2017

MIXTURE MODELING WITH APPLICATIONS IN ALZHEIMER'S DISEASE

Frank Appiah
University of Kentucky, frankappiah11@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.100

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Frank Appiah, Student

Dr. Richard J. Charnigo, Major Professor

Dr. Steven Browning, Director of Graduate Studies

MIXTURE MODELING WITH APPLICATIONS IN ALZHEIMER'S DISEASE

ABSTRACT OF DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy in
the College of Public Health at the
University of Kentucky

By
Frank Appiah
Lexington, Kentucky

Co-Directors: Dr. Richard Charnigo, Professor of Statistics and Biostatistics
and Dr. David Fardo, Associate Professor of Biostatistics
Lexington, Kentucky 2017

ABSTRACT OF DISSERTATION

MIXTURE MODELING WITH APPLICATIONS IN ALZHEIMER'S DISEASE

This dissertation involves an application of mixture of regression models to 114 individuals who are cognitively intact (from the Alzheimer's Disease and Neuroimaging Initiative-ADNI, data). The correct number of components in the model were estimated with the Singular BIC (SBIC), marking the first time it has been applied to such a problem. The smallest true model in conjunction with the approximation of SBIC was fixed at 1. The resulting posterior probabilities from the model were used to estimate the probability of a person transitioning and risk plots were obtained that could in principle be used by clinicians to identify patients at risk. This work also proposed a model selection criterion for mixture of regression models with application to the ADNI data. Finally simulation studies were conducted to compare the performance of the novel model selection and existing criteria.

KEYWORDS: Mixture models, cognitively intact, Alzheimer's disease, model complexity/component, selection criteria

Frank Appiah
Author's signature

April 25, 2017
Date

MIXTURE MODELING WITH APPLICATIONS IN ALZHEIMER'S DISEASE

By
Frank Appiah

Richard Charnigo
Co-Director of Dissertation

David Fardo
Co-Director of Dissertation

Steven Browning
Director of Graduate Studies

April 25, 2017
Date

This work is entirely dedicated to the memory of my late mother Paulina Appiah, my wife Leslie and my children Aiden and Naomi.

ACKNOWLEDGMENTS

I am indebted to my wife Leslie A. Appiah and my children Aiden O. Appiah and Naomi A. Appiah for their unwavering support and love throughout the years of my graduate education. I am also very appreciative of my advisor, Dr. Charnigo and co-chair Dr. Fardo, for providing great insights and motivation for this work. A similar heartfelt thanks go to my advising committee members Drs. Abner and Mays, for their patience, directions and thought provoking questions. This acknowledgement will be incomplete without giving a special thanks to the woman who with the help of God made this day a possibility. This woman is none but my late mother Mrs. Paulina Appiah, who passed away on March 19, 2014. I am very thankful for all the work and 'painful sacrifices' she made for me. Mom, I dedicate this work in its entirety to your memory. God bless.

TABLE OF CONTENTS

Acknowledgments iii

Table of Contents iv

List of Figures v

List of Tables vi

Chapter 1 Introduction 1
1.1 Definition of Mixture Models and Examples 1
1.2 Review of Applications 4
1.3 Review of Mixture Model Applications to Alzheimer's Disease 15
1.4 Review of Existing Model Selection Criteria and Mixture of Regression 19

Chapter 2 An Application of A Bivariate Normal Mixture Model 22
2.1 Motivation and Objectives 23
2.2 Methodologies, Cognitive Assessment and Review of Related Concepts 27
2.3 Results and Discussion 38
2.4 Limitations and Future Directions 47
2.5 Acknowledgements 51

Chapter 3 Application of Mixture of Linear Regressions Models And the Approximate Singular Bayesian Information Criterion 76
3.1 Introduction 76
3.2 General Overview of Mixture of Regression Models With An Illustration 77
3.3 Overview of Primary Objectives 82
3.4 Methodologies and Overview of AIC 84
3.5 Results and Discussions 89
3.6 Limitations and Future Directions 104
3.7 Illustrative Computations for A1-A3 in Drton and Plummer(2016) for Mixture of Regression Models 135

Chapter 4 A Singular Flexible Information Criterion From A Mixture of Linear Regressions Perspective 158
4.1 Introduction 158

4.2 Deduction of Within and Between Covariance Matrices 159
4.3 Definition and Derivation of SFLIC 163
4.4 Consistency of SFLIC 166
4.5 Application of SFLIC to the ADNI data 169

Chapter 5 Simulation Studies 170
5.1 Introduction 170
5.2 Overview of Approach 170
5.3 Simulation Design and Results 172

Chapter 6 Supplementary Chapter 178
6.1 Introduction 178
6.2 Review of Other Comparable Models Fitted As Part of This Work . . 178

Vita 188
Oral Presentations (Delivered by First author) 188
Poster Presentations (Delivered by First author) 189
External Funding 190
Awards 192
Professional Organizations 192
Leadership Skills 192

LIST OF FIGURES

2.1 Original Biomarkers 30
2.2 Derived Biomarker Histogram 31
2.3 BNM Component Membership Probability Distribution 53
2.4 BNM Predicted Components Scatterplot 54
2.5 BNM Component Kaplan Meier Plots 55
2.6 Component contour plots 56
2.7 The empty circles identify potential outliers among the eight participants 57
2.8 Joint Biomarker contour plot 58
2.9 Joint Biomarker density plot 59
2.10 ROC plot. AUC: Area under the curve 60
2.11 Risk Boundaries 61
2.12 Risk Boundaries 62
2.13 Risk Differentiation Boundaries 63
2.14 BNM Four Component Membership Probability Distribution 64
2.15 BNM Four Component Kaplan Meier Plots 65
2.16 BNM Predicted Four Component Plots 66
2.17 Comparison of rtaubeta between CN and MCI/AD groups 67
2.18 Comparison of rptaubeta between CN and MCI/AD groups 68
2.19 Contours embedded on risk components 69

3.1 Relationship between biomarkers ratios and race. + represents the grand mean 106
3.2 Relationship between biomarker ratios and Apoe4 + represents the grand mean 107
3.3 Relationship between biomarkers, race and Apoe4. + is grand mean. . . 108
3.4 Joint Distribution of rtaubeta and rptaubeta Given Race 109
3.5 Posterior probability plot with race as predictor in the mixture regression model indicates that the model is well separated 110
3.6 Individual biomarker ratios grouped into different risk regions. Here race is the predictor variable 111
3.7 The survivability of the two groups over time given in weeks. Here race is the predictor variable 112
3.8 Risk strata for Blacks from the posterior probabilities obtained from the mixture of linear regression with race as covariate 113

3.9 Risk strata for Whites from the posterior probabilities obtained from the mixture of linear regression with race as covariate 114
3.10 Joint Distribution of rtaubeta and rptaubeta Given Apoe4 115
3.11 Posterior probability plot indicates that the mixture of regression model with Apoe4 as predictor model is comparatively less well separated . . . 116
3.12 Individual biomarker ratios grouped into different risk regions. Here Apoe4 is the predictor variable. 117
3.13 The survivability of the two groups over time given in weeks. Here Apoe4 is the predictor variable. 118
3.14 This corresponds to the risk strata for Apoe4 carriers derived from the posterior probabilities obtained from the mixture of linear regression with Apoe4 as predictor variable. 119
3.15 This corresponds to the risk strata for none-Apoe4 carriers derived from the posterior probabilities obtained from the mixture of linear regression with Apoe4 as predictor variable. 120
3.16 Posterior probability plot indicates that the model with Apoe4 and race as predictors is comparatively less well separated 121
3.17 Individual biomarker ratios grouped into different risk regions. Here Apoe4 and race are the predictor variables. 122
3.18 The survivability of the two groups over time given in weeks. Here Apoe4 and race are the predictor variables. 123
3.19 Risk strata for black who are apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates 124
3.20 Risk strata for blacks who are none-apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates 125
3.21 Risk strata for whites who are apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates 126
3.22 Risk strata for whites who are none-apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates 127

5.1 Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the race mixture of regression model. Here the true mixture is k=2 175
5.2 Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the apoe4 mixture of regression model. Here the true mixture is k=2 176

5.3 Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the race and apoe4 mixture of regression model. Here the true mixture is k=2 177

LIST OF TABLES

2.1 Table of Demographics (n=3082), standard deviation(sd), lower and upper quartiles (Q1 and Q3), percentages may not add to 100 due to rounding 70
2.2 Table of Demographics (n=114), standard deviation(sd), lower and upper quartiles (Q1 and Q3), rptaubeta is the ratio of PTAU181P to ABETA142 and rtaubeta is the ratio of TAU to ABETA142 71
2.3 Selection of model complexity with three criteria 71
2.4 Medium/ high are respectively the component two/three rounded estimated membership probabilities for the hard and soft classification. c is the concordance. ** significant at 0.01 level and * significant at 0.05 level. Sample size is n=114. HR: estimated hazard ratio, SE: standard error of log(HR), 95% CI: 95% confidence interval, GPH Test: Global proportional hazard test 72
2.5 Assessing CN status of Participants. Bolded observations are potential outliers see Figure 2.7 72
2.6 Correlation Coefficient Matrix 73
2.7 Medium/ high are respectively the component two/three rounded estimated membership probabilities for the hard and soft classification. c is the concordance. ** significant at 0.01 level and * significant at 0.05 level. Sample size is n=106. HR: estimated hazard ratio, SE: standard error of log(HR), 95% CI: 95% confidence interval, GPH Test: Global proportional hazard test 74
2.8 Component estimated parameters. SE: standard error based on Bootstrap sampling with $B = 1000$ in R mixtools package,C1-C3 are components 1 through 3. 75
2.9 Contingency table to compare component predicted values to the true values 75
2.10 Sensitivity and Specificity 75

3.1 Selection of model complexity with three criteria. To calculate the learning coefficient sBIC, we are assuming the non-redundant one component. Numbers shown are differences between the information criteria at one component versus the information criteria at two component. 128

3.2 Estimates of the regression models within each component.Race is the only predictor variable in the model. Race is an indicator variable for Caucasian (coded as 1) and the referent group is black (coded as zero) *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level 129
3.3 High|Low risk is component one estimated probability for the soft and hard classification models. c is the concordance.HR is the hazard ratio. Race is the only predictor. When model was adjusted for education, MMSE, and age only prop.SBICO1 and rprop.SBICO1 were significant. This significance disappeared when Apoe4 was adjusted for (results not shown) 130
3.4 Estimates of the regression models within each component. Apoe4 is the only predictor variable in the model.Apoe4 is coded as 1 for carriers of the gene and 0 for non carriers. *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level 131
3.5 High|Low risk is component one estimated probability for the soft and hard classification models. c is the concordance. HR is the hazard ratio. Apoe4 is the only predictor variable. When model was adjusted for education, MMSE, race and age only prop.SBICO12 and race or rprop.SBICO12 were significant (results not shown) 132
3.6 Estimates of the mixture of regression models within each component. *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level 133
3.7 prop.SBIC013 is component one estimated probability and rprop.SBIC013 is component one hard classification. c is the concordance. HR is the hazard ratio. Race and Apoe4 are the predictors. When model was adjusted for education, MMSE, race and age only Apoe4 was significant (results not shown) 134

Chapter 1 Introduction

1.1 Definition of Mixture Models and Examples

Let $X_1, ..., X_n$ denote a random sample of size n, where X_j is a p-dimensional random vector with probability density function $f(x_j)$ on R^p. Then we define a k component finite mixture model:

$$f(x_j) = \sum_{i=1}^{k} p_i f(x_j|\theta_i) \tag{1.1}$$

where the functions $f(x_j|\theta_i)$ are called the component densities of the mixture and the quantities $p_1, .., p_k$ are the mixing proportions with $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$. The mixture model defined in this context assumes a known number k component(s). However, in reality the number of components is inferred from the data and so are the mixing proportions and the component- specific parameters. If we allow k to increase with the sample size n, then the resulting model is called a mixture sieve [19].

Example 1. Charnigo et. al.[27] modeled birthweight distribution of a population of white singleton infants born to heavily smoking mothers in the United States. In this study the number of components in the model was chosen with Flexible Information Criterion (FLIC), a model selection criterion that imposes a penalty based on sample size and data configuration. FLIC and Bayesian Information Criterion (BIC) chose a 4-component normal mixture as a good fit to the data. The resulting model structure

is given below and the plot corresponding to the model is shown in Figure 1 of Charnigo et. al. (2010):

$$0.009f(x,872,247)+0.231f(x,2890,726)+0.707f(x,3165,403)+0.054f(x,3821,365),$$

where the two numbers in each component-specific density are estimated mean and standard deviation. The first component of the model describes the distribution of extremely low and very low birthweight (ELBW and VLBW) infants, component 2 describes mostly moderate low birthweight (MLBW) and normal birthweight (NBW) infants with some VLBWs as well as high birthweight (HBW) cases. The third component is similar to the mean component in a contaminated[31] and a 2-component model[32]. The fourth component consists of NBW and HBW cases. The complexity underlying the birthweight distribution as outlined cannot be adequately captured with a single or perhaps fewer than four component Gaussian model. In the same vein, fixing the number of components a priori may not yield reasonable results because the appropriate complexity may vary across geographic and demographic boundaries.

Example 2. Santago et. al. [26] applied two versions of finite mixture models to automatically quantify single valued pixels of brain tissue types from Magnetic Resonance Imaging (MRI). The brain data consist of four adjacent images. In the first model, no partial volume effect was assumed; the errors were Gaussian with homoscedastic variance and the mixing parameters summed to 1. The model consists of three brain tissues; cerebrospinal fluid (CSF), white matter (WM) and gray matter(GM) and is thus referred to as the three tissue model. The three tissue model

was stated as:

$$p(v) = \sum_{t \epsilon T_3} Pr[t]_{P_{v|t}}(v|t), T_3 = \{CSF, WM, GM\}, \sum Pr[t] = 1$$

where v is the pixel intensity and $P_{v|t}$ is the component-specific density.

A second model called the six tissue model, assumed a partial volume effect. However, the error terms were normally distributed. The six tissue model was defined as:

$$p(v) = \sum_{t \epsilon T_6} Pr[t]_{P_{v|t}}(v|t), T_6 = \{CSF, WM, GM, CW, CG, GW\}, \sum Pr[t] = 1$$

where CW, CG and GW represent combinations of the aforementioned three tissue types. The resulting parameters in the models were estimated with tree annealing algorithm which minimizes $\|p(v) - h(v)\|^2$ where $h(v)$ is the histogram of the data and $p(v)$ is the model. Annealing algorithm is suitable for minimizing continuous functions with only the known form of the function and not its derivatives. The quantity of each brain material type was estimated with either parameter or Bayesian quantification method. The Bayesian approach quantifies each tissue type by relying on the normal model assumptions, the estimated mean and variance to optimize class decision boundaries. On the other hand, quantifying tissues directly with estimated parameters from the model was termed parameter quantification. The Bayesian approach was found to be more accurate and the three tissues model more consistent in its fidelity to data as shown in Figure 5a in Santago et' al. (1993).

1.2 Review of Applications

Mixture models have been used since the 19th century when they were applied by Karl Pearson[13] in the analysis of crab morphometry [4]. Many novel applications have been published in the fields of genetics, finance and engineering using mixture models. Extensive discussions on the application of mixture models have been well documented in Titterington et. al.[16] and Lindsay [17]. For instance a geneticist might be interested in knowing if a disease population is homogeneous in situations where a disease is caused in one group of individuals by one locus and in another group of individuals by another locus.[4] Since the incipience of mixture models, many methodological improvements have been proposed and justified. For example, Mclachlan [3] used a bootstrap method to estimate the number of components in normal mixture models. The bootstrap method was applied to a mixture based on yields from seven barley types grown in 6-blocks. The purpose was to determine the number of mixture components when barley yields were clustered. The method ultimately reduced the problem to choosing between two or three component mixtures. At K=19 bootstrap replications the p-value obtained suggested that the two-component model was more appropriate.

Chen and Chen[4] have shown that under the null hypothesis of homogeneity and under some regularity conditions including a compact parameter space, the likelihood ratio test statistic of a mixture model has an asymptotic distribution of $(\sup_\theta W^+(\theta))^2$, where W is a Gaussian process with mean 0 and variance 1. Other theoretical developments in the field of mixture models include the modification of the log likelihood ratio test by Chen et al. [5] and the D-testing [6]. Chen et al.[5] modified the log like-

lihood of the finite mixture model by adding a penalty $C\log(4\gamma(1-\gamma))$. They specified two related motivations motivations for this modification; lack of identifiability property in mixture models under the null hypothesis and the boundary issues regarding the mixing proportion γ possibly being zero. Under the null hypothesis, the mixing parameters are estimated as $1/2$ leading to no effective penalty. Thus the penalty only affects the alternative hypothesis model(heterogeneous model). The constant C in the penalty is used to control the modification so that for a bounded kernel density, one may choose C to be $C = \log(M)$ where M comes from the parameter space defined as $[-M, M]$. Under the null hypothesis and regularity conditions, Chen et al[5] obtained the asymptotic distribution of the modified likelihood ratio test (MLRT) as $0.5\chi_0^2 + 0.5\chi_1^2$ where χ_0^2 denotes a degenerate distribution at zero. Furthermore, a simulation study under the normal and the Poisson mixture models revealed that when the Kullback-Leibler information is small, MLRT and the Neyman Scott test [8] performed about the same and the method proposed by Mclachlan[3] performed poorly. However for a large Kullback -Leibler information, the modified LRT was preferable to the competing methods. Under the normal model assumption, Davies [7] method was precise in terms of p-value estimates but less powerful in comparison to the MLRT.

Charnigo et. al.[6] studied a new testing procedure for choosing the number of components in finite mixture models. Their method relies on the Euclidean L^2 distance between the competing models specified at the null and the alternative respectively. Appealing features of the test include the emphasis it places on wider differences between the density functions at the null and alternative hypotheses. In

addition it has a closed form expression with respect to the parameter estimates when the mixture components are from standard parametric families. Another strength of the test is its independence of the data given parameter estimates. As a result, testing can be performed in the absence of the original data if the parameter estimates are known.

Let $X_1 \ldots X_n$ be a simple random from the mixture distribution $\sum_{i=1}^{k} p_i f(x|\theta_i)$ where $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$, and $\{f(x|\theta)|\theta \epsilon \Theta \subset L^2\}$ is a family of probability density function associated with a scalar or vector parameter θ, then the D-test statistic can be defined as

$$d(k,n) = \int \left[\sum_{i=1}^{k} \hat{p}_i f(x|\hat{\theta}_i) - f(x|\hat{\theta}_0)\right]^2 dx = \int \left[\sum_{i=1}^{k} \hat{p}_i f(x|\hat{\theta}_i)\right]^2 dx$$

where $\hat{p_0} = -1$ and $\hat{\theta}_0$ estimates the single parameter under the null hypothesis.The corresponding closed form expressions for univariate and multivariate normal cases are presented below:

$$d(k,n) = \sum_{i=0}^{k} \sum_{j=0}^{k} \frac{\hat{p}_i \hat{p}_j}{\sqrt{2\pi(\hat{\sigma_i^2} + \hat{\sigma_j^2})}} \exp\left[-\frac{1}{2} \frac{(\hat{\mu_i} - \hat{\mu_j})^2}{\hat{\sigma_i^2} + \hat{\sigma_j^2}}\right]$$

$$d(k,n) = \sum_{i=0}^{k} \sum_{j=0}^{k} \frac{\hat{p}_i \hat{p}_j}{2^d \pi^{d/2}} \exp\left[-\frac{1}{2} ||\hat{\mu}_i - \hat{\mu}_j||^2\right]$$

assuming an identity covariance matrix within each component in the latter formula.

Let $X_1 \ldots X_n$ be iid under null hypothesis $H_0 : X_1 \sim f(x|\theta_0)$, for θ_0 an interior point in the compact parameter space Θ. Then under the five regularity conditions

assumed by Charnigo et. al. [6] the following convergence rates were obtained regarding maximum likelihood parameter estimation with $k = 2$ and $p_1 \geq p_2$:

$$\hat{p_1}(\hat{\theta_1} - \theta_0) + \hat{p_2}(\hat{\theta_2} - \theta_0) = Op(n^{-1/2})$$

$$\hat{p_1}(\hat{\theta_1} - \theta_0)^2 + \hat{p_2}(\hat{\theta_2} - \theta_0)^2 = Op(n^{-1/2})$$

Note that assuming the wrong model (that is two component when there is really only one) yields slower or no convergence: $\hat{\theta_1}$ is $n^{1/4}$-consistent while $\hat{\theta_2}$ is not consistent. To see this note that $p_1 \geq p_2, \rightarrow \hat{p_1} \geq \hat{p_2}$ and we have that

$$Op(n^{-1/2}) = \hat{p_1}(\hat{\theta_1} - \theta_0)^2 + \hat{p_2}(\hat{\theta_2} - \theta_0)^2 \geq \hat{p_1}(\hat{\theta_1} - \theta_0)^2 = Op(n^{-1/2}) \Rightarrow (\hat{\theta_1} - \theta_0)^2 = Op(n^{-1/2})$$

It follows that

$$\sqrt{(\hat{\theta_1} - \theta_0)^2} = \sqrt{Op(n^{-1/2})} \Rightarrow |\hat{\theta_1} - \theta_0| = O_p(n^{-1/4})$$

We also note that

$$Op(n^{-1/2}) = \hat{p_1}(\hat{\theta_1} - \theta_0)^2 + \hat{p_2}(\hat{\theta_2} - \theta_0)^2 \geq \hat{p_2}(\hat{\theta_2} - \theta_0)^2 = Op(n^{-1/2})$$

$$\Rightarrow (\hat{\theta_2} - \theta_0)^2 = Op(n^{-1/2})\frac{1}{\hat{p_2}}$$

But since $\hat{p_1} \geq \hat{p_2}$, we get stuck because $\hat{p_2} \leq 1/2 \Rightarrow \frac{1}{\hat{p_2}} \geq 2$ with no lower bound. Thus to find a bound for this expression we infer from what was established above that $(\hat{\theta_1} - \theta_0) = Op(n^{-1/4})$ and noting that $\hat{p_1} \leq 1$, it follows that

$$\hat{p_1}(\hat{\theta_1} - \theta_0) \leq 1 \times Op(n^{-1/4}) = Op(n^{-1/4})$$

It can be deduced that $X_n = Op(n^{-1/2}) \Rightarrow X_n = Op(n^{-1/4})$. Making use of the latter relation we further deduce that

$$Op(n^{-1/2}) = \hat{p}_1(\hat{\theta}_1 - \theta_0) + \hat{p}_2(\hat{\theta}_2 - \theta_0)$$

$$\Rightarrow Op(n^{-1/4}) = \hat{p}_1(\hat{\theta}_1 - \theta_0) + \hat{p}_2(\hat{\theta}_2 - \theta_0)$$

$$\Rightarrow Op(n^{-1/4}) - \hat{p}_1(\hat{\theta}_1 - \theta_0) = \hat{p}_1(\hat{\theta}_1 - \theta_0) + \hat{p}_2(\hat{\theta}_2 - \theta_0) - \hat{p}_1(\hat{\theta}_1 - \theta_0)$$

$$\Rightarrow \hat{p}_2(\hat{\theta}_1 - \theta_0) = Op(n^{-1/4})$$

Using Taylor expansion, the convergence rate $d(2,n) = Op(n^{-1})$ was then obtained. The authors also showed that the testing procedure was consistent against a fixed alternative.

The convergence rates above paved the way for properly rescaling critical values for $d(2,n)$ as elaborated below. Having that $d(2,n) = Op(n^{-1})$ under the null and $d_{\alpha;N(0,1)}$ the corresponding critical value of $d(2,n)$ for $N(0,1)$ then under the null hypothesis and $f(x|\theta_0) = N(0,1)$,

$$P(d(2,n) \geq d_{\alpha;N(0,1)}) \approx \alpha$$

and therefore conclude that $d_{\alpha;N(0,1)} = O(n^{-1})$. If we assume more than was in [6] and let $nd(2,n)$ converge in distribution to F under the null hypothesis and $f(x|\theta_0) = N(0,1)$, then we can have $P(nd(2,n) \geq F_{0.95}) \longrightarrow 0.05$, where $F_{0.95}$ is the 95^{th} percentile of F. It follows that:

$$P(d(2,n) \geq n^{-1}F_{0.95}) \approx 0.05$$

and

$$d_{\alpha;N(0,1)} \approx n^{-1} F_{0.95}$$

For example, if $d_{\alpha;N(0,1)} = 0.2$ when $n = 50$, then it follows from the previous set up that:

$$0.2 \approx 50^{-1} F_{0.95} \Rightarrow F_{0.95} \approx 10$$

Now using the estimated critical value at $n = 50$, we can estimate $d_{\alpha;N(0,1)}$ at another n, say $n = 100$ as follows:

$$d_{\alpha;N(0,1)} \approx n^{-1} F_{0.95} \approx \tfrac{1}{100} \times 10 = 0.1$$

Thus having estimated $d_{\alpha;N(0,1)}$ at one n, the $Op(n^{-1})$ convergence immediately estimates $d_{\alpha;N(0,1)}$ at another n.

Moreover, for any fixed n, let $d_{\alpha;N(\mu_0,\sigma^2)}$ and $d_{\alpha;exp(\beta_0)}$ denote the level α critical values of $d(2,n)$ based on null distributions of $N(\mu_0, \sigma^2)$ and $exp(\beta_0)$. Then $d_{\alpha;N(\mu_0,\sigma^2)} \approx d_{\alpha;N(0,1)}/\sigma$ and $d_{\alpha;exp(\beta_0)} \approx \beta_0 d_{\alpha;exp(1)}$ where the parameter spaces for the assumed models are; $[-M, M]$ and $[M^{-1}, M]$ for the normal and exponential distributions respectively. The D-test performed competitively with MLRT on two simulation studies including mixture of normals on one hand and mixture of exponentials on the other.

Gene differential expression testing presents new problems that have attracted the attention of researchers in the field of mixture modeling. As new methodologies are being uncovered to test for genes that are differentially expressed, it is worth

noting that the student t-test can be and has been used to test for genes that are differentially expressed. The t-test, although simple to implement, has the disadvantage of increasing the false positive rate of the tests because of the number of genes involved.Additional concerns about T testing are as follows:

1. With small number of subjects, within-group variances are poorly estimated and results of T testing may be very sensitive to this [33] as well as any underlying non-normality of expressions levels.

2. If the expression levels for different genes are correlated then the validity of omnibus testing (i.e. analyzing numerous T test statistics together through homogeneity testing in mixture modeling) may be compromised [30].

3. Differential expression may manifest not only in a change of mean level, which is measured by T testing, but also in a change of variability which is not assessed by T testing [34].

Newton et. al [14] developed a semi-parametric hierarchical mixture model to address the problem of detecting genes that are differentially expressed while accounting for complexities of microarray data. They considered two types of prior (mixing distribution) distributions on the mean gene specific expression: one parametric (gamma distribution) and the other non-parametric (defining the mean to have a probability distribution on an equally spaced grid). The former prior actually induced a parametric model intended as a comparator to the semi-parametric model induced by the latter. The semi-parametric model performed similarly to the competing parametric model when applied to data from the Gene-logic spike-in experiment. The poorest performance in the comparison was recorded by the gene-specific T test.

When the models were tested for robustness in a case study using a data leave-out approach, the semi-parametric model identified 80% of down regulated genes compared to 61% by the parametric method. The method proposed by Newton et. al. was limited to a simple two-group comparison and ignores dependencies among genes (conditional on gene-specific parameters). Also the non parametric prior can slow down the performance of the EM or other optimization algorithms.

Besides the semi parametric model approach in the aforementioned discussion, Bayesian models amongst others have also been used to obtain useful gene expression information in recent years. For example Zhou et. al [19] used Bayesian mixture models to partition gene expression data and Alexandridis et. al. [21] developed a multi-type classification method for gene selection and tissue sample separation.

In particular, Newton et. al. [28] developed EBarrays in R to compute dual character posterior probabilities for detecting patterns of genes and condition-specific expected values. This method was believed to capture relevant sources of variations in a high-dimensional expression profile and thus considered superior to some existing methods such as the paired sample T test. EBarray also require fewer replications of microarray data and does not require permutation. The undergirded assumptions of EBarray are as follows:

a) Parametric observations component (log-normal or gamma)

b) Parametric mean component (conjugate to observation component)

c) Constant coefficient of variation

d) Only marginal information (rather than among-gene dependence) is relevant

The log-normal-normal hierarchical (LNN) model in EBarray package when ap-

plied to mammary epithelial tissue from a rat model of breast cancer , identified 92.7% of the genes as equivalently expressed. EBarray models are however limited by the assumptions underlying its operations. For instance it will under perform if the constant variance assumption is violated or if the data deviate from the log-normal-normal assumption. To increase the flexibility of the model, Newton et. al. suggested using a nonparametric mean component approach. It is worth noting that Speed [29] used a similar approach as Newton et. al.[28] but did not assume constant variance of gene-specific expressions. As a result, Speed's flexible model may provide yet another avenue to model varying mixture components.

Omnibus tests [30] are also extensions of mixture models designed to overcome difficulties in simultaneous testing of differentially expressed genes. Suppose we decide to test for gene differential using t-test and adjust for multiplicity using the Scheffe, Kolmogorov and Tukey. Such a test will have low power for detecting differential expression due to the conservative nature of the pairwise comparison methods.

Omnibus tests combines the D-test [27] and modified likelihood ratio test [5] to determine whether p-values obtained from models used in testing differentially expressed genes come from uniform or beta contaminated distributions (uniform and beta mixture). If the $p - value \sim Unif(0, 1)$ then the batch of genes considered are not differentially expressed otherwise they are considered to be coming from a Beta contaminated distribution. The Beta contaminated model for p-values is defined as follows:

Let P_1, P_n be the random p-values from n hypothesis tests. For $i = 1, .., n$ define $Z_i = 1$

if a gene is differentially expressed and 0 otherwise. The conditional distribution are given as $(P_i|Z_i = 0) = uniform(0,1) = Beta(1,1)$ and $(P_i|Z_i = 1) = Beta(\alpha,\beta)$. Thus the marginal distribution of P_i for all $i = 1,,n$ is $P(P_i|Z_i) = (1-\pi)Beta(1,1) + \pi Beta(\alpha,\beta)$ for $0 \leq \pi \leq 1$, $\alpha > 0$ and $\beta > 0$. The corresponding posterior of $\tilde{p}_i = P(Z_i = 1|P_i,\alpha,\beta) = \frac{p_i f(p_i;\alpha,\beta)}{1-\pi+\pi f(p_i;\alpha,\beta)}$, where $\tilde{p}_i > T$, for some cut off T, suggests that the *ith* gene is differentially expressed.

The omnibus testing is applauded for the following strengths:

1) Ability to efficiently dispose of a batch of genes without alterations and thus increasing the power of the test.

2) Estimated parameters from the model can provide a frame of reference for multiple comparisons of the remaining batch.

3) Robustness in the sense that p-values from different distributions can be detected assuming the p-value distribution is uniform under the null hypothesis and Beta contaminated otherwise.

In addition, omnibus testing can reject the uniform(0,1) model even in the face of choosing $\alpha^* = 0$ and $\alpha^{**} = 1$ and it uses parameter estimates to determine the number of true positive, false positive and posterior differential expression probabilities to improve gene differential detection. In light of these strengths however, this testing procedure assumes independence of the p-values that may not be correct even though incorporating a covariance matrix of hundreds if not thousands of p-values may be challenging if not impossible and modeling p-values rather than the full data makes it impossible to recover information lost[28]. The assumption of a two-component

mixture is overly simplified but agrees with bootstrap studies [3] that two mixture assumption is appropriate in finite mixture problems. Furthermore, treating some parameters in the Beta distribution as known a priori may introduce biases into the testing procedure. These limitations call for an improvement to account for the identified problems.

1.3 Review of Mixture Model Applications to Alzheimer's Disease

Alzheimer's disease (AD) is another area where mixture models have been applied for diagnostic purposes. The Centers for Disease Control and Prevention defines AD as a 'progressive disease beginning with mild memory loss possibly leading to loss of the ability to carry on conversation and respond to the environment'. Three core criteria for identifying the predementia phase of AD have been proposed by a working group under the direction of the National Institute on Aging and the Alzheimer's Association. These include clinically based criteria for clinicians and healthcare providers, biomarker (cerebrospinal fluid measures) and brain imaging for research purposes and a combination of the clinical and biomarker evidence [25]. The prevalence of AD in the United States in 2013 was 5 million projected to be 14 million in 2050 [24]. Identified as the most common form of dementia, AD incidence is between 60 and 80 percent of all dementia cases [24]. In 2010, an estimated 600,000 (32% of all older adults death) adults 65 years and older with AD died in the United States. It is projected that mortality rates due to AD could top 43% of all older adults death by 2050. [22].

Presently, De Meyer et. al [9] are among few researchers who have used normal mixture components to separate patients with AD from those without using biomarkers. Using a mixture model framework, De Meyer et. al. classified cognitively normal elderly people into one of three categories; Alzheimer, mild cognitive impairment (MCI) and cognitively normal (NC) using biomarkers. Many studies have reported on the reliability of using biomarkers for detecting AD in its early stages. For instance Hampel et al.[10] compared the AD predictive potentials of

many existing biomarkers such as cererospinal fluid (CSF) tau protein ($p-tau_{199}$), threonine 231 ($p-tau_{231}$), threonine 231 and serine 235 ($p-tau_{231-235}$) threonine 181 ($p-tau_{181}$), and serine 396 and serine 404 ($p-tau_{396/404}$). They identified CSF total-tau (t-tau) and CSF beta-amyloid1-42 to have reasonable sensitivity and specificity rates for differentiating early and incipient AD groups from other age-associated disease such as Lewy body disease and some secondary dementia.

The procedure adapted by De Meyer et. al.[9] consists of three steps. First they applied mixture models to a data set from the US Alzheimer's Disease Neuroimaging Initiative (US ADNI) using a single biomarker (CSF Abeta1-42) to differentiate between Alzheimer's and none Alzheimer's cases. This resulted in a sensitivity of 91% and specificity of 62% when a cutoff value of $188pg/ml$ was used. A different decision criterion that balances the two arms of the ROC curve yields a comparable rate for the sensitivity (74%) and specificity(75%). Overall, 25.2% of the observations were misclassified.

In the second and third steps, De Meyer and colleagues extended their method to include the biomarkers CSF $p-tau_{188p}$ and or CSF tau. With an Akaike information Criterion (AIC) difference of 26 between the two competing models, the Abeta1-42 and CSF tau_{181p} (AIC= 4137) model was selected over the Abeta1-42 and CSF tau model ($AIC = 4163$). The selected model was validated with two independent data sets; an autopsy data set and a ADNI data set. The model detected 90% of AD signature in the AD group, 70% in the MCI group and 36% in the cognitively normal group of the ADNI data set. Out of 68 autopsy confirmed AD cases, 64 cases were correctly (94% sensitivity) classified as AD. The model also identified correctly all

patients on track to AD (100% sensitivity) when patients with MCI conditions were followed for 5 years.

The model's performance hinges on the functions of the two biomarkers; CSF Abeta1-42 as an initial biomarker and CSF p tau_{181p} as a subsequent stage biomarker associated with progression towards dementia. These intrinsic characteristics of the selected biomarkers are not new as documented in the literature by many authors including Montine [11], Albert et. al. [2], Stomrud et. al.[15] and Gustafson et. al. [12].

De Meyer and colleagues noted that AD signatures were present in 39% cognitively normal persons for the single biomarker model and 36% for the combined biomarker model. They concluded that these observations were consistent with neuropathological studies that healthy elderly individuals tend to have amyloid containing plaques and tau containing neurofibrillary tangles in their brains[9]. Thus their method was consistent with expected AD diagnosis and thus serve as a platform upon which future models may be developed.

De Meyer and colleagues' work breaks ground for further expansions in this area. Notably addition of Apolipoprotein (APOϵ4); the most robust genetic risk factor for sporadic AD known to be related to AD, MCI and NC [35] in the analysis may lead to improved results. We applaud the authors for their ground breaking application of mixture models to addressing AD related problems however they did not address all the potential stages of MCI as published by ADNI. Being able to detect candidates with early MCI (eMCI) may prove vital in delaying the development of AD using available therapeutic procedures. Although a hypothesis was not tested by the

authors, identifiability problems which are inherently associated with mixture models were not addressed by the authors. The use of AIC as the only model selection criteria raises concerns as it has been shown that AIC favors small sample size and more mixture components [27]. Thus the Flexible Information Criteria may be more appropriate in this case as the penalty involved considers the configuration of the data points in addition to the sample size [27]. Finally, the wide variations in the concentrations between different aliquots of Abeta1-42 [36] can significantly alter the conclusions drawn from the two models. Perhaps following the guidelines in [36] may help stabilize the variations in the concentrations of the analyte and thus improve the outcome from the models.

1.4 Review of Existing Model Selection Criteria and Mixture of Regression

Many model selection criteria span the field of mixture models. Popular amongst these selection criteria are MLRT[5], D-test[6],FLIC[27], BIC and AIC.We should observe that MLRT and D-test are hypothesis testing procedures compared to BIC, AIC and FLIC which are information theoretic criteria. However,the hypothesis test procedures may be used for model selection by determining the model complexity (or number of mixture components in the underlying population). Drton and Plummer(2016)[44] developed the singular Bayesian Information Criterion (sBIC) to address model selection problems arising from singular models (models whose Fisher information matrix is singular). The authors noted that sBIC differs from BIC in that although they both have Bayesian flavors, the regularity conditions underpinning the derivation of BIC are not satisfied by singular models. The authors proposed: $sBIC(M_i) = \log L'(M_i)$ where M_i is a finite set of candidate models and $\{L'(M_i) : i \in I\}$ is the unique solution to the equations $\sum_{j \leq i} [L'(M_i) - L'_{ij}]L'(M_j) = 0$, $i \in I$, that has all positive entries with $L'_{ij} = P(Y_n|\hat{\pi}, M_i)n^{-\lambda_{ij}}(\log n)^{m_{ij}-1}$ where λ_{ij} is the learning coefficient and m_{ij} is the multiplicity of λ_{ij}. Compared to BIC, the authors demonstrated that sBIC can achieves better frequentist model selection behavior, and allows more posterior mass to be assigned to larger models. When the models to be selected are regular, sBIC selects the same model as BIC; sBIC however does not rely on Monte-Carlo computation but rather on the information about the learning coefficient. For moderate number of models, sBIC and BIC have comparable computational burden. However there is a need for future work to address computational

burdens associated with the use of sBIC when larger models are involved.

Viele and Tong (2002)[46] proposed modeling with mixtures of linear regressions where the outcome of interest was modeled conditional on a set of covariates and the prior was implicitly data dependent. The key additions to the mixture models paradigm from this procedure include the ability to adjust for covariates, account for masked outliers and also ensure consistency of the posterior distribution using bracketing entropy. The likelihood of interest was defined as $g(y_i|x_{i1},...,x_{ip}) = \sum_{j=1}^{k} p_j N\left(\sum_{p=1}^{P} x_{ip}\beta_{jp}, \sigma_j^2\right)(y_i)$ where $\beta_{jp}\varepsilon R$ for $j = 1,...,k$, and $p = 1,...,P$ are regression coefficients, $\sigma_j^2\varepsilon R^+$ are the regression variances and $(p_1,...,p_k)\varepsilon S^k$ are the relative probabilities of the k components with S^k being a dimensional simplex $s = (s_1,...,s_k) : s_j > 0, \sum s_j = 1$. Their approach has a Bayesian flavor in that they placed priors on the mixing components, regression coefficients and regression variances (Dirichlet, normal and Gamma respectively [refer to section 2 page 317 details]). The posterior modes in the model were estimated with an EM algorithm and Gibbs sampling was used to sample from the identified modes.

Dai and Charnigo (2010) studied omnibus tests using Z or T statistics from multiple differential expression testing of genes assumed to arise from an underlying contaminated normal mixture model(CN). Prior to this study the authors developed the contaminated beta model(CB) for analyzing p-values arising from differential expression tests. The CN model with the corresponding hypothesis for the omnibus test were proposed as $(1-\gamma)N(0,\sigma^2) + \gamma N(\mu,\sigma^2)$ and $H_0 : \gamma\mu$ ver-

sus $H_1 : \gamma\mu \neq 0$ and the the penalized maximum modified likelihood was given as $l_n^*(\gamma, \mu, \sigma^2) = \sum_{i=1}^{n} log[(1-\gamma)f(Z_i; 0, \sigma^2) + \gamma f(A_i; \mu, \sigma^2)] + C \log[4\gamma(1-\gamma)]$ where Z_i is the resulting Z statistic from the i^{th} test, $\gamma\varepsilon[0, 1]$ is the proportion of genes in the batch that are deferentially expressed, μ and σ^2 are the mean and variance of Z_i given differential expression of the *ith* gene respectively. The hypothesis test was carried out with modified likelihood ratio test and D-test. Of note, the parameter estimates from the maximum modified log likelihood(MMLE) were utilized in the calculation of the D-test; an advantage of the D-test is that one only need to have the parameter estimates to use the test. In an empirical study to compare the performance of the new model (CN) to the old (CB), the authors noted that CN yields a more powerful test than CB when there's lack of symmetry between the over and under expressed gene batches; the ratio of $|\mu|$ to σ in CN is not too large; and when two sided test is of interest. To choose between the two models (CN or CB) one can apply a BIC type criterion on the estimated MMLEs from the two models.

Chapter 2 An Application of A Bivariate Normal Mixture Model

Introduction

Our research in this and subsequent chapters differs in many respects from existing analyses that used mixture modeling to analyze AD data[9]. Notable differences include: 1) we consider all three biomarkers simultaneously instead of investigating them pairwise. That is considering the $n \times 3$ outcome matrix $\mathbf{Y} := (Y_{1i}, Y_{2i}, Y_{3i})$, $i = 1, 2, 3, ..., n$,where n is the number of individuals, we derive the $n \times 2$ response matrix $\mathbf{Y}^* := (Y^*_{1i}, Y^*_{2i})$, $i = 1, 2, 3, ..., n$ corresponding to ratios of the original biomarkers where $Y^*_{1i} = \frac{Y_{1i}}{Y_{3i}}$ and $Y^*_{2i} = \frac{Y_{2i}}{Y_{3i}}$, $i = 1, 2, 3, ..., n$; 2) we prioritize placing people in groups based on biomarker data collected while they are still healthy and utilize an established mixture method to predict their future status, rather than placing people in groups based on biomarker data collected after they exhibit cognitive decline. At that point it will be relatively easier to separate groups, but such focus may lack prognostic relevance to those who are cognitively normal today but can potentially develop AD in the future; 3) since AIC is known to overestimate the number of components or groups [27] other statistical criteria such as singular Bayesian Information Criterion (sBIC) are used in addition to AIC to choose the number of groups or components; and, 4) more sophisticated statistical modeling is considered, to account for other covariates such as APOE4, age, gender, race, mini mental state exam score at baseline and level of education.

2.1 Motivation and Objectives

As we have already alluded to in our introductory section of this dissertation, AD is progressive in nature which means it worsens over time starting with a mild loss in cognition and later developing into dementia where the affected persons lose their ability to interact with or respond to their environment. Every 67 seconds, on average, someone living in the United States develops AD. One in three seniors dies with AD or another form of dementia [38].

In addition to the burden of this disease to families, the estimated related care cost paid by Medicare was about $11 billion in 2010 [52]. However, very little research in the literature has relied on mixture modeling to address the identifying candidates who are at high risk of transitioning from normal cognition at baseline. As far as we know, only **one** article ([9]) used mixture modeling to diagnose AD. This chapter differs from [9] and adds to the literature in four major ways as embodied in the objectives and explained here:

1. We apply mixture models to individuals who are cognitively normal as opposed to [9] where the participants were a combination of cognitively normal, mildly cognitively impaired and AD.

2. we use derived variables such as tau/ abeta and ptau/abeta as in [48-51] in which the authors showed that tau/abeta was a good predictor of future de-

cline in cognition.

3. We use sBIC in addition to the traditional model selection criteria (AIC and BIC) to select the number of components contrary to using only AIC as in [9]. We note that AIC may overestimate (too liberal) the model complexity and BIC may underestimate (too conservative) it. Hence AIC or BIC alone may be inadequate. However unlike AIC, sBIC operates on a Bayesian principle and in general neither overestimates or underestimates the model complexity as explained in [44]. If AIC and BIC disagree on the model complexity, sBIC may serve as a tie breaker.

4. We adjust for other well established covariates associated with cognitive decline whereas [9] did not.

Specifically in this chapter we use mixture modeling to achieve the following goals:

Objective 1:

To statistically determine the degree of heterogeneity within the population from which the data were drawn.This population constitutes all people who could potentially volunteer to participate in the ADNI study and are willing to undergo lumbar puncturing to test whether or not AD proteins are present in the spinal fluid. Objective one entails fitting various bivariate mixture models with differing

number of groups (components) and estimating the complexity of the smallest true model. Suppose once again that $\mathbf{Y} := (Y_{1i}, Y_{2i}, Y_{3i})$, $i = 1, 2, 3, ..., n$ we define the ratio of biomarkers as $\mathbf{Y}^* := (Y^*_{1i}, Y^*_{2i})$, $i = 1, 2, 3, ..., n$,where $Y^*_{1i} = \frac{Y_{1i}}{Y_{3i}}$ and $Y^*_{2i} = \frac{Y_{2i}}{Y_{3i}}$, $i = 1, 2, 3, ..., n$ represent the two ratios formed from the three biomarkers. Suppose moreover, that the response for a particular individual arises from the joint bivariate normal mixture distribution whose density is $g(y) = \sum_{k=1}^{m} \theta_k N(\mu_k, \sigma_k^2)(y)$ where θ_k is the mixing parameter for component k, m is the unknown number of components we wish to estimate, μ_k is a two vector of component means and σ_k^2 is a 2×2 covariance matrix for component k. So a two component mixture will take the form $\theta_1 N(\mu_1, \sigma_1^2) + \theta_2 N(\mu_2, \sigma_2^2)$ where $\theta_2 = 1 - \theta_1$ since in general $\theta_1 + \theta_2 + ... + \theta_m = 1$. We use various information-theoretic criteria (e.g., AIC, BIC and singular BIC[44]) to decide how many groups are suggested by the data.

Objective 2:

To investigate whether mixture components from objective 1 predict future disease status (i.e. cognitively normal, mild cognitive impairment, Alzheimer's Disease) and thereby estimate the hazard of future disease within each group. Suppose that T is a continuous random variable denoting the length of time(months) to event(AD) then T can be censored or uncensored depending on whether or not the event of interest was observed. The corresponding hazard rate can be generally defined as $\lambda(t) = \lim_{\delta \to 0} \frac{P(t \leq T < t+\delta | T \geq t)}{\delta} = \frac{f(t)}{1-F(t)}$ where F and f are the cumulative and density functions of T. Furthermore, the hazard rate is related to covariates through the functional form $\lambda(t|\mathbf{X}) = \lambda_0(t) \exp\{\mathbf{X}\beta\}$ where $X = (X_1, X_2, ...X_N)^T$ for some finite number N of covariates, consisting in our case, of estimated probabilities from the mixture model, age, race, APOE4 and or level of education. We define covariates to permit future adjustments.

2.2 Methodologies, Cognitive Assessment and Review of Related Concepts

Participant Characteristics

The data set used in this study is a subset of the original data from the ADNI study. ADNI was launched in 2003 and was spearheaded by Dr. Michael Weiner with the goal of testing magnetic resonance imaging (MRI), biomarkers and other modalities to measure progression of and to mild cognitive impairment and Alzheimer's disease. The ADNI project has three phases: ADNI 1, ADNI GO and ADNI 2. The first phase commenced in 2003 with a participant pool of 200 normal control, 400 with MCI and 200 with mild AD. The first phase ended in 2010 and the second phase began (2009) prior to phasing out the first. ADNI GO is made up of 200 newly recruited participants with early MCI (EMCI) and 500 normal controls and MCI inherited from ADNI 1, making up a total of 700 participants. This phase ended in early 2011. The third phase ADNI 2 started in early 2011 slightly overlapping the second phase. The participants included 150 new normal controls, 150 new EMCI, 150 new late MCI (LMCI) and 200 new mild AD. Approximately 450 to 500 participants with normal cognition and MCI came from ADNI 1 and approximately 200 participants with EMCI are included from ADNI GO[47]. The de-identified data are publicly available at adni.loni.usc.edu and can be obtained by completing a registration process.

From Table 2.1, 779(44.9%) of the participants are females and 956(55.1%) are males. There are 3(0.17%) American Indians, 29(1.67%) Asians, 77(4.44%) Blacks,

2(0.12%) Hawaiians, 21(1.21%) unknown and 1603(92.39%) Whites. Non-Hispanic/ Latinos make up 1666(96.2%) of the sample while Hispanic or Latino make up 58(3.34%). The minimum age at enrollment is 48.1 years and the maximum is 91.4 years old.

Our interest lies with the $n = 114$ subjects with known ages and biomarker levels classified by the Alzheimer's Disease Neuroimaging Initiative (ADNI) as cognitively normal at their baseline visits based on a mini mental state exam (MMSE) score between 24 and 30, a clinical dementia rating sum of boxes (CRDSB) score of zero and the absence of depression, MCI or dementia [9]. We arrived at 114 participants by removing all duplicates with nodupkey with SAS procedures. The demographics of the participants in this subsample are shown in Table 2.2 and Figure 2.1 illustrates the distribution of the biomarkers Abeta142, Ptau181p and Tau. The distribution of Abeta142 appears to be bimodal: one mode at lower Abeta142 values (less than 200) and the other mode at higher Abeta values(above 200). The Tau and Ptau distributions are arguably bimodal and most of the observations accumulate at the lower values. The distributions are skewed to the right. The derived biomarker measures are shown in Figure 2.2. The derivations are obtained by taking a ratio of Tau and Ptau with respect to Abeta142. Both distributions are also negotiably multimodal and skewed to the right akin to their counterparts in Figure 2.1. In addition the distributions appear to suggest three distinct groups: one large group on the extreme left another smaller group in the middle and a third on the extreme right which consists of few observations especially of Ptau181P/Abeta142. The three

groups apparent in the distributions of the derived variables are more visible to the untrained eye than the original variables (using the same breaks=20 in R package) in figure 2.1, which could provide a strong hint that the estimated number of mixture components $\hat{m}$ in this chapter's formal data analysis might be three.

The Institutional Review Board (IRB) at the University Of Kentucky approved an exemption for the use of this data set on conditions including: 1) that the data will be stored on a jump drive accessible to the two specific persons only and 2) that the IRB will be informed of any substantial future changes to the study described in the application.

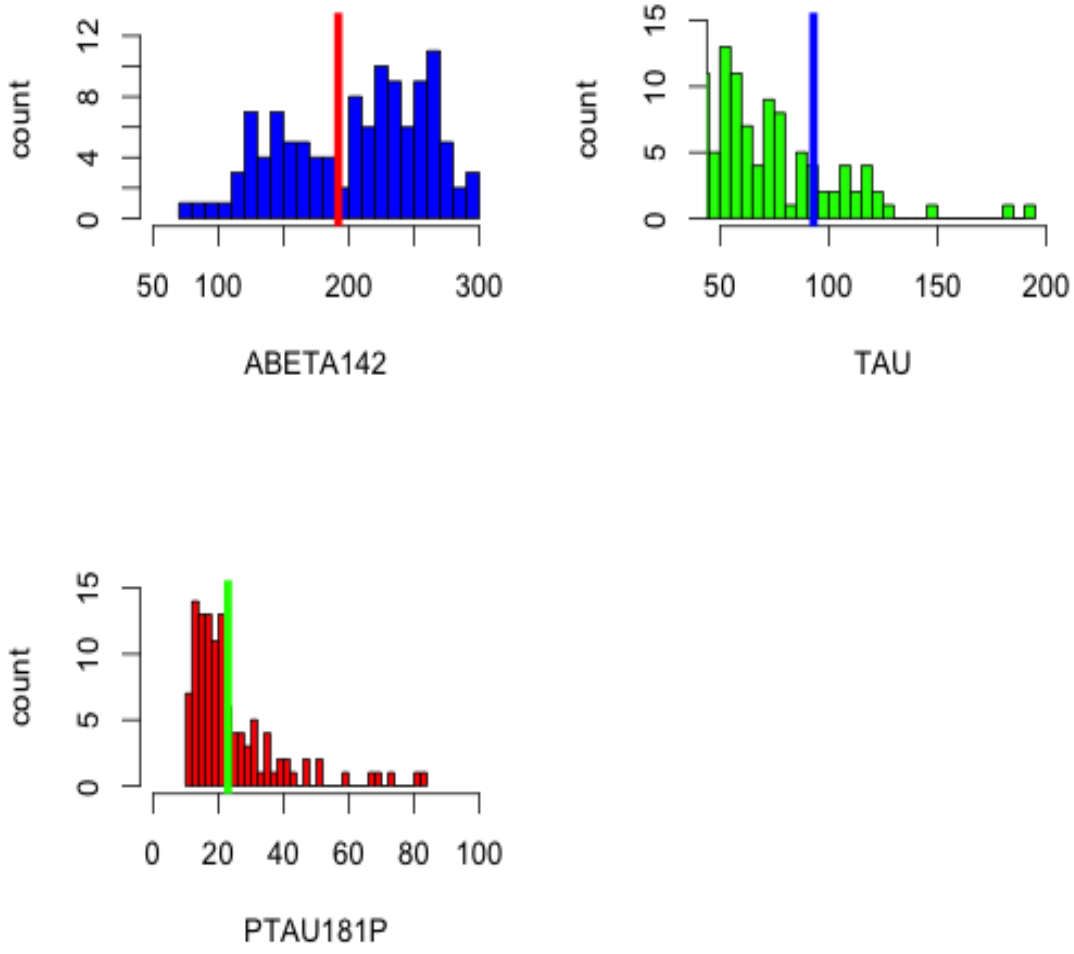

Figure 2.1: Original Biomarkers

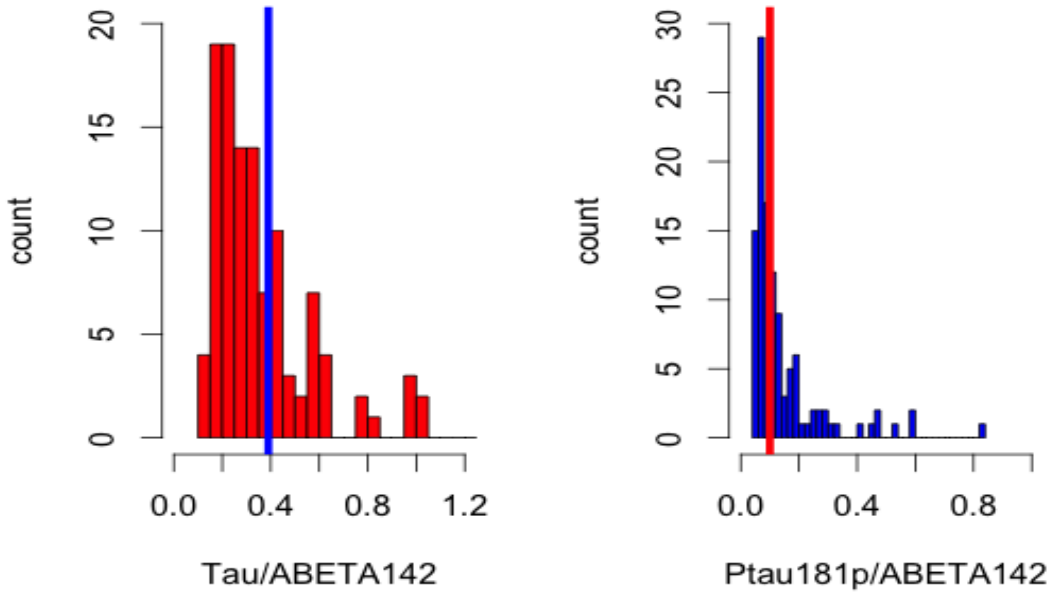

Figure 2.2: Derived Biomarker Histogram

Cognitive Assessment

To ascertain that the participants in our data are indeed cognitively normal at baseline, we examine a cross tabulation of the clinical dementia rating sum of boxes (CDRSB) and MMSE baseline to assess the degree of agreement between the two scores. Participants' scores at or below 24 on the MMSE scale or at 0.5 or higher on the CDRSB scale will counter the claim of normal cognition according to [39] and [40]. From Table 2.5 eight (8) participants in our data set who scored at and above 27 on the MMSE scale at baseline present a score at or greater than 0.5 on the CDRSB scale. These participants' CDRSB and MMSE baseline scores are compared with the Alzheimer's Disease Assessment Score (ADAS11) and the Rey's Verbal Auditory Learning Test percent forgetting (RAVLT) score.

Two of the eight with MMSE baseline scores of 27 and 30 respectively also have the same score of 6.67 for ADAS11 and their RAVLT forgetting scores are respectively 36% and 7.7%. Of those who scored 0.5 on CDRSB scale, two scored approximately 31% and 33% on the RAVLT forgetting scale with a corresponding 5.0 and 3.33 on the ADAS11 scale. The participant who scored 2.5 on the CDRSB scale also scored 71% on the RAVLT forgetting scale and 7 on ADAS11 scale with 30 on the MMSE baseline scale. One participant with a score of 1 on the CDRSB scale, 11 on ADAS11 scale, and 30 on MMSE baseline scale also scored 100% on RAVLT .

Based on these findings the participants that scored 0.5 and above on the CDRSB

scale will be included and excluded in two versions of the model fitting process to see if they influence our results. If dramatic changes such as a significant decrease or increase in the c-statistic, log rank test p-value, huge swings in the standard error estimates or covariates' p-values occur due to the absence of the eight participants' information in the modeling, then the output from both sets of the results will be presented: one with the participants excluded and the other with the participants included. If however there's no such dramatic changes from the inclusion or exclusion of the information from the eight participants then we shall present the output with information on all eight participants.

Statistical Modeling

As introduced in chapter 1 and re-emphasized earlier in this chapter a variety of mixture models will be fitted throughout this dissertation. In chapter 2 we fit the bivariate normal mixture model (BNM) without covariates using the mixtools package in R[42] and write and R code the sBIC function for model selection. The responses are collected in an $n \times 2$ matrix whose columns correspond to TAU/ABETA142 and PTAU181/ABETA where $n = 114$. The number of mixture components representative of the underlying heterogeneity in the data will also be selected using AIC, BIC and sBIC.

The aforementioned objective 2 is to predict the future cognition status of cognitively normal individuals into one of three groups: normal cognition (CN), mildly

impaired cognition (MCI) and Alzheimer's (AD).

Bayes method will be implemented to determine the respective (posterior) probability of individuals belonging to a given group given their biomarker information. We will classify individuals into one of two or three groups depending on the chosen number of mixture components. The characteristics of members in each component will be assessed based on current knowledge of the biomarker literature and visual representation; and the groups will be labeled as either (projected) cognitively normal (CN) or MCI/AD for the two component mixtures or (projected) CN, MCI and AD for the three component mixture. A high risk group will have higher scores overall on the TAU/ABETA and PTAU181/ABETA142 scales compared to a low risk group.

A multivariate Cox regression model for survival time of conversion from CN to MCI or AD will be used to assess the predictive utility of the model as expressed by the concordance (c) statistic. The logrank p-value associated with the hard classification will be used to test whether the group of survival time are statistically different. The raw (soft classified) and hard classified posterior probabilities of belonging to a component will be included in the Cox model such that a two component mixture model will yield one vector of posterior probabilities that represents the probability of belonging to the higher risk group. In this case, the group whose probability vector was not included in the model will be the baseline or referent risk group. For a three component mixture model we will have two probability vectors included in the Cox model: one vector for each of the two higher risk groups, and the third probability

vector not included in the Cox model will represent the baseline/referent risk group.

Review of related concepts

The singular BIC model selection criterion introduced in chapter one will be used in chapter two. A brief overview of the underlying principles is in order. The sBIC is approximated as $\exp(sBIC(M_i)) \approx P(Y_n|\hat{\pi}, M_i)n^{-\lambda_{ij}}(\log n)^{m_{ij}-1}$ where λ_{ij} is the learning coefficient and m_{ij} is the multiplicity of λ_{ij}[44]. In this chapter we assume a lowerbound for m_{ij} to be 1 consistent with [44] and deduce the upperbound for λ_{ij} as follows. Suppose the bivariate response of interest is $\mathbf{Y}^*$ as before and define the parameters

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{pmatrix}$$

$$\mu = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

and

$$\Theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \vdots \\ \theta_m \end{pmatrix}$$

; where m is the number of components (or model complexity), Θ is a vector of mixing coefficients, Σ is the component specific covariance matrix and μ the component specific vector of means.

Then to obtain an upperbound for the learning coefficient $\lambda_{i,j}$ we either fix the last $i-j$ entries of the mixing coefficient Θ or prior distribution for group membership (as regards to Bayesian methods) and estimate the number of free parameters (in this case component specific Σ and μ) or fix the parameters Σ and μ in the last $i-j$ components and allow Θ to vary.The primary goal here is to make the free parameters for the model estimable by overcoming the identifiability issues associated with mixture models.

Let us proceed by fixing Θ. Furthermore let's assume that i and j are respectively the indices two true models as in [44] with $i > j$ and j is the index of the smallest true model. Then the number of free parameters for this model can be generated as follows:

1. For $j = 1$ versus $i = 2$ we have $N(\mu_1, \Sigma_1) + 0N(\mu_2, \Sigma_2)$ which yields 10 free parameters.

2. For $j = 2$ versus $i = 3$ we have $\theta_1 N(\mu_1, \Sigma_1) + (1 - \theta_1)N(\mu_2, \Sigma_2) + 0N(\mu_3, \Sigma_3)$ thus giving rise to 16 free parameters.

3. For $j = 1$ versus $i = 3$ we have $N(\mu_1, \Sigma_1) + 0N(\mu_2, \Sigma_2) + 0N(\mu_3, \Sigma_3)$ thus giving rise to 15 free parameters.

The number of free parameters generated form a pattern that can be expressed as $5i + j - 1$. Accordingly we can bound the learning coefficient from above by $\lambda_{i,j} \leq \frac{1}{2}[5i + j - 1]$.

Notice that if we fix Σ and μ in the last $i - j$ component and allow Θ to vary, we get an upperbound of the form $\lambda_{i,j} \leq \frac{1}{2}[i + 5j - 1]$ (see [44] for similar derivations regarding the upper bound for the learning coefficient). Of note the smallest true model is fixed at $j = 1$ in subsequent work throughout this dissertation.

2.3 Results and Discussion

The BNM model reveals a three-component mixture underlying the distribution of the biomarker ratios as the correct model based on sBIC estimated values and as indicated in Table 2.3 and Figure 2.3 respectively. From Table 2.3 we notice that the AIC and sBIC are in agreement as opposed to the BIC. However when we obtain the probability of the correct model based on the biomarker information using the sBIC estimated values we notice that in fact the three component (with estimated probability of being correct = 1) model is narrowly preferred to a four component model (with estimated probability of being correct ≈ 0.9999). The estimated probabilities in each rectangle in the three histograms of Figure 2.3 are indicative of the individuals who apparently (rectangle near 1) do or do not (rectangle near 0) belong to the component under consideration given their biomarker information or who are not conclusively classified (rectangles between the two main peaks). That is individuals with $P(X = x|\mathbf{Y}^*) \approx 1$ almost surely belong to the component x and for those with $P(X = x|\mathbf{Y}^*) \approx 0$ almost surely do not belong to the component in question, where X and $\mathbf{Y}^*$ are the latent grouping variable and the biomarker ratios respectively. Those with $0 \ll P(X = x|\mathbf{Y}^*) \ll 1$ are comparatively fewer in number and have posterior probabilities identified between the two extreme rectangles and are illustrative of uncertainty about component membership.

Component three shows that about 10% of individuals in the sample belong to the component and about 90% do not. Components one and two on the other hand have comparable numbers of participants who belong with probability near 1 and

who do not with probability near zero. Figure 2.3 also suggests an overlap between components one and two due to the greater numbers of uncertain memberships in these two components. As a result component three is well separated from components one and two, an observation that is also well captured in Figure 2.4. In Figure 2.4 members of component one (as judged by hard classification) have comparatively the lowest risk ratios compared to components two and three. It also shows that two members of component three may be outliers (with respect to their rtaubeta values) in the sense that even though they have comparatively larger rptaubeta values their rtaubeta values are within the range of components one and two. The observed outlier may have been missed if only one biomarker has been used in the study.

The risk comparison plots in Figure 2.5 also shows the least in cognitive decline risk for members in component one (as judged by hard classification) compared to their counterparts in components two and three. Component three shows the steepest decline in survival compared to component two after the $50th$ month of follow-up. We also see the sharpest decline in survival for component three members at the $96th$ follow-up month whereas the sharpest cognitive decline for component two takes place around the $50th$ month of follow-up.

We compare the densities in each component using two dimensional contour plots displayed in Figure 2.6. The plot indicates that the densities in components one and two have the steepest contours compared to component three; an indication of lesser variability amongst members of these components (not a surprise as suggested by

Figure 2.4). An interesting observation from Figure 2.6 relates to the directions of the contour plots; components one and two exhibit similar directions whereas component three exhibits a direction opposite to that of components one and two. Thus in components one and two, the biomarker ratios are positively correlated whereas in component three they are negatively correlated.

Two Cox model outputs are generated from the hard classification and raw (for soft classification) estimated Bayes probabilities of belonging to a component given one's biomarker information. Each of the two Cox model outputs are adjusted for covariates in the data set and membership in component one is used as the reference as it presumably has the least risk.

In the hard classification output in Table 2.4, the posterior probability related to component two is significant at $\alpha = 0.001$ adjusting for the posterior probability related to component three. The reverse is also true as seen in Table 2.4. A unit increase in component one posterior probabilities increases the estimated hazard of developing AD by three ($HR = 3.02, 95\%CI = (1.36, 6.68)$) fold compared to posterior probabilities in component 2 whereas a 10% change in the posterior probabilities in component three increases ones estimated hazard by over 5 ($HR = 5.35, 95\%CI = (1.89, 15.15)$) fold. In other words,a person who transitions from component two to component one will experience an increase in hazard rate of about 3.02 whereas a transition from component two to component three will increase the hazard rate by 5.35. The concordance statistic is 63.3%. Adjusting for covariates led to an increase in the overall

concordance. We also observe that immediate RAVLT (Reys Auditory Verbal Learning Test) offers a significant ($HR = 0.94$, $95\%CI = (0.90, 0.98)$) protection for AD accounting for the other covariates in the model. Race was however not a significant predictor of the time to transition, however, including race in the adjusted model improved the c-statistic (from 0.73 to 0.76), hence we included race in the model. Education, gender, MMSE, and age were all included in the adjusted Cox model and eliminated by backward elimination method. The absence of latter covariates did not influence the c-statistic. The output from the soft classification model indicates that in addition to the posterior probabilities being significant predictors of the time to transition, race ($HR = 0.312$, $95\%CI = (0.11, 0.87)$) is now a statistically significant predictor of time to transition although it wasn't in the hard classification model. The soft classification model has a comparatively modest gain in c-statistic and a modest increase in standard errors which consequently led to wider confidence intervals. The global proportional hazard tests were not significant in both the hard and soft classification cases.

The BNM model output without the eight participants as mentioned above in the cognitive assessment section is shown in Table 2.7. The output with all participants included in the modeling process was displayed in Table 2.4. In Table 2.4 the estimated hazard rate and standard error for component three posterior probabilities is $5.35(SE = 0.53)$ which differs narrowly from the estimated hazard rate of $4.76(SE = 0.56)$ in Table 2.7. These are the most notable difference between the two mixture models. Apart from these differences, the other estimators are comparable

although the soft classification model in Table 2.7 seem to gain in the concordance statistic albeit with larger confidence intervals. Since the two outputs are similar we will henceforth discuss only the modeling including all participants. The component estimated parameters are shown in Table 2.8.

The output from the models and the preceding discussion align well with the following speculations: component three is most indicative of the individual who could potentially develop MCI/AD within eight years of follow up, component two is somewhat indicative of individuals on trajectory to developing AD or mild impairment and those in component one may be most likely to remain cognitively normal.

Discussion In this chapter of the dissertation we addressed two main goals: obtain $\hat{m}$ using statistical criteria other than AIC (and including AIC for comparison purposes) and determine if the mixture components obtained are predictive of future disease status using multivariate Cox modeling and Kaplan Meier plots as validation tools.

Derived variables from the biomarkers obtained from ADNI are used in the mixture modeling process. The raw biomarker ratios are not used in a straight forward Cox modeling approach because we wouldn't be contributing substantial novelty to the literature although existing studies did not exclusively focus on predicting future status of cognitively normal persons some examined biomarker ratios. Furthermore, we have statistical reasons to deliberately avoid the simple approach in favor of a mixture modeling approach. These reasons are:

1. Using the raw biomarker ratios in the Cox modeling assumes a linear relation between the log hazard function and the biomarkers which may not be the case.

2. Using raw biomarker ratios in Cox modeling will suppress the potential of uncovering any heterogeneity in the distribution of biomarker ratios.

We also used the posterior probabilities of each individual in the Cox model instead of the raw biomarkers because the latter is not a significant predictor of hazard rate

(results not shown). In both the soft and hard classification Cox models, the posterior probabilities were significant predictors of the survival of the participants whereas the use of raw biomarker ratios yielded insignificant results (not shown). This may suggest that the hazard function is more related to the posterior probabilities than a linear function of the biomarkers.

Mixture modeling was chosen for the following reasons:

1. It could capture the inherent heterogeneity in the population.
2. It does not impose a linear relationship between the hazard.
3. Patterns related via the components to which persons belong can be tracked.
4. The density function of each group can be estimated and visualized.
5. Cut offs are determined automatically and thus permits us to easily compare predicted component with true component using cross tabulation as in Table 2.9.

One,two and three component mixtures were fitted and the three statistical selection criteria AIC, BIC and (approximated)sBIC values were. All three criteria chose $\hat{m} = 3$ as the number of components that best describe the data. It should be noted that sBIC generally tends to choose $\hat{m}$ somewhere between the two extremes (AIC and BIC)[44] as BIC tends to be very conservative and often underestimates $\hat{m}$ whereas AIC is very liberal and may overestimate $\hat{m}$[27]. The approximated sBIC has an additional strength of choosing the correct model given the biomarker information.

In this particular application, all three criteria agreed on the number of components.

A multivariate Cox model with time to MCI/AD as the output and the posterior probabilities corresponding to the two highest risk mixture components as covariates were fitted. Both posterior probabilities were significant predictors of disease with or without adjusting for other covariates listed in the literature. We did not find any significant interaction between the estimated posterior probabilities and the covariates accounted for in the Cox modeling. The separation between the three KM curves is also significant by virtue of the log rank test, and the corresponding c statistics found are modestly high.

The study in this chapter has shown that the ratio of biomarkers may be key in diagnosing AD vulnerability among currently cognitively normal people. To the best of our knowledge this is the first study to apply mixture modeling techniques to the ratio of biomarkers obtained from cognitively normal individuals for diagnostic purposes. The model classifies individuals in the sample by separating them into one of three categories, depending on their risk of transitioning from normal cognition in the future.

Since this study did not incorporate potential confounders in the mixture modeling process, future studies will utilize mixture regression models that account for covariates. The rationale for such an investigation is the potential of improving the predictive performance of the model due to the additional information. Also includ-

ing covariates in the mixture modeling process could improve the model's ability to account for hidden outliers as in [46].

2.4 Limitations and Future Directions

The data used in this chapter is a non representative sample because it came from volunteers or people living in the US who may be concerned about the potential of developing AD. Thus the findings from this study cannot be generalized to all persons that can potentially develop AD or its related illnesses. For instance in Table 2.8 the estimated mixture proportions corresponding to components one and two may be unlikely to mimic those that might be derived from the general population. The reason is because we speculate that the proportion of people in the general population who could have been in component two were underestimated in our model because they do not have reasons to believe that they may have AD and thus are less likely to participate in a study like ADNI. Those who could have been in component one were over estimated by our model because they may have been concerned about their cognitive health and so most of them enrolled in the study. We speculate that the third component may be comparable to that of the general population that tend to be older but cognitively normal. Such individuals may be inclined to volunteer for a study like ADNI and we believe that the proportion of older cognitively normal people genuinely concerned with deteriorating cognition in the population may be low.

In the hard classification model an individual may have been classified into a component based on his/her posterior probability for that component being slightly higher than a competing component. For instance an individual with posterior probabilities 0.51, 0.49 and 0.0 associated with components 1,2, and 3 respectively will be

hard classified into component 1 although this individual could have plausibly been in component 2 as well. In our data set only two individuals had such profiles. If the posterior probabilities were like 34%, 33% and 33% probability of belonging is about the same for all three components, so we will classify the individual(s) into component 1 and make a note that they could also plausibly belong to the other two components. This case however did not arise in our data set.

Future Directions

In this study our interest was to primarily predict transition from cognitively normal state to either MCI or AD. Although the c-statistic (64%) obtained in our analysis is not demonstrably superior to that of using the raw biomarker ratios in Cox modeling (c = 63%), this study presents a paradigm to classify persons who are cognitively normal into one of three risks strata without knowledge of their future cognitive status.

The unsupervised learning nature of the mixture model for developing risk strata can be developed by clinicians whose interests may lie in persons at risk today so as to provide them with interventions; whereas such risk strata cannot be obtained from raw biomarker ratios in a Cox model since the latter cannot be fitted without knowing who experienced cognitive decline, except by reference to and extrapolation from historical strata. For instance Figure 2.11 shows the risk strata boundaries based on the participants in this study.Importantly, given ones biomarker ratios we can identify if this individual is at a lower, medium or high risk of developing MCI/AD when the individual is still cognitively normal and showing no apparent symptoms of MCI/AD by seeing where he/she falls in Figure 2.11.

Future studies will consider the transition from CN to early or late MCI and AD as distinct events. This will however require longer follow-up times to increase the pool of participants in the categories of MCI and AD separately. In the present study 34 ($\approx$ 30%) participants transition from CN to either MCI or AD. This suggests that

relatively few participants transitioned within a maximum period of eight years given the age distribution of the study participants.

Future studies will also appeal to mixture models that adjust for other covariates within mixture components in hope of improving on predicting MCI or AD and develop a new technique to estimate mixture complexity for such models. This will be the subject of chapter 3.

2.5 Acknowledgements

We are thankful to ADNI for the data used for this analysis, and we hereby reproduce their acknowledgement of their benefactors. "Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant $U01\ AG024904$) and DOD ADNI (Department of Defense award number $W81XWH-12-2-0012$). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.;Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University

of Southern California" [47].

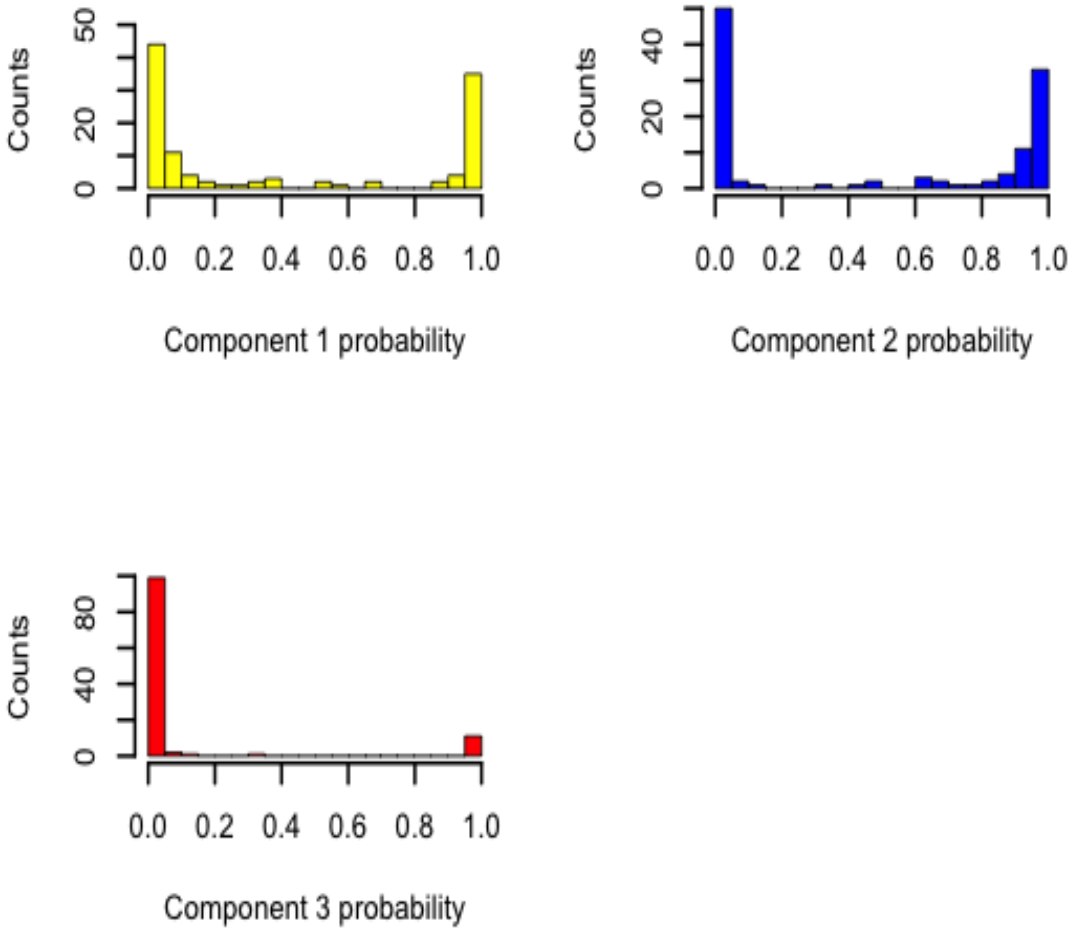

Figure 2.3: BNM Component Membership Probability Distribution

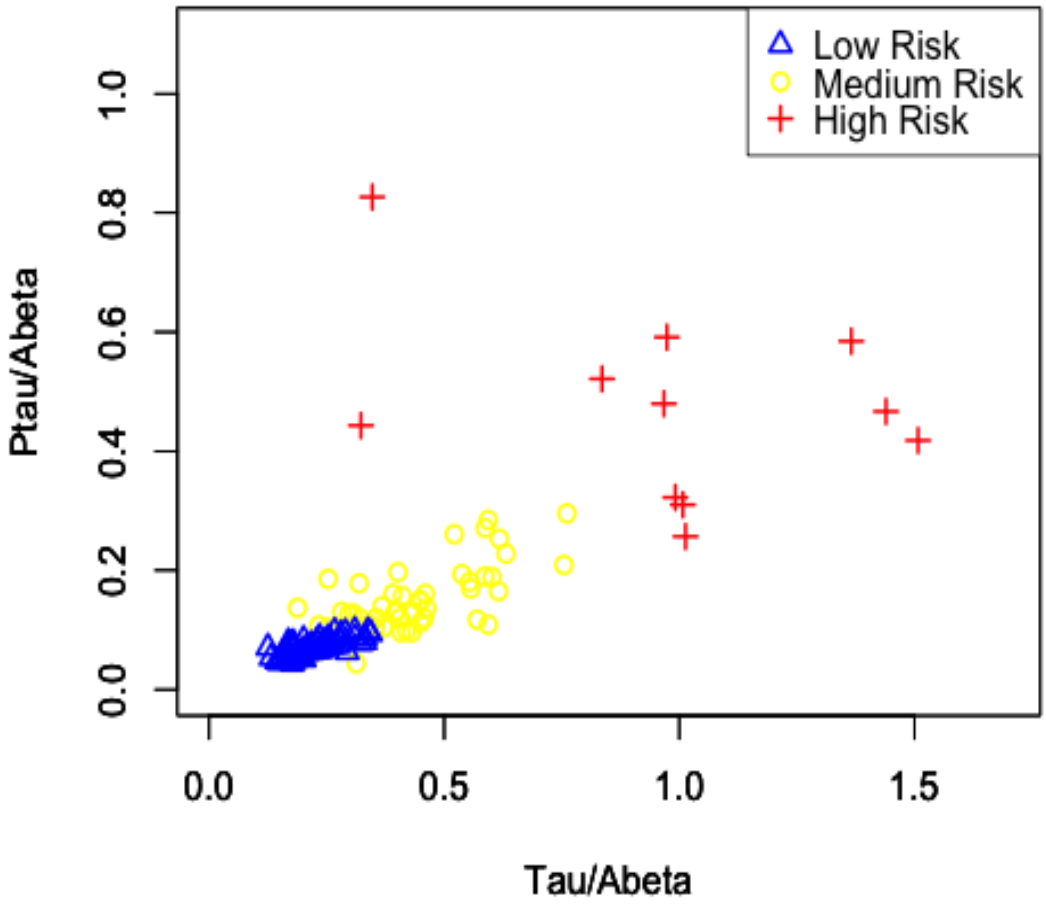

Figure 2.4: BNM Predicted Components Scatterplot

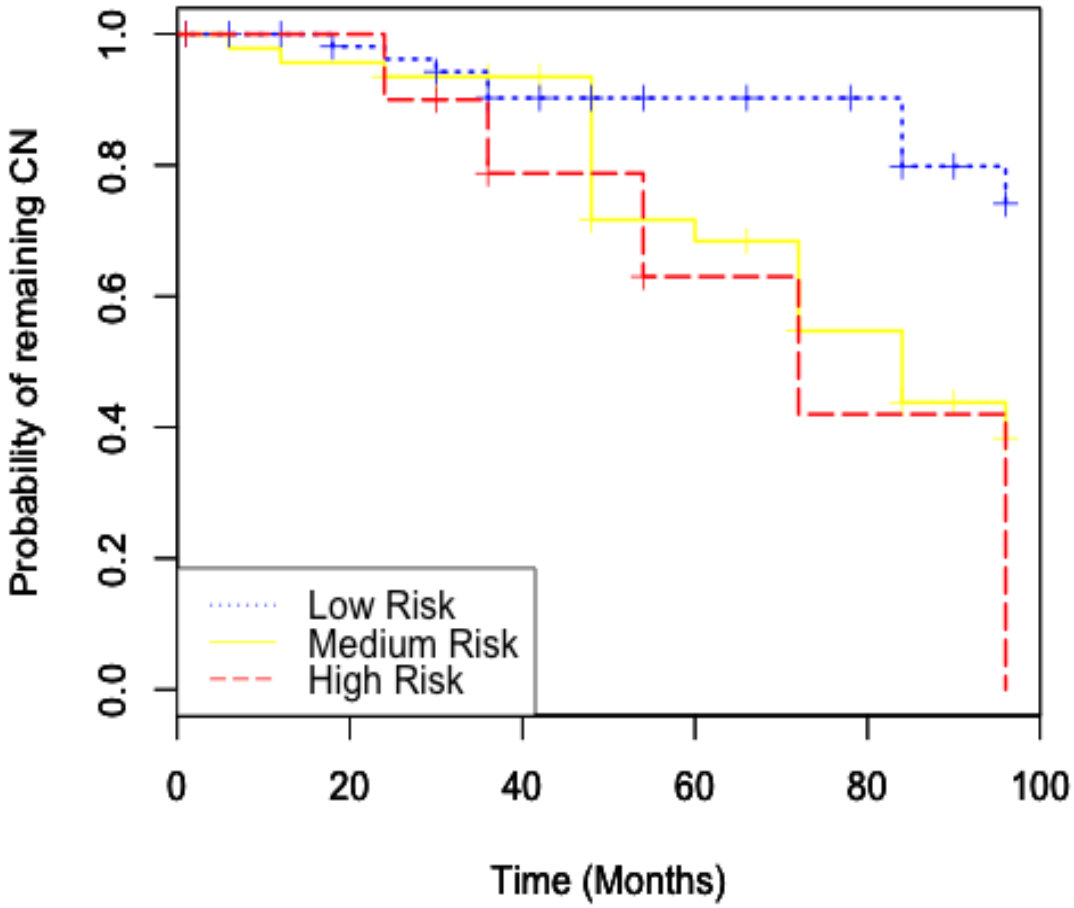

Figure 2.5: BNM Component Kaplan Meier Plots

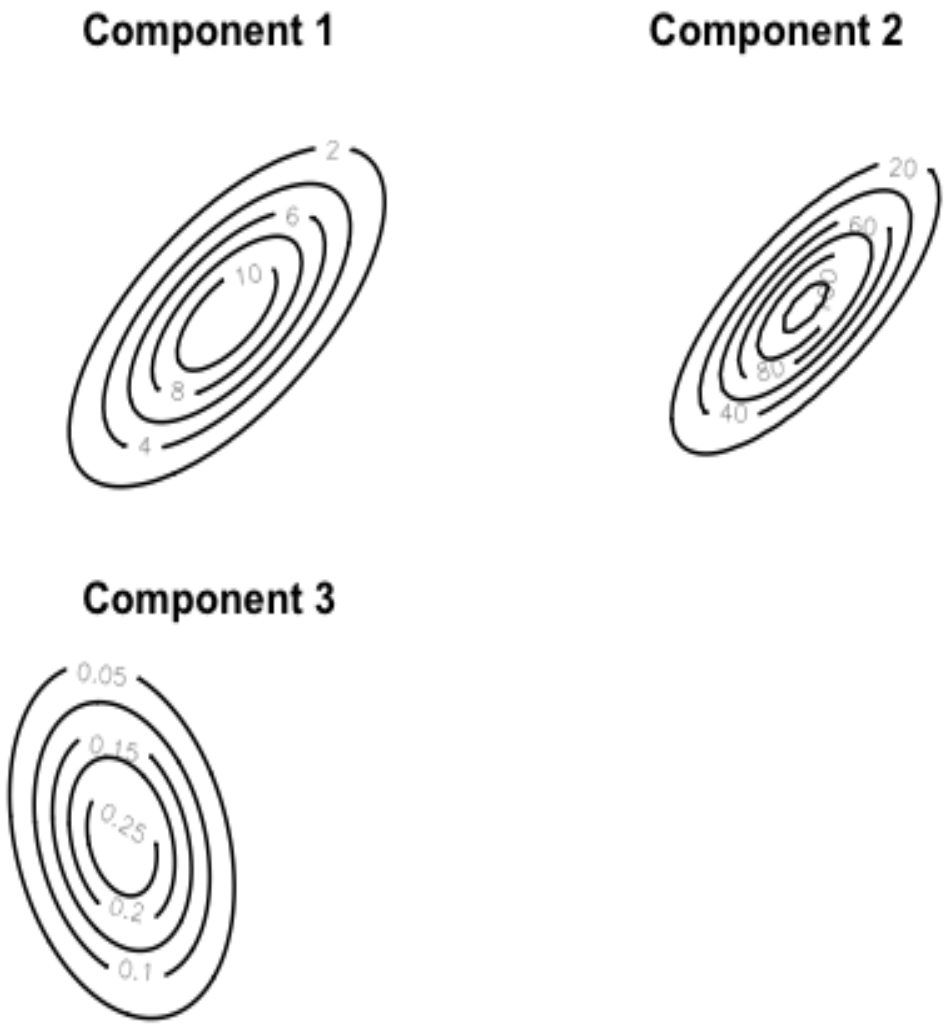

Figure 2.6: Component contour plots

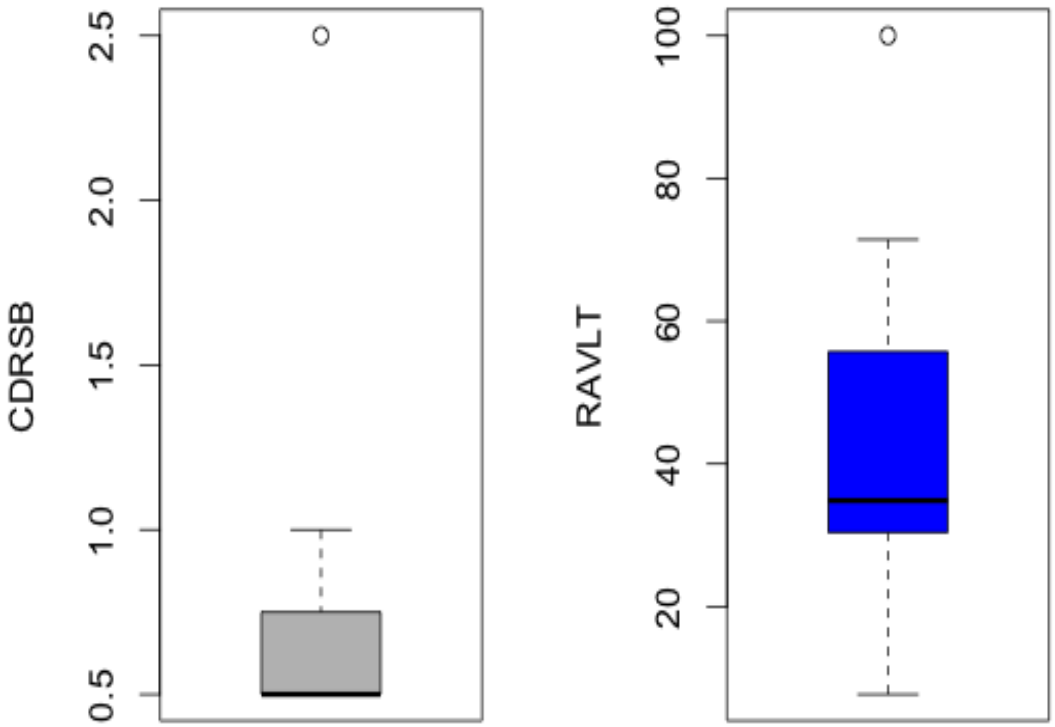

Figure 2.7: The empty circles identify potential outliers among the eight participants

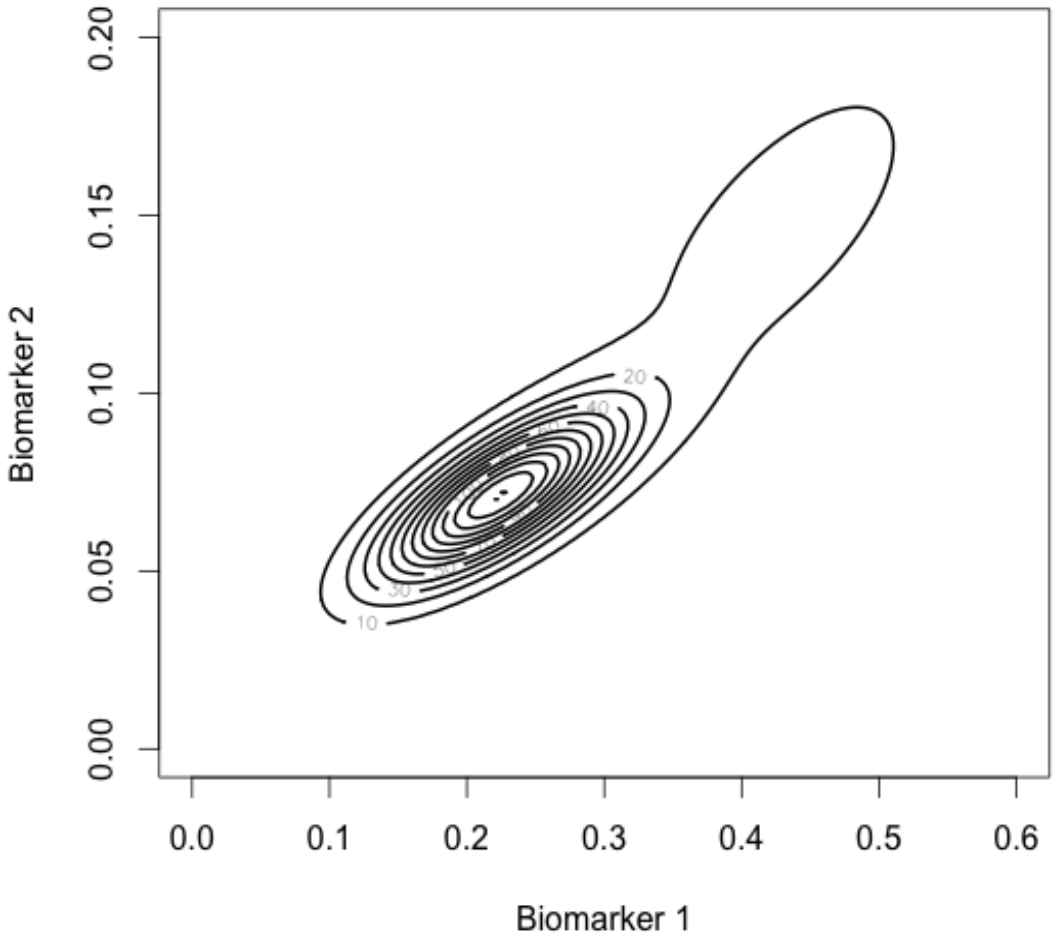

Figure 2.8: Joint Biomarker contour plot

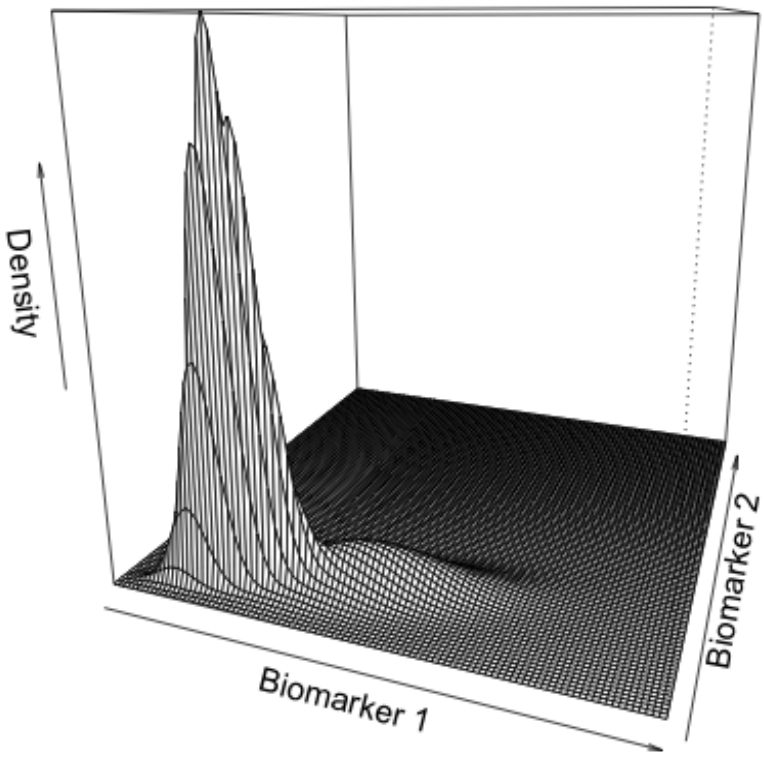

Figure 2.9: Joint Biomarker density plot

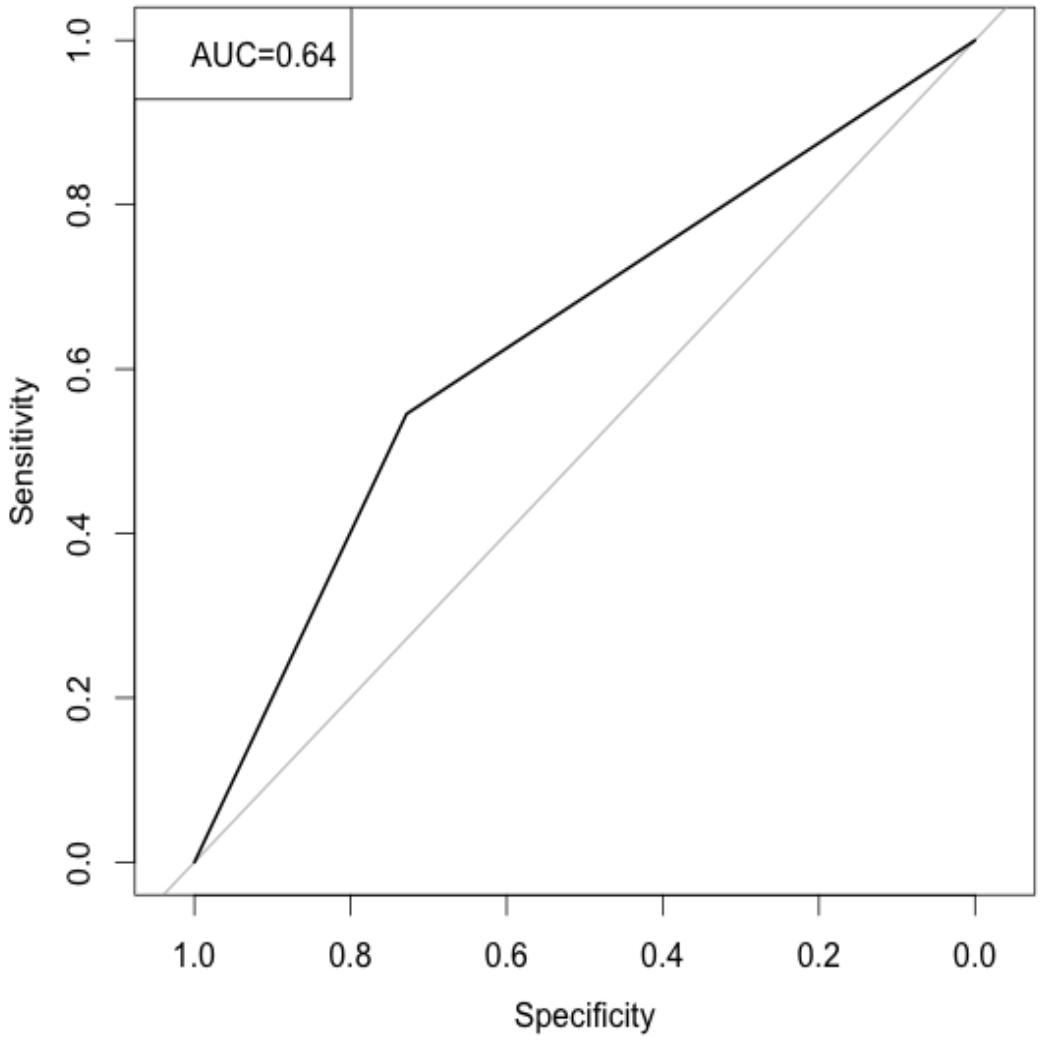

Figure 2.10: ROC plot. AUC: Area under the curve

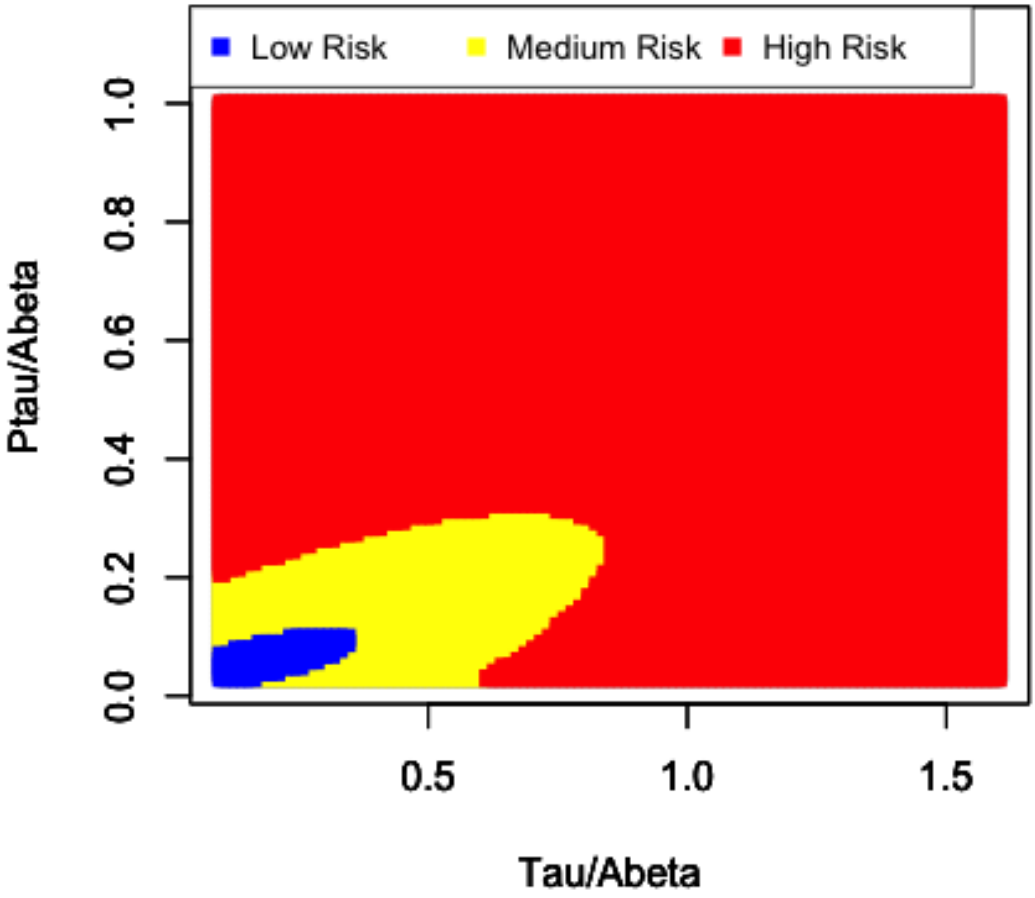

Figure 2.11: Risk Boundaries

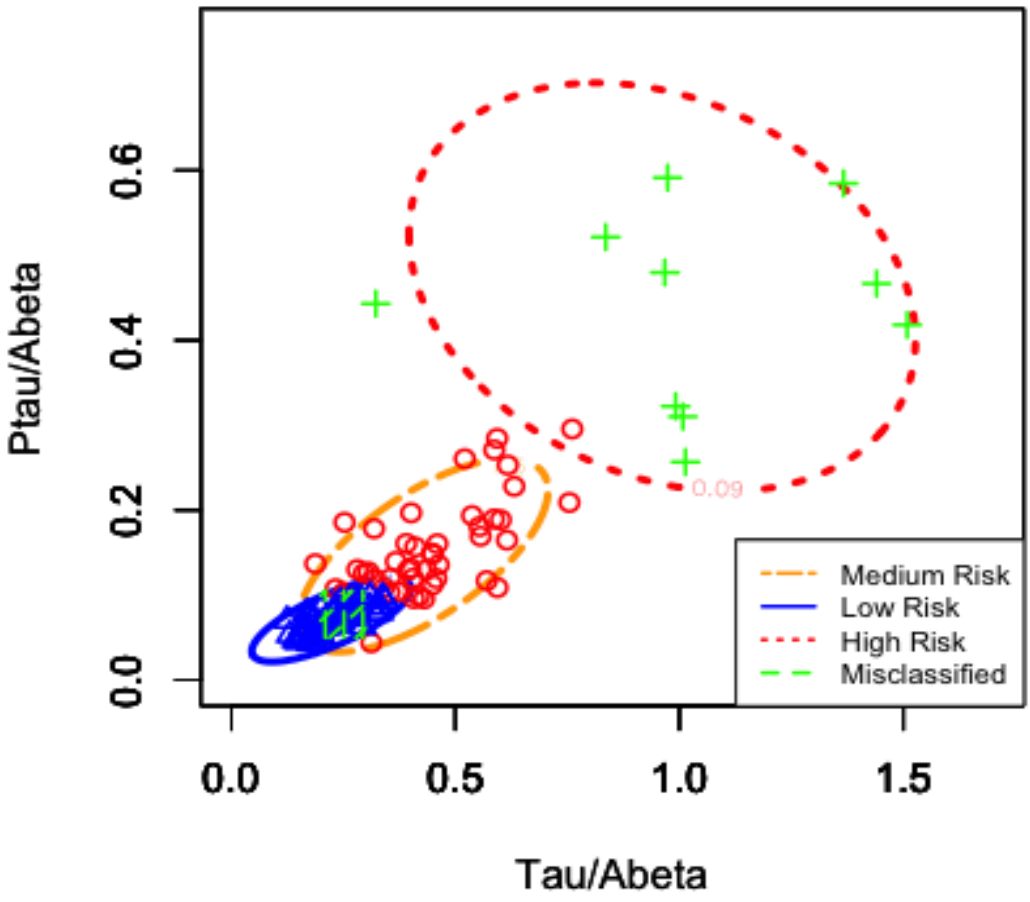

Figure 2.12: Risk Boundaries

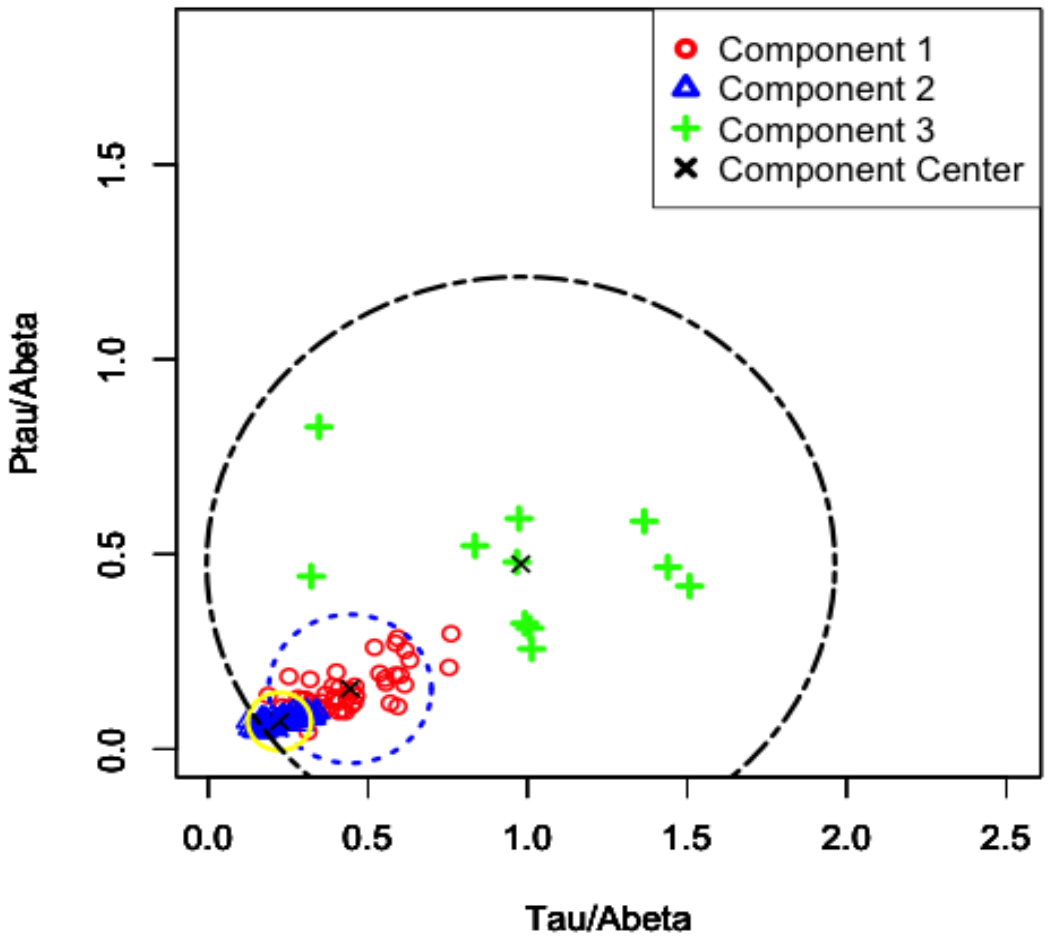

Figure 2.13: Risk Differentiation Boundaries

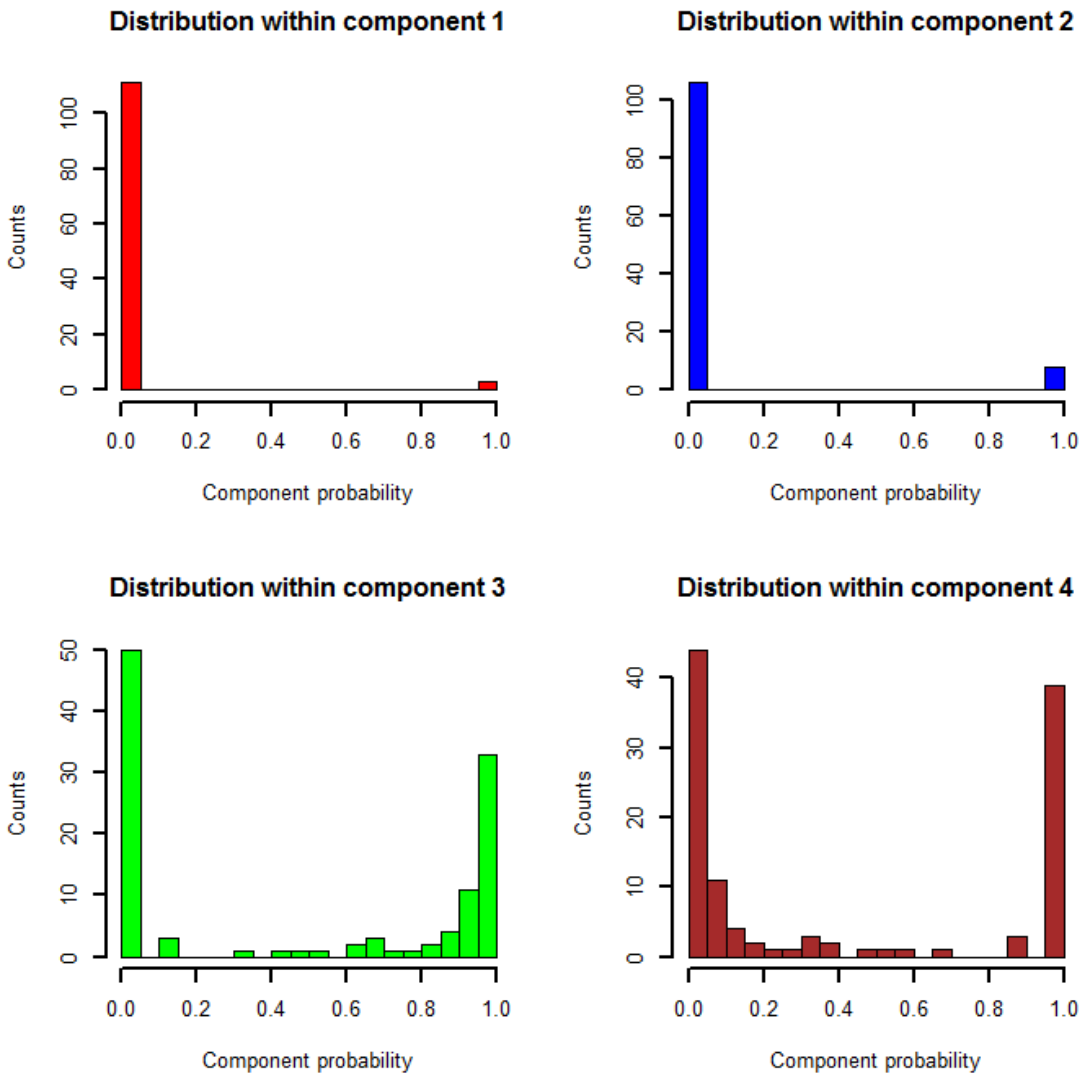

Figure 2.14: BNM Four Component Membership Probability Distribution

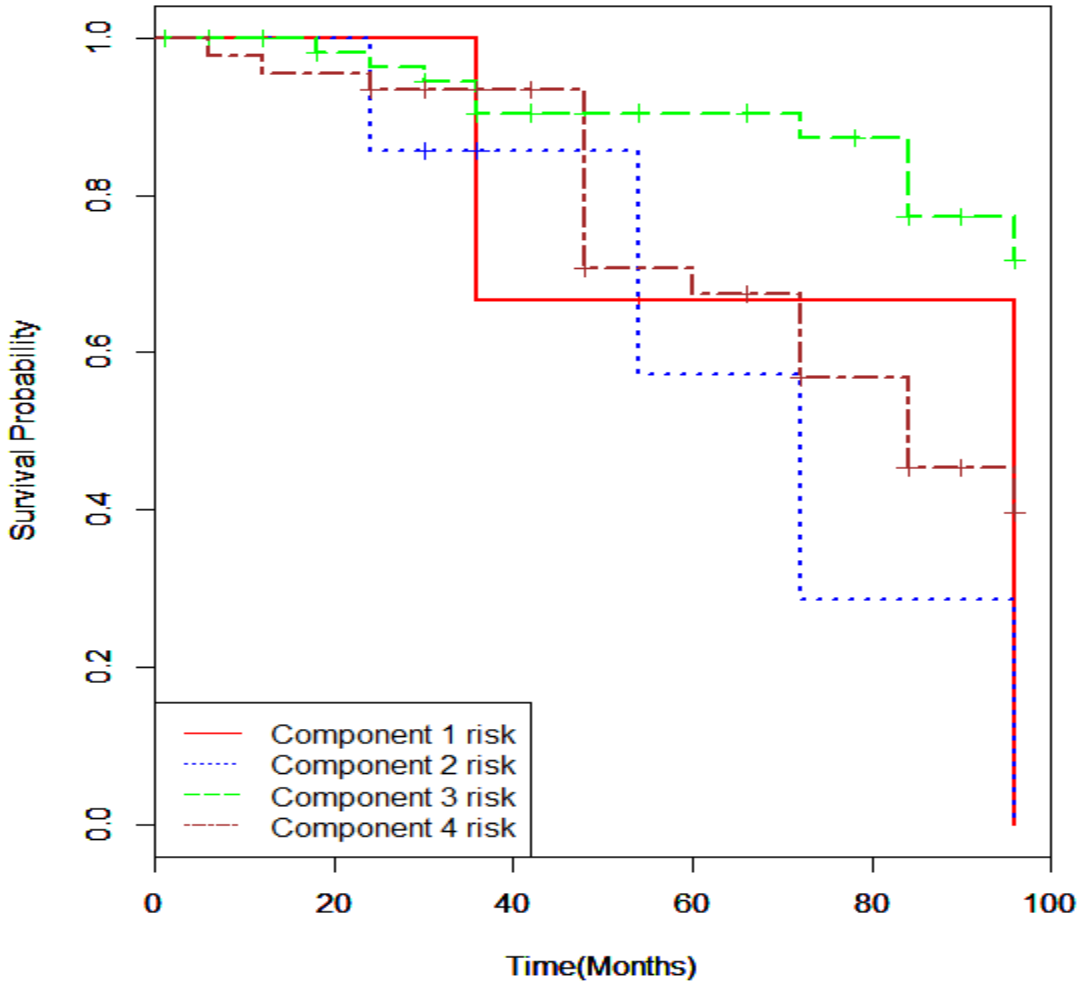

Figure 2.15: BNM Four Component Kaplan Meier Plots

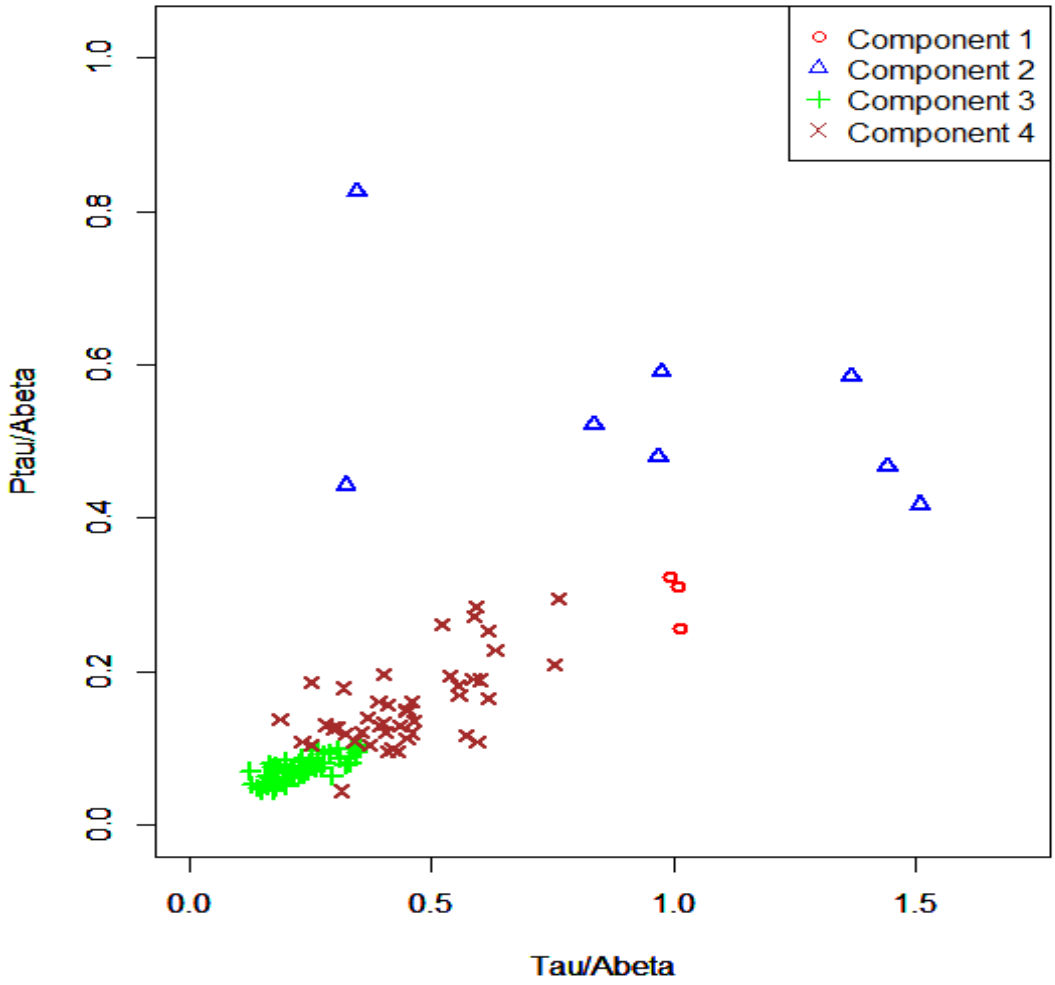

Figure 2.16: BNM Predicted Four Component Plots

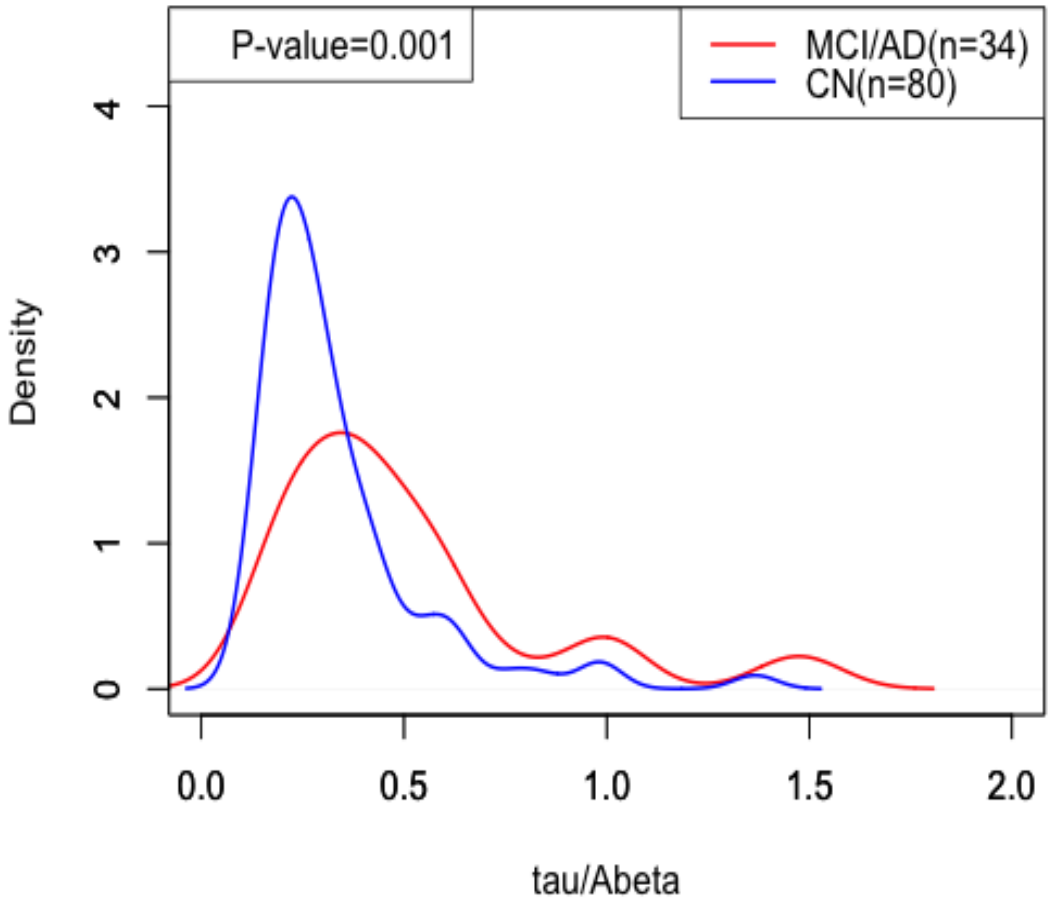

Figure 2.17: Comparison of rtaubeta between CN and MCI/AD groups

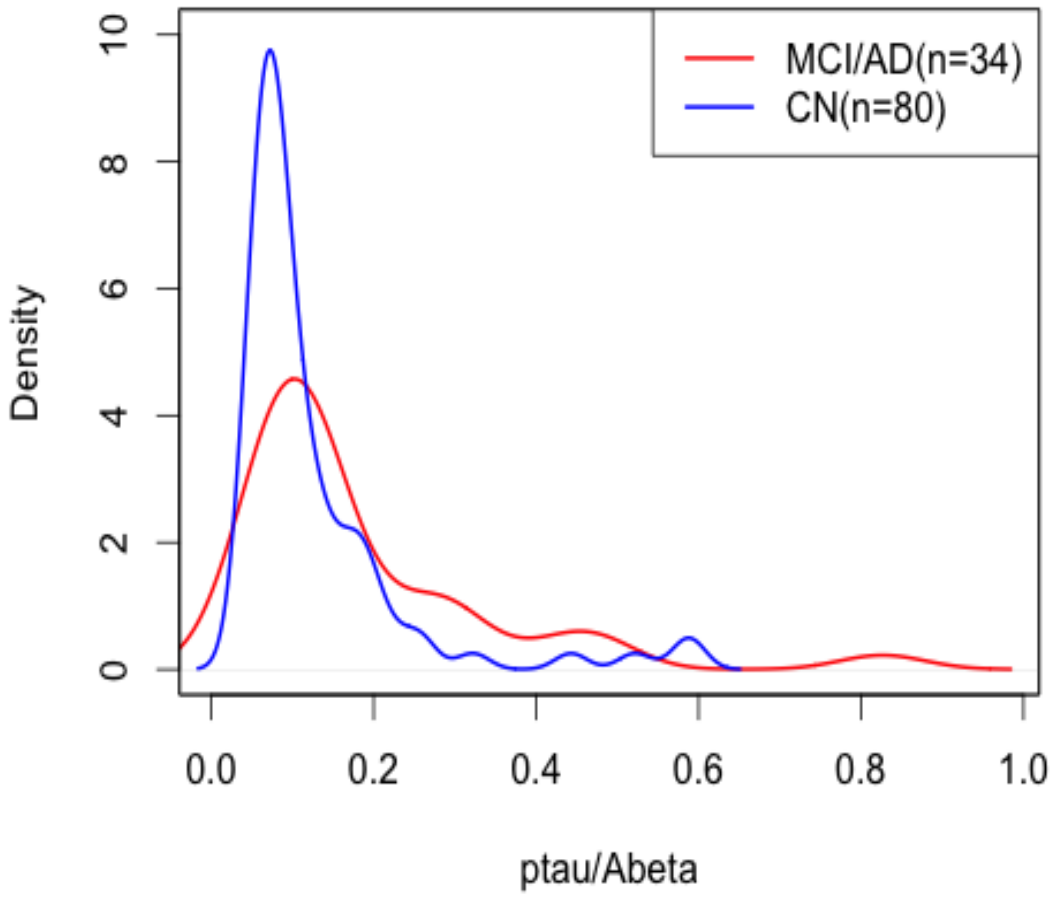

Figure 2.18: Comparison of rptaubeta between CN and MCI/AD groups

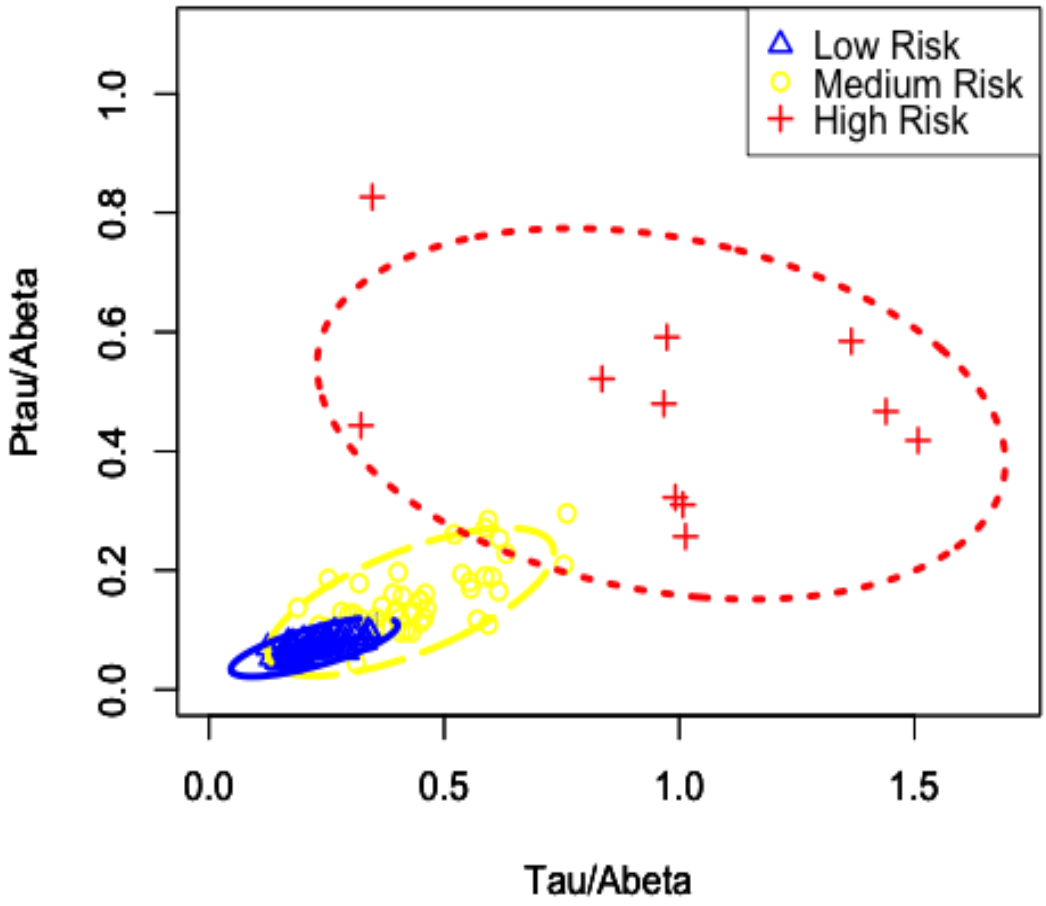

Figure 2.19: Contours embedded on risk components

Table 2.1: Table of Demographics (n=3082), standard deviation(sd), lower and upper quartiles (Q1 and Q3), percentages may not add to 100 due to rounding

Characteristics	**Median(Q1,Q3)**	**Min**	**Max**	**n(%)**
Gender				
Female	-	-	-	779(44.9)
Male	-	-	-	956(55.1)
Ethnicity				
Hispanic/Latino	-	-	-	58(3.3)
Non Hispanic/Latino	-	-	-	1666(96.0)
Unknown	-	-	-	11(0.6)
Race				
American Indian/Alaskan	-	-	-	3.00(0.17)
Asian	-	-	-	29.00(1.67)
Black	-	-	-	77.00(4.44)
Hawaiian	-	-	-	2.00(0.12)
More than one	-	-	-	18.00(1.04)
Unknown	-	-	-	3.00(0.17)
White	-	-	-	1603.00(92.39)
Other				
Age at entry	73.9(69.2,78.9)	48.1	91.4	-
Missing values per variable	-	-	-	1347.00(43.71)

Table 2.2: Table of Demographics (n=114), standard deviation(sd), lower and upper quartiles (Q1 and Q3), rptaubeta is the ratio of PTAU181P to ABETA142 and rtaubeta is the ratio of TAU to ABETA142

Characteristics	**Median(Q1,Q3)**	**Mean (sd)**	**Min**	**Max**	**n(%)**
Biomarkers (Units)					
Tau(pg/ml)	61.00(32.00, 85.25)	69.68(30.37)	32.00	194.00	-
Abeta142(pg/ml)	217.00(75.00, 252.80)	205.60(55.09)	75.00	300.00	-
Ptau181p(pg/ml)	20.00(10.00, 28.75)	24.86(14.58)	10	83	-
Derived Biomarkers					
rtaubeta	0.31(0.21,0.45)	0.39 (0.27)	0.13	1.51	-
rptaubeta	0.10(0.04,0.16)	0.14 (0.13)	0.04	0.82	-
Cognitive Test					
CDRSB	0.00(0.00, 0.00)	0.06	0.00	2.50	-
MMSE baseline	29.00(29.00,30.00)	29.09	25	30.00	-
ADAS11	6.33(4.00, 8.33)	6.53	1.67	15.33	-
Gene					
Apoe4	0.00(0.00,0.00)	0.24	0.00	1.00	27(24)
Race					
Black	-	-	-	-	10(9)
White	-	-	-	-	104(91)
Gender					
Female	-	-	-	-	56(49)
Male	-	-	-	-	58(51)
Other (Units)					
Age at entry(years)	75.55(71.85,78.51)	75.51(5.2)	62.0	89.6	-
Education(years)	16.00(14.00,18.00)	15.79	6.00	20.00	-
Time(month)	54.00(36.00,90.00)	59.33(29.54)	1.00	96.00	-

Table 2.3: Selection of model complexity with three criteria

Complexity/Criteria	AIC	BIC	sBIC
1	201.31	187.63	199.47
2	472.75	442.65	468.70
3	515.40	468.88	509.14
4	526.32	463.38	517.84

Table 2.4: Medium/ high are respectively the component two/three rounded estimated membership probabilities for the hard and soft classification. c is the concordance. ** significant at 0.01 level and * significant at 0.05 level. Sample size is n=114. HR: estimated hazard ratio, SE: standard error of log(HR), 95% CI: 95% confidence interval, GPH Test: Global proportional hazard test

Hard classification	HR	SE	P-value	95% CI	c	GPH Test
Medium|Low risk	3.02	0.41	$< 0.01^{**}$	(1.36, 6.68)	0.63	0.38
High|Low risk	5.35	0.53	$< 0.01^{**}$	(1.89, 15.15)		
With Adjustment						
Medium|Low risk	4.30	0.42	0.0005	(1.90, 9.76)	0.77	0.58
High|Low risk	4.49	0.55	0.006	(1.56, 13.50)		
White	0.45	0.51	0.12	(0.17, 1.23)		
RAVLT	0.94	0.02	0.003	(0.90, 1.98)		
Soft classification						
Medium|Low risk	3.32	0.46	$< 0.01^{**}$	(1.36, 8.13)	0.64	0.53
High|Low risk	6.15	0.54	$< 0.01^{***}$	(2.15, 17.57)		
With Adjustment						
Medium|Low	6.76	0.51	0.0002	(2.48, 18.40)	0.72	0.63
High|Low risk	6.45	0.65	0.0012	(2.09, 19.92)		
White	0.19	0.61	$< 0.01^{**}$	(0.11, 0.87)		
RAVLT	0.93	0.02	0.0.0017	(0.89, 0.98)		

Table 2.5: Assessing CN status of Participants. Bolded observations are potential outliers see Figure 2.7

CDRSB	ADAS11	MMSE	RAVLT	Tau	Abeta142	Ptau181p
0.50	6.67	27	36.36	44	173	15
0.50	8.00	29	30.00	97	216	32
0.50	10.33	29	40.00	48	265	13
0.50	3.33	30	33.33	53	300	13
0.50	5.00	30	30.77	37	201	11
2.50	7.00	30	71.43	119	123	59
0.50	6.67	30	7.69	86	165	43
1.00	11.00	30	**100.00**	61	235	18

Table 2.6: Correlation Coefficient Matrix

	Tau/Abeta	Ptau/Abeta	Tau	Abeta	Ptau
Tau/Abeta	1.00	0.74	0.85	−0.66	0.65
Ptau/Abeta	0.74	1.00	0.54	−0.65	0.92
Tau	0.85	0.54	1.00	−0.27	0.64
Abeta	−0.66	−0.64	−0.27	1.00	−0.40
ptau	0.65	0.92	0.64	−0.40	1.00

Table 2.7: Medium/ high are respectively the component two/three rounded estimated membership probabilities for the hard and soft classification. c is the concordance. ** significant at 0.01 level and * significant at 0.05 level. Sample size is n=106. HR: estimated hazard ratio, SE: standard error of log(HR), 95% CI: 95% confidence interval, GPH Test: Global proportional hazard test

Hard classification	HR	SE	P-value	95% CI	c	GPH Test
Medium|Low risk	2.68	0.41	0.02	(1.19, 6.02)	0.63	0.49
High|Low risk	4.76	0.56	$< 0.01^{**}$	(1.58, 14.30)		
With Adjustment						
Medium|Low risk	2.71	0.43	0.02^{*}	(1.11, 6.30)	0.73	0.58
High|Low risk	3.65	0.63	0.04^{*}	(1.06, 12.56)		
Apoe4	2.15	0.44	0.08	(0.91, 5.07)		
Male	1.37	0.43	0.48	(0.58, 3.17)		
White	0.23	0.60	0.02^{*}	(0.07, 0.76)		
Baseline MMSE	1.29	0.21	0.23	(0.85, 1.96)		
Education	0.99	0.07	0.87	(0.87, 1.13)		
Age	0.97	0.05	0.51	(0.88, 1.06)		
Soft classification						
Medium|Low risk	3.39	0.48	0.01	(1.32, 8.75)	0.66	0.67
High|Low risk	6.33	0.57	$< 0.01^{**}$	(2.06, 19.48)		
With Adjustment						
Medium|Low risk	3.93	0.52	$< 0.01^{**}$	(1.41, 10.99)	0.74	0.63
High|Low risk	5.66	0.67	$< 0.01^{**}$	(1.52, 21.12)		
Apoe4	1.88	0.45	0.16	(0.77, 4.56)		
Male	1.33	0.43	0.51	(0.57, 3.10)		
White	0.19	0.62	$< 0.01^{**}$	(0.06, 0.65)		
Baseline MMSE	1.28	0.21	0.24	(0.84, 1.94)		
Education	0.99	0.07	0.87	(0.87, 1.13)		
Age	0.97	0.05	0.55	(0.89, 1.06)		

Table 2.8: Component estimated parameters. SE: standard error based on Bootstrap sampling with $B = 1000$ in R mixtools package,C1-C3 are components 1 through 3.

	$\hat{\lambda}(SE_{\hat{\lambda}})$	$\hat{\mu}(SE_{\hat{\mu}})$	$\hat{\Sigma}$	$SE_{\hat{\Sigma}}$
C1	0.41(0.06)	0.2235(0.0096), 0.0684(0.0027)	$\begin{bmatrix} 0.0029 & 0.0005 \\ 0.0005 & 0.0002 \end{bmatrix}$	$\begin{bmatrix} 0.0009 & 0.0002 \\ 0.0002 & 0.0001 \end{bmatrix}$
C2	0.47(0.06)	0.3883(0.0264), 0.1282(0.0104)	$\begin{bmatrix} 0.0191 & 0.0049 \\ 0.0049 & 0.0028 \end{bmatrix}$	$\begin{bmatrix} 0.0052 & 0.0019 \\ 0.0019 & 0.0008 \end{bmatrix}$
C3	0.12(0.03)	0.9334(0.1179), 0.4589(0.0507)	$\begin{bmatrix} 0.1504 & -0.0303 \\ -0.0303 & 0.0274 \end{bmatrix}$	$\begin{bmatrix} 0.0547 & 0.0181 \\ 0.0181 & 0.0104 \end{bmatrix}$

Table 2.9: Contingency table to compare component predicted values to the true values

Actual (%)/Predicted(%)	Normal	AD/MCI	
Component 1	27(58.6)	19(41.3)	46
Component 2	48(84.2)	9(15.8)	57
Component 3	5(45.5)	6(54.5)	11
	80	34	114

Table 2.10: Sensitivity and Specificity

	Sensitivity	Specificity
Component 1 or 3	$\frac{25}{34}$(74%)	$\frac{48}{80}$(60%)
Component 3	$\frac{6}{34}$(18%)	$\frac{75}{80}$(94%)

Chapter 3 Application of Mixture of Linear Regressions Models And the Approximate Singular Bayesian Information Criterion

3.1 Introduction

The use of biomarkers to predict the potential of developing MCI/AD is not a new concept. However to make such a prediction while the individuals are still cognitively normal is rare in the literature. In the second chapter of this dissertation we attempted to address this problem using mixture models without covariates to see if the biomarker ratios by themselves can adequately predict future disease status.

In this chapter, we are interested in tapping into another form of mixture modeling; mixture of regressions, to address the same scientific problem within a more sophisticated analytical framework. Our approach is similar to [46] in that we will regress biomarkers on covariates within each of a finite set of components. Our analyses differ from existing ones in the following ways:

1. Unlike [46], our response variable is either trivariate or bivariate depending on whether we use the three biomarkers or two biomaker ratios in the mixture modeling.

2. We will also account for more than one covariate in our mixture modeling; as a result, we will be fitting a hyperplane in each of the finite number of components.

3. We will determine the complexity of the mixture modeling based on two existing model selection criteria, namely AIC and BIC,and a recently added criterion sBIC. If there are disagreements in the estimated model complexity we will revert to sBIC to select the correct model as sBIC has advantages over AIC and BIC; in particular AIC is generally inconsistent as it tends to over estimate the number of components and BIC though consistent often underestimates. Since neither AIC nor BIC uses the correct number of parameters for the respective penalties they impose for singular models such as mixtures, they cannot be used to used to estimate posterior probabilities and thus cannot be used to assess uncertainty regarding model's correctness.

3.2 General Overview of Mixture of Regression Models With An Illustration

We begin with a general overview of mixture of linear regression specific to our study and follow-up with an example for a single subject. First we define the general equations and the accompanying notations to be used throughout chapter three.

$$\mathbf{Y}_{n\times 2} = \mathcal{A}_x \mathcal{B}_x \mathcal{W} + \boldsymbol{\varepsilon}_{n\times 2} \boldsymbol{\Sigma}^{\frac{1}{2}}_{2\times 2} \tag{3.1}$$

where we assume that $\varepsilon \sim N(0, 1)$ and each of the entities in the model is explicitly defined as follows:

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \\ \vdots & \vdots \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{pmatrix} = \begin{pmatrix} Y_1^T \\ Y_2^T \\ Y_3^T \\ \vdots \\ Y_n^T \end{pmatrix}$$

$$\mathcal{W} = \begin{pmatrix} 1 & w_{11} & w_{12} & \dots & w_{1p} \\ 1 & w_{21} & w_{22} & \dots & w_{2p} \\ 1 & w_{31} & w_{32} & \dots & w_{3p} \\ 1 & w_{41} & w_{42} & \dots & w_{4p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & w_{n1} & w_{n2} & \dots & w_{np} \end{pmatrix} = \begin{pmatrix} W_1^T \\ W_2^T \\ W_3^T \\ \\ \vdots \\ W_n^T \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01x} & \beta_{02x} \\ \beta_{11x} & \beta_{12x} \\ \beta_{21x} & \beta_{22x} \\ \beta_{31x} & \beta_{32x} \\ \vdots & \vdots \\ \vdots & \vdots \\ \beta_{p1x} & \beta_{p2x} \end{pmatrix} = \begin{pmatrix} \beta_{0x}^T \\ \beta_{1x}^T \\ \beta_{2x}^T \\ \\ \vdots \\ \beta_{px}^T \end{pmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \\ \epsilon_{31} & \epsilon_{32} \\ \vdots & \vdots \\ \vdots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} \end{pmatrix} = \begin{pmatrix} \epsilon_1^T \\ \epsilon_2^T \\ \epsilon_3^T \\ \\ \vdots \\ \\ \epsilon_n^T \end{pmatrix}$$

Thus for a given individual the different matrix components with complexity $m = 3$ can be simplified as follows. Later in the chapter we will use these information to illustrate how a full model structure for an individual is formulated.

$$\mathcal{A}_{2\times 6} = \begin{pmatrix} a_1^T & 0_{1\times 3} \\ 0_{1\times 3} & a_1^T \end{pmatrix}, \mathcal{B}_{6\times 2} = \begin{pmatrix} B_{01} & B_{11} \\ B_{02} & B_{12} \end{pmatrix}, \mathcal{W}_{1\times 2} = \begin{pmatrix} 1 \\ w \end{pmatrix}$$

$$C_{2\times 6} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{21} & \sigma_{22} & \sigma_{23} \end{pmatrix}, \mathcal{Z}_{6\times 2} = \begin{pmatrix} a_1^T & 0_{1\times 3} \\ a_1^T & 0_{1\times 3} \end{pmatrix}$$

$$\mathbf{a}_1 = \begin{pmatrix} \mathbb{1}_{x_1=1} \\ \mathbb{1}_{x_1=2} \\ \mathbb{1}_{x_1=3} \end{pmatrix}$$

Notice from 3.1 that we have an $n \times 2$ bivariate matrix of biomarker ratios $\mathbf{Y}$ and an $n \times (p+1)$ matrix of covariates $\mathcal{W}$, where n is the sample size, p is the number of coefficients not including the intercept. We assume once again that $\mathbf{Y}|\mathcal{X} = x, \mathcal{W} \sim N((\mathcal{A}_x \mathcal{B}_x \mathcal{W}), \mathbf{\Sigma}_x)$ where $\mathbf{X}$ is a vector identifying the component membership and $\mathcal{B}_x$ is an $(p+1) \times 2$ matrix of coefficients.

To further clarify the set up above, we illustrate with the first set of outcomes (y_{11}, y_{12}) and how it relates to the regression coefficients and the covariates, taking into consideration the potential components that this individual could belong to. Here y_{11} and y_{12} are respectively the first and second biomarker ratios for the first individual. Furthermore a coefficient such as β_{123} will imply the slope (in the sense of linear regression) for the main effect of the first covariate on the second biomarker ratio in the third mixture component. Thus a general index of the form ijk will correspond to the i^{th} covariate effect on the j^{th} outcome in the k^{th} mixture component such that $i = 0, 1, 2, ..., p,\ j = 1, 2$ and $k = 1, 2, 3, ..., m$. For now let us assume that there are three components (i.e. $m = 3$) and assume one covariate ($p = 1$). Notice that we can present this concept in a matrix form as follows:

$$\begin{pmatrix} y_{11} & y_{12} \end{pmatrix} = \mathbf{Y}_1 = \begin{pmatrix} \mathbb{1}_{x_1=1}\{\beta_{011} + \beta_{111}w_1\} + \mathbb{1}_{x_1=2}\{\beta_{012} + \beta_{112}w_1\} + \mathbb{1}_{x_1=3}\{\beta_{013} + \beta_{113}w_1\} \\ \mathbb{1}_{x_1=1}\{\beta_{021} + \beta_{121}w_1\} + \mathbb{1}_{x_1=2}\{\beta_{022} + \beta_{122}w_1\} + \mathbb{1}_{x_1=3}\{\beta_{023} + \beta_{123}w_1\} \end{pmatrix} + \boldsymbol{\varepsilon_1}$$

Let

$$\boldsymbol{\beta_{01}} = \begin{pmatrix} \beta_{011} \\ \beta_{012} \\ \beta_{013} \end{pmatrix}, \boldsymbol{\beta_{11}} = \begin{pmatrix} \beta_{111} \\ \beta_{112} \\ \beta_{113} \end{pmatrix}, \boldsymbol{\beta_{02}} = \begin{pmatrix} \beta_{021} \\ \beta_{022} \\ \beta_{023} \end{pmatrix}, \boldsymbol{\beta_{12}} = \begin{pmatrix} \beta_{121} \\ \beta_{122} \\ \beta_{123} \end{pmatrix}, \mathbf{a}_1 = \begin{pmatrix} \mathbb{1}_{x_1=1} \\ \mathbb{1}_{x_1=2} \\ \mathbb{1}_{x_1=3} \end{pmatrix}$$

then we obtain the matrix expression

$$\begin{pmatrix} y_{11} & y_{12} \end{pmatrix} = \mathbf{Y}_1 = \begin{pmatrix} \mathbf{a}_1^T\boldsymbol{\beta_{01}} & \mathbf{a}_1^T\boldsymbol{\beta_{02}} \\ \mathbf{a}_1^T\boldsymbol{\beta_{11}} & \mathbf{a}_1^T\boldsymbol{\beta_{12}} \end{pmatrix} \begin{pmatrix} 1 \\ w \end{pmatrix} + \boldsymbol{\varepsilon_1}$$

which corresponds to the matrix in equation 3.1

Note also that from equation 3.1 the error term $\boldsymbol{\varepsilon_1} = \boldsymbol{\varepsilon}'\Sigma_x^{\frac{1}{2}}$ can be explicitly presented as follows:

$$\Sigma_{x_1=1} = \begin{pmatrix} \sigma_{11}^2 & \sigma_1 \\ \sigma_1 & \sigma_{21}^2 \end{pmatrix}, \Sigma_{x_1=2} = \begin{pmatrix} \sigma_{12}^2 & \sigma_2 \\ \sigma_2 & \sigma_{22}^2 \end{pmatrix}, \Sigma_{x_1=3} = \begin{pmatrix} \sigma_{13}^2 & \sigma_3 \\ \sigma_3 & \sigma_{23}^2 \end{pmatrix}$$

where σ_{jk}^2 is the variance of the j^{th} outcome in the k^{th} component and σ_k is the covariance between the j^{th} and $j^{th}+1$ outcome in the k^{th} component.

Letting x represent the mixture component we define the error term corresponding to the first observation as follows:

$$\begin{aligned} \boldsymbol{\varepsilon}_1 &= \left(\mathbb{1}_{x_1=1}\Sigma_1^{\frac{1}{2}} + \mathbb{1}_{x_1=2}\Sigma_2^{\frac{1}{2}} + \mathbb{1}_{x_1=3}\Sigma_3^{\frac{1}{2}} \right)\boldsymbol{\varepsilon}_1' \\ &= \left(\Sigma_1^{\frac{1}{2}} \quad \Sigma_2^{\frac{1}{2}} \quad \Sigma_3^{\frac{1}{2}} \right) \begin{pmatrix} \mathbb{1}_{x_1=1} \\ \mathbb{1}_{x_1=2} \\ \mathbb{1}_{x_1=3} \end{pmatrix} \boldsymbol{\varepsilon}_1' \end{aligned} \quad (3.2)$$

We re-write 3.2 as follows using the matrices defined above:

$$\boldsymbol{\varepsilon_1}_{2\times 1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{21} & \sigma_{22} & \sigma_{23} \end{pmatrix}_{2\times 6} \begin{pmatrix} a_1^T & 0_{1\times 3} \\ a_1^T & 0_{1\times 3} \end{pmatrix}_{6\times 2} \boldsymbol{\varepsilon}'_{\mathbf{1}2\times 1} = C\mathcal{Z}\boldsymbol{\varepsilon}'_{\mathbf{1}} \quad (3.3)$$

3.3 Overview of Primary Objectives

The primary objectives underscoring this chapter are as follows:

1. Use sBIC, AIC and BIC to determine the degree of complexity of a mixture of regression modeling while accounting for race and other covariates in the mixture. We will use the biomarker ratios and covariates in the mixture modeling akin to [46].

2. We will examine the future predictive ability of the chosen model. To accomplish this we will compare the predicted classes into which subjects were placed to their true future disease status so that we can address questions such as what a person's hazard or relative risk for developing MCI/AD is given what we know about their current age and other covariates. Furthermore we will be able to use their posterior probabilities to create a figure as in Figure 2.11 to make it easier for clinicians to adequately determine an individual's risk status.

3. Develop a risk strata plot for predictive purposes.

We begin with a graphical quick review of the biomarker ratios against selected covariates of interest. The rationale here is to get a sense of empirical support for the models we intend to fit. Figure 3.1 indicates that the mean biomarker ratios for Blacks are lower than that of Whites. However both means are below the grand mean of the biomarkers. The figure also reveals that the mean effect of race does not statistically influence the mean effect of the biomarkers which is why the line (a

degenerated ellipse because we have one degree of freedom) indicative of the hypothesis variation is within the error variation ellipse.

Figure 3.2 however shows that Apoe4 statistically significantly influences the mean biomarker ratios since the hypothesis variation line crosses the error variation ellipse. As shown in the plot, carriers of Apoe4 tend to express biomarker ratios higher than that of the grand mean whereas non-carriers tend to have lower average biomaker ratios than the grand mean.

Figure 3.3 examines the effect of both race and Apoe4 concurrently on the biomarker ratios. Again we notice that Apoe4 shows a significant effect whereas race does not. Also carriers and non carriers of Apoe4 have relatively higher biomarker ratios that are also above the grand mean than the two racial groups. Another interesting observation from Figure 3.3 is the lack of parallelism between the two hypothesis variations. Introducing an interaction term between apoe4 and race (plot not shown) resulted in a much shorter hypothesis variation plot deeply embedded in the error ellipse; an indication of an insignificant interaction between race and Apoe4.

The three plots in Figures 3.1-3.3 indicate that both race and Apoe4 influence the biomarker ratios in the same direction and in a linear fashion. However, Apoe4 will exhibit a greater slope hence comparatively more influence on the biomarker ratios than race as demonstrated by their respective degenerated hypothesis variation ellipses.

3.4 Methodologies and Overview of AIC

Participants and Cognitive Status

In chapter two we discussed the participants from the ADNI data and also explicitly assessed the cognitive status of the participants based on the information available in the data namely MMSE scores, CDRSB, ADAS11 and RAVLT percentage of forgetting. From our assessment we concluded that all but eight participants in the study can be considered cognitively normal as initially identified by ADNI. The eight participants had questionable CDRSB scores and thus were included and excluded in the model to measure any effect they my have. As seen in chapter two, the estimates from the fitted model and their corresponding standard errors were all reasonably similar and thus only the model with all the participants was reported.

In this chapter we operate on the foundation laid in chapter two; that the participants are indeed cognitively normal and thus we work with all 114 participants.

Statistical Modeling

We use flexmix[41] an R package to model a mixture of linear regression that will be used to classify subjects into one of a finite set of groups namely. These groups may will be labeled for example as high risk, medium risk, intermediate risk, low risk etc. The flexmix package assumes independence between the two response variable of interest and fits the mixture model providing estimates such as main/interactive effects, standard errors within each component. It also provides the variance of the response in each component in addition to overall component prior, ratio and posterior probability plot. Readers interested in knowing more about how flexmix operates can see [41] for details.

Similar to chapter two, we focus on using the two biomarker rations namely rtaubeta and rptaubeta as the response variables. We will assess the collective, individual and interactive effect of the following explanatory variables in this chapter: age, gender, race, education, APOE4 and baseline MMSE. In the interim we assume zero covariance between rtaubeta and rptaubeta after accounting for other covariates. This assumption will be duly adjusted if it fails the diagnostic test that is developed for assessing the model's efficacy. We also assume that the distributions of the responses are normal by virtue of the empirical evidence in the histogram plots.

We use sBIC to ascertain the number of components and the covariates to include in the model concurrently. As a result we consider all possible combinations of the covariates in the mixture model and then record their corresponding sBIC values. The model with the highest sBIC value will be noted and compared with the model with the highest change in sBIC in moving from a less complex model to a more complex one. The final phase ensures that the resulting model has estimable standard errors. Models that are selected by sBIC but exhibit lack of stability in their standard errors will be traded for those with slightly lower sBIC in addition to having stable and smaller standard errors.

The chosen model is used to obtain the posterior probability of each subject. The raw posterior probabilities will be the soft classification of the subjects into different risk classes given the covariates. The hard classification is obtained by selecting

the most probable cluster of belongingness for each subject from the set of components.The hard classification probabilities are used to obtain risk plots and survival curves.

Cox model is used to validate the model by determining how well the posterior probabilities predict future cognitive status of each subject. The posterior probabilities corresponding to low risk as determined by the survival plot will be the baseline in the Cox model. The performance of the prediction of the posterior probabilities will be estimated with the c statistic that is automatically generated as part of the Cox modeling output. Backward elimination method will be used to identify the most influential covariates in addition to the posterior probability. The final Cox model output is obtained for the selected covariate(s) and the corresponding c-statistic and p-values noted. The latter model is compared with the model with only the posterior probability as the covariate to determine which model fits the data best in terms of predicting subjects' transition from normal cognition. Thus we will decide between using only the posterior probabilities in the Cox model or the posterior probability with other significant covariates in the Cox model.

In-consistency of AIC

Suppose that $A\hat{I}C_j$ denotes the AIC selected model, then we would define the consistency of $A\hat{I}C_j$ as $P(A\hat{I}C_j = m_0) \to 1$ as $n \to \infty$. We wish to examine if this consistency hold for AIC without structural parameters as in Charnigo and Pilla (2007) and in a multivariate setting, given that m_0 is the true model.

Let $AIC_1 = 2\log L_1 - 2p_1$ and $AIC_2 = 2\log L_2 - 2p_2$ be the respective AIC's for models 1 and 2, where $2p$ is penalty (which depends on the number of free parameters). Then we note that AIC will choose model 1 over model 2 if $AIC_1 > AIC_2 \Rightarrow$

$$2\log L_1 - 2p_1 > 2\log L_2 - 2p_2 \tag{3.4}$$

$\Rightarrow$

$$2\log\frac{L_1}{L_2} > 2(p_1 - p_2) \tag{3.5}$$

For a multivariate normal mixture model with structural parameter we know that $2\log\frac{L_1}{L_2} \xrightarrow{L} \sup_{\theta\in\Theta}(W^+(\theta))^2$ as $n \to \infty$ when model 1 is correct where W^+ is a truncated Gaussian process as defined by Chen and Chen (2001)[4].

It follows that

$$P(A\hat{I}C_j = m_0) \leq P\left(2\log\frac{L_1}{L_2} > 2(p_1 - p_2)\right) \xrightarrow{L} \sup_{\theta\in\Theta}(W^+(\theta))^2 > 2(p_1 - p_2) \tag{3.6}$$

as $n \to \infty$ which is independent of n and thus regardless of how large n is the probability will not approach 1.Thus $P(A\hat{I}C_j = m_0) \nrightarrow 1$ if $m_0 = 1$ demonstrating

inconsistency.

3.5 Results and Discussions

The mixture of regressions with covariate race is preferred among more than eighty models considered. This is based on the chosen model scoring one of the highest sBIC values for a two component mixture, presenting stable (or estimable) standard errors and having well separated posterior plots as shown in Figure 3.4. The favored model also has estimable standard errors within each component as demonstrated in Table 3.2. Other two models with either apoe4 alone or Apoe4 and race as covariates share these desirable characteristics and will be included in the following discussion.

In Table 3.1, we present the three criteria used in the model selection procedure with our focus on the sBIC since it has ability to assess posterior probability of a singular correctness of model unlike AIC and BIC. Table 3.1 indicates that the two component model is preferred to a one component model by all three criteria. It should be noted that a four component model was preferred by sBIC with covariates age, gender and their interaction. However, this model and a similar three component model exhibited unstable standard errors, and thus the next tier model which has suitable characteristics in addition to high sBIC is considered instead.

Three models will now be discussed following from the latter considerations. The first model has race as a predictor variable, the second, Apoe4, and the last both race and Apoe4. Henceforth, these models will be referred to as race model, Apoe4 model and race and Apoe4 model respectively. Although these three models may share similar properties in terms of the stability of standard errors and future pre-

dictive capabilities, the model with race will be given more attention relative to the other two due to the following reasons:

1. The race model was chosen by sBIC as the better model among the three for consideration by assigning the model the highest value.

2. The race model has a slightly better estimated concordance statistic (67%) as shown in Table 3.3 than its competitors (66%, and 67% for Apoe4 alone as predictor and race+Apoe4 as predictors respectively in Tables 3.5 and 3.7) when the high risk posterior probability alone is used in the Cox model. Notice that the c-statistic corresponding to the model with apoe4 and race model is the same as that in the model with race without adjustment. This shows that using race alone is preferable if we appeal to parsimonious modeling procedures.

3. The race model exhibits comparatively better well separatedness characteristics as demonstrated in the plot of the posterior probabilities in Figure 3.4 compared to Figures 3.8 and 3.12 for the Apoe4 model alone and Apoe and race models respectively . This is key in knowing how well the clusters will be distinguished in later analysis.

4. When race alone is included in the Cox model, it was not a significant predictor of the event. This may be expected since 91% of all the participants being Caucasians creates an imbalance that leads to a low power of detecting any effect. Furthermore, this outcome also suggests that indeed the posterior probabilities are more statistically related to the estimated hazard ratio or ones propensity

of transitioning from the normal state of cognition independent of race. In the competing model, Apoe4 alone in the Cox model is a significant predictor of the event albeit with significantly lower c-statistic(62%). Although one may argue that 66% is indeed significantly different from 62% it still leaves room for skepticism about the use of mixture modeling with Apoe4 if including directly in the Cox model could produce a less complicated model that performs just as well.

5. The risk strata obtained from the posterior probabilities associated with the race model as shown in Figure 3.7 indicates that both biomarker ratios are key in determining the state of cognition when race is used as the predictor variable. Compared to the risk strata in Figure 3.11 obtained from the Apoe4 model, rtaubeta seem to be most influential in determining the strata as all the boundaries seem to be almost vertical (this may be peculiar to the nature of our study in that we are predicting future disease status from people who are presently cognitively normal). This further deepens our trust in the race model as we wish to explore the effects of both biomarker ratios in predicting future cognitive status. Figure 3.15 depicts the risk strata plot for the race and Apoe4 model. This plot looks similar to that of Figure 3.7 when only race was introduced to the model. Thus if we appeal to parsimonious modeling procedures we will prefer the race model to the race and Apoe4 model.

6. To the best of our knowledge Apoe4 has been well studied in the Alzheimer's Disease literature although not in the sense of mixture modeling, but little is known about the how race might increase ones risk of getting the disease.

Among the few existing studies Shadlen and colleagues(2006)[53] have noted no significant differences in the risk of dementia between Black and White subjects after accounting for confounders. Our study confirms their finding on one hand in the sense that adjusting for race in the Cox model results in race being a non-significant predictor of the event. However, our study also reveals an interesting fact that race actually plays a vital role in determining the posterior probabilities of the subjects which is a significant predictor of the event. Here we should recognize the unique role of the mixture modeling approach. As we have emphasized already, the mixture model approach helped identify other roles of race in predicting future cognition that otherwise would have been masked by using simpler analytical approaches.

We also note that the participants in the Shadlen et. al study consist of subjects who are either cognitively normal, have incident dementia, prevalent dementia or have MCI. Our study presents an approach to the same problem from a strikingly different perspective stretching from the make-up of the participants (all cognitively normal) to the methodological approaches adapted (mixture of linear models).

As we have demonstrated above, the mixture of regression model with race as covariate seem to embody favorable characteristics worthy of detailed study. The model output fitted to each of the two components is displayed in Table 3.2. Component 1 indicates that neither race nor the intercept are significant predictors of rtaubeta. We note that being of White race has a relatively larger effect on rtaubeta in component 1 (0.163 units) than in component 2 (0.036 units). In component two

being White increases ones rtaubeta by about 0.04 units albeit not statistically significant. The picture is not different when the model is fitted for rptaubeta. Once again, race is not a significant predictor of rptaubeta in both components one and two but its effects on rptaubeta are much greater in component one than two. A similar trend is observed in the intercept with component one intercept being greater than that of component two when rtaubeta or rptaubeta was the response.

Contrasting the above observation with the output from the competing models we note that the Apoe4 model shows that Apoe4 is indeed a highly significant predictor of both rtaubeta and rptaubeta in both components as seen in Table 3.4. Similar to the race model, the Apoe4 model shows a greater main effect (slope) in component one than two. When the two predictors were included in the model, Apoe4 was still a significant predictor of both biomarker ratios adjusting for race; and race was still not a significant predictor of the biomarker ratios in both components adjusting for Apoe4 as shown in table 3.6. However Table 3.6 also revealed the comparable trend of larger slopes in component one than two as indicated in the component plots in Figure 3.5.

Figure 3.5, reinforces what we observed from the fitted models within each component as it indicates that the relationship in component one follows a less linear pattern than that in component two. Another observation is clear; the intercepts in component one from Table 3.2 is much larger than that of component two for both rtaubeta and rptaubeta respectively. This follows from the fact that we would need

a larger intercept to fit the plot in component one than in component two.

In Figure 3.9 in many ways resembles that of Figures 3.5 and 3.13. However Figure 3.9 classifies participants with rtaubeta between 0.5 and 0.8 into component two whereas Figures 3.5 and 3.13 classify these participants into component one. Thus by virtue of the Apoe4 model, having a relatively higher rtaubeta but lower rptaubeta may not elevate ones risk of transitioning; whereas in the race model as well as the race and Apoe4 model, such a scenario will place the participant on the path to increased risk of transitioning.

The risk of transitioning from normal cognition in each group is shown in Figure 3.6. Component two members experience the earliest decline in cognition and the least risk as a result of comparatively gentler decline in cognition from month 5 to the 96^{th} months dotted with lots of censoring. Members in component one however experienced fewer censorship with comparatively sharper decline in cognition starting from around months 20 to months 96. Component one membership is indicative of comparatively latter decline in cognition around months 25 of follow-up and sharpest decline after month 25.

In comparison to the risk plot in Figure 3.10 associated with the Apoe4 model, the two risk components begin to separate rather earlier (about week 10) than that of the race model or the race and Apoe4 model as shown in Figures 3.6 and 3.14 respectively. The separation between the risk groups in Figures 3.6 and 3.14 appear

to be widening over time but this is not the case or at least not so obvious in the risk plot of the Apoe4 model.

The Cox proportional hazard model presented in Table 3.3 shows the predictive abilities of the soft and hard classified posterior probabilities of the race model. Based on the risk plot in Figure 3.6 we will reference component two as the baseline risk component. From Table 3.3 we notice from the soft classification model that when the raw posterior probabilities are entered into the Cox model alone, the estimated hazard ratio is $4.621(95\%CI = (1.661, 12.860))$. Thus the risk of transitioning from the normal cognitive state increases by four fold if one is in component one compared to being in component two. Adjusting for posterior probabilities for component one, being of White race is protective against ($\hat{HR} = 0.40, 95\%CI = (0.15, 1.08)$) transitioning from normal cognition compared to being of Black race safe that it is statistically insignificant. The corresponding c-statistic for the soft classification model with only the posterior probabilities of component one members is 67.3% and that of the adjusted model for immediate RAVLT is 77.6% respectively. We note here that when race was adjusted for the c-statistic was 76.7. This may suggest that race does not add any more information in distinguishing between those who will transition and those who will not when the posterior probabilities are included in the model. Of note Apoe4 was not a significant predictor ($\hat{HR} = 1.71, 95\%CI = (0.88, 3.30)$) of the outcome in the presence of the risk probabilities and immediate RAVLT.

In the Apoe4 model, Apoe4 was a significant predictor of the biomarker ratios

(Table3.4). The model also demonstrate a c-statistic of 66% when the only predictor is the posterior probabilities in the soft classification model(Table 3.5). The hazard of being in component one is $4.59, 95\%CI = (1.844, 11.420)$ times that of being in component two. Adjusting for Apoe4 increased the c-statistic modestly to 68% and adjusting for immediate RAVLT alone also resulted in an increase in the c-statistic. When both Apoe4 and RAVLT immediate were adjusted for, the posterior probability (high risk) was no longer a significant predictor of the hazard (results not shown). Notably both Apoe4 and RAVLT were significant predictors of the outcome when the risk probabilities were eliminated from the Cox model (results not shown). Table 3.7 presents the validation for Apoe4 and race model. As seen in Table 3.6, only Apoe4 was a significant predictor of the outcome with the two components. Table 3.7 shows that in the unadjusted model, the hazard risk associated with component two increase $4.099, 95\%CI = (1.49, 11.270)$ times that associated with component one with a corresponding c statistic of 60%. Adjusting for RAVLT increased the c-statistic to 75.9% (results not shown). Further improvements in the c-statistic (77.2) was gained by adjusting for Apoe4 although the latter was no longer a significant predictor of the outcome (results not shown). Thus in the presence of the risk probabilities derived from race alone or from race and Apoe4 and RAVLT, neither race nor Apoe4 was a significant predictor of time to transition. For instance race was a significant predictor of the outcome when RAVLT was absent from the race and Apoe4 Cox model albeit with a significantly lower c-statistic (67%) (results not shown).

In the hard classification model, belonging to component one referent to component two will significantly increase the risk of transitioning from the normal cognitive state by $3.026, 95\%CI = (1.24, 7.34)$. In the adjusted hard classification model, being of White race again decreases ones risk of transitioning from normal cognition by about 5.7% albeit statistically insignificant. We also note that the c-statistic for the hard classification and the adjusted hard classification models are respectively 55% and 75%. The hard classification model for Apoe4 shows also shows a reduced effect of the posterior probabilities on the hazard ($2.76, 95\%CI = (1.355, 5.623)$) as shown in Table 3.5. Adjusting for Apoe4 in the hard classification model resulted in both the posterior probability and Apoe4 being significant predictors of the hazard. A comparable observation can be made from the race and apoe4 hard classification model in Table 3.7. The posterior probability had a reduced effect ($3.026, 95\%CI = (1.247, 7.342)$) on the hazard of transitioning. When we adjusted for both race and Apoe4 both covariates are significant predictors of hazard ratio but the posterior probability is not.

Further more when only race was presented in the model, being of White race once again reduced the incidence of the event by $41\%(0.49, 95\%CI = (0.19, 1.26)$ (all details not shown) compared to being of Black race. The associated c-statistic was $54\%(SE = 0.026)$. This may suggest that the posterior probabilities from the mixture modeling is more significantly related to the hazard function for transitioning than the raw predictor (in this case race).

In addition to the aforementioned observations Figure 3.7 shows the risk strata associated with each model for diagnostic purposes. As seen in the Figure 3.7, high risk (H.Risk) individuals tend to have high biomarker ratios or one high and one relatively low biomarker ratio. It is clear that an approximate lower bound for the biomarker ratios for the hypothetically high risk individuals are about 0.7 for rptaubeta and about 1.25 for rtaubeta. The hypothetical intermediate risk (I.Risk) group will have an approximate rtaubeta value range of 1.05 to below 1.25 and 0.6 to below 0.7 for rptaubeta. For the medium risk (M.Risk) group, the range for the biomarker ratios are respectively 0.5 to below 0.6 for rptaubeta and about 0.8 to below 1.05 for rtaubeta. The minimal(least) risk group (L.Risk) will exhibit biomarker ratios below 0.5 for rptaubeta and below 0.8 for rtaubeta.
Furthermore since Figure 3.7 is a consequence of the raw posterior probabilities whereas Figure 3.5 is a consequence of the discretized (hard classified) posterior probabilities. The former provides a continuum (as opposed to the latter) of the progression of risk as one advances from a minimum risk region to a high risk, thus presenting a more plausible picture of what could be.

In contrast with Figure 3.15 is in many ways similar to that of Figure 3.5. This may be an indication that race is more influential in the race and Apoe4 model than Apoe4. In Figure 3.11 however, the risk strata shows that rtaubeta is more influential in classifying individuals into the various risk layers. The plot shows that slight changes in rtaubeta beyond 0.3 units could results in a transition with very little influence from the rptaubeta ratio.

Discussion

In this study we have examined the potential of adequately predicting future cognitive status of presently cognitively normal individuals using mixture of linear regressions. We used the ADNI database and with the help of further assessment of the cognitive measures such as MMSE, ADAS11, RVALT and CDRSB scores, 114 participants were confirmed as cognitively normal. We adapted the mixture of linear regressions method for the following reasons:

1. mixture of linear regression model is more adept in handling masked outliers[46].

2. mixture of linear regression model affords us the flexibility needed to understand the linear association between the biomarker ratios and the respective covariate(s) within each component.

3. mixture of linear regression model like in many mixture modeling procedures enables us to group participants that present identical characteristics and study their risks for transitioning from the normal cognitive state.

4. when we fit the Cox model with race alone it was not a significant predictor of the event. However the posterior probability belonging to component one is found to be a significant predictor of the event when presented in the Cox model. Thus the mixture modeling procedure uncovered a relationship between the event of interest and the covariates that otherwise would have been masked.

5. when we fit the Cox model with the posterior probability of belonging to component one and adjust for race, the posterior probability is found to be significant but race is not. This may suggest that all the information needed to understand the association between the event and the potential of transitioning is captured in the relationship between the event and the posterior probabilities.

6. the concordance statistics that measures the degree of agreement between predicted and actual event indicate that the posterior probability (using race indirectly through mixture modeling) from the mixture modeling procedure is preferable as a measure of predicting future cognition status than using race directly in a Cox modeling procedure.

More than eighty six models are examined with the aid of sBIC as the tool for model selection. The models with largest sBIC and more stable standard errors are considered for further analysis.

The risk of transitioning from normal cognition within the individual components indicate that participants in component two have the least risk with the highest risk going to the participants in component one. This may suggest that candidates who have low rtaubeta value and comparatively higher rptaubeta value or vice versa will be at a medium to high risk of transitioning from normal cognition. However, if the rptaubeta and rtaubeta values are relatively low, then the risk of transitioning is minimal.

Figures 3.7 and 3.15 generalize the risk plot in Figures 3.5 and 3.13 and show that having a high rptaubeta value with low rtaubeta or having a high rtaubeta keeping

rptaubeta low both lead to an increased risk of transitioning from normal cognition. The generalization of Figure 3.9 by Figure 3.11 is slightly different in that rtaubeta tends to be the primary decider of the risk strata.

Furthermore, the risk strata described above can be used as a predictive mechanism by physicians and other health practitioners based on an individual's biomarker measures. Candidates who fall in the blue colored region based on their biomarker ratios will have about 25% or less chance of transitioning given their biomarker ratios and race (or biomarker ratio and Apoe4 or biomarker ratios, race and Apoe4 depending on the model of choice). Subjects in the yellow region will be above 25% risk but below 50% risk of transitioning given their information. Being in the brown region increases ones risk above 50% but below 75% chance of transitioning given his/her information. The riskiest region is denoted by red in which one has at least 75% chance of transitioning given their information.

We speculate that the blue region may be indicative of individuals who will most likely remain cognitively normal regardless of their race or Apoe4. Region yellow or the intermediate region may be also indicative of a subject who will exhibit potential signals of transitioning but may not transition and thus more likely to remain cognitively normal. In the brown region we speculate that subjects are more likely to transition than they are to remain cognitively normal. We also cautiously entertain the possibility of the brown region being indicative of the region within which a well targeted intervention may yield desired results since the associated risk in that region is about the toss of a fair coin (in the neighborhood of 50%). Finally the red zone

may be representative of subjects who will most likely transition with probabilities well above a simple toss of a fair coin. Again we are cautiously optimistic that persons who fall within this region based on their information if given the most targeted interventions could reduce or slow their probability of transitioning out of normal cognition.

We observe that the risk strata defined for participants who are black or white are comparable with respect to their joint biomarker ratios. However, black participants tend to be at a slightly higher risk for the rptaubeta ratio than the white participants (Figures 3.9-3.10). In terms of Apoe4 carriers and non-carriers, the risk strata related to the latter seem to be driven by the rptaubeta ratio. In comparison, the risk strata for carriers of Apoe4 are about equally influenced by the two biomarker ratios (Figures 3.16-3.17). Given that a participant is black the risk of transitioning associated with being non-Apoe4 carrier is primarily influenced by the rptaubeta biomarker whereas the risk of transitioning associated with Apoe4 carriers are about equally influenced by the two biomarkers (Figure 21-22). Given that a participant is white the risk of transitioning for Apoe4 carriers is almost always determined by their rptaubeta biomarker whereas the risk of transitioning for non-Apoe4 carriers although in favor of rptaubeta biomarker is not so at values of rptaubeta lower than 0.4 (Figure 3.23-2.24).

Based on the foregone discussion, we speculate that in principle if we know a person is white and an apoe4 carrier, the risk of transitioning in the future can solely be determined using their baseline rptaubeta biomarker. This may not be so for non-Apoe4 whites /blacks and apoe4 blacks in that these scenarios require knowledge on

both biomarker ratios at baseline. Indeed these group specific diagnoses sounds the bell that in principle the findings here may lend themselves to group specific interventions.

3.6 Limitations and Future Directions

Limitations

In our analysis we assumed that biomarker ratios follow a normal distribution based on the empirical evidence provided by the data via a histogram plot. Theoretically we may be inclined to use Cauchy distribution as the biomarker ratios arise from the ratio of two biomarkers which are assumed to be normally distributed. However in this case the empirical evidence strongly favored normality. In addition to the empirical evidence, the Gaussian distribution has nicer properties such as mean and standard deviation which are both key in studying the properties of the biomarker ratios presented here. Cauchy distribution is very limited in this sense.

We also assumed independence between the two biomarkers in our models. This assumption arose from the fact that if we adjust for a covariate the correlation between the biomarkers will dissipate. Indeed we didn't find any evidence that trumps our assumption so much so that we had to change course. It is worth mentioning that conditioning on the covariates such as race or Apoe4 or both reduced the correlation between the outcomes.

Furthermore, only 10 of the 114 participants were Black which may bias the race model outcome due to lack of power as the proportion of Black and White will be very different. However the standard errors produced by the race model exhibited stable properties which is non-indicative of biasedness.

Future Direction

In future studies we hope to access more events due to longer follow up periods in the ADNI studies. In this study we had 34 events thus far. Longer follow-up could yield more events which will increase the power of the study.

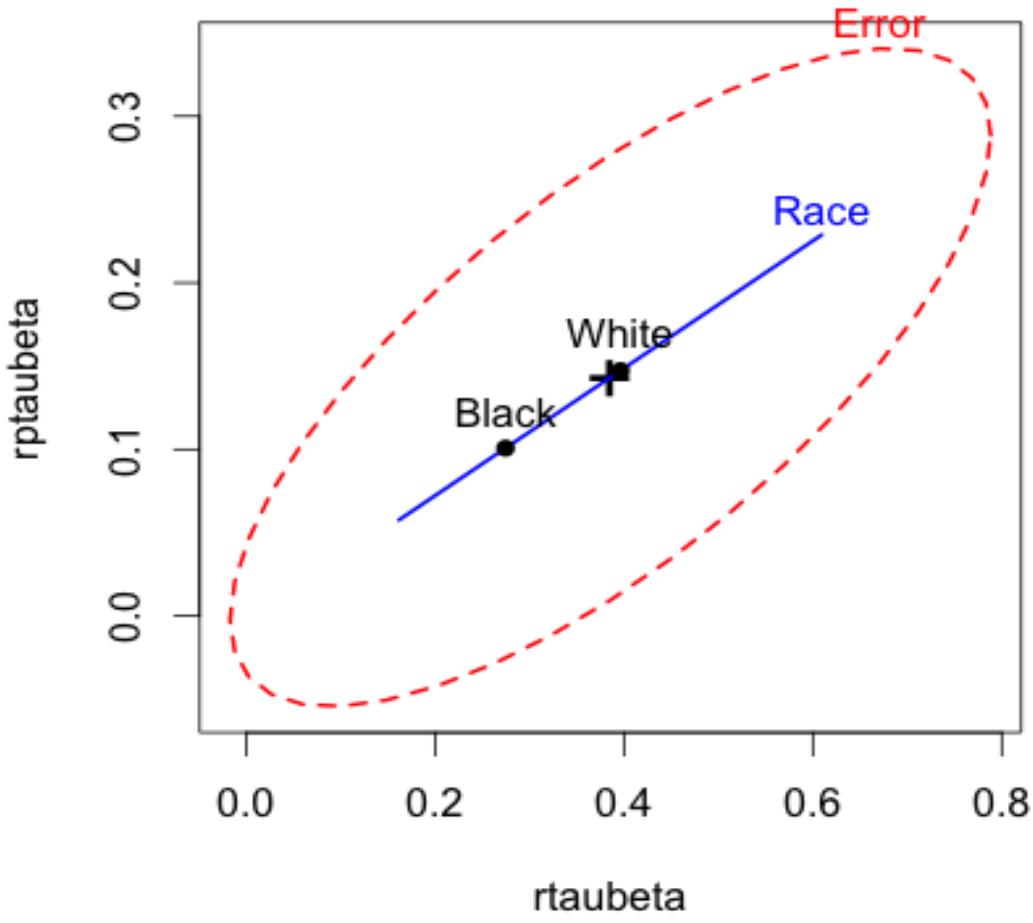

Figure 3.1: Relationship between biomarkers ratios and race. + represents the grand mean

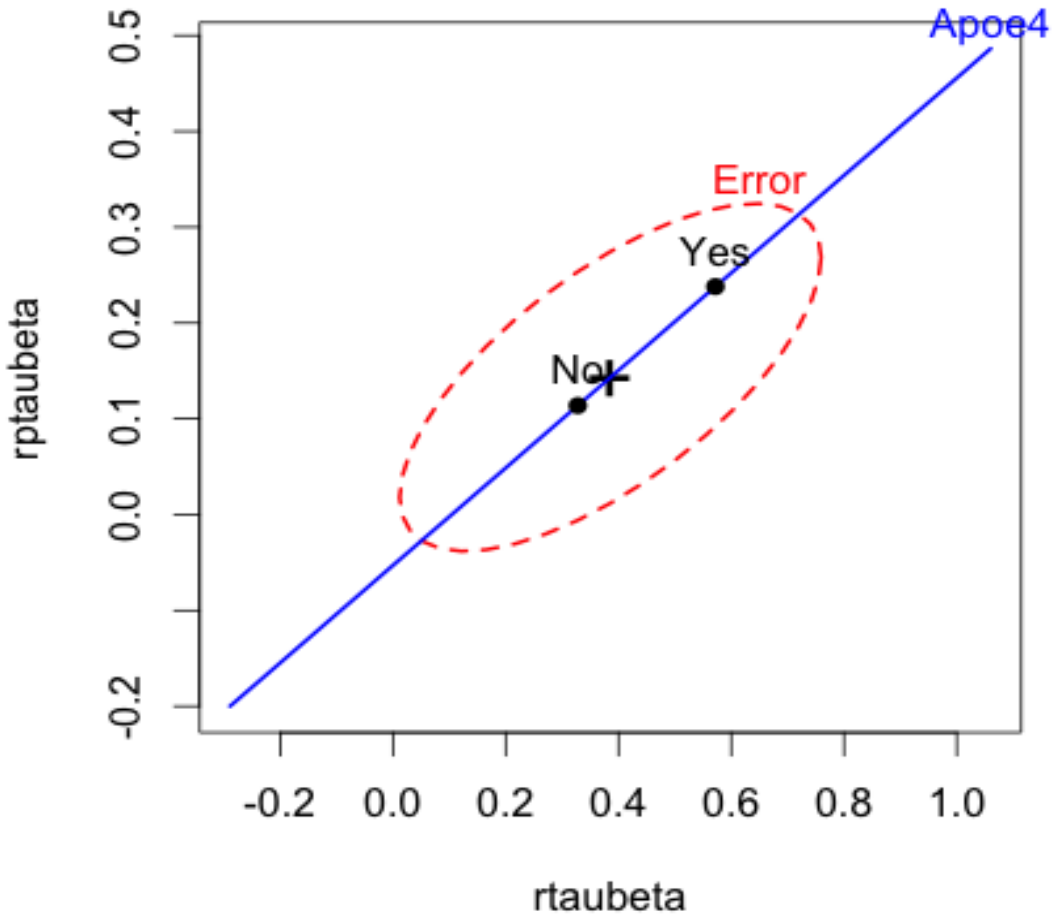

Figure 3.2: Relationship between biomarker ratios and Apoe4 + represents the grand mean

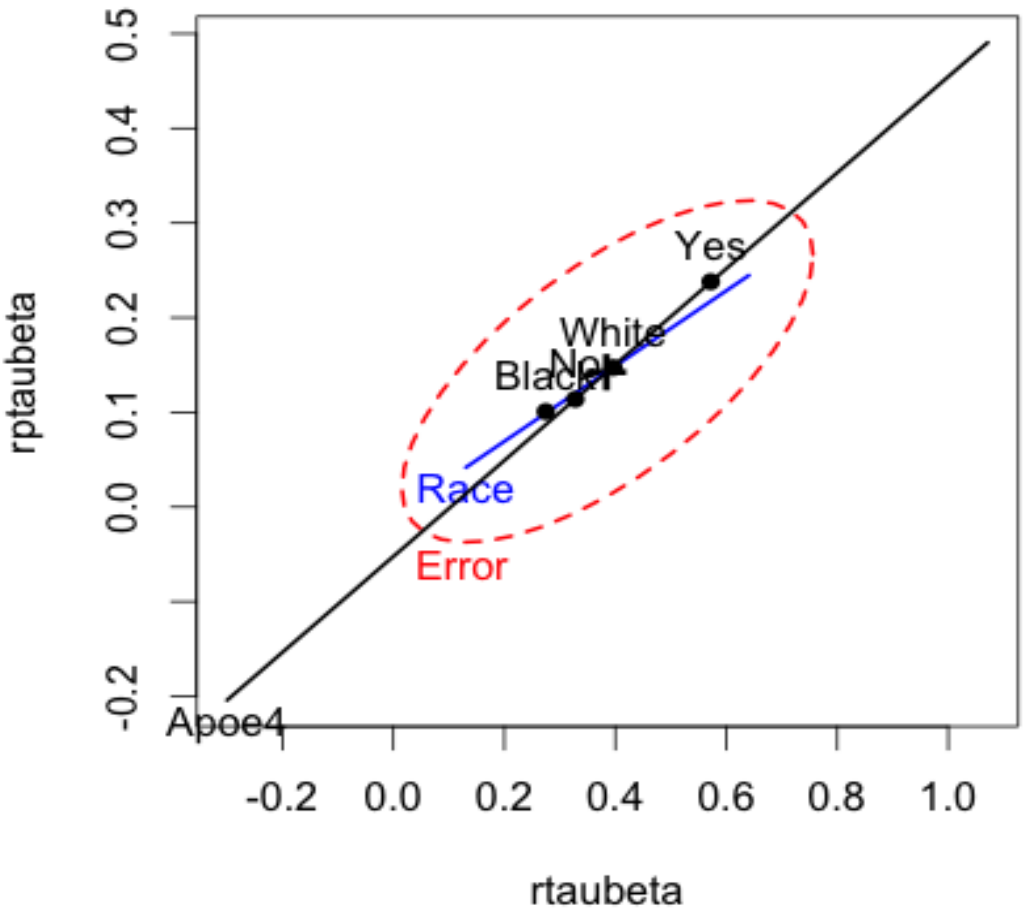

Figure 3.3: Relationship between biomarkers, race and Apoe4. + is grand mean.

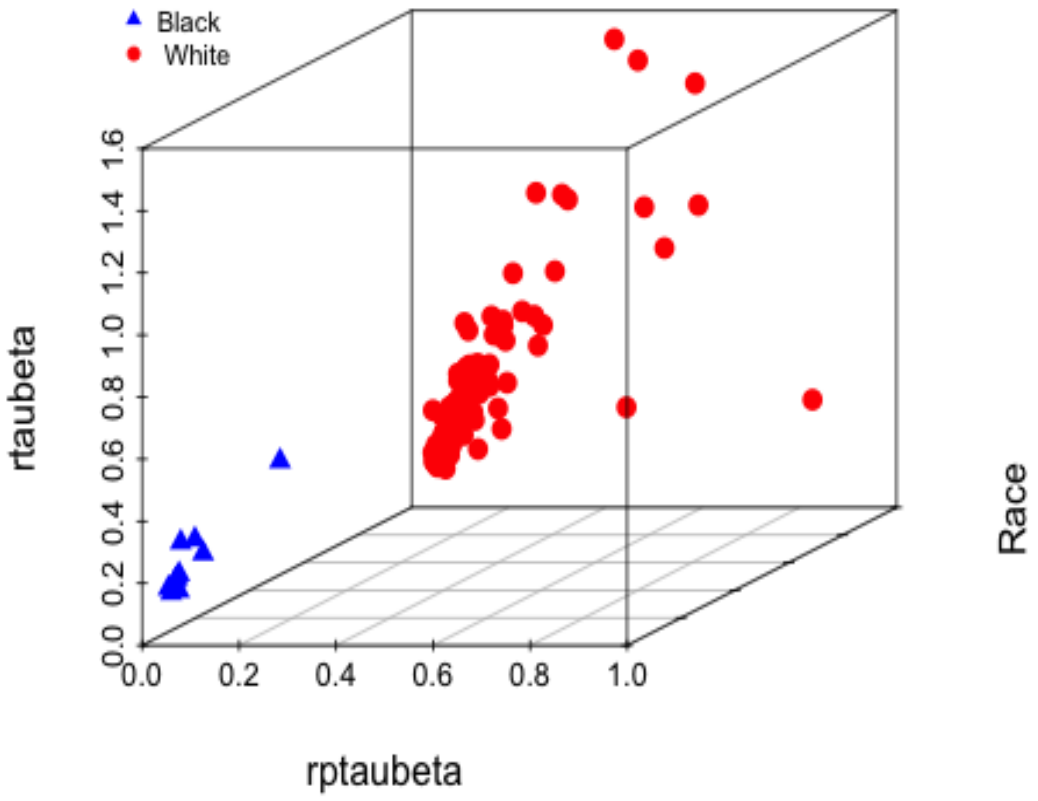

Figure 3.4: Joint Distribution of rtaubeta and rptaubeta Given Race

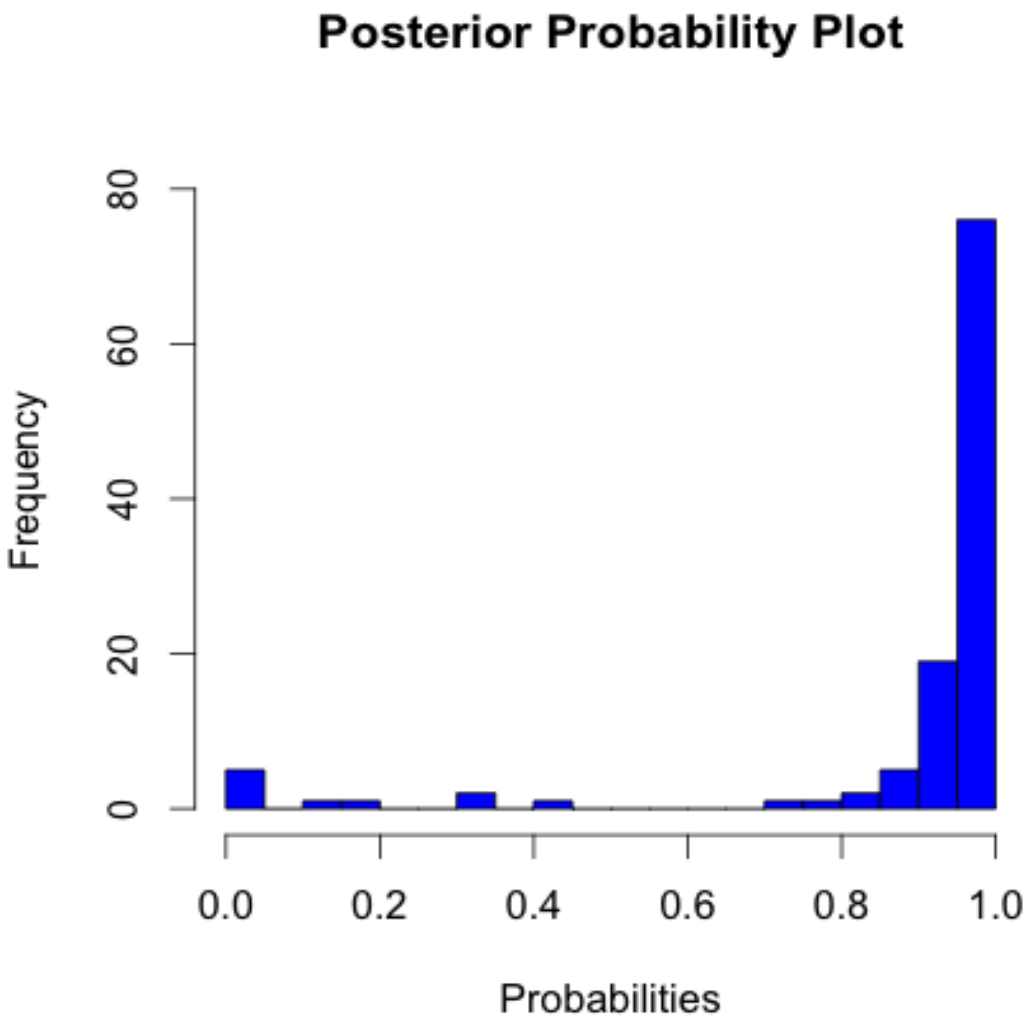

Figure 3.5: Posterior probability plot with race as predictor in the mixture regression model indicates that the model is well separated

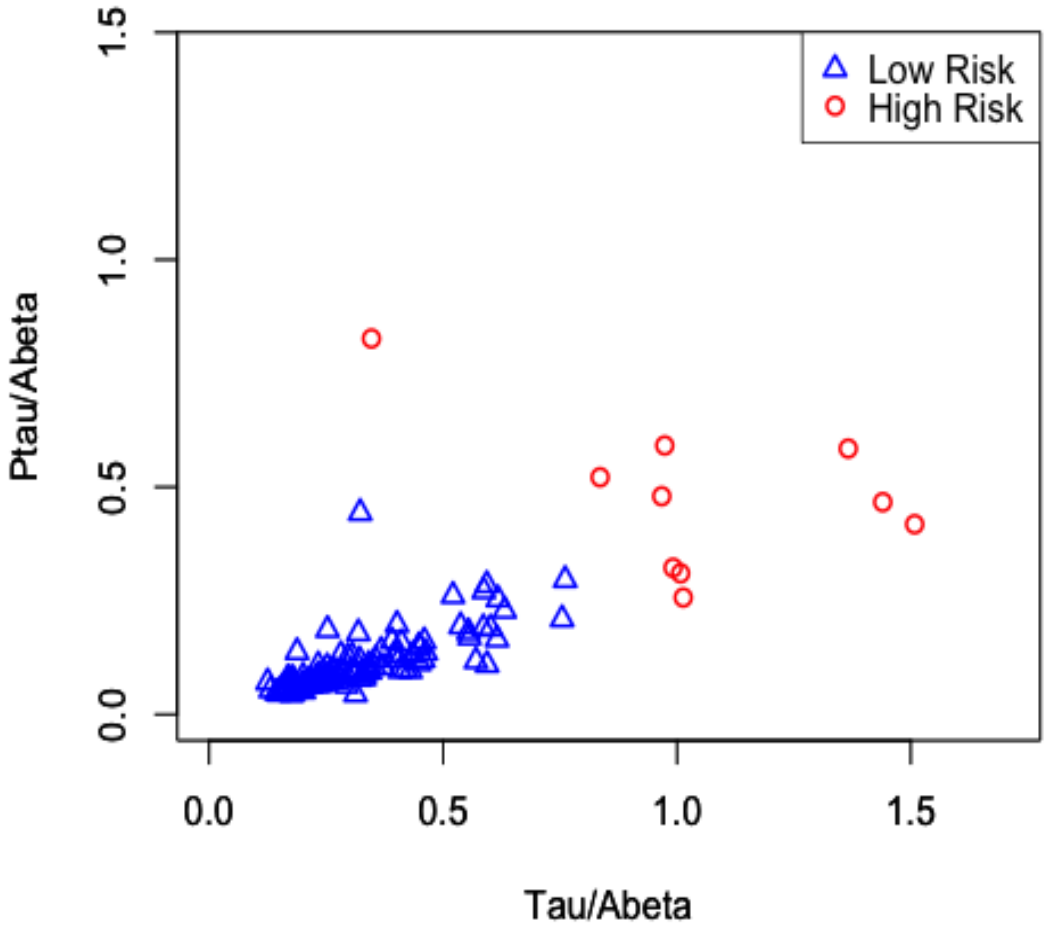

Figure 3.6: Individual biomarker ratios grouped into different risk regions. Here race is the predictor variable

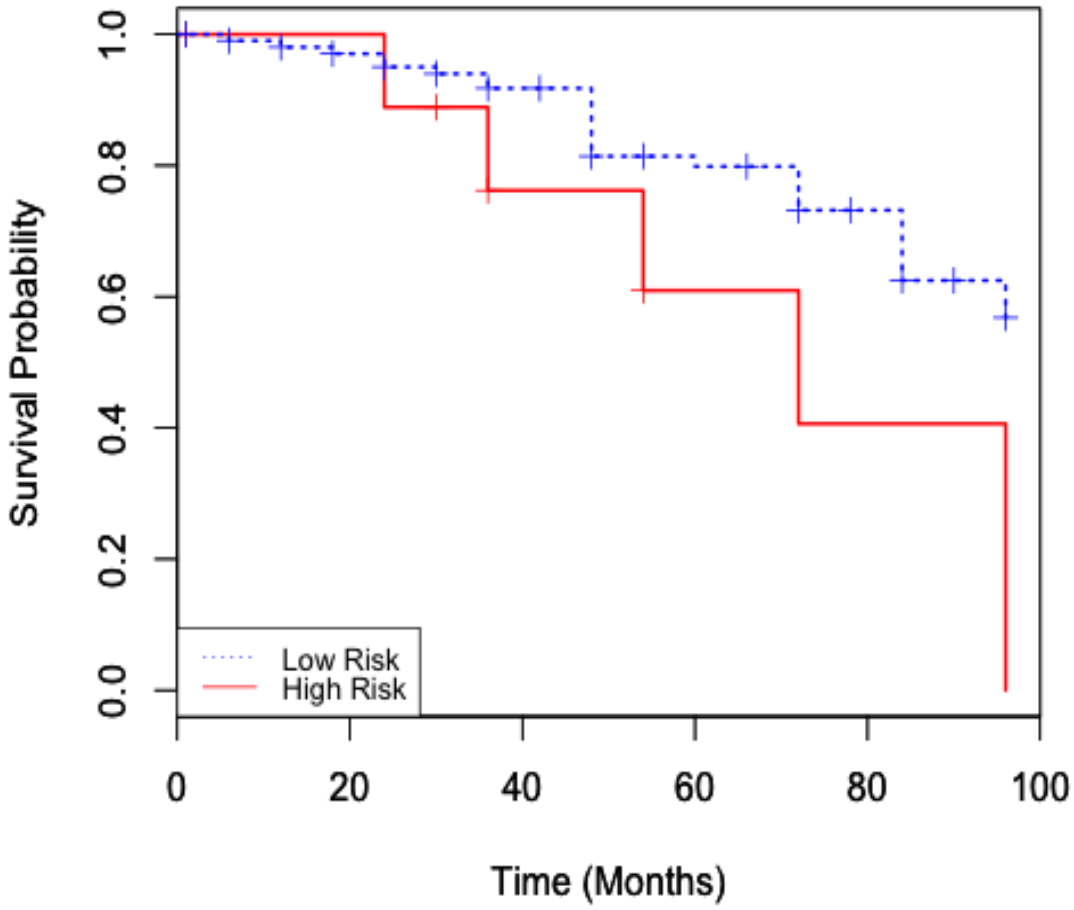

Figure 3.7: The survivability of the two groups over time given in weeks. Here race is the predictor variable

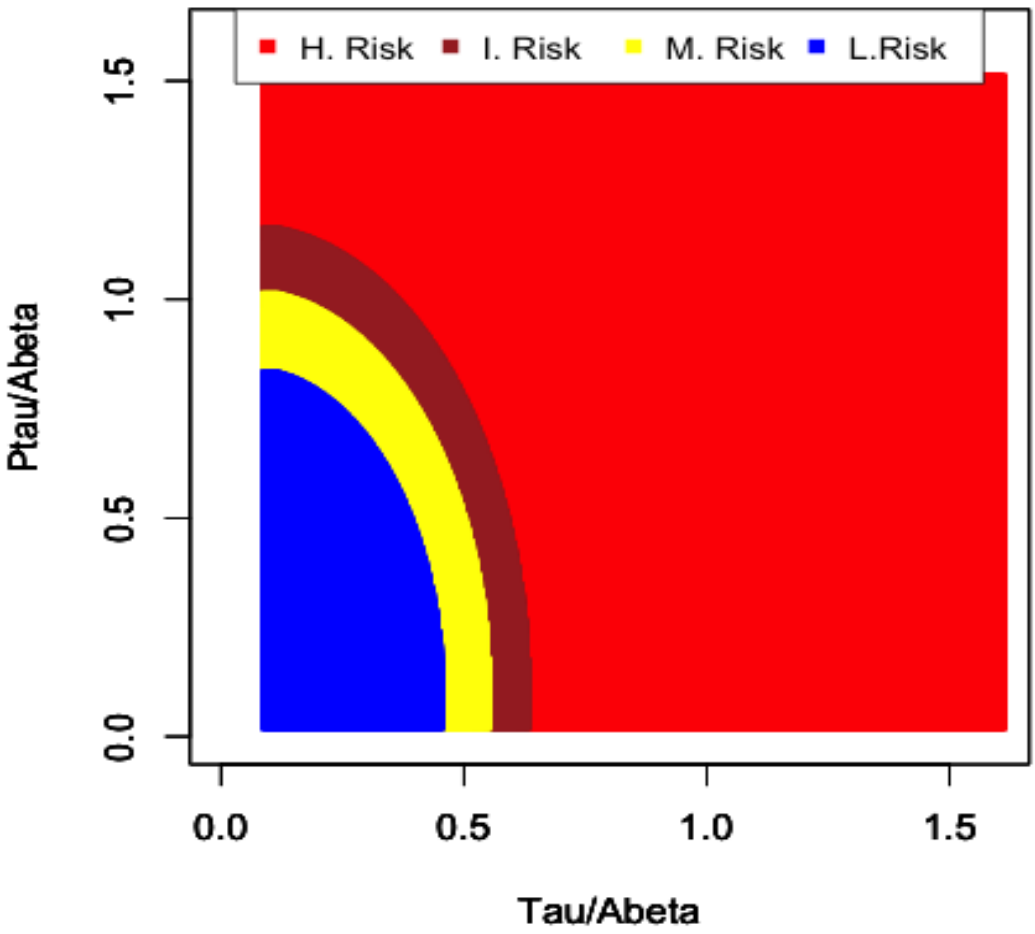

Figure 3.8: Risk strata for Blacks from the posterior probabilities obtained from the mixture of linear regression with race as covariate

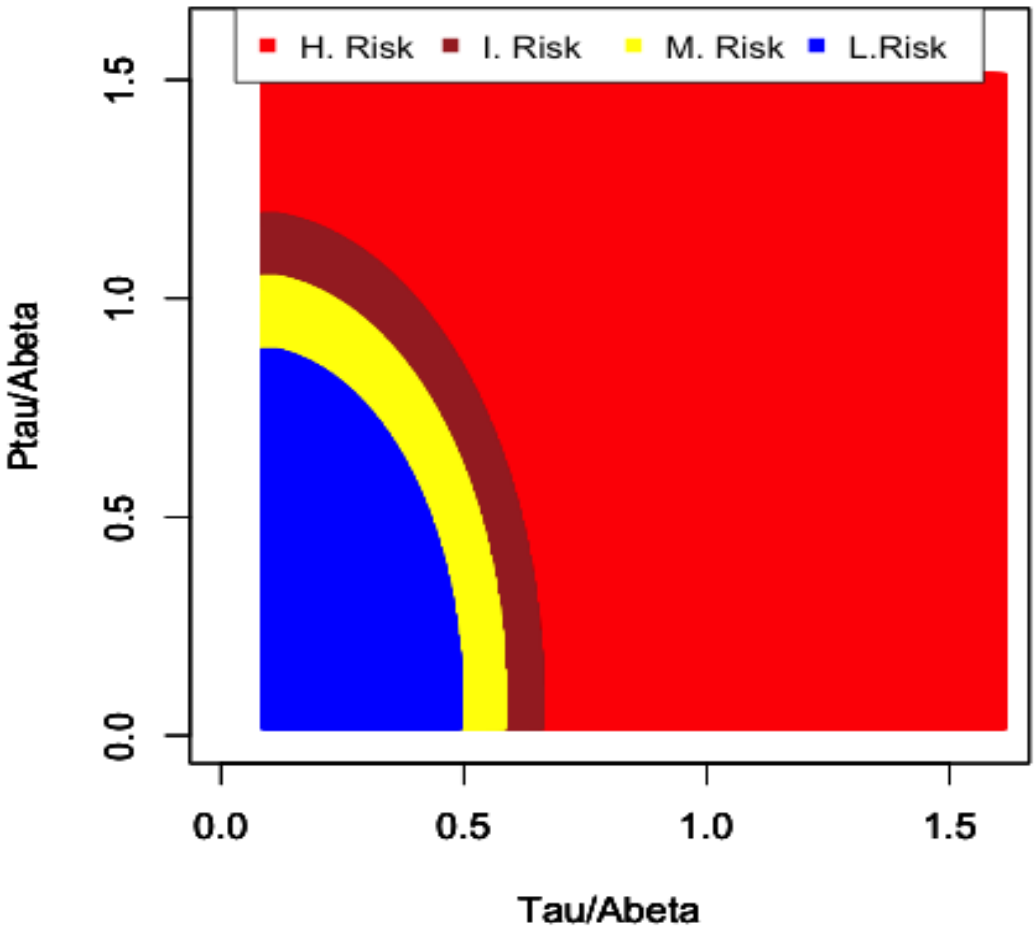

Figure 3.9: Risk strata for Whites from the posterior probabilities obtained from the mixture of linear regression with race as covariate

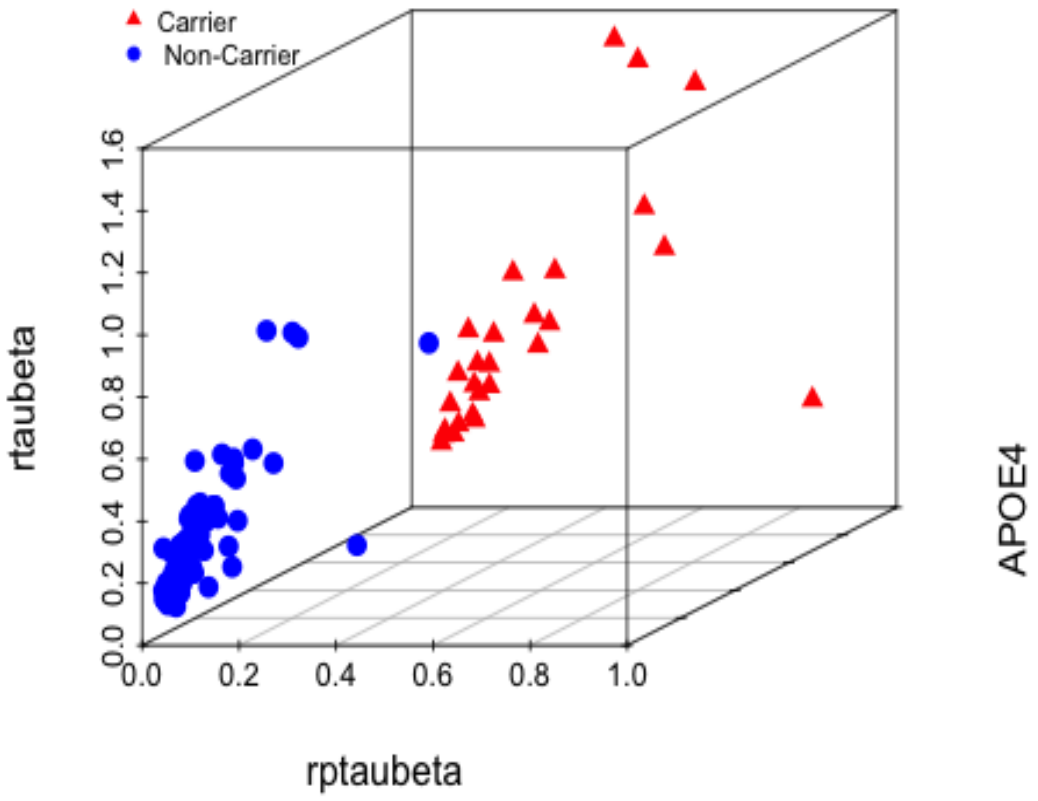

Figure 3.10: Joint Distribution of rtaubeta and rptaubeta Given Apoe4

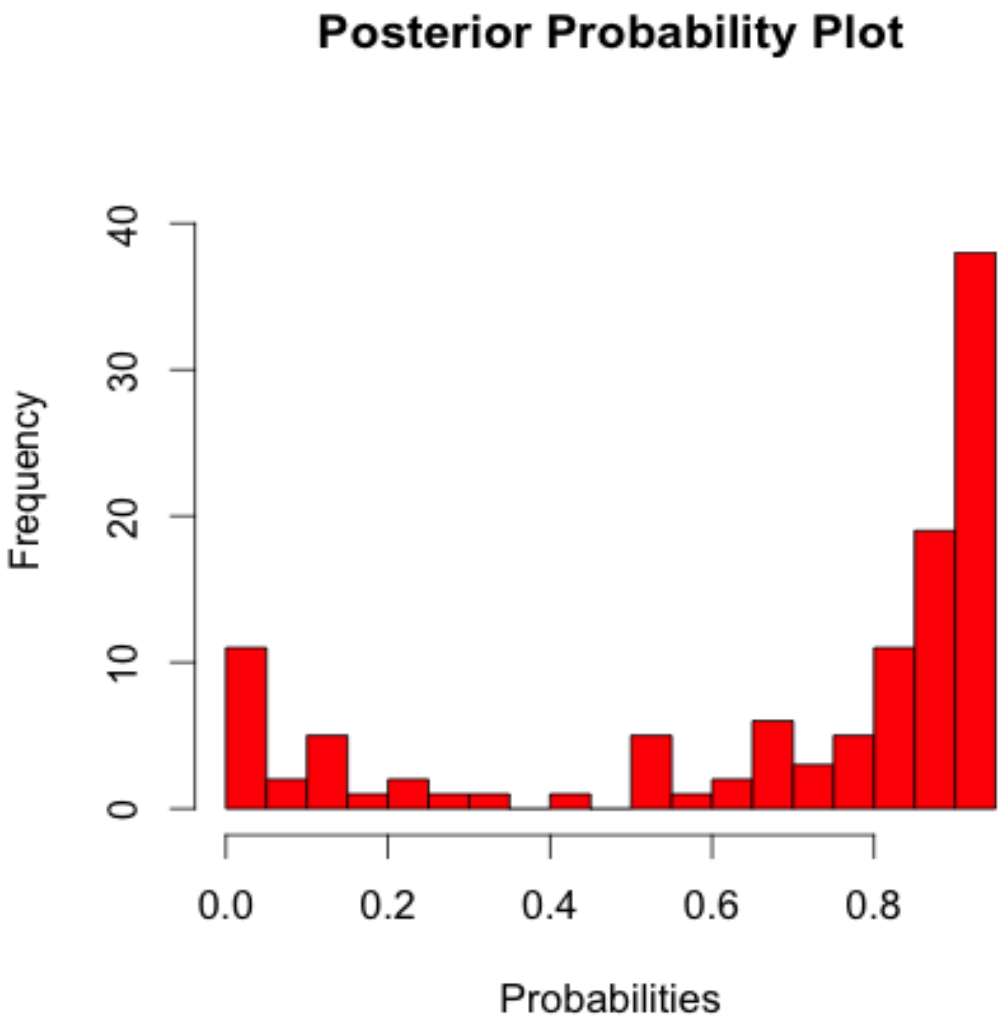

Figure 3.11: Posterior probability plot indicates that the mixture of regression model with Apoe4 as predictor model is comparatively less well separated

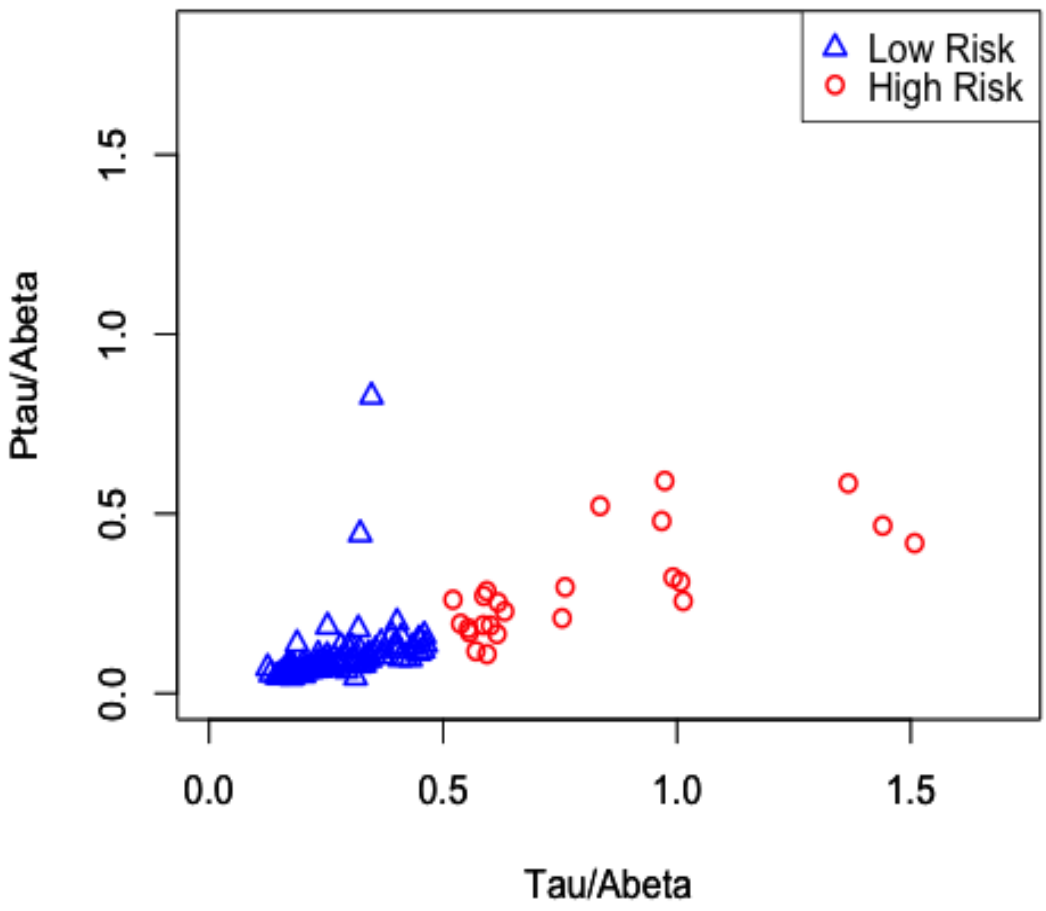

Figure 3.12: Individual biomarker ratios grouped into different risk regions. Here Apoe4 is the predictor variable.

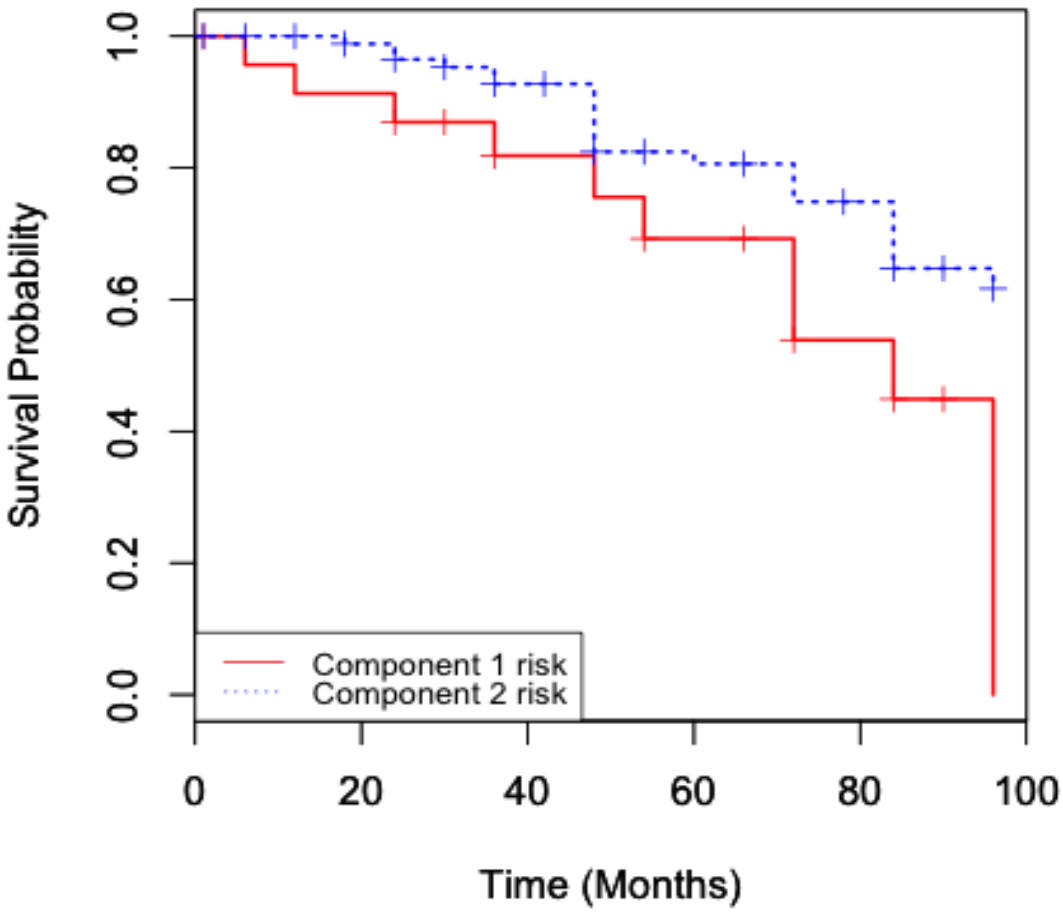

Figure 3.13: The survivability of the two groups over time given in weeks. Here Apoe4 is the predictor variable.

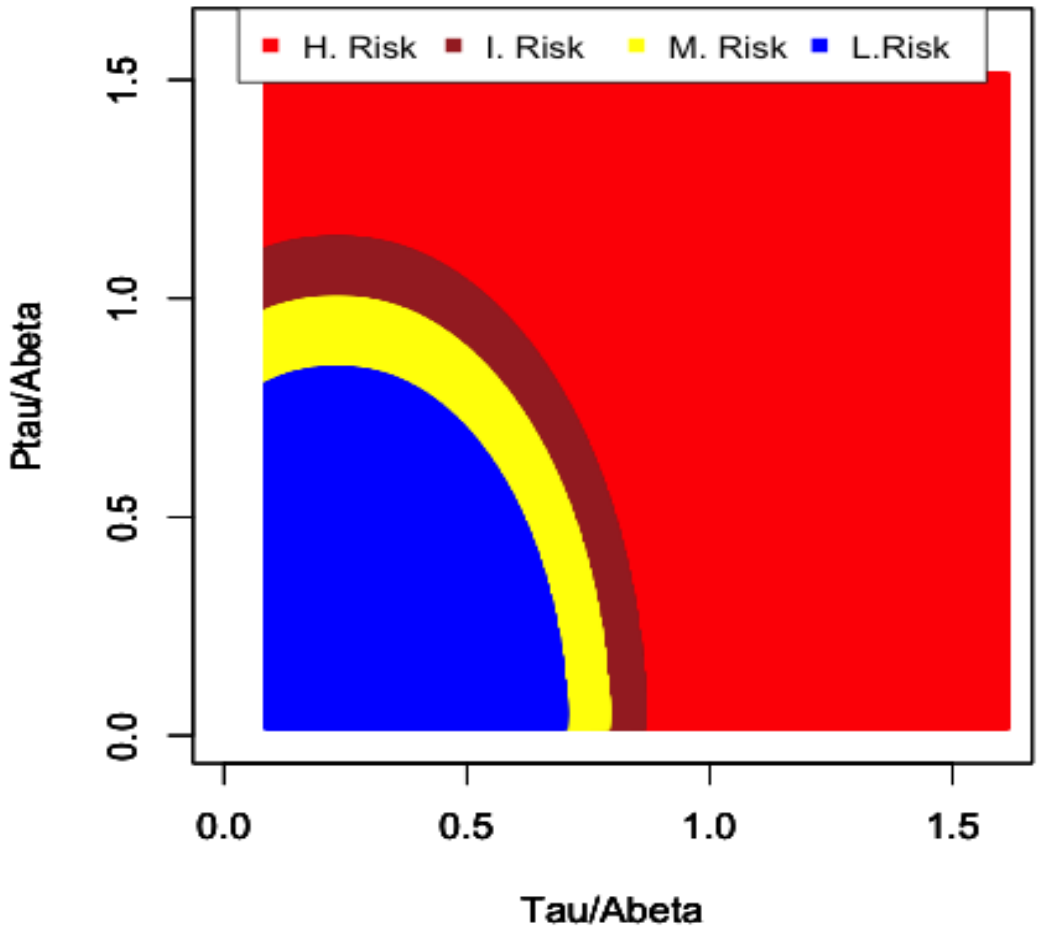

Figure 3.14: This corresponds to the risk strata for Apoe4 carriers derived from the posterior probabilities obtained from the mixture of linear regression with Apoe4 as predictor variable.

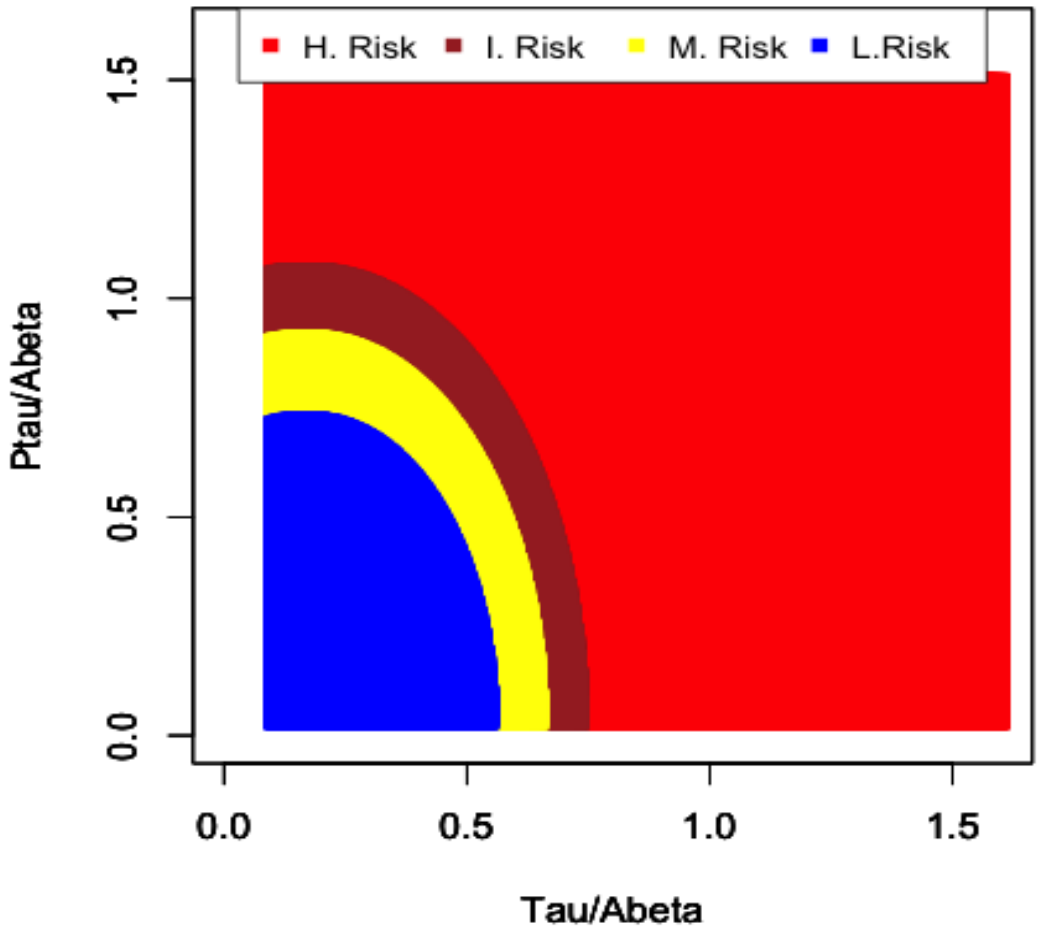

Figure 3.15: This corresponds to the risk strata for none-Apoe4 carriers derived from the posterior probabilities obtained from the mixture of linear regression with Apoe4 as predictor variable.

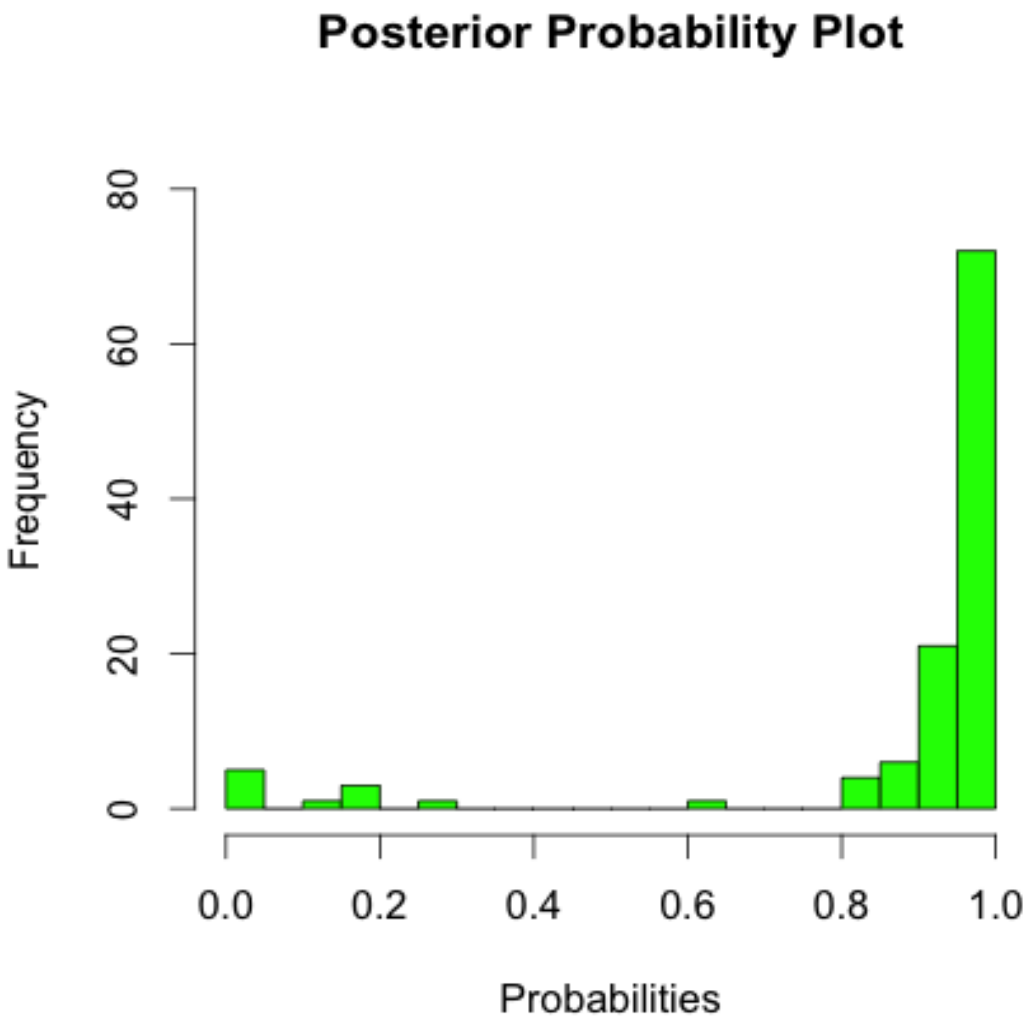

Figure 3.16: Posterior probability plot indicates that the model with Apoe4 and race as predictors is comparatively less well separated

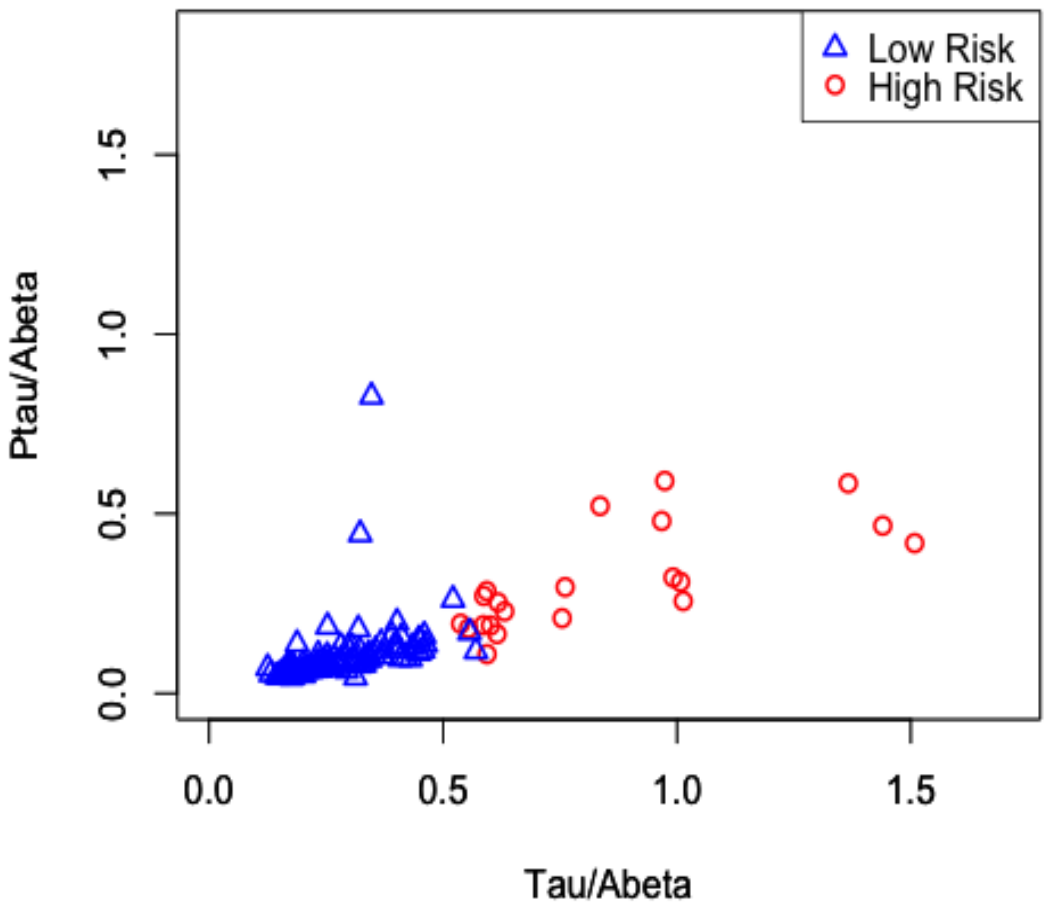

Figure 3.17: Individual biomarker ratios grouped into different risk regions. Here Apoe4 and race are the predictor variables.

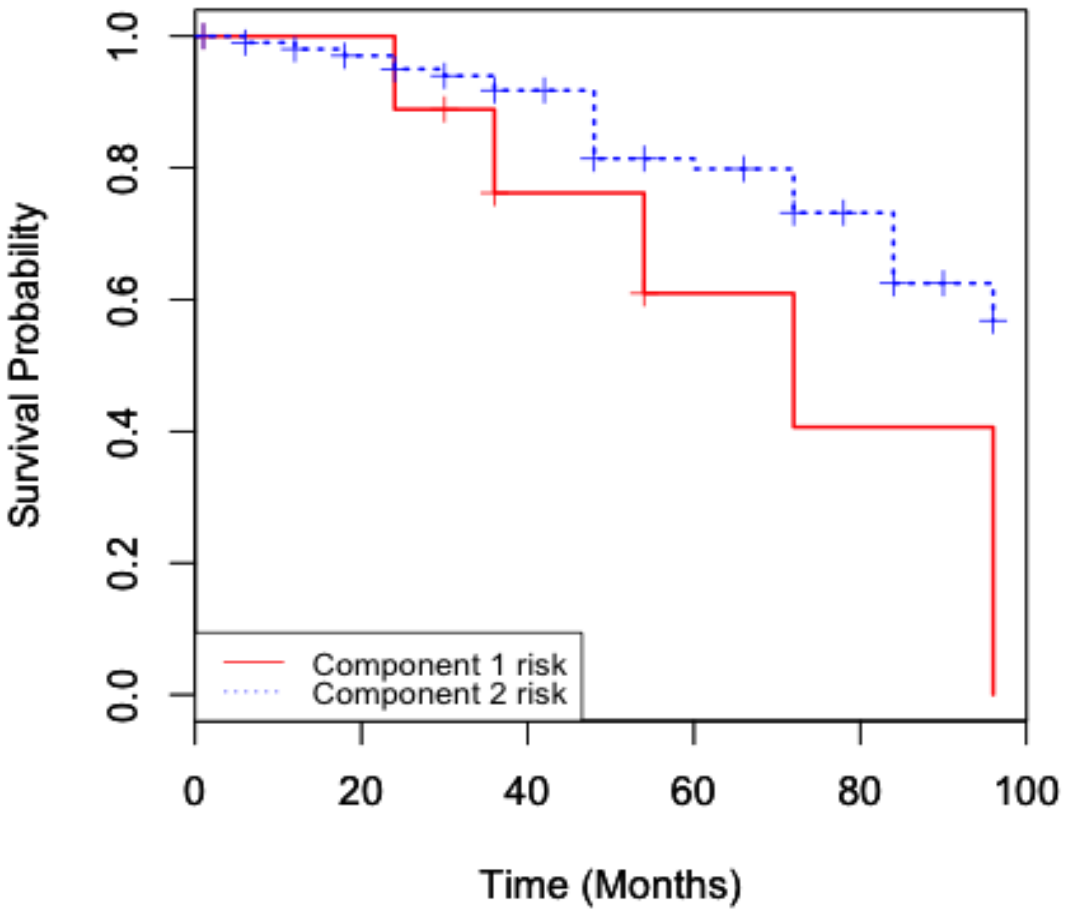

Figure 3.18: The survivability of the two groups over time given in weeks. Here Apoe4 and race are the predictor variables.

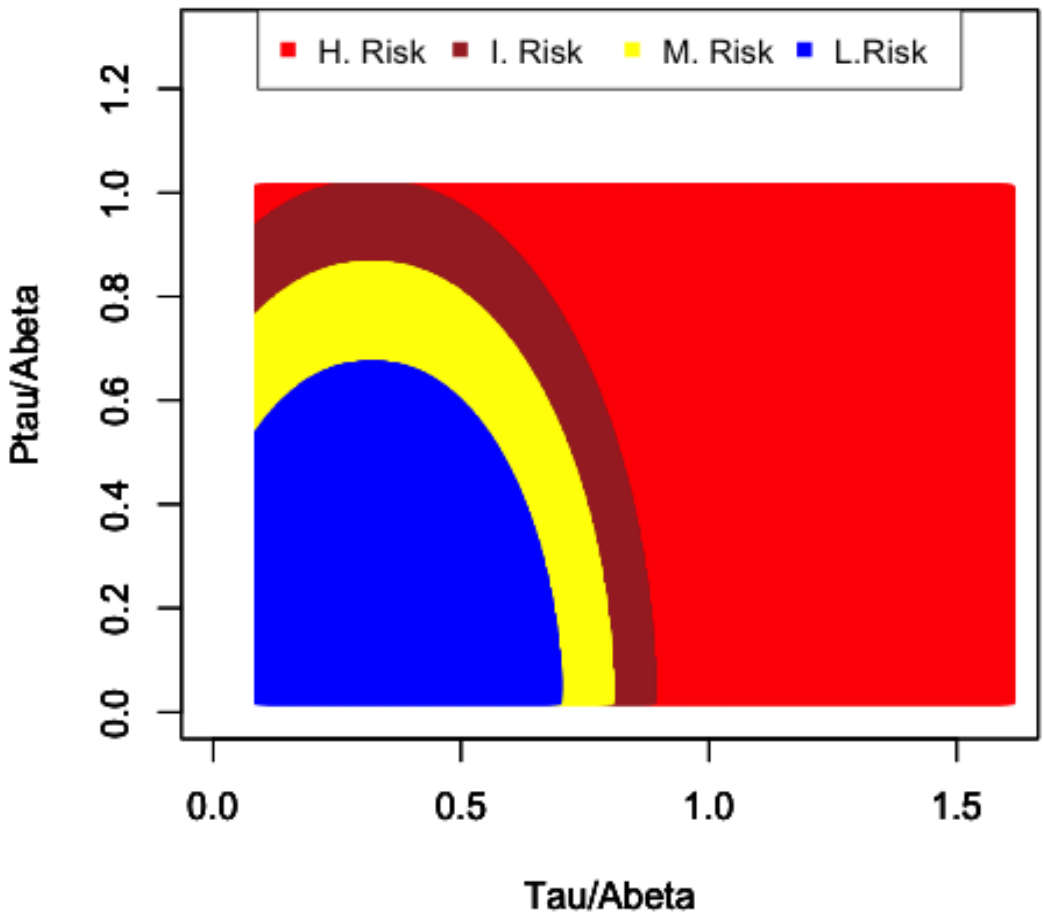

Figure 3.19: Risk strata for black who are apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates

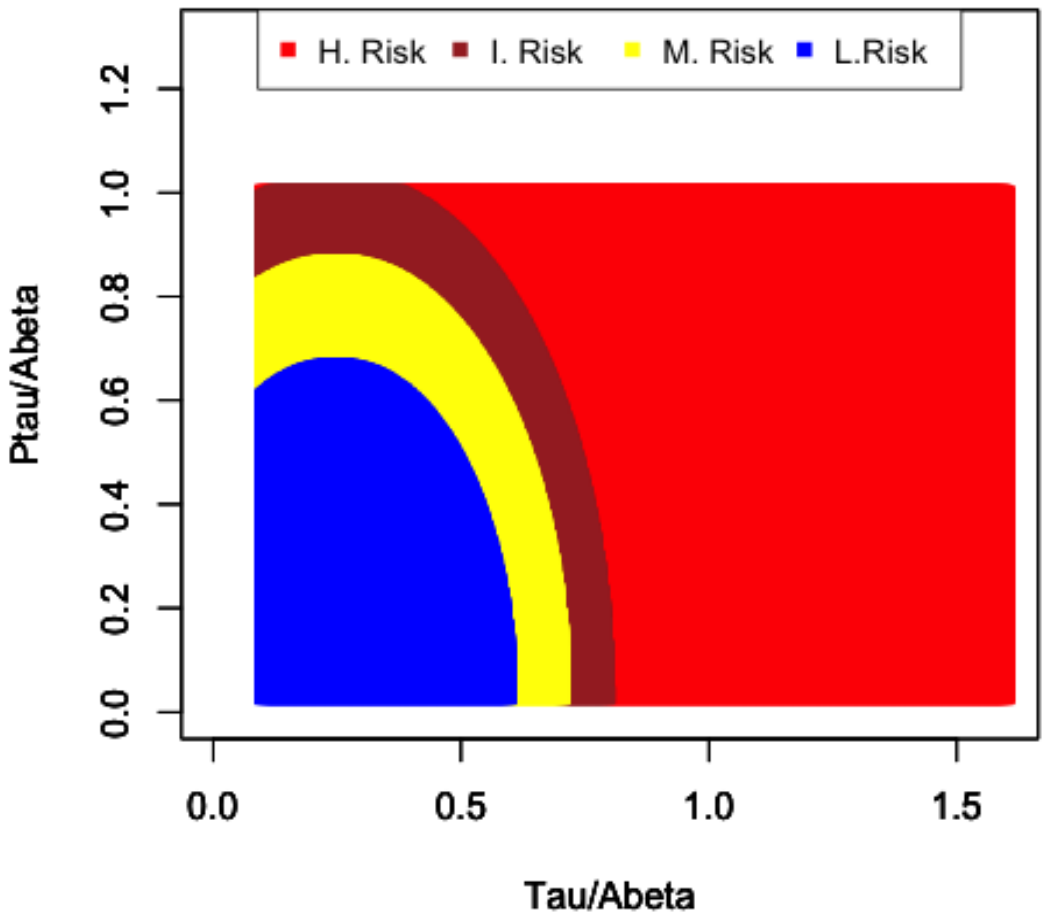

Figure 3.20: Risk strata for blacks who are none-apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates

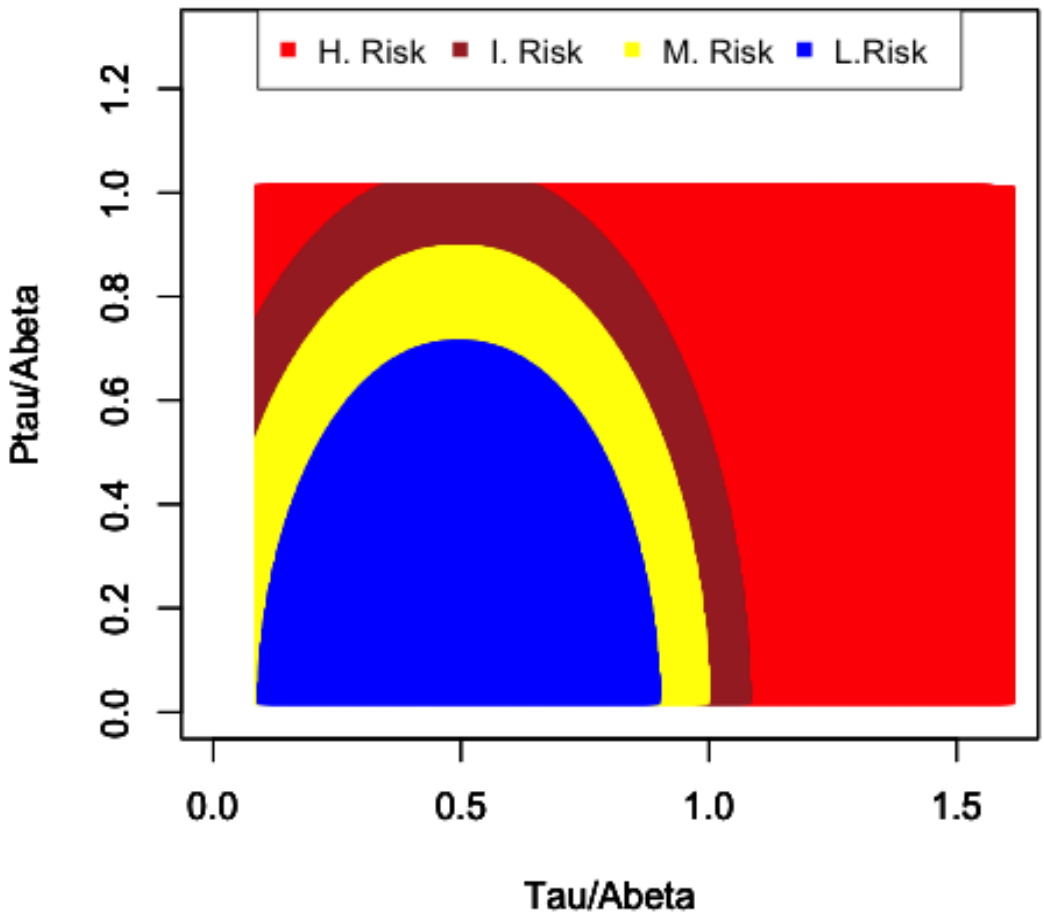

Figure 3.21: Risk strata for whites who are apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates

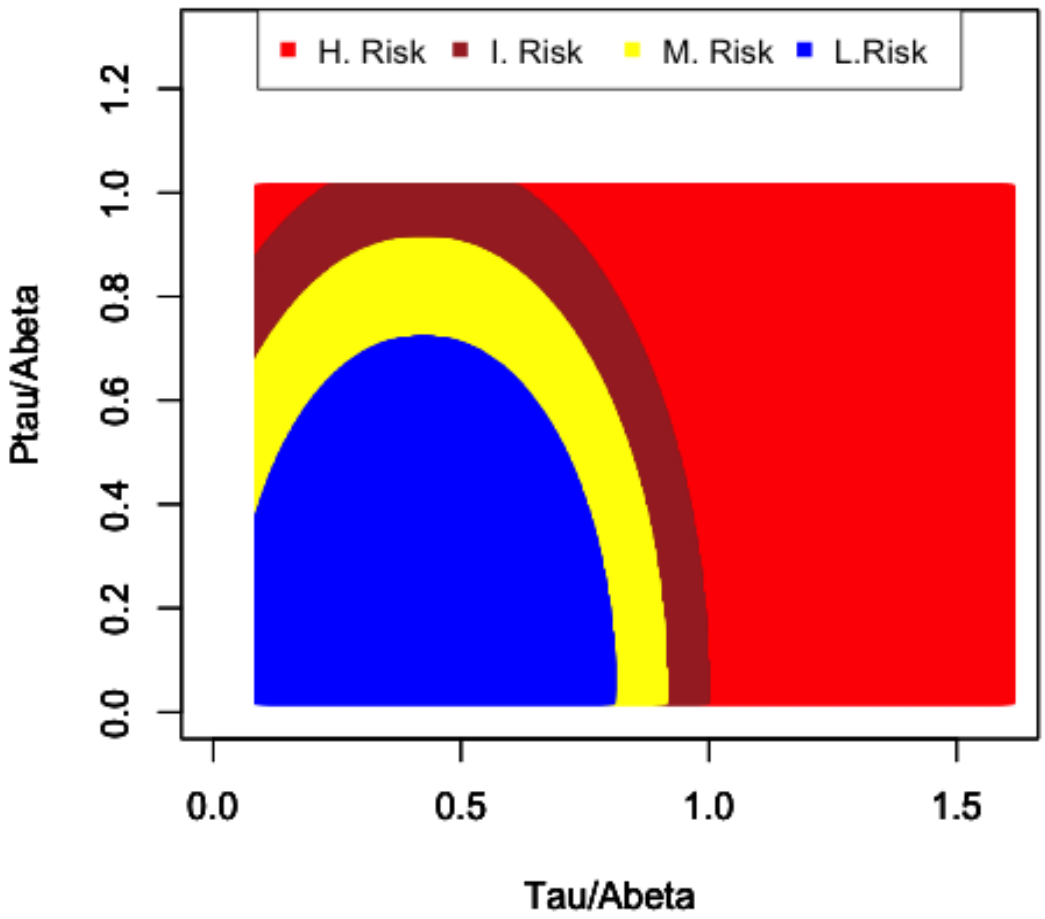

Figure 3.22: Risk strata for whites who are none-apoe4 carriers obtained from the mixture of linear regression with Race and Apoe4 as covariates

Table 3.1: Selection of model complexity with three criteria. To calculate the learning coefficient sBIC, we are assuming the non-redundant one component. Numbers shown are differences between the information criteria at one component versus the information criteria at two component.

Complexity/Criteria	AIC	BIC	sBIC
1-2	535.35	502.52	530.94

Table 3.2: Estimates of the regression models within each component.Race is the only predictor variable in the model. Race is an indicator variable for Caucasian (coded as 1) and the referent group is black (coded as zero) *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level

Component 1 for rtaubeta	Estimate	SE	P-value
Intercept	0.563	0.315	0.074
Race	0.163	0.322	0.613
Component 2 for rtaubeta			
Intercept	0.238	0.031	$< 0.001^{***}$
Race	0.036	0.034	0.282
Component 1 for rptaubeta	Estimate	SE	P-value
Intercept	0.266	0.169	0.115
Race	0.039	0.172	0.819
Component 2 for rptaubeta			
Intercept	0.080	0.011	$< 0.001^{***}$
Race	0.009	0.012	0.443

Table 3.3: High|Low risk is component one estimated probability for the soft and hard classification models. c is the concordance.HR is the hazard ratio. Race is the only predictor. When model was adjusted for education, MMSE, and age only prop.SBICO1 and rprop.SBICO1 were significant. This significance disappeared when Apoe4 was adjusted for (results not shown)

Soft classification	Estimated HR	P-value	CI	c(SE)
High|Low risk	4.621	0.001	(1.661, 12.860)	0.680(0.058)
With Adjustment				
High|Low risk	3.98	0.001	(1.693, 9.369)	0.776(0.058)
RAVLT	0.95	0.007	(0.912, 0.986)	
Hard classification				
High|Low risk	3.026	0.014	(1.24, 7.34)	0.547(0.026)
With Adjustment				
High|Low risk	3.016	0.003	(1.460, 6.228)	0.745(0.057)
Race	0.943	0.003	(0.908, 0.980)	

Table 3.4: Estimates of the regression models within each component. Apoe4 is the only predictor variable in the model.Apoe4 is coded as 1 for carriers of the gene and 0 for non carriers. *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level

Component 1 for rtaubeta	Estimate	SE	P-value
Intercept	0.562	0.065	$< 0.001^{***}$
Apoe4	0.307	0.108	0.005^{**}
Component 2 for rtaubeta			
Intercept	0.247	0.012	$< 0.001^{***}$
Apoe4	0.106	0.026	$< 0.001^{***}$
Component 1 for rptaubeta	Estimate	SE	P-value
Intercept	0.217	0.032	$< 0.001^{***}$
Apoe4	0.192	0.054	$< 0.001^{***}$
Component 2 for rptaubeta			
Intercept	0.078	0.004	$< 0.001^{***}$
Apoe4	0.034	0.008	$< 0.001^{***}$

Table 3.5: High|Low risk is component one estimated probability for the soft and hard classification models. c is the concordance. HR is the hazard ratio. Apoe4 is the only predictor variable. When model was adjusted for education, MMSE, race and age only prop.SBICO12 and race or rprop.SBICO12 were significant (results not shown)

Soft classification	Estimated HR	P-value	CI	c(SE)
High|Low risk	4.590	0.001^{**}	(1.844, 11.420)	0.662(0.058)
With Adjustment				
High|Low risk	4.758	0.005	(1.976, 11.58)	0.764(0.058)
RAVLT	0.942	0.002	(0.907, 0.979	
Hard classification				
High|Low risk	2.760	0.001	(1.355, 5.623)	0.586(0.039)
With Adjustment				
High|Low risk	2.108	0.110	(0.844, 5.263)	0.69(0.057)
RAVLT	0.951	0.013	(0.914, 0.990)	

Table 3.6: Estimates of the mixture of regression models within each component. *** significant at 0.001 level, ** significant at 0.01 level and * significant at 0.05 level

Component 1 for rtaubeta	Estimate	SE	P-value
Intercept	0.255	0.264	0.333
Race	0.312	0.260	0.230
Apoe4	0.335	0.109	$< 0.01^{**}$
Component 2 for rtaubeta			
Intercept	0.213	0.031	$< 0.001^{***}$
Race	0.039	0.032	0.221
Apoe4	0.108	0.025	$< 0.001^{***}$
Component 1 for rptaubeta	Estimate	SE	P-value
Intercept	0.079	0.127	0.537
Race	0.141	0.125	0.263
Apoe4	0.206	0.053	$< 0.001^{***}$
Component 2 for rptaubeta			
Intercept	0.072	0.009	$< 0.001^{***}$
Race	0.008	0.010	0.432
Apoe4	0.035	0.008	$< 0.001^{***}$

Table 3.7: prop.SBIC013 is component one estimated probability and rprop.SBIC013 is component one hard classification. c is the concordance. HR is the hazard ratio. Race and Apoe4 are the predictors. When model was adjusted for education, MMSE, race and age only Apoe4 was significant (results not shown)

Soft classification	Estimated Hazard	P-value	CI	c(SE)
prop.SBIC013	4.099	0.006	(1.490, 11.270)	0.599(0.058)
With Adjustment				
prop.SBIC013	2.863	0.058	(0.967, 8.480)	0.670(0.058)
Race	0.380	0.052	(0.143, 1.009)	
Apoe4	2.312	0.036	(1.056, 5.066)	
Hard classification				
rprop.SBIC013	3.026	0.014	(1.247, 7.342)	0.547(0.026)
With Adjustment				
rprop.SBIC013	2.345	0.083	(0.895, 6.145)	0.634(0.046)
Race	0.373	0.048	(0.140, 0.993)	
Apoe4	2.366	0.029	(1.093, 5.119)	

3.7 Illustrative Computations for A1-A3 in Drton and Plummer(2016) for Mixture of Regression Models

Adapting some of the notations from Dacunha-Castelle and Gassiat(1999)[55] we define a family of mixture of regression models as:

$$G_k = \left\{ g = g(\mathbf{y}_i|\mathbf{x}_i, \mathbf{\Theta}_j) = \sum_{j=1}^{k} \pi_j f_j(\mathbf{y}_i|\mathbf{x}_i, \mathbf{\Theta}_j) \right\} \tag{3.7}$$

where $\mathbf{x}_i$ is the vector of covariates for subject i, $\mathbf{\Theta}_j$ is the component specific parameters (i.e. $\mathbf{\Theta}_j = \left(\boldsymbol{\beta_j}, \boldsymbol{\Sigma_j}\right)$) and π_j are the mixing proportions or weights such that $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^{k} \pi_j = 1$. Furthermore we assume that the $\boldsymbol{\Sigma_j}$ are positive definite and their eigenvalues are bounded away from zero. That is $\exists\ \epsilon > 0$ such that $\min\{ev\left(\Sigma_j\right)\} \geq \epsilon > 0$.

Assume that

$$\boldsymbol{X} \sim N(\mu, \tau^2)$$

and

$$\boldsymbol{Y|X=x}, \beta, \sigma^2 \sim N(\beta_0 + \beta_1 x, \sigma^2).$$

Then $var(Y) = var(E(Y|X)) + E(var(Y|X)) = \beta_1^2\tau^2 + \sigma^2$. The covariance is also given as $cov(X,Y) = corr(X,Y) * \sqrt{var(X)var(Y)} = R\tau\sqrt{\beta_1^2\tau^2 + \sigma^2}$ noting from the standard regression slope formula that $\beta_1 = R\frac{s_y}{s_x} = R\frac{\sqrt{\beta_1^2\tau^2+\sigma^2}}{\tau} \Rightarrow R = \frac{\hat{\beta}_1\tau}{\sqrt{\beta_1^2\tau^2+\sigma^2}}$. Here forward we will adapt the notation $r = R$. The joint distribution is given as:

$$\boldsymbol{X,Y}|\tau^2, \sigma^2, \beta_0, \beta_1, \mu \sim N\left(\begin{pmatrix} \mu \\ \beta_0 + \beta_1\mu \end{pmatrix}, \begin{pmatrix} \tau^2 & r\tau\sqrt{\beta_1^2\tau^2 + \sigma^2} \\ r\tau\sqrt{\beta_1^2\tau^2 + \sigma^2} & \sigma^2 + \tau^2\beta_1^2 \end{pmatrix}\right). \tag{3.8}$$

$$\boldsymbol{X},\boldsymbol{Y}|\tau^2,\sigma^2,\beta_0,\beta_1,\mu \sim N\left(\begin{pmatrix} \mu \\ \beta_0+\beta_1\mu \end{pmatrix}, \begin{pmatrix} \tau^2 & \beta_1\tau^2 \\ \beta_1\tau^2 & \sigma^2+\tau^2\beta_1^2 \end{pmatrix}\right). \tag{3.9}$$

The rationale for the above assumptions and the subsequent derivation of the joint density of X and Y is to be able to invoke assumptions P0 and P1 in Dacunha-Castelle and Gassiat (1999) which assumed a parametric family of marginal densities. Our original density is conditional on X and so to arrive at the joint distribution we notice by elementary conditional probability that $f(y|x) = \frac{f(x,y)}{f(x)}$.

Note that

$$G_k = \left\{ g = \sum_{j=1}^{k} \pi_j \frac{f(x_i, y_i|\theta_j)}{f(x_i|K)} \right\},$$

where $\theta_j = (\beta_j, \tau^2, \sigma^2)$ and $K = (\mu, \tau^2)$ The corresponding log likelihood of the family of conditional densities defined above is:

$$ln(g) = \sum_{i=1}^{n} \ln \sum_{j=1}^{k} \pi_j f(y_i|x_i,\theta_j) = \sum_{i=1}^{n} \ln \sum_{j=1}^{k} \pi_j \frac{f(x_i,y_i|\theta_j)}{f(x_i|K)} = \sum_{i=1}^{n} \ln \frac{1}{f(x_i|K)} \sum_{j=1}^{k} \pi_j f(x_i,y_i|\theta_j)$$

Assuming a q mixture component under the null hypothesis we have that

$$f_0 = \sum_{j=1}^{q} \pi_{j_0} \frac{f(x_i,y_i|\theta_{j_0},K_0)}{f(x_i|K_0)},$$

for some parameters K_0, θ_{j_0} and π_{j_0} respectively and θ_{j_0} is the true value of θ_j. Define a statistic based on the log likelihood as: $T_n(k) = \sup_{g \in G_k} ln(g) - ln(f_0)$, then the LRT statistic for testing a q mixture component versus a k mixture component can be defined based on T_n as follows: $V_n = T_n(k) - T_n(q)$. In essence V_n is:

$$V_n = \sup_{g\in G_k} \sum_{i=1}^{n} \ln \frac{1}{f(x_i|K)} \sum_{j=1}^{k} \pi_j f(x_i, y_i|\theta_j) - \sup_{g\in G_q} \sum_{i=1}^{n} \ln \frac{1}{f(x_i|K)} \sum_{j=1}^{q} \pi_j f(x_i, y_i|\theta_j)$$

$$= \sum_{i=1}^{n} \left(\ln \frac{1}{f(x_i|\hat{K}_0)} \sum_{j=1}^{k} \hat{\pi}_j f(x_i, y_i|\hat{\theta}_j) - \ln \frac{1}{f(x_i|\hat{K}_0)} \sum_{j=1}^{q} \hat{\pi}_{j_0} f(x_i, y_i|\hat{\theta}_{j_0}) \right)$$

$$= \sum_{i=1}^{n} \left(-\ln f(x_i|\hat{K}_0) + \ln \sum_{j=1}^{k} \hat{\pi}_j f(x_i, y_i|\hat{\theta}_j) + \ln f(x_i|\hat{K}_0) - \ln \sum_{j=1}^{q} \hat{\pi}_{j_)} f(x_i, y_i|\hat{\theta}_{j_0}) \right)$$

$$= \sum_{i=1}^{n} \left(\ln \sum_{j=1}^{k} \hat{\pi}_j f(x_i, y_i|\hat{\theta}_j) - \ln \sum_{j=1}^{q} \hat{\pi}_{j_0} f(x_i, y_i|\hat{\theta}_{j_0}) \right)$$

Based on the latter results, testing on the conditional densities is equivalent to testing on the marginal densities.

Recall that:

$$f(\boldsymbol{w}|\boldsymbol{\Sigma}, \boldsymbol{\Gamma}) = \frac{1}{(2\pi)^{dim(w)/2}|\boldsymbol{\Sigma}|^{0.5}} \exp\left\{-0.5(w-\Gamma)^T \Sigma^{-1}(w-\Gamma)\right\}$$

where

$$\boldsymbol{w} = \begin{pmatrix} x \\ y \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mu \\ \beta_0 + \beta_1 \mu \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \tau^2 & \beta_1 \tau^2 \\ \beta_1 \tau^2 & \sigma^2 + \tau^2 \beta_1^2 \end{pmatrix}$$

We now want to show that $\exists\, h(x,y)$ and $\epsilon \in (0,1)$ such that $|\ln f(x,y|\Sigma,\Gamma)| \leq h(w) = h(x,y)$ where $Eh(X,Y) < \infty$ assuming that $\frac{1}{\epsilon} \geq \tau^2 \geq \epsilon > 0$, $\frac{1}{\epsilon} \geq \sigma^2 \geq \epsilon > 0$, $-\frac{1}{\epsilon} \leq \mu \leq \frac{1}{\epsilon}$, $-\frac{1}{\epsilon} \leq \beta_0 \leq \frac{1}{\epsilon}$, $-\frac{1}{\epsilon} \leq \beta_1 \leq \frac{1}{\epsilon}$ and $-\frac{1}{\epsilon} \leq \beta_0 + \beta_1\mu \leq \frac{1}{\epsilon}$.

$$2|\ln f(\boldsymbol{w}|\boldsymbol{\Sigma},\boldsymbol{\Gamma})| = \left|\ln 2\pi - \ln|\Sigma| - (w-\Gamma)^T\Sigma^{-1}(w-\Gamma)\right| \leq |\ln 2\pi| + |\ln|\Sigma|| + ||(w-\Gamma)^T\Sigma^{-1}(w-\Gamma)||$$
$$\leq c_1 + c_2 + ||(w-\Gamma)^T\Sigma^{-1}(w-\Gamma)|| \quad (3.10)$$

where $c_1 = \ln 2\pi$ and

$$|\ln|\Sigma|| = |\ln((\tau^2\sigma^2 + \tau^4\beta_1^2) - \beta_1^2\tau^4)| = |\ln \sigma^2\tau^2| \leq |\ln \epsilon^4| = 4|\ln \epsilon| = c_2. \quad (3.11)$$

Furthermore by applying the Cauchy-Schwartz inequality and the induced norm we have that

$$||(w-\Gamma)^T\Sigma^{-1}(w-\Gamma)|| \leq ||(w-\Gamma)^T\Sigma^{-1}||\ \ ||(w-\Gamma)|| \leq ||(w+(-\Gamma))^T||\ \ ||\Sigma^{-1}||\ \ ||(w+(-\Gamma))||$$
$$\leq (||w^T|| + ||\Gamma^T||)||\Sigma^{-1}||(||w|| + ||\Gamma||) \quad (3.12)$$

where

$$||\Gamma^T|| = ||\Gamma|| = \sqrt{\mu^2 + (\beta_0 + \beta_1\mu)^2} \leq \sqrt{\frac{1}{\epsilon^2} + \left(\frac{1}{\epsilon} + \frac{1}{\epsilon^2}\right)^2} = c_3 \quad (3.13)$$

also applying the Frobenius norm (norm of a matrix) we get

$$||\Sigma^{-1}|| \leq ||\Sigma^{-1}||_F = \sqrt{tr({\Sigma^{-1}}^T\Sigma^{-1})} = \sqrt{tr(\Sigma^{-1}\Sigma^{-1})}. \quad (3.14)$$

We note that

$$\Sigma^{-1}\Sigma^{-1} = \begin{pmatrix} \frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2} & \frac{-\beta_1}{\sigma^2} \\ \frac{-\beta_1}{\sigma^2} & \frac{1}{\sigma^2} \end{pmatrix}\begin{pmatrix} \frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2} & \frac{-\beta_1}{\sigma^2} \\ \frac{-\beta_1}{\sigma^2} & \frac{1}{\sigma^2} \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2}\right)^2+\frac{\beta_1^2}{\sigma^4} & \frac{-\beta_1}{\tau^2\sigma^2}-\frac{\beta_1^3+\beta_1}{\sigma^4} \\ \frac{-\beta_1}{\tau^2\sigma^2}-\frac{\beta_1^3+\beta_1}{\sigma^4} & \frac{\beta_1^2+1}{\sigma^4} \end{pmatrix}$$

$$\Rightarrow tr(\Sigma^{-1}\Sigma^{-1}) = \frac{1}{\tau^4}+\frac{2\beta_1^2}{\tau^2\sigma^2}+\frac{\beta_1^4+2\beta_1^2+1}{\sigma^4}$$

$$\Rightarrow ||\Sigma^{-1}||_F = \sqrt{tr(\Sigma^{-1^T}\Sigma^{-1})} = \sqrt{\frac{1}{\tau^4}+\frac{2\beta_1^2}{\tau^2\sigma^2}+\frac{\beta_1^4+2\beta_1^2+1}{\sigma^4}}$$

$$\leq \sqrt{\frac{1}{\epsilon^4}+\frac{2/\epsilon^2}{\epsilon^4}+\frac{1/\epsilon^4+2/\epsilon^2+1}{\epsilon^4}}$$

$$= \sqrt{\frac{1}{\epsilon^8}+\frac{4}{\epsilon^6}+\frac{2}{\epsilon^4}} = c_4 \quad (3.15)$$

Thus from above we let

$$h(w) = c_1 + c_2 + c_4(||w^T||+c_3)(||w||+c_3) = c_1 + c_2 + c_4(||w^T||\,||w||+2c_3||w||+c_3^2), \quad (3.16)$$

which depends only on w and let $E_{f_1}(.)$ denote the expectation with respect to say $f_1(w)$. Assuming the true parameters at the null are respectively $\gamma_0 = (\mu_0, \tau_0, \beta_{00}, \beta_{10}, \sigma_0)$ and $f_o = \sum_{i=1}^{q} \pi_i f_{i0}$ it follows that

$$\iint h(w)f_0(w|\gamma_0)dw = \pi_1 \iint h(w)f_{10}(w|\gamma_0)dw + ... + \pi_q \iint h(w)f_{q0}(w|\gamma_0)dw$$

$$= \pi_1 E_{f_1}(h(w)) + ... + \pi_q E_{f_q}(h(w)) \quad (3.17)$$

Then the first expectation can be evaluated as follows:

$$\pi_1 E_{f_1}(h(w)) = \pi_1\left\{c_1 + c_2 + c_4 E(W_1^2+W_2^2) + 2c_3c_4E\left(\sqrt{W_1^2+W_2^2}\right) + c_4c_3^2\right\} < \infty, \quad (3.18)$$

since in a general setting as in $E(W_1^T W_1 + ... + W_p^T W_p) = E(W_1^T W_1) + ... + E(W_p^T W_p)$ and in particular $W_1 = [X, Y]$, then $E(W_1^T W_1) = E(X_1^2 + Y_1^2) = var(X_1) + E(X_1)^2 + var(Y_1) + E(Y_1)^2 = \tau^2 + \mu^2 + \sigma^2 + \tau^2\beta_1^2 + (\beta_0 + \beta_1\mu)^2$ and p is the number of rows in W.

It also follows by Jensen's inequality and the concavity of the square root function that $E\left(\sqrt{W_1^T W_1 + ... + W_p^T W_p}\right) \leq \sqrt{E(W_1^T W_1 + ... + W_p^T W_p)} = \sqrt{E(W_1^T W_1) + ... + E(W_p^T W_p)}$. Specifically for $p = 2$,

$$E\left(\sqrt{W_1^T W_1}\right) \leq \sqrt{E(W_1^T W_1)}$$
$$= \sqrt{\tau^2 + \mu^2 + \sigma^2 + \tau^2\beta_1^2 + (\beta_0 + \beta_1\mu)^2} \leq \sqrt{\frac{4}{\epsilon^2} + \frac{1}{\epsilon^4}} < \infty \quad (3.19)$$

by virtue of our assumptions on the parameters. So we have that,

$$2|\ln f(w|\Sigma, \Gamma)| \leq h(w) \Longrightarrow |\ln f(w|\Sigma, \Gamma)| \leq h(w), \quad (3.20)$$

where $E(h) \leq \infty$

Furthermore, we note that $f(w|\Sigma, \Gamma)$ possesses partial derivatives up to the order 5 and that $\forall z \leq 5$,

$$\frac{D^z_{i_1 \ldots i_z} f(w|\Sigma_0, \Gamma_0)}{f_0} \in L^3(f_0 \nu), \quad (3.21)$$

where $i_1 ... i_z$ indexes the densities of the function $f(w|\Sigma_0, \Gamma_0)$ and z the order of derivative being taken with respect to the parameters in question. So if for instance $i_1 = i_2$ and $z = 2$ then the numerator of the expression above will yield the second derivative with respect to one parameter. In particular we will have for example, that

$$\frac{D^2_{\gamma_0\gamma_0} f(w|\Sigma_0\Gamma_0)}{f_0} = \frac{\frac{\partial^2 f(w|\gamma_0)}{\partial \gamma_0^2}}{f_0}, \quad (3.22)$$

which is the second derivative of the function with respect to γ_0. On the other hand if $i_1 \neq i_2$ and $z = 2$ then we take the partial derivative with respect to say β_1 first and then μ second getting a mixed partial derivative.

To see why eq(11) holds (here we suppress the dependence on component l in our notation), consider the following expansions:

$$\begin{aligned}(w-\Gamma)^T\Sigma^{-1}(w-\Gamma) &= \begin{pmatrix} x-\mu \\ y-(\beta_0+\beta_1\mu)\end{pmatrix}^T \begin{pmatrix} \frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2} & \frac{-\beta_1}{\sigma^2} \\ \frac{-\beta_1}{\sigma^2} & \frac{1}{\sigma^2}\end{pmatrix}\begin{pmatrix} x-\mu \\ y-(\beta_0+\beta_1\mu)\end{pmatrix} \\ &= (x-\mu)^2\left(\frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2}\right)+\frac{-2\beta_1}{\sigma^2}(x-\mu)(y-(\beta_0+\beta_1\mu))+\frac{1}{\sigma^2}(y-(\beta_0+\beta_1\mu))^2 \quad (3.23)\end{aligned}$$

Thus the density can be expressed as:

$$\begin{aligned}&f(w|\Sigma,\Gamma) \\ &\propto \frac{1}{\tau\sigma}\exp\left\{-0.5\left((x-\mu)^2\left(\frac{1}{\tau^2}+\frac{\beta_1^2}{\sigma^2}\right)+\frac{-2\beta_1}{\sigma^2}(x-\mu)(y-(\beta_0+\beta_1\mu))+\frac{1}{\sigma^2}(y-(\beta_0+\beta_1\mu))^2\right)\right\} \quad (3.24)\end{aligned}$$

We verify that the general derivative (D) of the density as defined above is of the form $Df = f(x,y|\sigma,\tau,\mu,\beta_1,\beta_0) * polynomial$ where the plynomial is in terms of $x-\mu$ and y where all parameters are expressed as rational functions. This will be clarified below.

To see why we let $v = 1/\sigma$, $s = 1/\tau$ and notice that f can be written as:

$$f(x, y|\sigma, \tau, \mu, \beta_1, \beta_0) \propto$$
$$vs \exp\left\{ -0.5\left[(x-\mu)^2(s^2 + v^2\beta_1^2) - 2v^2\beta_1(x-\mu)(y-(\beta_0+\beta_1\mu)) + v^2(y-(\beta_0+\beta_1\mu))^2\right]\right\}$$

We begin with the first derivative to form the foundation of the derivatives and then proof by induction that the proposed pattern holds true for all higher order and mixed derivatives. For simplicity we shall use f to represent the function defined above and D to represent the derivative with respect to the parameter of interest expressed as a subscript. In that respect we will have that:

$$D_{\beta_1} f = f * \left[-v^2\beta_1(x-\mu)^2 + (x-\mu)\left\{v^2(y-(\beta_0+\beta_1\mu)) - v^2\beta_1\mu\right\} + \mu(y-(\beta_0+\beta_1\mu))\right]$$

which is of the form $f * \textit{polynomial}$ in $x - \mu$.

$$D_{\beta_0} f = f * \left[-v^2\beta_1(x-\mu) + v^2\mu(y-(\beta_0+\beta_1\mu))\right]$$

which is of the form $f * \textit{polynomial}$ in $x - \mu$.

$$D_{\mu} f = f * \left[(x-\mu)\left\{(s^2 + v^2\beta_1^2) - v^2\beta_1^2\right\}\right]$$

which is of the form $f * \textit{polynomial}$ in $x - \mu$.

$$D_{\frac{1}{\sigma}} f = D_v f = f * \left[-v\beta_1^2(x-\mu)^2 + 2v\beta_1(x-\mu)(y-(\beta_0+\beta_1\mu)) - v(y-(\beta_0+\beta_1\mu))^2 + \frac{1}{v}\right]$$

which is of the form $f * polynomial$ in $x - \mu$.

$$D_{\frac{1}{\tau}} f \quad = \quad D_s f \quad = \quad f \quad * \quad \left[\quad - \quad s(x \quad - \quad \mu)^2 \quad + \quad \frac{1}{s} \right]$$

which is of the form $f * polynomial$ in $x - \mu$. Without loss of generality we shall define the polynomial as follows:

$$\sum_{j=0}^{\lambda} \sum_{r=0}^{t} c_r(\gamma) x^{t-r} y^{\lambda - j} \tag{3.25}$$

The foregone derivatives above are all of the form $f * polynomial$ in x, y, where $c(\gamma)'s$ are polynomial collections and rational functions of γ. Next we assume that this observation holds true for all derivatives in the sense that the derivative of the form $f * polynomial$ regardless of the parameter of interest yields a similar format $\tilde{f} * \tilde{polynomial}$ in that taking the derivative of a polynomial yields a polynomial and the derivative of f gives the product of f and a polynomial.

Mathematically we assume that

$$D_{i_1,\ldots,i_z-1} f = f * polynomial$$

where the polynomial is in terms of x and y. To maximize the polynomial we note that γ is restricted to a compact region in the sense of $\|\gamma - \gamma_0\| \leq \epsilon$ and also recall that $C_r(\gamma)$ is a rational function of the form

$$C_r(\gamma) = \frac{P_r(\gamma)}{v^{\delta}, s^{\Delta}}, \Delta \geq 0, \delta \geq 0,$$

where $P_r(\gamma)$ is a polynomial function in γ. However as we have mentioned above γ is in a compact region and since the polynomial function $P_r(\gamma)$ is continuous with

respect to γ vis a vis $\|\gamma - \gamma_0\| \leq \epsilon$, $\exists\ M(\epsilon) \geq 0$ for which $|P_r(\gamma)| \leq M(\epsilon)$. In addition we recall from previous assumptions that $\frac{1}{s} \leq \frac{1}{\epsilon}$ and $\frac{1}{v} \leq \frac{1}{\epsilon}$ and it follows that:

$$C_r(\gamma) = \frac{P_r(\gamma)}{v^{\delta},\, s^{\Delta}} \leq \frac{M(\epsilon)}{\epsilon^{\delta+\Delta}} < \infty,$$

which in turn ensures that the polynomial

$$\sum_{j=0}^{\lambda} \sum_{r=0}^{t} c_r(\gamma) x^{t-r} y^{\lambda-j} < \infty$$

In a broader sense if we let $g = x$ and $\gamma = \beta_0, \beta_1, \sigma, \tau, \mu$ then we are specifically assuming that

$$D_{i_1,\ldots,i_z-1} f = f * [c_1(\gamma) g^t y^{\lambda} + c_2(\gamma) g^{t-1} y^{\lambda-1} + \ldots + c_r(\gamma)]$$

for an r^{th} term polynomial with degree $t + \lambda$.

We proceed to show that all the derivatives (both mixed and otherwise) of $D_{i_1,\ldots,i_z-1} f = f * \textit{polynomial}$ are also of the similar form $f^{**} * \textit{polynomial}^{**}$

$$D_{i_z}(D_{i_1,\ldots,i_{z-1}} f) = D_{i_z}\Big[f * [c_1(\gamma) g^t + c_2(\gamma) g^{t-1} + \ldots + c_r(\gamma)\Big]$$

$$= D_{iz}\Big(f * c_1(\gamma) g^t y^{\lambda}\Big) + \ldots + D_{iz}\Big(f * c_r(\gamma)\Big)$$

$$= \Big\{\Big([D_{iz} f] * c_1(\gamma) g^t y^{\lambda} + f * [D_{iz} c_1(\gamma) g^t y^{\lambda}]\Big) + \ldots + [D_{i_z} f] * c_r(\gamma) + f * [D_{i_z} c_r(\gamma)]\Big)\Big\}$$

$$= f * \textit{polynomial} * c_1(\gamma) g^t + f[D_{i_z}(c_1(\gamma)) g^t y^{\lambda} + (D_{i_z} g^t y^{\lambda}) c_1(\gamma) + \ldots + f * \textit{polynomial} * c_r(\gamma) + f * D_{i_z}(c_r(\gamma))$$

$$= f\Big[\textit{polynomial} * c_1(\gamma) g^t y^{\lambda} + [D_{i_z}(c_1(\gamma)) g^t y^{\lambda} + (D_{i_z} g^t y^{\lambda}) c_1(\gamma) + \ldots + \textit{polynomial} * c_r(\gamma) + \frac{\partial}{\partial i_z} c_r(\gamma)\Big]$$

$$= f * \textit{polynomial} = f(x, y|\sigma, \tau, \mu, \beta_1, \beta_0) \sum_{j=0}^{\lambda} \sum_{r=0}^{t} c_r(\gamma) x^{t-r} y^{\lambda-j}$$

Based on the derivations above we have that

$$\frac{D^z_{i_1,\ldots,i_z} f|_{\gamma_0}}{f_0} = \frac{f_{l0} * polynomial}{\pi_1 f_{10} + \ldots + \pi_q f_{q0}} \leq \frac{f_{l0} * polynomial}{\pi_l f_{l0}} = \frac{1}{\pi_{l0}} * polynomial$$
$$= \frac{1}{\pi_{l0}} \sum_{j=0}^{\lambda} \sum_{r=0}^{t} c_r(\gamma) x^{t-r} y^{\lambda-j}$$

with respect to x, y and coefficients $c(\gamma)$ which are rational functions of γ_0. Since $E(|X|^p) < \infty$ and $E(|Y|^p) < \infty$ it follows that

$$\int\int \left|\frac{D^z_{i_1,\ldots,i_z} f|_{\gamma_0}}{f_0}\right|^{1/3} f(x,y|\gamma_0)dxdy \leq \Big(\int\int \left|\frac{1}{\pi_{l0}} \sum_{j=0}^{\lambda} \sum_{r=0}^{t} c_r(\gamma_0) x^{t-r} y^{\lambda-j}\right|^3 f(x,y|\gamma_0)\Big)^{1/3} dxdy < \infty$$

Next we show that there exists an $H_5(x,y)$ such that

$$\sup_{||\gamma-\gamma_0||\leq\epsilon} \left|\frac{D^5_{i_1,\ldots,i_5} f_\gamma}{f_0}\right| \leq H_5(x,y),$$

and $E_{f_0}(H_5^3(X,Y)) < +\infty$.

Recall that $f(x,y) = vs \exp\{-0.5[(x-\mu)^2(s^2+v^2\beta_1^2) - 2v^2\beta_1(x-\mu)(y-\bar{\beta}) + v^2(y-\bar{\beta})^2]\}$ where $\bar{\beta} = \beta_0 + \beta_1\mu$.

Also recall that $f_0 = \sum_{i=1}^{q} \pi_i f_i(x,y)$ which we shall represent as $f_0 = \pi_1 f_1 + \ldots + \pi_q f_q$

To show the P0, we notice that $\left|\frac{D^5 f_{\gamma r}}{f_0}\right|^3 = \left|\frac{D^5 f_{\gamma r}}{\pi_1 f_1 + \ldots + \pi_q f_q}\right|^3 \leq \left|\frac{D^5 f_{\gamma r}}{\pi_r f_r}\right|^3$, for $r \leq q$, where D is an arbitrary partial derivative, f_{γ_r} is the nearby and $f_r = f_{\gamma_{r0}}$ is the true density for component r. In the subsequent proof, expressions without subscripts are the nearby quantities and those with subscripts are the true quantities.

But

$$\left|\frac{Df_{\gamma_r}}{\pi_r f_r}\right|$$

$$\leq \frac{vs\exp\{-0.5[(x-\mu)^2(s^2+v^2\beta_1^2)-2v^2\beta_1(x-\mu)(y-\bar{\beta})+v^2(y-\bar{\beta})^2]\}M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t}x^{t-r}y^{\lambda-j}\right|}{\pi_r v_r s_r\exp\{-0.5[(x-\mu)^2(s_r^2+v_r^2\beta_{1r}^2)-2v^2\beta_{1r}(x-\mu_r)(y-\bar{\beta}_r)+v^2(y-\bar{\beta}_r)^2]\}}$$

$$= \frac{vs}{\pi_r v_r s_r}\exp\Bigg(-0.5\Big[(x-\mu)^2(s^2+v^2\beta_1^2)-(x-\mu_r)^2(s_r^2+v_r^2\beta_{1r}^2)-2v^2\beta_1(x-\mu)(y-\bar{\beta})$$

$$+2v_r^2\beta_{1r}(x-\mu_r)(y-\bar{\beta}_r)+v^2(y-\bar{\beta})^2-v_r^2(y-\bar{\beta}_r)^2\Big]\Bigg)M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t}x^{t-r}y^{\lambda-j}\right|$$

$$= \frac{vs}{\pi_r v_r s_r}\exp\Bigg(-0.5\Big[x^2\{(s^2+v^2\beta_1^2)-(s_r^2+v_r^2\beta_{1r}^2)\}+x\{-2\mu(s^2+v^2\beta_1^2)+2\mu_r(s_r^2+v_r^2\beta_{1r}^2)\}$$

$$\mu^2(s^2+v^2\beta_1^2)-\mu_r^2(s^2+v^2\beta_1^2)+xy[-2v^2\beta_1-2v_r^2\beta_{1r}]$$

$$x[-2v^2\beta_1\bar{\beta}+2v^2\beta_{1r}\bar{\beta}_r]+y[2v^2\beta_1\mu-2v_r^2\beta_{1r}\mu_r]$$

$$\mu[-2v^2\beta_1\bar{\beta}+2v_r^2\beta_{1r}\bar{\beta}_r]+y^2[v^2-v_r^2]+y[-2v^2\bar{\beta}+2v_r^2\bar{\beta}]+(v^2\bar{\beta}^2-v_r^2\bar{\beta}_r^2)$$

$$+2v_r^2\beta_{1r}(x-\mu_r)(y-\bar{\beta}_r)+v^2(y-\bar{\beta})^2-v_r^2(y-\bar{\beta}_r)^2\Big]\Big)*M(\epsilon)\Big|\sum_{j=0}^{\lambda}\sum_{r=0}^{t}x^{t-r}y^{\lambda-j}\Big|$$

$$=\frac{vs}{\pi_r v_r s_r}\exp\Big(-0.5\Big[x^2\{(s^2+v^2\beta_1^2)-(s_r^2+v_r^2\beta_{1r}^2)\}+x\{-2\mu(s^2+v^2\beta_1^2)+2\mu_r(s_r^2+v_r^2\beta_{1r}^2)-2v^2\beta_1\bar{\beta}+2v^2\beta_{1r}\bar{\beta}_r\}$$

$$xy[-2v^2\beta_1-2v_r^2\beta_{1r}]+y^2[v^2-v_r^2]+y[2v^2\beta_1\mu-2v_r^2\beta_{1r}\mu_r-2v^2\bar{\beta}+2v_r^2\bar{\beta}]$$

$$\mu^2(s^2+v^2\beta_1^2)-\mu_r^2(s^2+v^2\beta_1^2)+\mu[-2v^2\beta_1\bar{\beta}+2v_r^2\beta_{1r}\bar{\beta}_r]$$

$$+(v^2\bar{\beta}^2-v_r^2\bar{\beta}_r^2)\Big]\Big)*M(\epsilon)\Big|\sum_{j=0}^{\lambda}\sum_{r=0}^{t}x^{t-r}y^{\lambda-j}\Big|$$

$$=\exp\Big(-0.5\Big[ax^2+bx+cxy+dy^2+ey+f\Big]\Big)*M(\epsilon)\Big|\sum_{j=0}^{\lambda}\sum_{r=0}^{t}x^{t-r}y^{\lambda-j}\Big|,$$

where

$$a=-0.5\Big[(s^2+v^2\beta_1^2)-(s_r^2+v_r^2\beta_{1r}^2)\Big],\ \ b=-0.5\Big[-2\mu(s^2+v^2\beta_1^2)+2\mu_r(s_r^2+v_r^2\beta_{1r}^2)-2v^2\beta_1\bar{\beta}+2v^2\beta_{1r}\bar{\beta}_r\Big]$$

$$c=-0.5\Big[-2v^2\beta_1-2v_r^2\beta_{1r}\Big],\ \ d=-0.5\Big[v^2-v_r^2\Big],\ \ e=-0.5\Big[2v^2\beta_1\mu-2v_r^2\beta_{1r}\mu_r-2v^2\bar{\beta}+2v_r^2\bar{\beta}\Big]$$

$$f = -0.5\Big[\mu^2(s^2+v^2\beta_1^2)-\mu_r^2(s_r^2+v_r^2\beta_{1r}^2)+\mu[-2v^2\beta_1\bar{\beta}+2v_r^2\beta_{1r}\bar{\beta}_r]+(v^2\bar{\beta}^2-v_r^2\bar{\beta}_r^2)+\log\left(\frac{vs}{\pi_r v_r s_r}\right)\Big]$$

Now assuming that $\|\mu-\mu_r\|\leq\epsilon$, $\|\tau^2-\tau_r^2\|\leq\epsilon$, $\|\sigma^2-\sigma_r^2\|\leq\epsilon$, $\|\beta_1-\beta_{1r}\|\leq\epsilon$, $\|\beta_0-\beta_{0r}\|\leq\epsilon$, $\|\bar{\beta}-\bar{\beta}_r\|\leq\epsilon$, $\|s^2-s_r^2\|\leq\epsilon$, $\|v^2-v_r^2\|\leq\epsilon$, $\|s-s_r\|\leq\epsilon$, and $\|v-v_r\|\leq\epsilon$ then we can find the respective maximum and minimum expressions for $a, b, \ldots, f$.

For we notice that

$$\begin{aligned}\left|0.5\Big[(s^2+v^2\beta_1^2)-(s_r^2+v_r^2\beta_{1r}^2)\Big]\right| &\leq \left|s^2-s_r^2\right|+\left|v^2\beta_1^2-v_r^2\beta_{1r}^2\right| \\ &\leq \left|(s_r^2+\epsilon)-s_r^2\right|+\left|(v^2+\epsilon)(\beta_{1r}^2+\epsilon)-v_r^2\beta_{1r}^2\right| \\ &= \epsilon+\left|\epsilon v_r^2+\epsilon\beta_{1r}^2+\epsilon^2\right| = a_{max,\epsilon}\end{aligned}$$

similarly we can obtain a maximum for b as follows:

$$\left|0.5\Big[-2\mu(s^2+v^2\beta_1^2)+2\mu_r(s_r^2+v_r^2\beta_{1r}^2)-2v^2\beta_1\bar{\beta}+2v^2\beta_{1r}\bar{\beta}_r\Big]\right|$$

$$\leq \left|\mu s^2-\mu_r s_r^2\right|+\left|\mu_r v^2\beta_1^2-\mu_r v_r^2\beta_{1r}^2\right|+\left|v^2\beta_1\bar{\beta}-v^2\beta_{1r}\bar{\beta}_r\right|$$

$$\begin{aligned}\leq \left|(\mu_r\pm\epsilon)(s_r^2+\epsilon)-\mu_r s_r^2\right|+\left|(\mu_r\pm\epsilon)(v_r^2+\epsilon)(\beta_{1r}^2+\epsilon)-\mu_r v_r^2\beta_{1r}^2\right| \\ +\left|(v_r^2+\epsilon)(\beta_{1r}\pm\epsilon)(\bar{\beta}_r\pm\epsilon)-v^2\beta_{1r}\bar{\beta}_r\right|\end{aligned}$$

$$= \left|(\mu_r \pm s_r + \epsilon)\epsilon\right| + \left|\epsilon(\mu_r\beta_{1r}^2 \pm v_r^2\beta_{1r}^2 \pm \beta_{1r}^2\epsilon + \mu_r v_r^2 + \epsilon\mu_r \pm \epsilon v_r^2 \pm \epsilon^2)\right|$$
$$+\left|\epsilon(\bar{\beta}_r\beta_{1r\pm v_r^2\bar{\beta}_r} \pm \epsilon\bar{\beta}_r \pm v_r^2\beta_{1r} + \epsilon v_r^2 \pm \epsilon\beta_{1r} + \epsilon^2)\right|$$
$$\leq b_{max,\epsilon}$$

Following similar derivations we obtain the following maximums:

$$\left|0.5(-2v^2\beta_1 - 2v_r^2\beta_{1r})\right| \leq |(v_r^2 + \epsilon)(\beta_{1r} \pm \epsilon) - v_r^2\beta_{1r}| \leq \left|\epsilon(v_r^2 + \beta_{1r} + \epsilon)\right| \leq c_{max,\epsilon}$$

$$|0.5(v^2 - v_r^2)| \leq |(v_r^2 + \epsilon) - v_r^2| = |\epsilon| = d_{max,\epsilon}$$

$$\left|0.5\left[2v^2\beta_1\mu - 2v_r^2\beta_{1r}\mu_r - 2v^2\bar{\beta}_r + 2v_r^2\bar{\beta}\right]\right| \leq \left|(v_r^2 + \epsilon)(\beta_{1r} \pm \epsilon)(\mu_r \pm \epsilon) - v_r^2\beta_{1r}\mu_r\right| + \left|(v_r^2 + \epsilon)(\bar{\beta}_r \pm \epsilon) - v_r^2(\bar{\beta}_r \pm \epsilon)\right|$$
$$= \left|\epsilon\left(\mu_r\beta_{1r} \pm v_r^2\beta_{1r} \pm \epsilon\beta_{1r} \pm \mu_r v_r^2 + \epsilon v_r^2 \pm \epsilon\mu_r + \epsilon^2\right)\right| + \left|\epsilon\left(v_r^2 \pm v_r^2 - \bar{\beta}_r + \epsilon\right)\right| \leq e_{max,\epsilon}$$

and

$$\left|0.5\left[\mu^2(s^2 + v^2\beta_1^2) - \mu_r^2(s^2 + v^2\beta_1^2) + \mu[-2v^2\beta_1\bar{\beta} + 2v_r^2\beta_{1r}\bar{\beta}_r] + (v^2\bar{\beta}^2 - v_r^2\bar{\beta}_r^2) + \log\left(\tfrac{vs}{\pi_r v_r s_r}\right)\right]\right|$$

$$\leq \left|0.5(\mu^2(s^2 + v^2\beta_1^2) - \mu_r^2(s^2 + v^2\beta_1^2))\right| + \left|\mu\left(v^2\beta_1^2\bar{\beta} - v_r^2\beta_r\bar{\beta}_r\right)\right| + \left|0.5(v^2\bar{\beta}^2 - v_r^2\bar{\beta}_r^2)\right| + \left|\log\left(\frac{vs}{\pi_r v_r s_r}\right)\right|$$

$$\leq \left|0.5((\mu_r^2+\epsilon)(s_r^2+\epsilon)+(v_r^2+\epsilon)(\beta_{1r}^2+\epsilon)-\mu_r^2 s_r^2-\mu_r^2 v_r^2\beta_r^2))\right| + \left|(\mu_r \pm \epsilon)\Big((v_r^2+\epsilon)(\beta_{1r}^2+\epsilon)(\bar{\beta}_r \pm \epsilon)-v_r^2\beta_r\bar{\beta}_r\Big)\right|$$

$$+\left|0.5((v_r^2+\epsilon)(\bar{\beta}_r^2+\epsilon)-v_r^2\bar{\beta}_r^2)\right| + \left|\log\tfrac{1}{\pi_r}\right| + \left|\log\tfrac{v_r+\epsilon}{v_r}\right| + \left|\log\tfrac{s_r+\epsilon}{s_r}\right|$$

$$\leq \left|\epsilon\Big(\mu_r^2+s_r^2+\epsilon+v_r^2\mu_r^2+\mu_r^2\beta_{1r}^2+v_r^2\beta_{1r}^2+v_r^2+\epsilon\beta_{1r}^2+\epsilon\Big)\right| + \left|\epsilon(v_r^2+\bar{\beta}_r^2+\epsilon)\right|$$

$$+\left|\epsilon\Big(\mu_r v_r^2\bar{\beta}_r \pm \beta_{1r}^2\mu_r v_r^2 \pm \mu_r v_r^2\epsilon+\mu_r\beta_{1r}^2\bar{\beta}_r \pm \epsilon\mu_r\beta_{1r}^2\epsilon\mu_r\bar{\beta}_r \pm \epsilon^2\mu_r \pm v_r^2\beta_{1r}^2\bar{\beta}_r+\epsilon v_r^2\beta_{1r}^2 \pm \epsilon v_r^2\bar{\beta}_r\epsilon^2 v_r^2 \pm \beta_{1r}^2\bar{\beta}_r + \epsilon^2\beta_{1r}^2 \pm \epsilon^2\bar{\beta}_r+\epsilon^3\Big)\right|$$

$$+\left|\log\tfrac{1}{\pi_r}\right| \pm \tfrac{\epsilon}{v_r}(1+O(\epsilon)) + \pm\tfrac{\epsilon}{s_r}(1+O(\epsilon)) = f_{max,\epsilon}$$

Thus we have for $x > 0$ and $y > 0$ that

$$\exp\Big(-0.5\Big[ax^2+bx+cxy+dy^2+ey+f\Big]\Big) * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|$$

$$\leq \exp\Big[a_{max,\epsilon}x^2+b_{max,\epsilon}x+c_{max,\epsilon}xy+d_{max,\epsilon}y^2+e_{max,\epsilon}y+f_{max,\epsilon}\Big] * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|$$

The afore-derived expression is wholly in terms of x and y, the true parameters and ϵ. Thus we may define $H_5(x, y)$ for the different values that x and y can assume as follows:

$$H_5(x,y) = \left|\frac{D^5 f_{\gamma r}}{\pi_r f_r}\right|^3 =$$

$$\begin{cases} \left(e^{[a_{max}x^2+b_{max}x+c_{max}xy+d_{max}y^2+e_{max}y+f_{max}]} * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|\right)^3 & x \geq 0, y \geq 0 \\ \left(e^{[a_{max}x^2+b_{min}x+c_{min}xy+d_{max}y^2+e_{max}y+f_{max}]} * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|\right)^3 & x \leq 0, y \geq 0 \\ \left(e^{[a_{max}x^2+b_{max}x+c_{min}xy+d_{max}y^2+e_{min}y+f_{max}]} * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|\right)^3 & x \geq 0, y \leq 0 \\ \left(e^{[a_{max}x^2+b_{min}x+c_{max}xy+d_{max}y^2+e_{min}y+f_{max}]} * M(\epsilon)\left|\sum_{j=0}^{\lambda}\sum_{r=0}^{t} x^{t-r}y^{\lambda-j}\right|\right)^3 & x \leq 0, y \leq 0 \end{cases}$$

Now we show that $E(H_5(X,Y)) < \infty$ as follows:

$E(H_5(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_5(x,y)f_0 dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_1 H_5(x,y)f_1 dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_2 H_5(x,y)f_2 dxdy \ldots + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_q H_5(x,y)f_q dxdy$

However by appealing to the linearity of the integration above we have for the first part that:

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_1 H_5(x,y)f_1 dxdy < \infty$ since $E(X^p) < \infty$, $E(Y^p) < \infty$ and $E(X^{t-r}Y^{\lambda-j}) < \infty$ and the exponential functions $H_5(x,y)f_1 \propto \bar{\bar{f}}_1$ which resembles Gaussian densities and thus the kernel method is applicable. Moreover since these facts hold for each of the q integrations above we conclude that $E(H_5(X,Y)) < \infty$ which concludes the proof of P0.

To prove $P1$ we appeal to Yakowitz's and Spragins' (1968)[61] characterizations of identifiability theorem that a finite mixture from a family $\mathcal{F}$ of cdf's is identifi-

able iff $\mathcal{F}$ is linearly independent over the real numbers. Since the identifiability has been checked by Yakowitz's and Spragins' (1968)[61] proposition 2, the proof of P1 follows forthwith as shown below noting yet again that f_0 is the q mixture at the null.

Lemma 1. Suppose that

$$\sum_{i=1}^{m_1} \gamma_i f(x, y, \theta_i) = \sum_{i=1}^{m_2} \alpha_i f(x, y, \Phi_i)$$

then we see that taking integrals on both sides yield

$$\begin{aligned}
\int_{-\infty}^{r}\int_{-\infty}^{s}\sum_{i=1}^{m_1}\gamma_i f(x,y,\theta_i)dxdy &= \int_{-\infty}^{r}\int_{-\infty}^{s}\sum_{i=1}^{m_2}\alpha_i f(x,y,\Phi_i)dxdy \\
\iff \sum_{i=1}^{m_1}\int_{-\infty}^{r}\int_{-\infty}^{s}\gamma_i f(x,y,\theta_i)dxdy &= \sum_{i=1}^{m_2}\int_{-\infty}^{r}\int_{-\infty}^{s}\alpha_i f(x,y,\Phi_i)dxdy \\
\iff \sum_{i=1}^{m_1}\gamma_i F(r,s,\theta_i) &= \sum_{i=1}^{m_2}\alpha_i F(r,s,\Phi_i),
\end{aligned}$$

which is identifiable according to proposition 2 in Yakowitz and Spragins[61]. We conclude that identifiability of a family of Gaussian cdf's implies identifiability of the corresponding pdf's. This means that $m_1 = m_2$, $\gamma_i = \alpha_i$, and $\theta_i = \Phi_i$ for all $1 \leq i \leq m_1$.

Lemma 2. Suppose that $\alpha_1 f_1(x, y, \theta_1) + \ldots + \alpha_k f_k(x, y, \theta_k) = \sum_{l=1}^{k} \alpha_l f_l(x, y, \theta_l) = 0$, then we have by integrating both sides that

$$\int_{-\infty}^{u}\int_{-\infty}^{v}\sum_{l=1}^{k}\alpha_l f_l(x,y,\theta_l)dxdy = 0 \iff \sum_{l=1}^{k}\int_{-\infty}^{u}\int_{-\infty}^{v}\alpha_l f_l(x,y,\theta_l)dxdy = 0 \iff \sum_{l=1}^{k}\alpha_l F_l(u,v,\theta_l) = 0,$$

which implies that $\alpha_1 = \ldots \alpha_k = 0$ by Yakowitz and Spragins [61]. Thus linear independence of a family of Gaussian cdf's implies the linear independence of the corresponding pdf's.

Define $\bar{f}(x,y,\gamma,\gamma_0) = \left\{\left(\frac{f_{\gamma^l}}{f_0}\right)_{l=1,\ldots,p}, \left(\frac{f_{\gamma^{l0}}}{f_0}\right)_{l=1,\ldots,q}, \left(\frac{D_i^1 f_{\gamma^{l0}}}{f_0}\right)_{l=1,\ldots,q,i=1,\ldots,k}, \left(\frac{D_{ij}^2 f_{\gamma^{l0}}}{f_0}\right)_{l=\sigma(1),\ldots,\sigma(p_2),i,j=1,\ldots,k}\right\} = \left\{\left(\frac{f_{\gamma^1}}{f_0}, \frac{f_{\gamma^l}}{f_0}, \ldots, \frac{f_{\gamma^p}}{f_0}\right), \left(\frac{f_{\gamma^{10}}}{f_0}, \frac{f_{\gamma^{20}}}{f_0}, \ldots, \frac{f_{\gamma^{q0}}}{f_0}\right), \left(\frac{D_i^1 f_{\gamma^{10}}}{f_0}, \frac{D_i^1 f_{\gamma^{20}}}{f_0}, \ldots, \frac{D_i^1 f_{\gamma^{q0}}}{f_0}\right)_{i=1,\ldots,k}, \left(\frac{D_{ij}^2 f_{\gamma^{10}}}{f_0}, \frac{D_{ij}^2 f_{\gamma^{20}}}{f_0}, \ldots, \frac{D_{ij}^2 f_{\gamma^{\sigma(p2)0}}}{f_0}\right)_{l=\sigma(1),\ldots,\sigma(p_2),i,j=1,\ldots,k}\right\}$ according to the notations in [55].

Suppose for some constants $\eta_{1l}, \ldots, \eta_{4b}$ we have that

$$\sum_{l=1}^{p} \eta_{1l}\left(\frac{f_{\gamma^l}}{f_0}\right) + \sum_{a=1}^{q} \eta_{2a}\left(\frac{f_{\gamma^{a0}}}{f_0}\right) + \sum_{a=1}^{q} \eta_{3a}\left(\frac{D_i^1 f_{\gamma^{a0}}}{f_0}\right)_{i=1,\ldots,k} + \sum_{b=\sigma(1)}^{\sigma(p2)} \eta_{4b}\left(\frac{D_{ij}^2 f_{\gamma^{b0}}}{f_0}\right)_{i,j=1,\ldots,k} = 0. \quad (3.26)$$

Recall from Lemma 1 and 2 that $\sum_{l=1}^{p} \eta_{1l}\left(\frac{f_{\gamma^l}}{f_0}\right) = 0 \iff \sum_{l=1}^{p} \eta_{1l} f^* = 0$ implies that $\eta_{11}, \ldots, \eta_{1q} = 0$ and similarly, $\sum_{a=1}^{q} \eta_{2a}\left(\frac{f_{\gamma^{a0}}}{f_0}\right) = 0 \iff \sum_{a=1}^{q} \eta_{2a} f^*_{\gamma^{a0}} = 0$ implies that $\eta_{21}, \ldots, \eta_{2q} = 0$. Recall again that the partial derivatives of the density results in a product of a polynomial (as described in equation 3.25) and the density. Moreover, the polynomial is continuous on a compact set of parameters and thus bounded by some $M(\epsilon)$ for $\epsilon \in (0,1)$. As a result we note from $\bar{f}(x,y,\gamma,\gamma_0)$ suppressing all other indices that:

$$\sum_{l=1}^{p_1}\eta_{1l}\frac{f_{\gamma^l}}{f_0}+\sum_{a=1}^{q}\eta_{2a}\frac{f_{\gamma^{a0}}}{f_0}+\sum_{a=1}^{q}\eta_{3a}\frac{D^1 f_{\gamma^{a0}}}{f_0}+\sum_{b=\sigma(1)}^{\sigma p_2}\eta_{4b}\frac{f_{\gamma^{b0}}}{f_0}$$

$$=\sum_{l=1}^{p_1}\eta_{1l}f^*_{\gamma^l}+\sum_{a=1}^{q}\eta_{2a}f^*_{\gamma^{a0}}+\sum_{a=1}^{q}\eta_{3a}\sum_{j=0}^{\lambda}\sum_{r=0}^{t_1}c_{r_1}(\gamma)x^{t_1-r_1}y^{\lambda_1-j}f^*_{\gamma^{a0}}+\sum_{b=\sigma(1)}^{\sigma p_2}\eta_{4b}\sum_{j=0}^{\lambda}\sum_{r=0}^{t}c_r(\gamma)x^{t-r}y^{\lambda-j}f^*_{\gamma^{b0}}$$

$$\leq\sum_{l=1}^{p_1}\eta_{1l}f^*_{\gamma^l}+\sum_{a=1}^{q}\eta_{2a}f^*_{\gamma^{a0}}+\sum_{a=1}^{q}\eta_{3a}\sum_{j=0}^{\lambda}\sum_{r=0}^{t_1}M_1(\epsilon)x^{t_1-r_1}y^{\lambda_1-j}f^*_{\gamma^{a0}}+\sum_{b=\sigma(1)}^{\sigma p_2}\eta_{4b}\sum_{j=0}^{\lambda}\sum_{r=0}^{t}M(\epsilon)x^{t-r}y^{\lambda-j}f^*_{\gamma^{b0}}$$

$$=\sum_{l=1}^{p_1}\eta_{1l}f^*_{\gamma^l}+\sum_{a=1}^{q}\eta_{2a}f^*_{\gamma^{a0}}+\sum_{a=1}^{q}\eta^*_{3a}f^*_{\gamma^{a0}}+\sum_{b=\sigma(1)}^{\sigma p_2}\eta^*_{4b}f^*_{\gamma^{b0}}$$

$$=\eta_{11}f^*_{\gamma^1}+\ldots+\eta_{1p}f^*_{\gamma^{p_1}}+\eta_{21}f^*_{\gamma^{10}}+\ldots+\eta_{2q}f^*_{\gamma^{q0}}+\eta^*_{31}f^*_{\gamma^{10}}+\ldots+\eta^*_{3q}f^*_{\gamma^{q0}}+\eta^*_{41}f^*_{\gamma 10}+\ldots+\eta^*_{4\sigma(p_2)}f^*_{\gamma^{\sigma(p_2)}}$$

$$=\eta_{11}f^*_{\gamma^1}+\ldots+\eta_{1p}f^*_{\gamma^{p_1}}+(\eta_{21}+\eta_{31}+\eta_{41})f^*_{\gamma^{10}}+\ldots+(\eta_{2q}+\eta_{3q}+\eta_{4q})f^*_{\gamma^{q0}}+\eta_{4\sigma(p_2)}f^*_{\gamma^{\sigma(p_2)}}$$

$$=e_1f_1+\ldots e_rf_r,$$

where $\eta^*_{4b}=\eta_{4b}\sum_{j=0}^{\lambda}\sum_{r=0}^{t}M(\epsilon)x^{t-r}y^{\lambda-j}$, $\eta^*_{3a}=\eta_{3a}\sum_{j=0}^{\lambda}\sum_{r=0}^{t_1}M_1(\epsilon)x^{t_1-r_1}y^{\lambda_1-j}$, $e_1=\eta_{11}$, $e_r=\eta_{4\sigma(p_2)}$, $f_1=f^*_{\gamma^1}$ and $f_r=f^*_{\gamma^{\sigma(p_2)}}$.

Suppose now that $e_1f_1+\ldots e_rf_r=0$ then by virtue of Lemma 1 and 2, $e_1=\ldots e_r=0$ since $\eta_{11},\ldots,\eta_{4\sigma(p_2)}=0$ by identifiability and illustrations above. Thus making the set of functions in $\bar{f}(x,y,\gamma,\gamma_0)$ linearly independent which establishes P1 as required.

Assuming P0 and P1 we invoke theorem 3.2 in [55] as follows by first recalling that $T_n(k)=\sup_{g\in G_k}\ln(g)-\ln(f_0)$ and noting that:

$$T_n(k)\xrightarrow{d}\frac{1}{2}\sup_{d\in D}\xi_d^2 1_{\xi_d\geq 0}$$

and

$$T_n(q) \xrightarrow{d} \frac{1}{2} \sup_{d_0 \in D_0} \xi^2_{d_0} 1_{\xi_{d_0} \geq 0},$$

where D is the subset of the unit sphere of H (Hilbert space) of functions of the form $\frac{1}{N(\theta)}\Big(\sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{D_i^l f_{\gamma^l,0}}{f_0} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\gamma^i}}{f_0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^l,0}}{f_0} \Big)$ and ξ_d is the Gaussian process indexed by D as defined in [55].

So following from theorem 3.6 in [55] we have that:

$$V_n \quad = \quad \sup_{g \in G_k} \ln \frac{g}{f_0} \quad - \quad \sup_{g \in G_q} \ln \frac{g}{f_0} \quad = \quad \ln \frac{\sup_{g \in G_k} \frac{g}{f_0}}{\sup_{g \in G_q} \frac{g}{f_0}} \quad \xrightarrow{d} \quad \frac{1}{2} \sup_{\mu \in \mathcal{U}} \xi^2_{\mu} 1_{\xi_{\mu} \geq 0}$$

It follows from above that $V_n = O_p(1)$, which gives us assumption A1 in Drton and Plummer (2016).

Now we show that A2 in Drton and Plummer (2016) is also satisfied by the mixture of regression model. We begin by adopting the simplistic forms $g = g(x_i, y_i)$ and $f = f(x_i, y_i)$ respectively unless otherwise defined.

We further define the following three models; the first two are considered true models and the last, false model:

$T_n(k) = \sup_{g \in G_k} \sum_{i=1}^{n} \ln(g) - \sum_{i=1}^{n} \ln(f_0)$, $T_n(q) = \sup_{g \in G_q} \sum_{i=1}^{n} \ln(g) - \sum_{i=1}^{n} \ln(f_0)$

$T_n(p) = \sup_{g \in G_p} \sum_{i=1}^{n} \ln(g) - \sum_{i=1}^{n} \ln(f_0)$.

We notice from above that $T_n(k) - T_n(q) = O_p(1)$. In the same spirit of LRT and following the approach adopted in [45] we may state the comparison between a false model $T_n(p)$ and a true model $T_n(k)$ as follows:

$$T_n(p) - T_n(k) = T_n(p) - T_n(q) + T_n(q) - T_n(k) = T_n(p) - T_n(q) - O_p(1). \tag{3.27}$$

We also note by the strong law of large numbers that:

$\frac{1}{n}\sum_{i=1}^{n} \ln(g) \xrightarrow{a.s} E(\ln(g))$ and

$\frac{1}{n}\sum_{i=1}^{n} \ln(f) \xrightarrow{a.s} E(\ln(f))$, under the assumption that $\hat{\pi} \xrightarrow{a.s} \pi$, $\hat{\theta} \xrightarrow{a.s} \theta$ and $\hat{K} \xrightarrow{a.s} K$ (and the parameters spaces are compact). It follows by Slut sky's theorem that:
$\frac{1}{n}T_n(p) - \frac{1}{n}T_n(q) - \frac{1}{n}O_p(1) \xrightarrow{a.s} \sup_{g\in G_p} E\Big(\ln\frac{g}{f}\Big) - \sup_{g\in G_q} E\Big(\ln\frac{g}{f}\Big) = \sup_{g\in G_p} E\Big(\ln(g)\Big) - \sup_{g\in G_q} E\Big(\ln(g)\Big)$.
Recalling from equation 3.28 above we have that:

$\frac{1}{n}T_n(p) - \frac{1}{n}T_n(k) \xrightarrow{a.s} \sup_{g\in G_p} E\Big(\ln(g)\Big) - \sup_{g\in G_q} E\Big(\ln(g)\Big)$

$\leq -\Big[\inf_{g\in G_p}\Big(-\ln\Big(E(g)\Big)\Big) + \sup_{g\in G_q}\ln\Big(E(g)\Big)\Big] = -\Delta$, by Jensen's inequality where $\Delta > 0$ and $p < q$.

Thus we conclude that

$P\Big(\frac{1}{n}T_n(p) - \frac{1}{n}T_n(q) \leq -\frac{\Delta}{2}\Big) \to 1$ for $n \to \infty$, and so

$T_n(p) - T_n(q) = \sup_{g\in G_p}\sum_{i=1}^{n}\ln(g) - \sup_{g\in G_q}\sum_{i=1}^{n}\ln(g) = \ln\frac{\sup_{g\in G_p}\prod_{i=1}^{n} g}{\sup_{g\in G_q}\prod_{i=1}^{n} g} \leq -n\frac{\Delta}{2}$. It follows that

$$\frac{\sup_{g\in G_p}\prod_{i=1}^{n} g}{\sup_{g\in G_q}\prod_{i=1}^{n} g} \leq \exp^{\{-n\frac{\Delta}{2}\}}$$

which is akin to A2 in Drton and Plummer 2016.

Finally A3 in Drton and Plummer 2016 follows immediately from [44] and [45] by recalling that the learning coefficient pertaining to the mixture of regression as discussed earlier is bounded as follows: $\lambda_{ij} \leq \frac{1}{2}[6i + j - 1]\ \forall j < i$. We assumed that the multiplicity $m_{ij} = m_l = m_{eek} = 1$ for some $i, j, k, l \in I$. Denote the lexicographic order on R^2 by $\prec$ and note that for any model indexed by $j, k \in I$ and sub models indexed by $i, l \in I$ for all $i < j < k$ and $l < i < j$ we can easily check that $(\lambda_{ij}, -m_{ij}) \prec (\lambda_{eek}, -m_{eek})$ and likewise $(\lambda_l, -m_l) \prec (\lambda_{ij}, -m_{ij})$. For instance for any model with indices $j = 3, k = 4 \in I$ and sub models with indices $l = 1, i = 2 \in I$, it is obvious that $(\lambda_{32}, -1) \prec (\lambda_{42}, -1) \iff (18/2, -1) \prec (25/2, -1)$ and similarly $(\lambda_{31}, -1) \prec (\lambda_{32}, -1) \iff (9, -1) \prec (19/2, -1)$ respectively.

Chapter 4 A Singular Flexible Information Criterion From A Mixture of Linear Regressions Perspective

4.1 Introduction

In this chapter we will establish a novel model selection criterion for mixture of regression models called the SFLIC (Singular Flexible Information Criterion). In the more basic setting of an ordinary mixture, SFLIC is a hybrid between FLIC[27] and sBIC [44]. The SFLIC developed in this work is methodologically different from that in [45], in that the modeling framework here is a mixture of regression model contrary to that of [45] which is a heirarchical mixture model setting, making the derivation approaches very different. The following steps will be traversed to fully develop this new criterion.

1. First we will identify a penalty that will increase or decrease depending on whether there is apparent homogeneity or heterogeneity in the mixture problem at hand. The chosen penalty will also possess the ability to sandwich the criterion between some singular versions of AIC and BIC (sAIC and sBIC if you will); so that when the criterion is very liberal it will bear the mark of sAIC and when it is very conservative it will resemble the sBIC. We may define the sAIC as the AIC for a singular model (i.e. AIC with a learning coefficient instead of the number of parameters). In other words sAIC may be defined as

$2 * loglik - 2 * \lambda$ and sBIC may be defined as $2 * loglik - 2 * \lambda * \log(n)$. Both sAIC and sBIC, although slightly different from AIC and BIC, account for singularity of the model.

2. Secondly, we will adapt Pilla and Charnigo's(2007) bivariate function for a vector outcome (contrary to the scalar outcome in Pilla and Charnigo(2007)). This function was chosen because as part of the penalty, it will exhibit the characteristics described above.

3. Thirdly, since our goal is to create a criterion that works in the general settings work as described by Drton and Plummer(2016), we will adopt that general setting but replace the penalty in Drton and Plummer (2016) with the penalty established here and call it SFLIC. Importantly Drton and Plummer(2016) did not consider specifically a mixture regression model.

4. Fourthly, we examine the statistical properties of the new criterion in regards to determine if it converges in probability to the correct order. That is, does $\hat{m} := argmax_{m \in \{1,2,..,M\}} SFLIC_m \xrightarrow{p} m_0$ where m_0 is the true order of the model?

4.2 Deduction of Within and Between Covariance Matrices

To address objective one regarding penalty formulation, we will appeal to the underlying concept of ANOVA and compare component specific fitted models to component specific outcomes on one hand; and on the other, compare the component specific fitted models to a weighted model fitted to all the components. The former comparison will be the within variance covariance structure and the latter, the between

variance covariance structure.

Within Variance Covariance Matrix Derivation

By definition the residual of a fitted regression model is the difference between the observed and the predicted outcomes. For the purposes of our mixture regression problem we more explicitly represent the residual for rtaubeta on race as: $\widehat{e_{i1j}} = Y_{i1j} - (\widehat{\beta_{01j}} + \widehat{\beta_{11j}} X_i)$ conditional on being in component j and similarly represent the residual resulting for rptaubeta on race as $\widehat{e_{i2j}} = Y_{i2j} - (\widehat{\beta_{02j}} + \widehat{\beta_{12j}} X_i)$. (This can be generalized to multiple covariates.)

In essence $\widehat{e_{i1j}}$ will measure the distance from Y_{i1j} to $\widehat{Y_{i1j}}$ if we know that subject i is in component j, where $i = 1, 2, ..., n$ and $j = 1, 2, .., m$, m being the number temporarily assumed known mixture components in the model. Thus $i1j$ will correspond to the i^{th} observation for outcome one (in this case rtaubeta) if assumed in the j^{th} mixture component.

Using the definition above for residuals we can estimate the hard and soft classification aggregated within variances (similar to generalized variance in a multivariate setting) as follows:

Hard Classification

$$\mathbf{W}_{hm} := \sum_{j=1}^{m} \left[\sum_{i}^{n} 1_{ij} \widehat{e_{i1j}^2} \sum_{i}^{n} 1_{ij} \widehat{e_{i2j}^2} - \left(\sum_{i}^{n} 1_{ij} \widehat{e_{i1j}} \widehat{e_{i2j}} \right)^2 \right] \tag{4.1}$$

where $1_{ij} = 1_{\widehat{P_{ij}} = \max_{1 \le k \le m} \widehat{P_{eek}}}$. In words, the indicated condition says that for person

i, the conditional probability of belonging to component j is greater than or equal to the conditional probability of belonging to component k for any k. $\mathbf{W}_{hm}$ is the summation of component specific determinants.

Soft Classification

$$\mathbf{W}_{sm} := \sum_{j=1}^{m} \left[\sum_{i}^{n} \widehat{P_{ij}} \widehat{e_{i1j}^2} \sum_{i}^{n} \widehat{P_{ij}} \widehat{e_{i2j}^2} - \left(\sum_{i}^{n} \widehat{P_{ij}} \widehat{e_{i1j}} \widehat{e_{i2j}} \right)^2 \right] \tag{4.2}$$

where $\widehat{P_{ij}}$ is the posterior probability of individual i belonging to component j given all the information we know about this individual.

Between Variance Covariance Matrix Derivation

Here we aim to examine the variation between each component's fitted model and the weighted average of all the models fitted to the various components. This will quantify how different the model fitted to a given component compares to the weighted average of the rest. We begin by drawing analogy to one way ANOVA and define the subject specific contribution to between variance in the context of mixture of linear regressions as follows:

The between variance resulting from assignment of subject i to component j regarding rtaubeta is:

$$\widehat{B_{i1j}^2} := \left[(\widehat{\beta_{01j}} + \widehat{\beta_{11j}} X_i) - \sum_{l}^{m} \widehat{\pi_l} (\widehat{\beta_{01l}} + \widehat{\beta_{11l}} X_i) \right]^2 \tag{4.3}$$

where $\widehat{\pi_l}$ is the proportion of membership in component l and the between variance from rptaubeta is

$$\widehat{B^2_{i2j}} := \left[(\widehat{\beta_{02j}} + \widehat{\beta_{12j}} X_i) - \sum_l^m \widehat{\pi_l} (\widehat{\beta_{02l}} + \widehat{\beta_{12l}} X_i) \right]^2 \tag{4.4}$$

where m is the number of components.

Verbally we may interpret $\widehat{B_{i1j}}$ as the distance from $\widehat{Y_{i1j}}$ knowing that subject i is in component j from $\widehat{Y_{i1.}}$ if we don't know to which component subject i belongs, where $\widehat{Y_{i1.}} = \sum_l^m \widehat{\pi_l} (\widehat{\beta_{02l}} + \widehat{\beta_{12l}} X_i)$.

The corresponding hard and soft classification summation of component specific determinants are as follows:

Hard Classification

$$\mathbf{B}_{hm} := \sum_{j=1}^{m} \left[\sum_i^n 1_{ij} \widehat{B^2_{i1j}} \sum_i^n 1_{ij} \widehat{B^2_{i2j}} - \left(\sum_i^n 1_{ij} \widehat{B_{i1j}} \widehat{B_{i2j}} \right)^2 \right] \tag{4.5}$$

Soft Classification

$$\mathbf{B}_{sm} := \sum_{j=1}^{m} \left[\sum_i^n \widehat{P_{ij}} \widehat{B^2_{i1j}} \sum_i^n \widehat{P_{ij}} \widehat{B^2_{i2j}} - \left(\sum_i^n \widehat{P_{ij}} \widehat{B_{i1j}} \widehat{B_{i2j}} \right)^2 \right], \tag{4.6}$$

where $\widehat{P_{ij}}$ is as previously defined.

4.3 Definition and Derivation of SFLIC

Having established aggregated within and between variances for the mixture of linear regression model, we proceed to define a statistic that will be used to later describe the degree of heterogeneity in the fitted mixture models.

Define

$$\tau(\mathbf{Y}) := \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbf{W}_k(\mathbf{Y})}{\mathbf{B}_k(\mathbf{Y}) + \mathbf{W}_k(\mathbf{Y})} \tag{4.7}$$

where $\boldsymbol{W}_k(Y)$ is either $\boldsymbol{W}_{kh}(Y)$ or $\boldsymbol{W}_{ks}(Y)$ and likewise for $\boldsymbol{B}_k(Y)$.

We are thus averaging $\frac{\boldsymbol{W}_k}{\boldsymbol{B}_k+\boldsymbol{W}_k}$ ratio over all models under consideration. It follows that large(small) values of $\tau(\mathbf{Y})$ may be indicative of more homogeneity(heterogeneity) in the data.

According to Pilla and Charnigo (2007) we can define a bivariate function such that:

$$B(n, \tau(\mathbf{Y})) = \frac{\Phi((\log(\sqrt{n})^{\tau(\mathbf{Y})})) - \Phi(1)}{1 - \Phi(1)} \tag{4.8}$$

Using the bivariate function above we define the singular flexible information criterion (SFLIC) as follows inspired by Drton(2016):

$$SFLIC_k := 2\log P[Y_n|\pi_0, M_k] - 2\lambda_k(\pi_0)\log(n)^{B(n,\tau(\mathbf{Y}))} \tag{4.9}$$

also assuming a multiplicity factor of 1 as in Drton (2016) and Fan(2014) where the terms in 4.11 are similar to those defined in chapter two under 'Review of related concepts'.

We deduce the following observations from the SFLIC:

1) If $n \to \infty$ then $B(n, \tau(\mathbf{Y})) \overset{a.s}{\to} 1$ because as $n \to \infty$

$$B(n, \tau(\mathbf{Y})) = \frac{\Phi((\log(\sqrt{n})^{\tau(\mathbf{Y})})) - \Phi(1)}{1 - \Phi(1)} \geq \frac{\Phi((\log(\sqrt{n})^{\frac{1}{M}})) - \Phi(1)}{1 - \Phi(1)} \overset{a.s}{\to} 1$$

and so SFLIC becomes

$$SFLIC_k = 2\log P[Y_n|\pi_0, M_k] - 2\lambda_k(\pi_0)\log(n)$$

which is akin to sBIC.

2) If $\tau(\mathbf{Y})$ is small then $B(n, \tau(\mathbf{Y})) \approx 0$ because

$$\Phi((\log(\sqrt{n})^{\tau(\mathbf{Y})})) \approx \Phi(1) \implies \frac{\Phi(1) - \Phi(1)}{1 - \Phi(1)} = 0$$

. Thus SFLIC will be approximately

$$SFLIC_k = 2\log P[Y_n|\pi_0, M_k] - 2\lambda_k(\pi_0),$$

which is akin to sAIC or how the AIC might be defined for a singular model.

Thus the SFLIC is sandwiched between sBIC and sAIC and drifts to the former as n tends to infinity.

4.4 Consistency of SFLIC

In a similar analogy to Drton (2016) we consider a finite set of true models $M_i : i \in I$ and a fixed data generating distribution $\pi_0 \in \bigcup_{i \in IM_i}$. M_i is true if $\pi_i \in M_i$ else M_i is false. A smallest true model M_i is a true model whose sub models are all false model. That is if $j \prec i \Rightarrow \pi_0 \notin M_j$.

M_i is said to have a smaller Bayes complexity than M_j if $(-\lambda_i(\pi_0), M_i(\pi_0)) < (-\lambda_j(\pi_0), M_j(\pi_0))$ for $\pi_0 \in M_i$. This is equivalent to $\lambda_i(\pi_0) > \lambda_j(\pi_0)$. Of note the Bayes factor is defined as $n^{\lambda_i(\pi_0)}(\log n)^{m_i(\pi_0)-1}$ where $\lambda_i(\pi_0)$ is the learning coefficient and $m_i(\pi_0)$ is its corresponding multiplier. The former and latter together describes the complexity of model M_i under the data-generating distribution π_0.

The following assumptions have been shown to be consistent with the proposed mixture of regression models.

Assumptions proposed by Drton (2016)

A1) for any two true models M_i and M_j

$$\frac{P(Y_n|\hat{\pi}_k, M_k)}{P(Y_n|\hat{\pi}_i, M_i)} = Op(1)$$

A2) For any pair of a true model M_i and false model M_k $\exists$ a constant $\delta_{eek} > 0$ such that

$$\frac{P(Y_n|\hat{\pi}_k, M_k)}{P(Y_n|\hat{\pi}_i, M_i)} \leq e^{-\delta_{eek} n}$$

as $n \to \infty$

A3) Let M_i and M_k be any two true models such that $j \preceq i$ and $l \preceq k$ index any two respective sub models. Then the Bayes complexity is monotonically increasing in the sense $(-\lambda_{ij}, m_{ij}) < (-\lambda_{ij}, m_{kl})$ if $i \prec k$ and $j \preceq l$.

Theorem 4.1 (Consistency): Ler M_i be the model selected by maximizing the SFLIC, that is

$$\hat{i} = argmax_{i \in I} SFLIC(M_i). \tag{4.10}$$

Then under assumptions A1-A3, the probability that M_i is a true model of minimal Bayes complexity and thus the smallest true model tends to one as n goes to infinity.

it suffices to show that:

1. The SFLIC of any true model is asymptotically larger than that of any false model.

2. SFLIC of a true model can be asymptotically maximal only if the model minimizes Bayes complexity among the true models.

Proposition 4.1

Under assumption (A2) above, if model M_i' is true and model M_k' is false then the probability that

$$SFLIC(M_i') > SFLIC(M_k') \to 1, n \to \infty \tag{4.11}$$

To show prove proposition 4.1 we fix $j' \preceq i'$ and $l' \preceq k'$ and let M_k' and M_i' be respectively false and true models. Then according to assumption A2 in Drton(2016)

we have that

$$\frac{P(Y_n|\hat{\pi}'_k, M'_k)}{P(Y_n|\hat{\pi}'_i, M'_i)} < e^{-\delta_{i'k'n}}.$$

Thus there exist $\epsilon > 0$ such that

$$P\left[\left|\frac{P(Y_n|\hat{\pi}'_k, M'_k)}{P(Y_n|\hat{\pi}'_i, M'_i)} - e^{-\delta_{i'k'n}}\right|\right] \to 0, n \to \infty \tag{4.12}$$

Hence following the definition of consistency we have that

$$\begin{aligned}
&P(|SFLIC_k - SFLIC_i| < \epsilon) \\
&\quad = P(|P(Y_n|\hat{\pi}'_k, M'_k) - P(Y_n|\hat{\pi}'_i, M'_i) - \log(n)^{B(n,\tau(y))}(\lambda_k(\pi_0) - \lambda_i(\pi_0))| < \epsilon) \\
&\quad \leq P(|P(Y_n|\hat{\pi}'_k, M'_k) - P(Y_n|\hat{\pi}'_i, M'_i)| < \epsilon/2) + P(|\log(n)^{B(n,\tau(y))}(\lambda_k(\pi_0) - \lambda_i(\pi_0)| < \epsilon/2)
\end{aligned}$$

But

$$P(|\log(n)^{B(n,\tau(y))}(\lambda_k(\pi_0) - \lambda_i(\pi_0)| < \epsilon/2) = P\left(\left|\log(n)^{B(n,\tau(y))}\right| < \frac{\epsilon/2}{\lambda_k(\pi_0) - \lambda_i(\pi_0)}\right) \to 0, n \to \infty \tag{4.13}$$

and

$$\begin{aligned}
&P(|P(Y_n|\hat{\pi}'_k, M'_k) - P(Y_n|\hat{\pi}'_i, M'_i)|\epsilon/2) = P(|Op(P(Y_n|\hat{\pi}'_i, M'_i)) - P(Y_n|\hat{\pi}'_i, M'_i)| < \epsilon/2) \\
&= P(|P(Y_n|\hat{\pi}'_i, M'_i)(Op(1) - 1)| < \epsilon/2) = P\left(|P(Y_n|\hat{\pi}'_i, M'_i)| < \frac{\epsilon/2}{(Op(1) - 1)}\right) \to 0, n \to \infty
\end{aligned} \tag{4.14}$$

As a result

$$P(|SFLIC_k - SFLIC_i| < \epsilon) \to 0, n \to \infty \tag{4.15}$$

4.5 Application of SFLIC to the ADNI data

The SFLIC was applied to the ADNI data, specifically to the mixture of regression model with race as covariate. The SFLIC favored a 2 component ($SFLIC \approx 746.8$) mixture model slightly over a three component ($SFLIC \approx 743.5$) and in the apoe4 mixture model, SFLIC selected two components ($SFLIC \approx 747.5$) as opposed to three components ($SFLIC \approx 744.6$). When the SFLIC was applied to the race and apoe4 mixture of regression model, it once again favored a two component ($SFLIC \approx 746.5$) to a three component ($SFLIC \approx 743.4$). Thus in all three models SFLIC (similar to SBIC considering the model with more stable standard errors) selected the number of components that produced stable standard errors as seen in chapter 3.

Chapter 5 Simulation Studies

5.1 Introduction

Mixture modeling applications have been well received in many fields for identifying subgroups underlying a given population in a non-parametric manner. For instance mixture modeling has been applied in market response models and multidimensional scaling Sarstedt and Schaiger(2008). Andrews and colleagues (2002) also identified finite mixture modeling as a comparable model to the well received hierarchical Bayes conjoint analysis models in terms of model fit, prediction and robustness with regards to individuals decision making in market research. Crawford et. al(2012) recently applied mixture models to classify lake chemistry distributions into lake sub population.

5.2 Overview of Approach

The foregone background suggests that indeed mixture modeling is widely used as a tool in many fields to address varied problems and to make important decisions. But to be able to make a well informed decisions based on this modeling approach, it is imperative for one to identify the correct number of heterogeneity underlying the population on interest. To this end, and as we have already elaborated in the previous chapters, AIC and BIC are popular in this regard. Drton and colleagues (2016) have also added sBIC which is both suitable for modeling in the presence of

identifiability issues and tends to be less extreme in comparison to AIC and BIC. Furthermore, SFLIC was developed specifically to be able to address model selection problems in mixture of regressions in the presence of identifiability issues. However no study to the best of our knowledge has compared the performance of these four criteria in regards to their ability to correctly identify the heterogeneity in the data. The importance of a correct identification of the different segments underlying the population from which the data are obtained, is invaluable to reach reasonable decisions from any analysis (Sarstedt et. al (2008).

As a result, the purpose of this chapter is to conduct simulation study to compare the performance of the novel model selection criterion developed in chapter four to AIC, BIC and SBIC. In our quest we will further compare the performance of each of the criteria to random chance, proportional chance and maximum chance criteria as suggested by Sarstedt and Schaiger(2008). In particular our study will seek to address the following goals:

1. For a known number of mixing component in a mixture of regression model how well does SFLIC perform in comparison to AIC, BIC and sBIC.

2. For varying sample sizes and known number of mixing component in a mixture of regression model how well does SFLIC perform in comparison to AIC, BIC and sBIC.

5.3 Simulation Design and Results

Three simulations were conducted namely; the race mixture of regression model simulation, the Apoe4 mixture of regression model simulations and the race and Apoe4 mixture model simulations. Each simulations followed the steps outlined below:

1. The covariates were drawn from Bernoulli distribution with a prespecified probability of 0.9 for race and 0.27 for Apoe4. These probabilities were determined from the original data.

2. The biomarker ratios were also drawn from two component normal mixture of regression.

3. FlexMix package[41] in R was utilized to obtain a finite mixture of regression with one or two covariates depending on the model. The number of components k was varied from 2 to 4 and SFLIC, AIC, BIC and sBIC were used to select the correct number of components which are called their success rates.

4. The simulation exercise was repeated for 7 sample sizes from 500 to 5000 in an uneven increment and 8000.

5. The performances of the model selection criteria were displayed in a success rate by sample size graph. The simulation size was fixed at $B = 50$. This simulation size was chosen to make the process less expensive regarding computational memory.

Simulation Results: race mixture model

The results of the race mixture model simulation shows the following:

1. All the model selection criteria performed better than the random chance criterion (0.33) and the proportional chance ($0.3^2 + 0.3^2 + 0.4^2 = 0.34$).

2. Overall BIC was sub optimal in comparison.

3. SFLIC and AIC performed slightly better than sBIC.

4. For sample sizes less than or equal to 4000 but greater than 2000, AIC and SFLIC performed about the same.

5. For sample sizes below 3000 SFLIC performed the slightly better than sBIC and AIC.

6. For sample larger than 4000, AIC performed slightly better than SFLIC.

Simulation Results: Apoe4 mixture model

The results of the Apoe4 mixture model simulation shows the following:

1. All the model selection criteria performed better than the random chance criterion (0.33) and the proportional chance ($0.3^2 + 0.3^2 + 0.4^2 = 0.34$).

2. Overall BIC was sub optimal in comparison.

3. Overall SFLIC and AIC performed slightly better than sBIC.

4. For sample sizes from 2000 to 4500, the SFLIC performed and AIC performed about the same.

5. For sample larger than 4500, AIC performed slightly better than SFLIC.

Simulation Results: race and Apoe4 mixture model

The results of the race and Apoe4 mixture model simulation shows the following:

1. All the model selection criteria performed better than the random chance criterion (0.33) and the proportional chance ($0.3^2 + 0.3^2 + 0.4^2 = 0.34$).

2. Overall BIC was sub optimal in comparison.

3. Overall SFLIC and AIC performed slightly better than sBIC for smaller sample sizes.

4. For sample sizes between to 2000 and 2500 SFLIC performed slightly better than AIC.

5. For sample sizes above 2500 SFLIC performed the slightly better than sBIC and AIC.

6. For sample larger than 2500, SFLIC, AIC and sBIC all performed about the same.

7. All four criteria had a success rate at or better than 37% and they all seem to perform compartively better in the race only simulations.

8. None of the criteria achieved 100% success rate partly because the sample size increases, the penalty grows logarithmically (refer to chapter 4) and thus SFLIC and SBIC become very conservative behaving more like BIC.

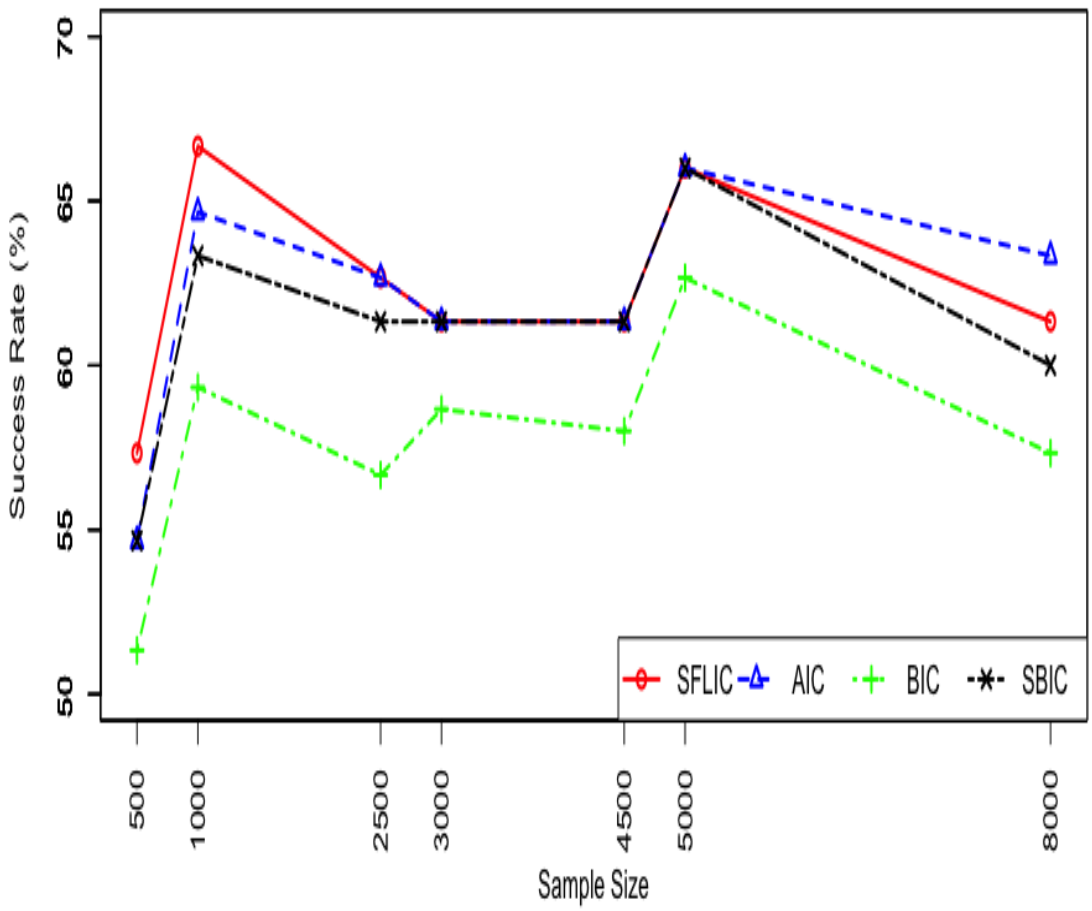

Figure 5.1: Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the race mixture of regression model. Here the true mixture is k=2

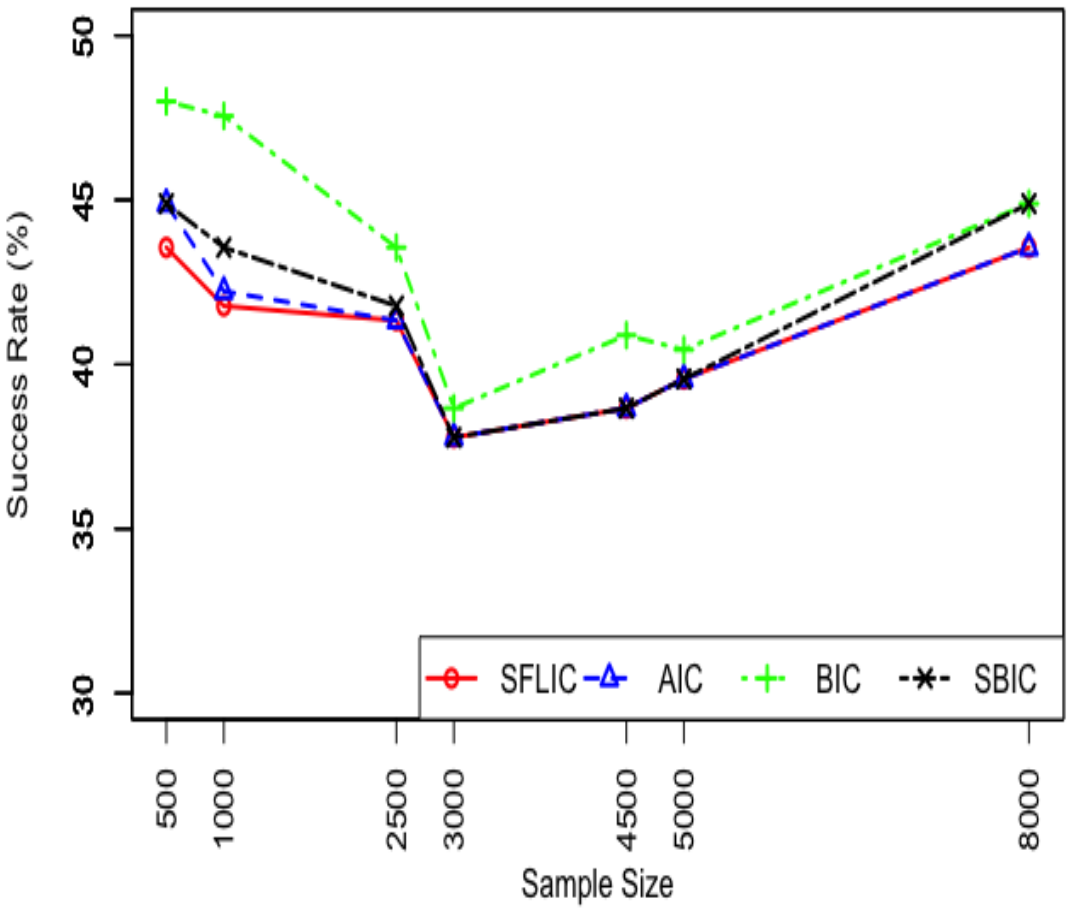

Figure 5.2: Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the apoe4 mixture of regression model. Here the true mixture is k=2

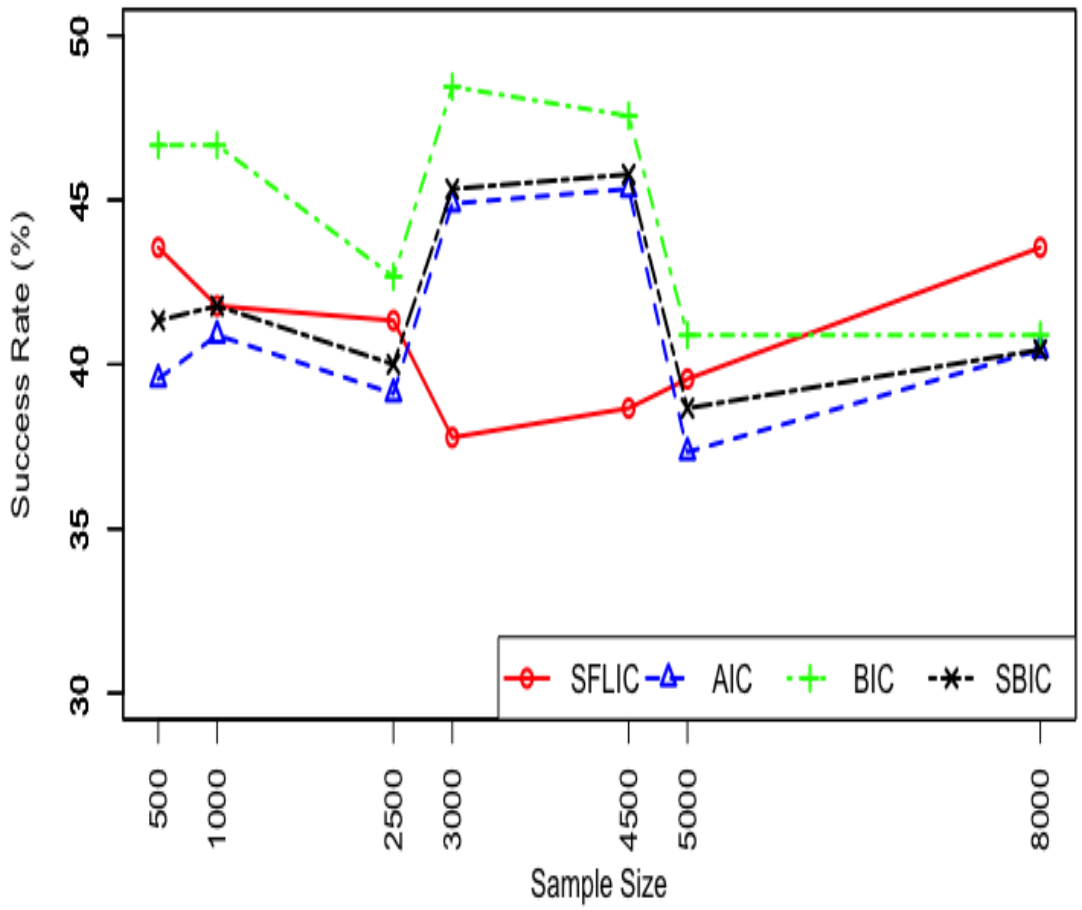

Figure 5.3: Simulation comparing the success rates of SFLIC, AIC, BIC and SBIC with respect to the race and apoe4 mixture of regression model. Here the true mixture is k=2

Chapter 6 Supplementary Chapter

6.1 Introduction

This section addresses models that were fitted that were comparable to those discussed in this work however fell short on some key elements. Below we addressed each model and provide substantive reasons for their exclusion in the main work.

6.2 Review of Other Comparable Models Fitted As Part of This Work

1. The use of the raw CSF biomarkers have been used by De Meyer and colleagues (2010) to predict AD. They also used Abeta/Ptau ratio. The difference however lies in their objective, which was to identify AD patterns in an independent, and unsupervised way. This also influenced the data they used.

2. When we fitted Abeta and ptau, AIC and sBIC choose three components model against a four component model selected by BIC. When the grouping probabilities were fitted in the Cox reg model we obtained an unadjusted $c-stats = 66\%$ and adjusted $c = 69.5\%$ (results not shown). In this case the unadjusted c-statistic is the c-statistic resulting from fitting the Cox model with only the grouping probabilities. Adjusting the Cox model for covariates results in the adjusted c-statistic. So comparatively, this model is sub optimal to the one created with the ratios in terms of the c-statistic. We noted also that just as in the ratio model, this model predicts being of white race as protective

with a reduced risk of transitioning. Also the posterior plots indicated lots of misclassifications in the model. In addition the medium risk was no longer significant after accounting for other risk factors. This suggests that the posterior probabilities from the raw biomarker model may not possess the same predictive abilities as that of the ratios. Indeed it is possible that accounting for an appropriate risk could wipe the effect of the grouping probabilities entirely. If this were the case, then mixture modeling apporach may not be worthwhile, however, the identified appropriate risk factors will serve the interest of the clinician whose key interest is to provide cure without endulging in complicated metholodolgies such as mixture modeling.

3. Putting all three biomarkers in the model resulted in a c-statistic of 58.4% (unadjusted) and 69.4% (adjusted) for the hard classification model (adjusted 67.5%) and (unadjusted 62.4%) for the soft classification model. When we adjusted both hard and soft classification model for RAVLT, the c-statistics were respectively 71% and 72% (results not shown). The estimated group probabilities also had larger standard errors, which resulted in wider confidence intervals. Again this model is sub optimal in comparison to the ratio model. The group probability failed the proportional hazard test in the presence of the other risk factors.

4. We conclude that the raw biomarkers may not possess the optimal grouping probabilities needed for predicting future cognitive status of people who are cognitively intact.

5. Using pca was not helpful in predicting either. In addition pca loses meaning due to the fact that it is a linear combination of the predictors (in this case the biomarkers) and we're not sure what the linear combination of the biomarkers really means.

6. The ratio biomarkers are enhanced to identify risks in the sense that if we keep the numerators tau and ptau fixed and reduce the denominator abeta, then the entire fraction will be enhanced and thus provide an indicator of high risk of transitioning. On the other hand, using just the raw biomarkers may not be enhanced enough to capture the potential transitioning. In essence the ratio is accounting for the effect of abeta indirectly and incorporating it in the modeling procedures.

7. Also the use of ratios affords us the flexibility of having two derived quantities that move in the same direction in terms of low , medium or high risks. That is the ratios if low then low risk and if high then high risk is preferred to an inverse one in using something like abeta and tau or abeta and ptau.

1. Blalock, E.M., et al., Gene Microarrays in hippocampal aging: Statistical profiling identifies novel processes correlated with cognitive impairment. Journal of Neuroscience, 2003. 23(9): p. 3807-3819.

2. Albert, M.S., et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement, 2011. 7(3): p. 270-9.

3. Mclachlan, G.J., On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture. Applied Statistics-Journal of the Royal Statistical Society Series C, 1987. 36(3): p. 318-324.

4. Chen, H.F. and J.H. Chen, The likelihood ratio test for homogeneity in finite mixture models. Canadian Journal of Statistics-Revue Canadienne De Statistique, 2001. 29(2): p. 201-215.

5. Chen, H.F., J.H. Chen, and J.D. Kalbfleisch, A modified likelihood ratio test for homogeneity in finite mixture models. Journal of the Royal Statistical Society Series B-Statistical Methodology, 2001. 63: p. 19-29.

6. Charnigo, R. and J.Y. Sun, Testing homogeneity in a mixture distribution via the L-2 distance between competing models. Journal of the American Statistical Association, 2004. 99(466): p. 488-498.

7. Davies, R.B., Hypothesis Testing When a Nuisance Parameter Is Present Only under Alternative. Biometrika, 1977. 64(2): p. 247-254.

8. Neyman, J. and E. Scott, On the Use of C(Alpha) Optimal Tests of Composite Hypotheses. Bulletin of the International Statistical Institute, 1965. 41(1): p. 477-

497.

9. De Meyer, G., et al., Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. Arch Neurol, 2010. 67(8): p. 949-56.

10. Hampel, H., et al., Core biological marker candidates of Alzheimer's disease - perspectives for diagnosis, prediction of outcome and reflection of biological activity. J Neural Transm, 2004. 111(3): p. 247-72.

11. Montine, T.J., et al., National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. Acta Neuropathol, 2012. 123(1): p. 1-11.

12. Gustafson, D.R., et al., Cerebrospinal fluid beta-amyloid 1-42 concentration may predict cognitive decline in older women. J Neurol Neurosurg Psychiatry, 2007. 78(5): p. 461-4.

13. Pearson, k., Contributions to the mathematical theory of evolution. Philosophical Transactions, 1893-1895(A186, A185): p. 342-414, 71-110.

14. Newton, M.A., et al., Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics, 2004. 5(2): p. 155-76.

15. Stomrud, E., et al., Cerebrospinal fluid biomarkers predict decline in subjective cognitive function over 3 years in healthy elderly. Dement Geriatr Cogn Disord, 2007. 24(2): p. 118-24.

16. D. M. Titterington, A.F.M.S.a.U.E.M., Statistical Analysis of Finite Mixture Distribution. 1985.

17. Lindsay, B.G., Mixture Models: Theory, Geometry and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics, 1995. 5.

18. Brown, L.D., Fundamental of Statistical Exponential Families with Application in Statistical Decision Theory. 1986.

19. Peel, G.M.a.D., Finite Mixture Models. Applied Probabiity and Statistics, ed. W. series. 2001, Canada: John Wily and Sons Inc

20. Zhou, C. and J. Wakefield, A Bayesian mixture model for partitioning gene expression data. Biometrics, 2006. 62(2): p. 515-25.

21. Alexandridis, S.L.a.R., Classification of Tissue Sample Using Mixture Modeling of Microarray Gene Expression Data. institute of mathematical Statistics, 2003: p. 419-435.

22. Weuve, J., et al., Deaths in the United States among persons with Alzheimer's disease (2010-2050). Alzheimer's Dement, 2014. 10(2): p. e40-6.

23. Schonknecht, P., et al., Increased tau protein differentiates mild cognitive impairment from geriatric depression and predicts conversion to dementia. Neurosci Lett, 2007. 416(1): p. 39-42.

24. Hebert, L.E., et al., Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. Neurology, 2013. 80(19): p. 1778-83.

25. Association, T.A.s., Alzheimer's Disease-Facts and Figures. 2014, The Alzheimer's Association.

26. Peter Santago, H.D.G., Quantification of MR Brain Images by Mixture Density and Partial Volume Modeling. IEE TRANSACTION MEDICAL IMAGING, 1993. 12.

27. Charnigo, R., et al., Thinking outside the curve, part I: modeling birthweight distribution. Bmc Pregnancy and Childbirth, 2010. 10.

28. Michael A Newton, P.W.a.C.K., Hierarchical Mixture Models for Expression Profiles. 2006.

29. Speed, I.L.a.T., Replicated Microarray Data. Statistica Sinica, 2002. 12: p. 21-46.

30. Dai, H.Y. and R. Charnigo, Omnibus testing and gene filtration in microarray data analysis. Journal of Applied Statistics, 2008. 35(1): p. 31-47.

31. Umbach, D.M. and A.J. Wilcox, A technique for measuring epidemiologically useful features of birthweight distributions. Stat Med, 1996. 15(13): p. 1333-48.

32. Gage, T.B. and G. Therriault, Variability of birth-weight distributions by sex and ethnicity: analysis using mixture models. Hum Biol, 1998. 70(3): p. 517-34.

33. Cui, X. and G.A. Churchill, Statistical tests for differential expression in cDNA microarray experiments. Genome Biol, 2003. 4(4): p. 210.

34. Bar, H.Y., J.G. Booth, and M.T. Wells, A Bivariate Model for Simultaneous Testing in Bioinformatics Data. Journal of the American Statistical Association, 2014. 109(506): p. 537-547.

35. Shaw, L.M., et al., Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol, 2009. 65(4): p. 403-13.

36. Toombs, J., et al., Identification of an important potential confound in CSF AD studies: aliquot volume. Clin Chem Lab Med, 2013. 51(12): p. 2311-7.

37. Center for Disease Control and Prevention (2015). "Healthy Aging." 2015, from http://www.cdc.gov/aging/aginginfo/Alzheimer's.htm.

38. Alzheimer's Association, 2015. "Alzheimer's Facts and Figures." from http://www.alz.org

39. Folstein, M.F., L.N. Robins, and J.E. Helzer, The Mini-Mental State Examination. Arch Gen Psychiatry, 1983. 40(7): p. 812.

40. O'Bryant, Sid E., et al. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. Archives of neurology 65.8 (2008): 1091-1095.

41. Grun, B. and F. Leisch, FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters. Journal of Statistical Software, 2008. 28(4): p. 1-35.

42. Benaglia, T., et al., mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software, 2009. 32(6): p. 1-29.

43. Shadlen, M.F., et al., Education, cognitive test scores, and black-white differences in dementia risk. J Am Geriatr Soc, 2006. 54(6): p. 898-905.

44. Drton, M. and Plummer, M. (to appear), A Bayesian Information Criterion for Singular Models. JRSS-B, 2016.

45. Fan, Qian, "Normal Mixture and Contaminated Model With Nuisance Parameter and Applications" (2014). Theses and Dissertations–Statistics.

46. Viele, K. and Tong, B. (2002) Modeling with mixtures of linear regressions, Statistics and Computing, 12: p. 315 – 330

47. ADNI. Retrieved January, 2015, 2015, from www.ida.loni.usc.edu.

48. Craig-Schapiro, R., et al., Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimer's disease diagnosis and prognosis. PLoS One, 2011. 6(4): p. e18850.

49. Ewers, M., et al., Prediction of conversion from mild cognitive impairment to

Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging, 2012. 33(7): p. 1203-14.

50. Landau, S.M., et al., Comparing predictors of conversion and decline in mild cognitive impairment. Neurology, 2010. 75(3): p. 230-8.

51. Vemuri, P., et al., MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology, 2009. 73(4): p. 294-301.

52. Hurd, M.D., et al., Monetary costs of dementia in the United States. N Engl J Med, 2013. 368(14): p. 1326-34.

53. Shadlen, M. F., et al. (2006). "Education, cognitive test scores, and black-white differences in dementia risk." J Am Geriatr Soc 54(6): 898-905.

54. Glymour, M. M., et al. (2008). "Lifecourse social conditions and racial disparities in incidence of first stroke." Ann Epidemiol 18(12): 904-912.

55. Dacunha-Castelle, D. and E. Gassiat, Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. Annals of Statistics, 1999. 27(4): p. 1178-1209.

56. Gassiat, E., Likelihood ratio inequalities with applications to various mixtures. Annales De L Institut Henri Poincare-Probabilites Et Statistiques, 2002. 38(6): p. 897-906.

57. Sarstedt, Marko, and Manfred Schwaiger. "Model selection in mixture regression analysis-a Monte Carlo simulation study." Data analysis, machine learning and applications. Springer Berlin Heidelberg, 2008. 61-68.

58. Andrews, Rick L., Asim Ansari, and Imran S. Currim. "Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and part-

worth recovery." Journal of Marketing Research 39.1 (2002): 87-98.

59. Crawford, Sybil L., et al. "Modeling lake-chemistry distributions: Approximate Bayesian methods for estimating a finite-mixture model." Technometrics 34.4 (1992): 441-453.

60. Keribin, Christine. "Consistent Estimation Of The Order Of Mixture Models". the Indian Journal of Statistics 1.49-66 (2000): n. pag. Print.

61. Yakowitz, Sidney J. and John D. Spragins. "On The Identifiability Of Finite Mixtures". Ann. Math. Statist. 39.1 (1968): 209-214. Web.

Vita

PhD. Candidate, Epidemiology and Biostatistics, University of Kentucky, May 2017

MS. Statistics, University of Kentucky, May 2012

Grad. Cert. Data Mining, University of Louisville, January 2010

MS. Mathematics, Youngstown State University, May 2007

B.Ed. Mathematics, University of Cape Coast, July 2001

Minors: Physics and Education

Oral Presentations (Delivered by First author)

K.Obeng, P. Ignaciuk, J. Kim, **Frank Appiah**, N. Darboe, E. Escott, Evaluation of CT the Laryngopharyngeal Structures Using Quiet Respiration Versus Dynamic 'eee' Phonation, European Congress of Radiology, Vienna, Austria, 2016.

Appiah Frank, Abner E, Fardo D, Mays Glen, Charnigo R, 'Predicting Alzheimer's Disease with Bivariate Mixture Modeling, An Application to ADNI Data', ENAR, Austin, TX, March 2016.

Appiah Frank, Abner Erin, 'Trajectories Identifying Robust-Normal, Prodromal-Normal and Not-Normal Participants Using CERAD T-scores and Premorbid Estimates of Cognitive Ability', ASA-KY Branch, Lexington, KY, March 2016.

Appiah Frank, Abner Erin, 'Trajectories Identifying Robust-Normal, Prodromal-Normal and Not-Normal Participants Using CERAD T-scores and Premorbid Estimates of Cognitive Ability', 11th Annual CCTS Spring Conference- Lexington, KY, April 2016.

Appiah Frank, Fardo D, Abner E, Mays G, Charnigo R, 'Predicting Alzheimer's Disease with Mixture of Regression Modeling', Joint Statistical Meetings, Chicago, IL, August 2016.

Li X, Chakraborty AK, Landwehr KP, **Appiah F**, Kingsbury AR, Hobbs SB, Winkler MA, 'Effects of Different Intravenous Access Sites for Power Injection of Iodinated Contrast for CT Angiography of the Thoracic Aorta', Annual Meeting of Society of Cardiovascular Computed Tomography, Los Vegas, NV, December 2015.

Poster Presentations (Delivered by First author)

Appiah Frank, Fardo D, Abner E, Glen M, Charnigo R, 'Predicting Alzheimer's Disease with Mixture of Regression Modeling', SRCOS, Bentonville, AR, June 2016.

Appiah Frank, Jing L, Bush H, Glen M, Williams M, Stromberg A, 'A Review of Methodology for Analyzing Translational Care Data' Southern Regional Council on Statistics, Wilmington, NC, July 2015.

Kingsbury A, **Appiah F**, Woodward C, Landwehr K, Chakraborty A, Pittman A,

Lowry C, Winkler M, 'Can Chest Width Be Used as a Surrogate for Weight for Selection of Contrast Injection Rate for Computed Tomographic Angiography', Society of Cardiovascular Computed Tomography. Orlando, FL, June 2016. 2016.

Woodward C, Kingsbury A, **Appiah F**, Landwehr K, Chakraborty A, Li X, Pittman A, Hobbs S, Winkler M, 'A Comparison of Intravenous Access Sites For Contrast Administration For Thoracic Computed Tomographic Angiography', Society of Cardiovascular Computed Tomography, Orlando, FL, June 2016.

Kingsbury A, Woodward C, **Appiah F**, Talley C, Li X, Fleischmann D, Winkler M, 'The Use of Intraosseous Needles For Injection Of Contrast Media For Computed Tomographic Angiography Of The Thorax', Society of Cardiovascular Computed Tomography. Orlando, FL, June 2016.

Woodward, C; Kingsbury, A; **Appiah, F**, Landwehr, K, Chakraboryty, A, Hobbs, S, Winkler, M, 'Differences in sites of intravenous access for Contrast Administration for Thoracic Computed Tomographic Angiography', Center for Clinical and Translational Science, Lexington, KY, April 2016.

External Funding

Michael Winkler (PI) et. al. 'The Effect of the Use of Proximal vs. Distal Intravenous Access Devices for Power Injection of Iodinated Contrast Media on the Safety, Quality and Rapidity of Computed Tomography Angiography, (3048112696), Teleflex Medi-

cal Incorporated, $104,008.00'. Role: Statistician.

2015-present

Awards

Fellow Award, Graduate Student Fellow: workshop to learn to create multimodal student learning outcomes and rubrics, August 2016-present.

Best Summer Student Award, Outstanding Summer Student Award, The Operations Research, Modeling & Simulation Office, Department of Defense, August 2016.

Best Student Presentation Award, Oral Presentation Winner Award, College of Public Health Research Day at the Center for Clinical And Translational Science Conference, Lexington, KY, April 2016.

Poster Travel Award, Boyd Harshbarger Travel Award for Poster Presentation, Southern Regional Council on Statistics, July 2015 & June 2016.

Fellowship Award,L.T. Johnson Fellowship Award, August 2010-August 2013.

Professional Organizations

American Statistical Association (ASA), ASA Kentucky Chapter

Leadership Skills

President, Graduate Student Congress University of Kentucky. *May 2013 & 14*
Achievements:

Led the congress in negotiating a fair deal with administration to compensate students previously living in graduate housing before demolition exercise was undertaken.

Led the congress to strengthen its standing nationally by working with the National Association of Professional Graduate Students.

Helped organize annual graduate students' appreciation day.

Help expand organization's umbrella by working with other groups and affiliations on campus.

Directed the annual ice cream social for new graduate students.

Taught a three-day course to initiate new TA's and provided a platform for further discussion and collaboration.

International Graduate Teaching Assistant's Orientation University of Kentucky. *May 2013 & 2014*

Achievements:

Mentored new graduate teaching assistants (TA) through one-on-one meetings and supervision

Participated in experienced TA panel discussion on what new TA's should expect and how they can succeed.

Led a discussion on cultural shocks and language barriers for international TA's.

Founder, Math & Science Association (MSA) Maysville Community & Technical College *2012*

Achievements:

Assisted students to draft by-laws for the organization.

Worked with other faculty members to designate mentors to MSA

Helped students with field event planning

Used MSA as a good recruiting tool for the Science, Technology, Engineering and Mathematics program

Founder, Statistics Students' Association (SSA) University of Kentucky. *2012*

Achievements:

Drafted the by-laws to describe the objectives of the organization, how to conduct elections and how it will be funded.

Coordinated with the DGS and chair of the department to get faculty members assigned to the organization as mentors.

Assisted in planning periodic students' presentation throughout the academic year.

Assisted in annual fund raising for underprivileged children in the Lexington, KY area.