis and verified our conjecture stated in Section 3.1 about the null distribution, $0.5\chi_1^2 + 0.5\chi_2^2$. Therefore, this asymptotic null distribution, $0.5\chi_1^2 + 0.5\chi_2^2$, was used to infer the critical values for the tests defined in this dissertation. For our power study, we considered the following parameter setting for sample size n = 200, 400, and 1,000:

β_0 = -2, -1, and 0;

β_1 = 0.5, 1.0, and 1.5;

π = 0.1, 0.3, 0.5, 0.7, and 0.9.

For each configuration we simulated one thousand samples and found the power of the LRT. Each power was based on the 95th percentile of the asymptotic null distribution, 5.1, which was found in Chapter 3.

Additionally, we investigated the precision of the estimates under mixture populations with two components logistic regression models. To compare the precision in the ordinary logistic regression models with logistic regression mixture models, mean squared errors (MSEs) were calculated from these two approaches based on one thousand replicates. For simplicity we set the true values of β_0 and β_1 equal zero and one, respectively. Five sample sizes (n = 100, 200, 400, 1,000, and 2,000; i.e., n_x = 25, 50, 100, 250, and 500) and five mixing proportions were also considered (π = 0.1, 0.3, 0.5, 0.7, and 0.9).

4.2 Power Study Based on Asymptotic Null Distribution of the LRT Statistics

The purpose of this thesis is to compare the power of the two hypothesis tests which detect the effect of the quantitative explanatory variable X on the binary response variable Y. One test is conducted in terms of the ordinary logistic regression model and the other detects the effect in terms of a logistic regression mixture model having equal intercept and two unequal slopes as defined in Section 1.3. We compared the power of testing with these two different alternatives, H_a^h for the ordinary logistic regression model and H_a^m for the logistic regression mixture model.

In addition, we used McNemar's test to determine whether there is a significant difference in the performance of the two LRT for each configuration and also carried out logistic regression on the findings for each case (1) to determine whether there is an overall increase in power associated with one method and (2) to identify the conditions affecting the power difference.

The various configurations of parameters settings used to generate the data in this power study are shown in Table 4.1. This table contains the power of the test based on the simulated data for each configuration. The results of the McNemar's test are also shown in this table with a significance level of 0.05 and 0.01. The power comparisons between two models according to the values of β_0, β_1, and π are shown in Figure 4.1 for sample size of 1,000. The patterns were almost identical for sample size of 200 and 400 (See Appendix C).

The power of test, taken as a whole, increased as the value of β_1 increases or the mixing proportion π increases. It also appears that the power of the logistic regression mixture model is slightly greater than the power of the ordinary logistic regression model, especially for smaller mixing proportion π.

Table 4.1 – Power of the LRT using the asymptotic 95th percentile, calculated from 1,000 replicates: comparison the power of ordinary logistic regression models (H_a^h) with logistic regression mixture models (H_a^m)

			Power of the LRT					
			n = 200		n = 400		n = 1,000	
β_0	β_1	π	Ordinary[1]	Mixture[2]	Ordinary	Mixture	Ordinary	Mixture
-2.0	0.5	0.1	0.07	0.07	0.07	0.07	0.10	0.16**
		0.3	0.18	0.21*	0.30	0.34	0.62	0.74**
		0.5	0.37	0.45**	0.65	0.71**	0.96	0.98*
		0.7	0.61	0.67**	0.90	0.92	1.00	1.00
		0.9	0.78	0.85**	0.98	0.99*	1.00	1.00
	1.0	0.1	0.13	0.17*	0.20	0.25*	0.45	0.55**
		0.3	0.62	0.69**	0.91	0.94*	1.00	1.00
		0.5	0.95	0.98**	1.00	1.00	1.00	1.00
		0.7	1.00	1.00	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00
	1.5	0.1	0.21	0.24	0.33	0.38*	0.69	0.77**
		0.3	0.85	0.87	0.99	0.99	1.00	1.00
		0.5	1.00	1.00	1.00	1.00	1.00	1.00
		0.7	1.00	1.00	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00
-1	0.5	0.1	0.07	0.07	0.07	0.11**	0.11	0.17**
		0.3	0.17	0.23**	0.34	0.40**	0.68	0.78**
		0.5	0.44	0.50**	0.73	0.81**	0.99	0.99
		0.7	0.71	0.77**	0.95	0.96	1.00	1.00
		0.9	0.89	0.94**	1.00	1.00	1.00	1.00
	1.0	0.1	0.09	0.11	0.14	0.17	0.29	0.37**
		0.3	0.47	0.55**	0.80	0.83	0.99	0.99
		0.5	0.91	0.92	1.00	1.00	1.00	1.00
		0.7	1.00	1.00	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00
	1.5	0.1	0.11	0.13	0.19	0.21	0.39	0.45**
		0.3	0.62	0.65	0.91	0.90	1.00	1.00
		0.5	0.97	0.97	1.00	1.00	1.00	1.00
		0.7	1.00	1.00	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00

Note 1. Test in terms of the ordinary regression model, H_a^h : $\log(Odds) = \beta_0 + \beta_1 x$

2. Test in terms of the logistic regression mixture model,
H_a^m : $\log(Odds) = \beta_0 + \beta_1 x$ with probability of π; $\log(Odds) = \beta_0$ with $1 - \pi$

3. The 95% margin of error is ±0.03 for each configuration.

4. Significantly different in power compared to ordinary logistic regression using McNemar's Test (* 0.05; ** 0.01)

Table 4.1 (Continued) – Power of the LRT using the asymptotic 95th percentile, calculated from 1,000 replicates: comparison the power of ordinary logistic regression models (H_a^h) with logistic regression mixture models (H_a^m)

			Power of the LRT					
			n = 200		n = 400		n = 1,000	
β_0	β_1	π	Ordinary[1]	Mixture[2]	Ordinary	Mixture	Ordinary	Mixture
0.0	0.5	0.1	0.06	0.08*	0.05	0.09**	0.09	0.12*
		0.3	0.13	0.19**	0.25	0.29	0.53	0.61**
		0.5	0.34	0.40**	0.61	0.69**	0.94	0.96
		0.7	0.62	0.68**	0.91	0.94*	1.00	1.00
		0.9	0.85	0.89**	0.99	1.00	1.00	1.00
	1.0	0.1	0.06	0.08	0.08	0.11*	0.14	0.20**
		0.3	0.28	0.31	0.50	0.54	0.88	0.90
		0.5	0.67	0.71*	0.93	0.93	1.00	1.00
		0.7	0.94	0.96	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00
	1.5	0.1	0.07	0.08	0.10	0.12	0.17	0.21*
		0.3	0.35	0.34	0.60	0.59	0.94	0.94
		0.5	0.79	0.79	0.97	0.96	1.00	1.00
		0.7	0.98	0.98	1.00	1.00	1.00	1.00
		0.9	1.00	1.00	1.00	1.00	1.00	1.00

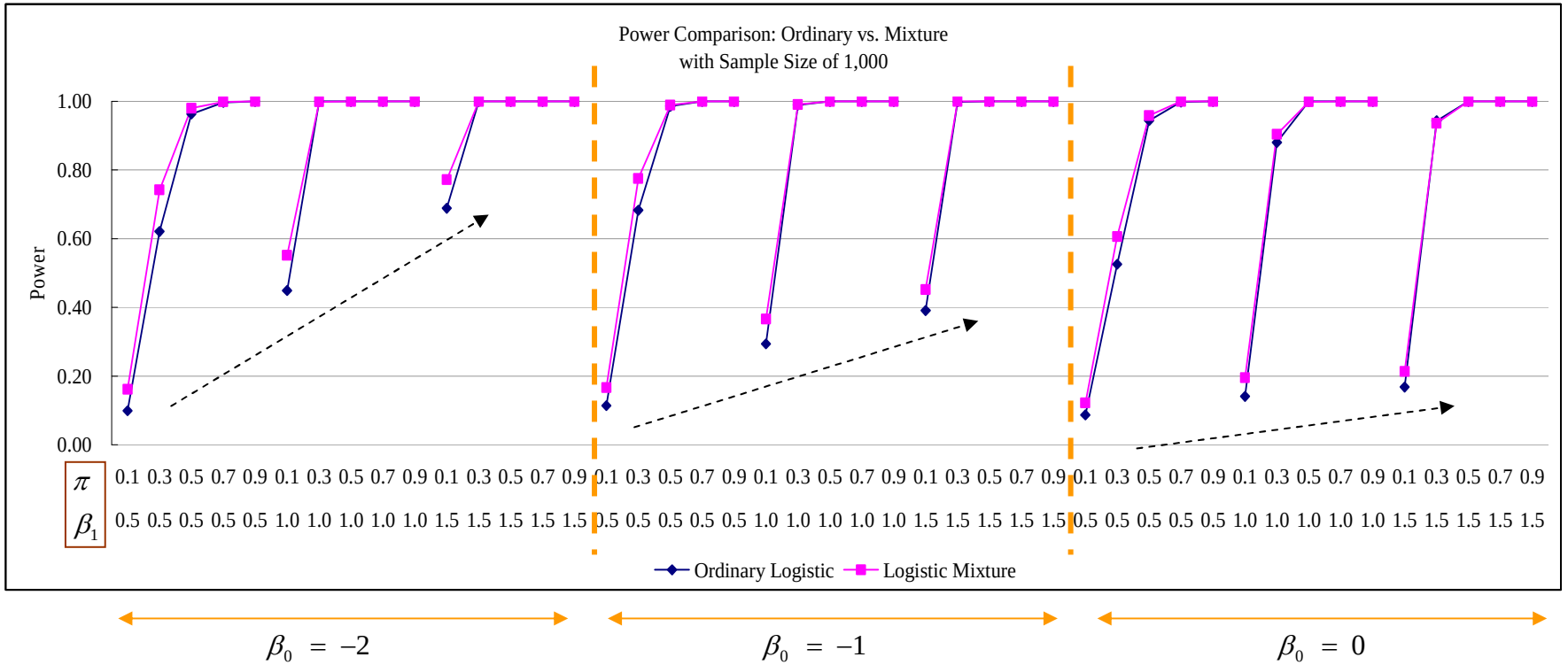
Note 1. Test in terms of the ordinary regression model, H_a^h : log(Odds) = $\beta_0 + \beta_1 x$

2. Test in terms of the logistic regression mixture model,
H_a^m : log(Odds) = $\beta_0 + \beta_1 x$ with probability of π; log(Odds) = β_0 with $1 - \pi$

3. The 95% margin of error is ±0.03 for each configuration.

4. Significantly different in power compared to ordinary logistic regression using McNemar's Test (* 0.05; ** 0.01)

Figure 4.1 – Comparison of the Power of ordinary logistic and logistic mixture models for various values of the intercept β_0, the slope β_1 and mixing proportion π for sample size of 1,000: β_0 = -2, -1, and 0; β_1 = 0.5, 1.0, and 1.5; π = 0.1, 0.3, 0.5, 0.7, and 0.9



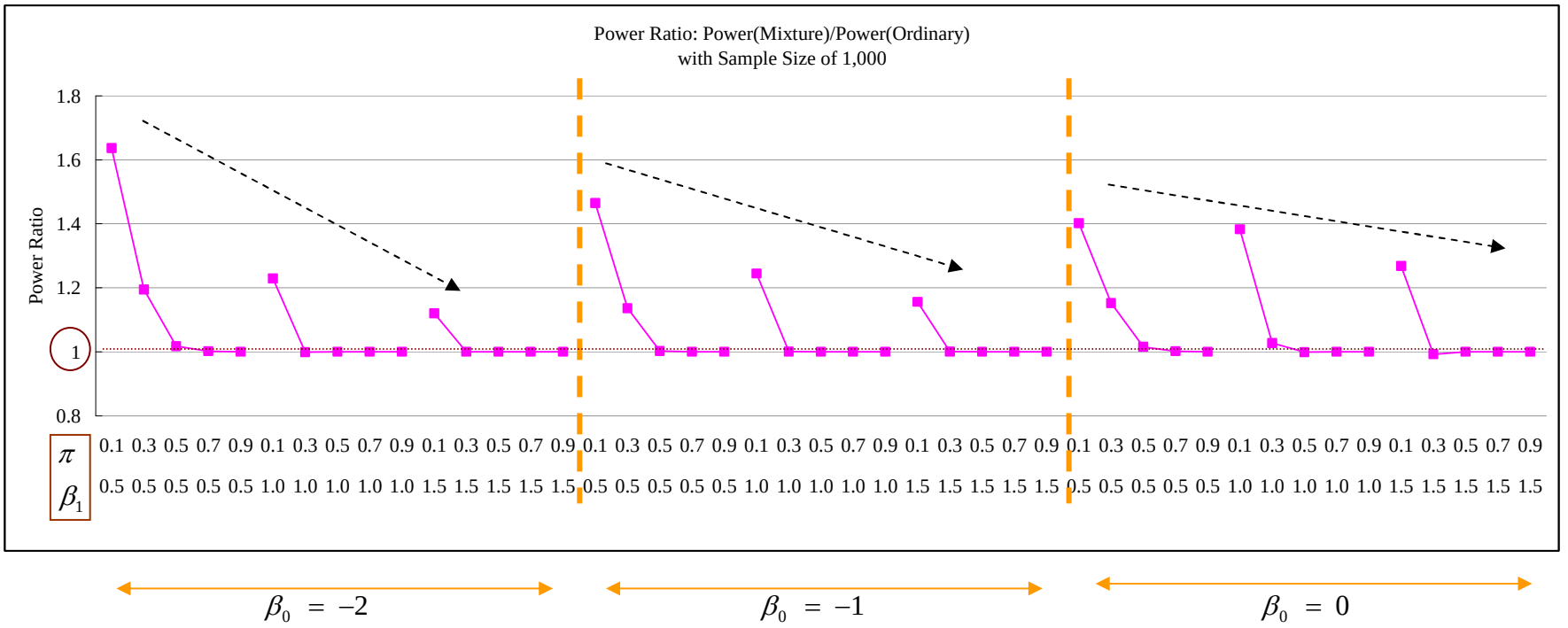
Note 1. The power results are based on 1,000 replicates with sample size of 1,000.

2. The dotted arrows represent the patterns of the power by the parameter settings given the same value of β_0.

Additionally, to investigate if there is a relationship between the difference in power and various parameters, we computed the power ratio as the power of the logistic regression mixture model divided by the power of the ordinary logistic regression model for each configuration. Figure 4.2 illustrates this power ratio for various configurations of parameters with sample size of 1,000. We also computed the power ratio for sample size of 200 and 400 in the same way (See Appendix D).

As one can see by looking at Figure 4.2, it appears that the power ratio decreases as the value of β_1 and the mixing proportion π increase given equal values of β_0 for sample size of 1,000. Since this power ratio is related to the improvement in power, we can expect the performance of the mixture models to detect the effect of the quantitative variable X on the response variable Y to be improved when both the value of β_1 and the mixing proportion π are small. However, this reasonable trend is more marked indeed as sample size increases. We will verify our findings obtained from this simulation in the following section.

Figure 4.2 – Power Ratio of logistic regression mixture models compared with ordinary logistic regression models for various values of the intercept β_0, the slope β_1 and mixing proportion π for sample size of 1,000: β_0 = -2, -1, and 0; β_1 = 0.5, 1.0, and 1.5; π = 0.1, 0.3, 0.5, 0.7, and 0.9



Note 1. The power results are based on 1,000 replicates with sample size of 1,000.

2. Power Ratio was calculated as the power of logistic regression mixture models divided by the power of ordinary logistic model

3. The dotted arrows represent the trend of the power ratio by the parameter settings given the same value of β_0.

4.3 Modeling the Difference in Power

Based on the visual examination in the previous section, we expected that the β_1 and π affect the difference in power of two approaches to detect the relationship between Y and X. In this section we find the model of the improvement including all possible factors by a general linear model. Our interest in this model is the difference in power between ordinary logistic regression models and logistic regression mixture models. Therefore, we considered the odds ratio of improvement as the response variable and all of the parameters in the logistic regression mixture model as factors in this model:

$$\textit{Odds Ratio}(\text{improvement}) = \alpha + \gamma\beta_0 + \delta\beta_1 + \eta\pi + \psi n. \qquad (4.1.1)$$

The interaction of the factors will be included in the above model if the above model (Equation 4.1.1) does not fit.

Since the improvement in power means the power in the context of the mixture model is greater than the ordinary logistic regression model, the odds ratio of improvement can be computed from the number of different decisions (N_{12} and N_{21}) by the two models in the simulated *N* replicates as follows (See Table 4.2).

$$\textit{Odds Ratio}(\text{improvement}) = N_{12} / N_{21}. \qquad (4.1.2)$$

We cannot obtain the odds ratio in the case of $N_{12} = N_{21} = 0$ or $N_{12} \neq N_{21} = 0$, thus these cases were excluded in our model. Table 4.3 reports the odds ratio for each configuration. If this odds ratio is greater than one, we conclude that there is an improvement in power of the test by using logistic mixture models.

Table 4.2 – Matched Pairs Data Structure

Ordinary	Mixture: Accept H_0	Mixture: Reject H_0	Total
Accept H_0	N_{11}	N_{12}	N_{1+}
Reject H_0	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	N

Note. The total number of *N* is the number of replicates per configuration

We first conducted the ANOVA (analysis of variance) including all three main factors and all two-way interactions in order. The results indicated that the interactions were not significant and the value of β_1 and the mixing proportion π were significant for each sample size. In addition, the sample size *n* was not significant for the overall observed samples. These results are consistent with the findings in the previous section. The significance level of 0.05 was used for these conclusions. Detailed SAS output can be found in Appendix E.

Next, we fit a general linear model with only these significant factors – the value of slope β_1 and mixing proportion π – to the odds ratio for each sample size. The regression coefficients of the corresponding values for each factor are summarized in Table 4.4.

Table 4.3 – Summary of the Odds Ratio of the power with ordinary logistic and logistic regression mixture models for each configuration: n = 200, 400, and 1,000

			Odds Ratio[1]						Odds Ratio		
β_0	β_1	π	n=200	n=400	n=1,000	β_0	β_1	π	n=200	n=400	n=1,000
		0.1	1.13	1.05	1.75			0.1	1.43	1.66	1.46
		0.3	1.26	1.17	1.73			0.3	1.47	1.20	1.39
-2	0.5	0.5	1.41	1.30	1.89	0	0.5	0.5	1.31	1.43	1.38
		0.7	1.29	1.24	3.00			0.7	1.27	1.42	-
		0.9	1.51	2.30	-			0.9	1.41	2.67	-
		0.1	1.31	1.30	1.51			0.1	1.32	1.40	1.48
		0.3	1.37	1.46	0.00			0.3	1.18	1.17	1.28
	1.0	0.5	2.09	-	-		1.0	0.5	1.21	1.00	0.00
		0.7	-	-	-			0.7	1.29	0.00	-
		0.9	-	-	-			0.9	2.00	-	-
		0.1	1.19	1.24	1.54			0.1	1.27	1.26	1.34
		0.3	1.23	1.43	-			0.3	0.96	0.95	0.89
	1.5	0.5	-	-	-		1.5	0.5	0.99	0.66	-
		0.7	-	-	-			0.7	1.11	-	-
		0.9	-	-	-			0.9	-	-	-
		0.1	1.02	1.58	1.55						
		0.3	1.41	1.30	1.62						
-1	0.5	0.5	1.28	1.51	1.30						
		0.7	1.38	1.48	-						
		0.9	1.73	1.33	-						
		0.1	1.23	1.19	1.39						
		0.3	1.36	1.25	1.11						
	1.0	0.5	1.17	0.25	-						
		0.7	1.33	-	-						
		0.9	-	-	-						
		0.1	1.15	1.13	1.30						
		0.3	1.17	0.87	-						
	1.5	0.5	0.87	-	-						
		0.7	-	-	-						
		0.9	-	-	-						

Note 1. The odds ratio is computed by N_{12} / N_{21}, where

(1) N_{12} represents the number of cases that the null hypothesis was rejected based on the mixture model and not rejected based on the ordinary logistic regression model.

(2) N_{21} represents the number of cases that the null hypothesis was not rejected based on the regression mixture model and rejected based on the ordinary logistic regression model. (See Table 4.2)

2. The missing value of odds ratio occurs when $N_{12} = N_{21} = 0$ or $N_{12} \neq N_{21} = 0$.

Table 4.4 – Estimates of Regression Coefficients in the fitted General Linear Models of the Odds Ratio for sample size

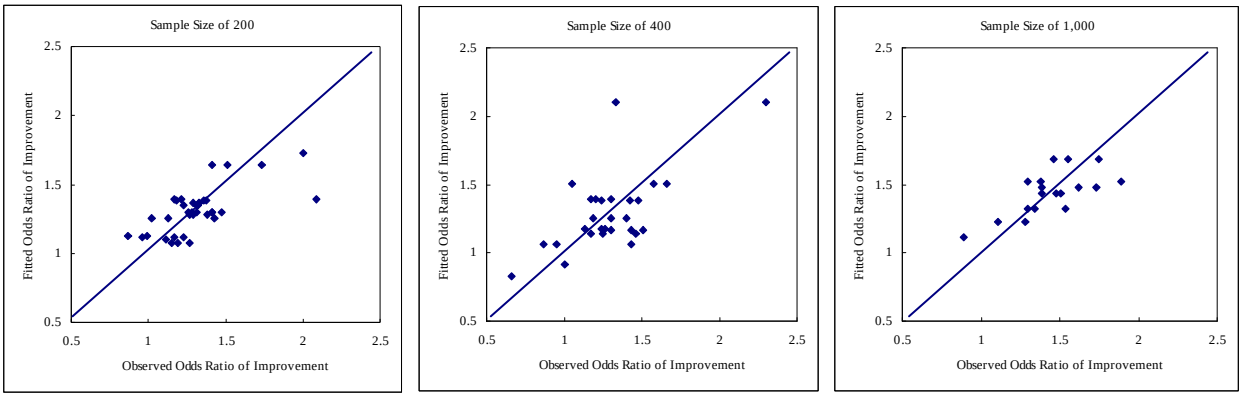
Parameter		Estimates of Coefficients n[1] = 200	n = 400	n = 1,000
Intercept		0.3297	0.3759	0.1262
β_1	0.5	0.1561	0.3241	0.2808
	1.0	0.2097	0.0151	0.1048
	1.5[2]	0.0000	0.0000	0.0000
π	0.1	-0.2509	-0.2274	0.1338
	0.3	-0.2216	-0.3192	-0.0358
	0.5	-0.2424	-0.6713	0.0000
	0.7	-0.2344	-0.3808	-[3]
	0.9[2]	0.0000	0.0000	-[3]

Note 1. The number of observations per x values $n_x = n / 4$ for each sample size.

2. The Baseline of the factor β_1 set to be 1.5 and the baseline of the factor π is 0.9.
3. For sample size of 1,000 the case of the mixing proportion π of 0.7 and 0.9 are not included in the fitted model.

By using the fitted general linear models we can obtain the fitted values for the odds ratio of improvement. Then we can compare the observed odds ratio and the fitted odds ratio. The results are illustrated in Figure 4.3. Based on the results of these analyses, we see that these models mostly fit and the fitted models are fairly reasonable within our expectation.

Figure 4.3 – Scatter plots of Observed Odds Ratio of improvement vs. Fitted Odds Ratio of improvement obtained by using the fitted model for each sample size



Note. The blue line is a diagonal line for each case.

4.4 The Precision of Estimates

As a measure of the accuracy of the estimates we used the MSE of the estimates taken about the true population values and compared the MSEs obtained by two different models, ordinary logistic regression models and logistic regression mixture models. The MSE of an estimate $\hat{\theta}$ with respect to the parameter θ is defined as

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[(\hat{\theta} - \theta)^2\right] \\ &= Var(\hat{\theta}) + (Bias(\hat{\theta}))^2. \end{aligned}$$

The bias term measures how far the mean estimates is from the true value and the variance term measures how far each estimator is from the mean estimates. Since the MSE decomposes into a sum of the bias and variance of the estimator, both quantities are important and need to be as small as possible to achieve good estimation performance. It is common to have a situation where (1) for simple models, the bias increases very quickly, while (2) for complex models, the variance increases very quickly. This basic tradeoff arises in a wide variety of settings, as it seems to be fundamental to the various nature of generalization of any data that involve an unknown mixture of regular and random elements.

From the previous simulation results in Section 2.3.2, we can expect that the MSE of the estimate $\hat{\beta}_1$ decreases as sample size increases in logistic regression mixture models because the variance and the expected bias of the estimates decreased as sample size increases (Table 2.2). The MSE, the variance, and the bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ based on one thousand replicates are summarized in Table 4.5 and Table 4.6, respectively.

Comparing the precision of ordinary logistic regression and logistic regression mixture model, in particular, the variance of the estimates $\hat{\beta}_1$ is larger in logistic mixture models than ordinary models. It is because the variance depends on the estimate of π in the mixture models. In the same context of this dependence, the variance of $\hat{\beta}_1$ is larger as the corresponding mixing proportion π is smaller. Meanwhile, the estimates in mixture models have smaller biases for each sample size. Thus, we can conclude that the large MSE in the mixture models is mainly caused by the variance instead of the bias of estimates for most of the configuration considered. However, we can see the mixture models have smaller MSE values than ordinary logistic regression models when sample size is sufficiently large (n > 400) and the mixing proportion is greater than 0.3. The bold lines in Table 4.6 show that there is any improvement in terms of the MSE of $\hat{\beta}_1$ under the mixture models in these configurations (n > 400 and π > 0.3). As well, Figure 4.4 illustrates the precision of $\hat{\beta}_1$ (MSE, variance, and bias) according to the generated population given mixing proportions for each sample size. From these results, we conducted our power studies based on the relatively large sample sizes (n = 200, 400, and 1,000; n_x = 50, 100, 250).

Table 4.5 – Summary of the MSE of $\hat{\beta}_0$ with the standard error (in parentheses) of estimates under the logistic regression mixture population in each case: five sample sizes and five mixing proportions are considered (n = 100, 200, 400, 1,000, and 2,000; π = 0.1, 0.3, 0.5, 0.7, and 0.9), $\beta_0 = 0$

Sample Size (n)[3]	π	MSE (SE) of $\hat{\beta}_0$		Var($\hat{\beta}_0$)		Bias($\hat{\beta}_0$)	
		Ordinary [1]	Mixture [2]	Ordinary	Mixture	Ordinary	Mixture
100	0.1	0.11 (0.01)	0.18 (0.01)	0.11	0.18	0.00	-0.01
	0.3	0.12 (0.01)	0.17 (0.01)	0.12	0.17	0.04	-0.03
	0.5	0.12 (0.01)	0.17 (0.01)	0.12	0.17	0.06	-0.07
	0.7	0.13 (0.01)	0.16 (0.01)	0.13	0.16	0.06	-0.06
	0.9	0.14 (0.01)	0.16 (0.01)	0.14	0.15	0.03	-0.06
200	0.1	0.06 (0.00)	0.07 (0.00)	0.06	0.07	0.01	0.00
	0.3	0.06 (0.00)	0.07 (0.00)	0.06	0.07	0.04	0.00
	0.5	0.06 (0.00)	0.07 (0.00)	0.06	0.07	0.07	-0.04
	0.7	0.07 (0.00)	0.08 (0.00)	0.06	0.07	0.07	-0.05
	0.9	0.07 (0.00)	0.08 (0.00)	0.06	0.07	0.04	-0.05
400	0.1	0.03 (0.00)	0.04 (0.00)	0.03	0.04	0.02	-0.01
	0.3	0.03 (0.00)	0.04 (0.00)	0.03	0.04	0.05	-0.02
	0.5	0.03 (0.00)	0.04 (0.00)	0.03	0.04	0.07	-0.02
	0.7	0.04 (0.00)	0.04 (0.00)	0.03	0.04	0.08	-0.02
	0.9	0.03 (0.00)	0.04 (0.00)	0.03	0.04	0.05	-0.03

Note 1. The results of the column are obtained by fitting ordinary logistic regression models

2. The results of the column are obtained by fitting logistic regression mixture models

3. The number of observations per x values $n_x = n / 4$ for each sample size.

4. Simulation results are based on 1,000 replicates per line.

Table 4.5 (Continued) – Summary of the MSE of $\hat{\beta}_0$ with the standard error (in parentheses) of estimates under the logistic regression mixture population in each case: five sample sizes and five mixing proportions are considered (n = 100, 200, 400, 1,000, and 2,000; π = 0.1, 0.3, 0.5, 0.7, and 0.9), $\beta_0 = 0$

Sample Size (n)[3]	π	MSE (SE) of $\hat{\beta}_0$		Var($\hat{\beta}_0$)		Bias($\hat{\beta}_0$)	
		Ordinary [1]	Mixture [2]	Ordinary	Mixture	Ordinary	Mixture
1,000	0.1	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.01	0.00
	0.3	0.01 (0.00)	0.02 (0.00)	0.01	0.01	0.04	-0.01
	0.5	0.02 (0.00)	0.01 (0.00)	0.01	0.01	0.07	-0.01
	0.7	0.02 (0.00)	0.01 (0.00)	0.01	0.01	0.07	-0.01
	0.9	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.04	-0.01
2,000	0.1	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.02	0.01
	0.3	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.04	0.00
	0.5	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.07	0.00
	0.7	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.07	0.00
	0.9	0.01 (0.00)	0.01 (0.00)	0.01	0.01	0.04	-0.01

Note 1. The results of the column are obtained by fitting ordinary logistic regression models
2. The results of the column are obtained by fitting logistic regression mixture models
3. The number of observations per x values $n_x = n / 4$ for each sample size.
4. Simulation results are based on 1,000 replicates per line.

Table 4.6 – Summary of the MSE of $\hat{\beta}_1$ with the standard error (in parentheses) of estimates under the logistic regression mixture population in each case: five sample sizes and five mixing proportions are considered (n = 100, 200, 400, 1,000, and 2,000; π = 0.1, 0.3, 0.5, 0.7, and 0.9), $\beta_1 = 1$

Sample Size (n)[3]	π	MSE (SE) of $\hat{\beta}_1$		Var($\hat{\beta}_1$)		Bias($\hat{\beta}_1$)	
		Ordinary [1]	Mixture [2]	Ordinary	Mixture	Ordinary	Mixture
100	0.1	0.90 (0.01)	3.47 (0.39)	0.03	3.11	-0.93	-0.60
	0.3	0.68 (0.01)	3.19 (0.39)	0.04	3.18	-0.80	-0.13
	0.5	0.47 (0.01)	2.93 (0.35)	0.04	2.86	-0.65	0.26
	0.7	0.26 (0.01)	2.01 (0.30)	0.05	1.88	-0.46	0.36
	0.9	0.11 (0.00)	1.08 (0.17)	0.08	0.96	-0.17	0.34
200	0.1	0.89 (0.01)	2.47 (0.28)	0.02	2.12	-0.94	-0.60
	0.3	0.68 (0.01)	3.42 (0.37)	0.02	3.40	-0.81	0.12
	0.5	0.47 (0.01)	2.48 (0.31)	0.02	2.35	-0.67	0.35
	0.7	0.25 (0.00)	1.83 (0.27)	0.02	1.67	-0.48	0.40
	0.9	0.07 (0.00)	0.56 (0.09)	0.02	0.48	-0.20	0.29
400	0.1	0.89 (0.01)	2.54 (0.31)	0.01	2.26	-0.94	-0.53
	0.3	0.67 (0.00)	3.66 (0.32)	0.01	2.64	-0.82	0.14
	0.5	0.46 (0.00)	1.83 (0.27)	0.01	1.76	-0.67	0.27
	0.7	0.25 (0.00)	0.59 (0.12)	0.01	0.57	-0.49	0.14
	0.9	0.06 (0.00)	0.15 (0.01)	0.02	0.14	-0.21	0.12

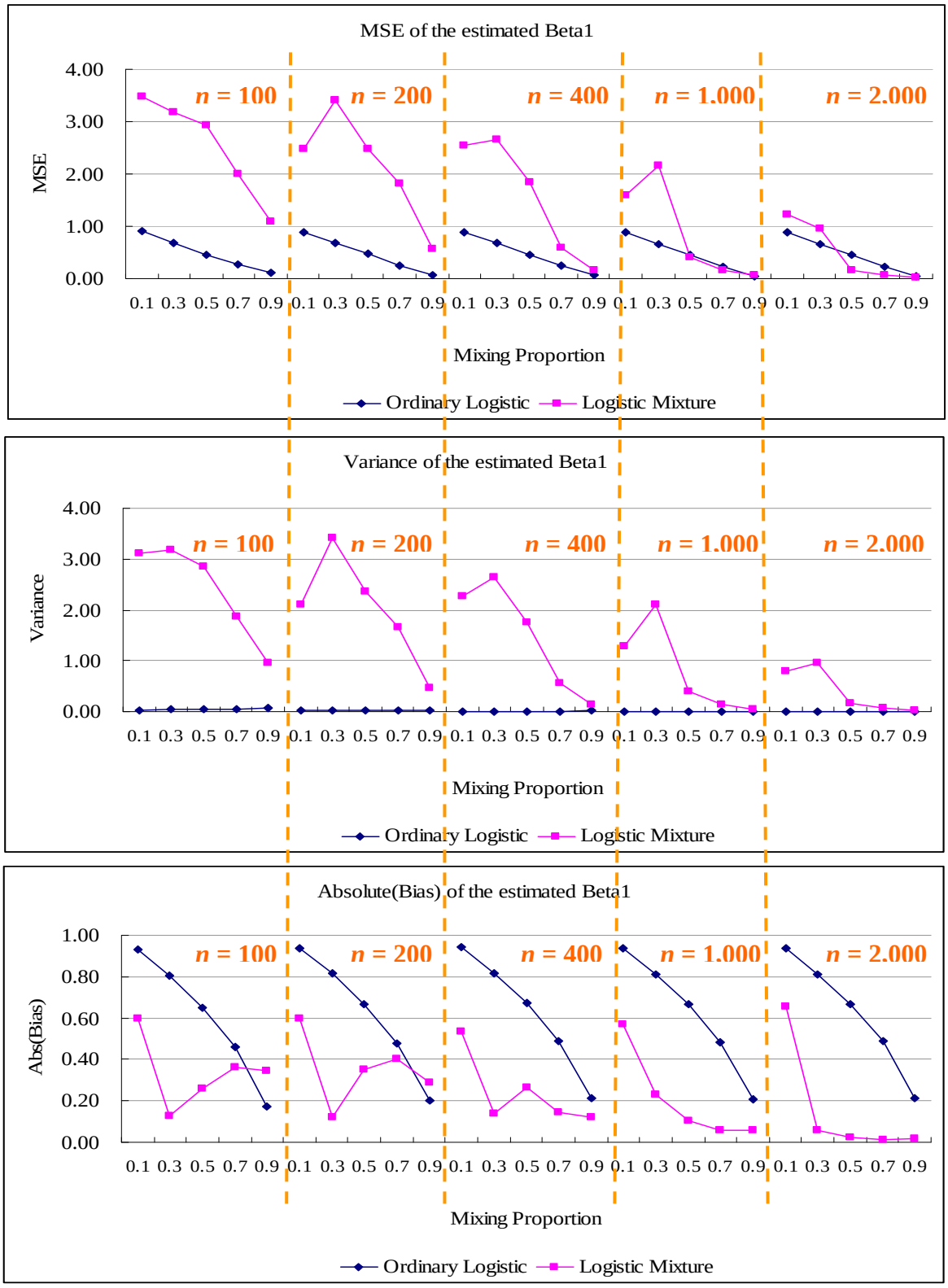
Note 1. The results of the column are obtained by fitting ordinary logistic regression models
2. The results of the column are obtained by fitting logistic regression mixture models
3. The number of observations per x values $n_x = n / 4$ for each sample size.
4. Simulation results are based on 1,000 replicates per line.

Table 4.6 (Continued) – Summary of the MSE of $\hat{\beta}_1$ with the standard error (in parentheses) of estimates under the logistic regression mixture population in each case: five sample sizes and five mixing proportions are considered (n = 100, 200, 400, 1,000, and 2,000; π = 0.1, 0.3, 0.5, 0.7, and 0.9), $\beta_1 = 1$

Sample Size (n)[3]	π	MSE (SE) of $\hat{\beta}_1$		Var($\hat{\beta}_1$)		Bias($\hat{\beta}_1$)	
		Ordinary [1]	Mixture [2]	Ordinary	Mixture	Ordinary	Mixture
1,000	0.1	0.88 (0.00)	1.60 (0.19)	0.00	1.28	-0.94	-0.57
	0.3	0.66 (0.00)	2.15 (0.29)	0.00	2.10	-0.81	0.23
	0.5	**0.45 (0.00)**	**0.41 (0.05)**	**0.00**	**0.40**	**-0.67**	**0.10**
	0.7	**0.24 (0.00)**	**0.15 (0.01)**	**0.00**	**0.15**	**-0.48**	**0.06**
	0.9	0.05 (0.00)	0.06 (0.00)	0.01	0.06	-0.21	0.05
2,000	0.1	0.88 (0.00)	1.22 (0.15)	0.00	0.78	-0.94	-0.66
	0.3	0.66 (0.00)	0.96 (0.16)	0.00	0.96	-0.81	0.06
	0.5	**0.45 (0.00)**	**0.17 (0.01)**	**0.00**	**0.17**	**-0.67**	**0.02**
	0.7	**0.24 (0.00)**	**0.07 (0.00)**	**0.00**	**0.07**	**-0.49**	**0.01**
	0.9	**0.05 (0.00)**	**0.03 (0.00)**	**0.00**	**0.03**	**-0.21**	**0.02**

Note 1. The results of the column are obtained by fitting ordinary logistic regression models
2. The results of the column are obtained by fitting logistic regression mixture models
3. The number of observations per x values $n_x = n / 4$ for each sample size.
4. Simulation results are based on 1,000 replicates per line.
5. Bold lines represent the cases having an improvement in fitting mixture models.

Figure 4.4 – The MSE, Variance, and Bias of the estimate $\hat{\beta}_1$: compared in the context of ordinary logistic regression models and logistic regression mixture models



Chapter 5

Discussion and Conclusions

In this dissertation two approaches were used to compare the performance of the LRT to evaluate an association between a quantitative predictor and a dichotomous response. One is the ordinary logistic regression model, and the other is the logistic regression mixture model defined in Section 1.3. We developed the LRT to detect the relationship between a quantitative explanatory variable and a dichotomous response variable based on these two methods. The EM algorithm was utilized to find the MLEs of the parameters in the mixture model.

Before we conducted our power analyses, we investigated the null distribution of LRT statistics to infer the critical value for the test. To verify the conjecture that the asymptotic null distribution reduces to $0.5\chi_1^2 + 0.5\chi_2^2$, we obtained the empirical null distribution and the fitted null distribution of the statistics by simulation studies. Based on the simulation results, we found that our conjecture was correct and concluded to use the asymptotic null distribution for power study.

From the power study we simulated a situation where the population consists of a mixture for whom there is an association and a fraction of individuals for whom there is no association between the quantitative predictor and the binary response. We thus evaluated the power to detect the association in the ordinary logistic regression and the logistic regression mixture

models. The mixture model resulted in the improvement of approximately 20% (on average) in power over ordinary logistic models. The improvement increases as the value of β_1 and the mixing proportion π decrease, especially for sample size of 1,000. This may due to the fact that a bigger value of β_1 and a larger mixing proportion π resulted in a greater power in both approaches, In the context of this, one would expect that the performance of the test will be good enough even in the ordinary logistic regression model when the value of β_1 is large and/or the mixing proportion of the population for whom there is an association is large. Additionally, we obtained the fitted model for the difference in power between ordinary logistic regression and logistic regression mixture models. As we expected, the slope β_1 and the mixing proportion π were significant in terms of the relative difference of the performance (i.e., the odds ratio of improvement in terms of power).

From the view of precision of the estimates using the logistic regression mixture models, we found that a very large sample size is needed to obtain a substantial improvement in precision of estimation under heterogeneous populations even though these estimates had less bias under the mixture models. Therefore, we could apply the logistic regression mixture model on the larger sample size than we used in this dissertation if obtaining precise estimates of the slope is the objective.

There are several limitations of this research some of which could be seen as interesting directions for future research. The first is the fact that we consider here only a special case of switching regression, namely a situation where the slope is 0.0 in one component and non zero in the other.

$$H_0 : \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \quad \text{vs. } H_a : \begin{cases} \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x & \text{with probability } \pi \\ \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_2 x & \text{with probability } 1 - \pi \end{cases}$$

A second point is that we consider only a sampling design where we have fixed values of the predictor. This would be the case in a dose response study. However, we could instead have an observational study where the predictor variable X, is a random variable. This would be the case perhaps in a study where disease susceptibility is a function of some quantitative variables in a subset of the population and unrelated to this factor in the remainder of the population. Both of these above limitations require straightforward extensions of our methodology that we used in this dissertation.

In addition, the improvement in fit obtained using the mixture model could be evaluated through the following null and alternative hypotheses. This is equivalent to the test of

$H_0 : \pi = 1$ vs. $H_1 : \pi < 1$.

$$H_0 : \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \quad \text{vs.} \quad H_a : \begin{cases} \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x & \text{with probability } \pi \\ \log\left(\frac{p}{1-p}\right) = \beta_0 & \text{with probability } 1 - \pi \end{cases}$$

One would expect that the null distribution of the LRT statistics would be $0.5\chi_0^2 + 0.5\chi_1^2$. However, upon investigating this, we noted that the null distribution was not invariant to generating models under the null hypothesis, i.e., the values of β_0 and β_1. Thus, we need to use a bootstrap sampling method to obtain P-values for each simulated sample. This is indeed an interesting but different problem that could be the basis of a future research.

References

[1] Agresti, A. *Introduction to Categorical Data Analysis*. Second Edition, Wiley.

[2] Boyles, R. A. (1983). On the convergence of the EM algorithm, *Journal of the Royal Statistical Society*, **45B**, 47-50.

[3] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, **28**, 591-605.

[4] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

[5] Day, N. E. (1969). Estimating the components of a mixture of normal distributions, *Biometrika*, **56**, 463-474.

[6] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, **39B**, 1-38.

[7] Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Annals of Statistics*, **7**, 1-26.

[8] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.

[9] Fienberg, S. E., Bromet, E. J., Follmann, D., Lambert, D., and May, S. M. (1985). Longitudinal analysis of categorical epidemiological data: A study of Three Mile Island, *Environmental Health Perspectives*, **63**, 241-248.

[10] Fisher, R. A. (1921). On the ‘probable error’ of a coefficient of correlation deduced from a small sample, *Metron*, **1**, 1-32.

[11] Good, P. I. (1979). Detection of a treatment effect when not all experimental subjects will respond to treatment, *Biometrics*, **35**, 483-489

[12] Hartley, M. J. (1978). Comment (on Quandt and Ramsey [27]), *Journal of American Statistical Association*, **73**, 738-741.

[13] Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated measures models with structural covariance matrices, *Biometrics*, **42**, 805-820.

[14] Kiefer, N. M. (1980). A note on switching regressions and logistic discrimination, *Econometrica*, **48**, 1065-1069.

[15] Lansky, D., Casella, G., McCulloch, C., and Lansky, D. (1992). Convergence and invariance properties of the EM algorithm, Proceedings of the Statistical Computing Section, *American Statistical Association*, 28-33.

[16] Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *Journal of American Statistical Association*, **83**, 1014-1022.

[17] McCulloch, C. E. and Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

[18] McLachlan, G. J. and Krishnan, T. (1996). *The EM Algorithm and Extensions*, New York: Wiley.

[19] McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

[20] Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inferences. *Biometrika*, **20A**, 175-240, 263-294.

[21] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Transactions of the Royal Society of London Series*, **A**, 231, 289-337.

[22] Pearson, K. (1894). Contributions to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society of London A*, **185**, 71–110

[23] Press, W. H. *et al*. (1995). *Numerical Recipes in C*: *The Art of Scientific Computing* (Cambridge: Cambridge University Press).

[24] Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes, *Journal of American Statistical Association*, **53**, 873-880.

[25] Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of American Statistical Association*, **55**, 324-330.

[26] Quandt, R. E. (1972). A new approach to estimating switching regressions, *Journal of American Statistical Association*, **67**, 306-310.

[27] Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regression, *Journal of American Statistical Association*, **73**, 730-738.

[28] Ramsey, J. B. (1975). Mixtures of distributions and maximum likelihood estimation of parameters contained in finitely bounded compact spaces, Econometrics Workshop Paper No. 7501, Michigan State University.

[29] Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of American Statistical Association*, **82**, 605-610.

[30] Snedecor, G. W. and Cochran, W. G. (1967). *Statistical Methods*. Ames: Iowa State University Press.

[31] Stokes, M. E., Davis, C. S., and Koch, G. G., *Categorical Data Analysis Using the SAS System*, Second Edition, Wiley.

[32] Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed model, *Biometrics*, **50**, 1171-1177.

[33] Weldon, W. F. R. (1893), On certain correlated variations in carcinus maenas, *Proceedings of the Royal Society of London*, **54**, 318–329

[34] Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Natl. Acad. Sci.* USA, **17**, 684-688.

[35] Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, **11**, 95-103.

Appendices

Appendix A. Summary of the probabilities of Y = 1 given X values for each configuration

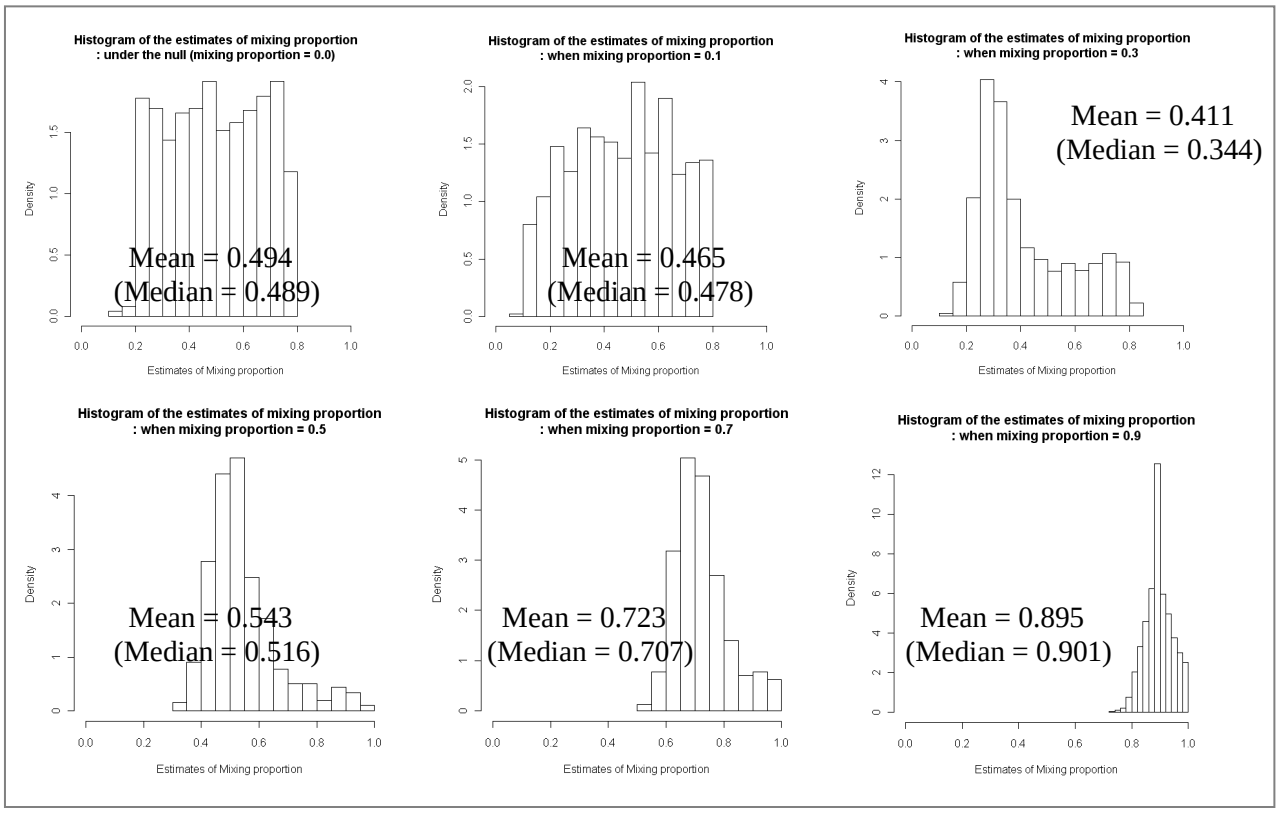
Table A.1 – Summary of the probability of Y = 1 given the value of the quantitative explanatory variable X for each parameter setting: β_0 = -2, -1, and 0; β_1 = 0.0, 0.5, 1.0, 1.5, and 2.0

		The Values of X			
β_0	β_1	X = 0	X = 1	X = 2	X = 3
	0.0[1]	0.12	0.12	0.12	0.12
	0.5	0.12	0.18	0.27	0.38
-2	1.0	0.12	0.27	0.50	0.73
	1.5	0.12	0.38	0.73	0.92
	2.0	0.12	0.50	0.88	0.98
	0.0[1]	0.27	0.27	0.27	0.27
	0.5	0.27	0.38	0.50	0.62
-1	1.0	0.27	0.50	0.73	0.88
	1.5	0.27	0.62	0.88	0.97
	2.0	0.27	0.73	0.95	0.99
	0.0[1]	0.50	0.50	0.50	0.50
	0.5	0.50	0.62	0.73	0.82
0	1.0	0.50	0.73	0.88	0.95
	1.5	0.50	0.82	0.95	0.99
	2.0	0.50	0.88	0.98	1.00

Note 1. The cases of β_1 = 0.0 represent the null hypothesis defined in this paper, i.e., there is no association between the explanatory variable X and the response variable Y

Appendix B. The distribution of the estimates of mixing proportions

Figure B.1 – The distribution of the estimates of mixing proportion according to the true value of the mixing proportion (π = 0.0, 0.1, 0.3, 0.5, 0.7, and 0.9) with the mean and median (in parenthesis) for sample size of 2,000

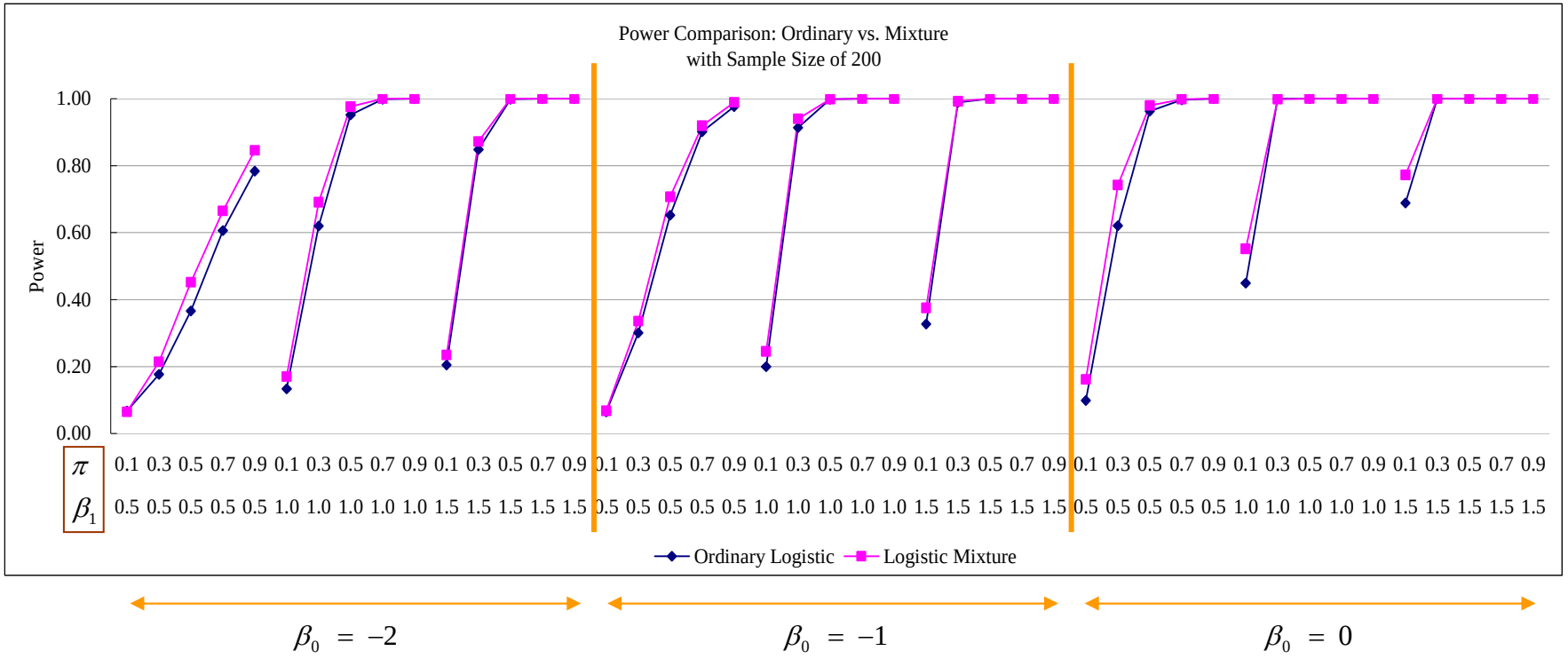


Note 1. The distributions are based on 1,000 replicates for each case.

2. As the true value of the mixing proportion increases the estimate approaches the true value, while the distribution of the estimates approximately follows a uniform distribution under the null hypothesis or small mixing proportions (π = 0.0 and 0.1).

Appendix C. Comparison of Power for Sample Size of 200 and 400

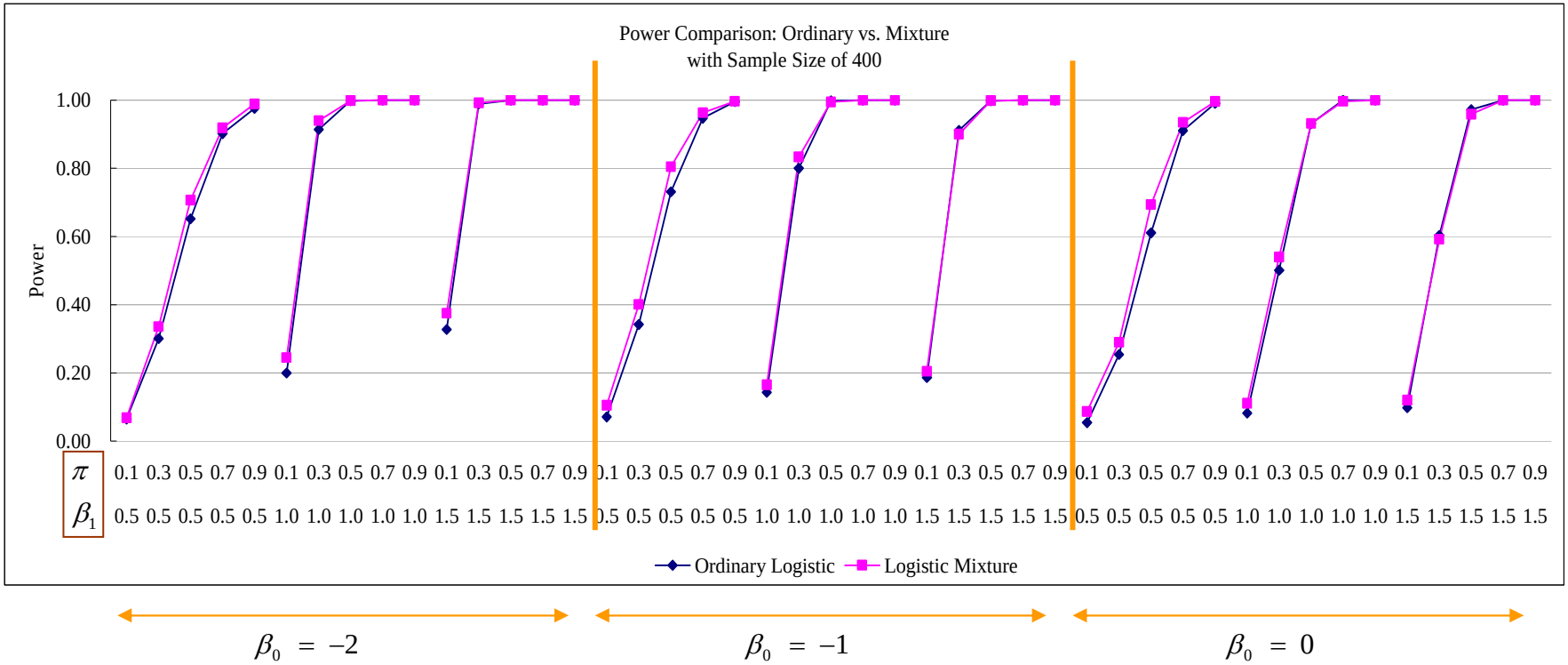
Figure C.1 – Comparison of the Power of ordinary logistic and logistic mixture models for various values of the intercept β_0 and the slope β_1 and mixing proportion π for sample size of 200.



Note. The power results are based on 1,000 replicates with sample size of 200.

Appendix C (Continued). Comparison of Power for Sample Size of 200 and 400

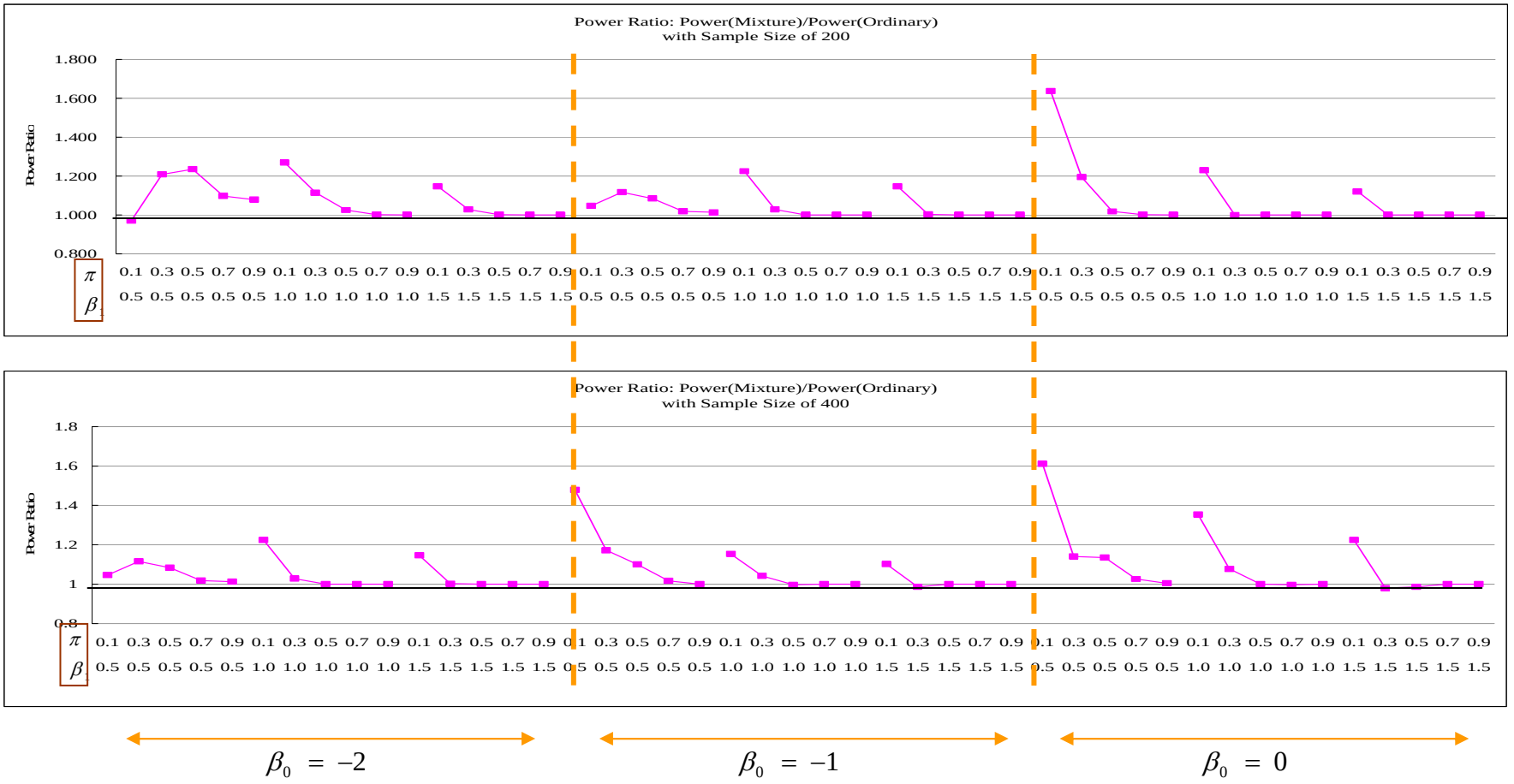
Figure C.2 – Comparison of the Power of ordinary logistic and logistic mixture models for various values of the intercept β_0 and the slope β_1 and mixing proportion π for sample size of 400.



Note. The power results are based on 1,000 replicates with sample size of 400.

Appendix D. Power Ratio of the Improvement in Power for Sample Size of 200 and 400

Figure D.1 – Power Ratio between logistic regression mixture models and ordinary logistic regression models for various values of the intercept β_0 and the slope β_1 and mixing proportion for sample size of 200 and 400



Note 1. The power results are based on 1,000 replicates with sample size of 400.

2. Power Ratio was calculated as the power of logistic regression mixture models divided by the power of ordinary logistic model

Appendix E. ANOVA output from SAS

---------- Sample Size = 200 ----------

The ANOVA Procedure

Dependent Variable: odds_ratio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	1.48769611	0.07829980	2.10	0.0694
Error	16	0.59593444	0.03724590		
Corrected Total	35	2.08363056			

R-Square	Coeff Var	Root MSE	odds Mean
0.713992	14.74785	0.192992	1.308611

Source	DF	Anova SS	Mean Square	F Value	Pr > F
beta0	2	0.08060246	0.04030123	1.08	0.3625
beta1	2	0.51754833	0.25877417	6.95	0.0067
prop	4	0.58267361	0.14566840	3.91	0.0212
beta0*beta1	4	0.13086310	0.03271577	0.88	0.4986
beta1*prop	7	0.17600861	0.02514409	0.68	0.6908

---------- Sample Size = 400 ----------

The ANOVA Procedure

Dependent Variable: odds_ratio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	4.22868619	0.26429289	3.17	0.0208
Error	13	1.08418048	0.08339850		
Corrected Total	29	5.31286667			

R-Square	Coeff Var	Root MSE	odds Mean
0.795933	22.10112	0.288788	1.306667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
beta0	2	0.21580293	0.10790146	1.29	0.3073
beta1	2	1.24168048	0.62084024	7.44	0.0070
prop	4	2.48296111	0.62074028	7.44	0.0024
beta0*beta1	4	0.48199993	0.12049998	1.44	0.2747
beta1*prop	4	0.00000000	0.00000000	0.00	1.0000

Appendix E (Continued). ANOVA output from SAS

---------- Sample Size = 1,000 ----------

The ANOVA Procedure

Dependent Variable: odds_ratio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.81834214	0.13639036	12.45	0.0002
Error	11	0.12055230	0.01095930		
Corrected Total	17	0.93889444			

R-Square	Coeff Var	Root MSE	odds Mean
0.871602	7.272714	0.104687	1.439444

Source	DF	Anova SS	Mean Square	F Value	Pr > F
beta0	2	0.42614825	0.21307413	19.44	0.0002
beta1	2	0.29289944	0.14644972	13.36	0.0011
prop	2	0.09929444	0.04964722	4.53	0.0367

---------- Overall Sample Sizes ----------

The ANOVA Procedure

Dependent Variable: odds_ratio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	21	4.59716490	0.21891261	3.41	<.0001
Error	62	3.98365891	0.06425256		
Corrected Total	83	8.58082381			

R-Square	Coeff Var	Root MSE	odds Mean
0.535749	18.97380	0.253481	1.335952

Source	DF	Anova SS	Mean Square	F Value	Pr > F
sample_size	2	0.24543214	0.12271607	1.91	0.1567
beta0	2	0.46186726	0.23093363	3.59	0.0334
beta1	2	1.51161291	0.75580646	11.76	<.0001
prop	4	2.15752536	0.53938134	8.39	<.0001
beta0*beta1	4	0.33454740	0.08363685	1.30	0.2793
beta1*prop	7	0.00000000	0.00000000	0.00	1.0000