Frequentist-Bayesian Hybrid Tests in Semi-parametric and Non-parametric Models with Low/High-Dimensional Covariate

Yangyi Xu

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Inyoung Kim, Chair
Feng Guo
Yili Hong
Leanna L. House

December 03, 2014
Blacksburg, Virginia

KEYWORDS: Bayes Factor; Bayes Sum Test; Discrete Fourier Transform; Hybrid; Laplace approximation; Neyman Test; Omnibus; Resampling; Score; Single index; Spline Approximation.

Frequentist-Bayesian Hybrid Tests in Semi-parametric and Non-parametric Models with Low/High-Dimensional Covariate

Yangyi Xu

ABSTRACT

We provide a Frequentist-Bayesian hybrid test statistic in this dissertation for two testing problems. The first one is to design a test for the significant differences between non-parametric functions and the second one is to design a test allowing any departure of predictors of high dimensional X from constant. The implementation is also given in construction of the proposal test statistics for both problems.

For the first testing problem, we consider the statistical difference among massive outcomes or signals to be of interest in many diverse fields including neurophysiology, imaging, engineering, and other related fields. However, such data often have nonlinear system, including to row/column patterns, having non-normal distribution, and other hard-to-identifying internal relationship, which lead to difficulties in testing the significance in difference between them for both unknown relationship and high-dimensionality. In this dissertation, we propose an Adaptive Bayes Sum Test capable of testing the significance between two nonlinear system basing on universal non-parametric mathematical decomposition/smoothing components. Our approach is developed from adapting the Bayes sum test statistic by Hart (2009). Any internal pattern is treated through Fourier transformation. Resampling techniques are applied to construct the empirical distribution of test statistic to reduce the effect of non-normal distribution. A simulation study suggests our approach performs better than the alternative method, the Adaptive Neyman Test by Fan & Lin (1998). The usefulness of our approach is demonstrated with an application in the identification of electronic chips as well as an application to test the change of pattern of precipitations.

For the second testing problem, currently numerous statistical methods have been developed for analyzing high-dimensional data. These methods mainly focus on variable selection approach, but are limited for purpose of testing with high-dimensional data, and often are required to have explicit derivative likelihood functions. In this dissertation, we propose "Hybrid Omnibus Test" for high-dimensional data testing purpose with much less requirements. Our Hybrid Omnibus Test is developed under semi-parametric framework where likelihood function is no longer necessary. Our Hybrid Omnibus Test is a version of Freqentist-Bayesian hybrid score-type test for a functional generalized partial linear single index model, which has link being functional of predictors through a generalized partially linear single index. We propose an efficient score based on estimating equation to the mathematical difficulty in likelihood derivation and construct our Hybrid Omnibus Test. We compare our approach with a empirical likelihood ratio test and Bayesian inference based on Bayes factor using simulation study in terms of false positive rate and true positive rate. Our simulation results suggest that our approach outperforms in terms of false positive rate, true positive rate, and computation cost in high-dimensional case and low-dimensional case. The advantage of our approach is also demonstrated by published biological results with application to a genetic pathway data of type II diabetes.

Acknowledgments

I would like to express my sincere gratitude and thanks to my advisor Dr. Inyoung Kim for her support, guidance, encouragement and trust throughout my graduate studies. She has guided me all the way to earn my PhD through challenges and opportunities; she has helped me to understand the importance of thinking beyond limitation and solid foundation in work and research; and also, she has made me become a true Statistician, and with the spirit of a Statistician to observe, to inference, and to think.

It is also my honor to be advised by a group of prominent and brilliant committee members, Dr. Feng Guo, Dr. Yili Hong, and Dr. Leanna L. House, who offered value advice and support.

I would like to extend my deepest thanks to all group members, who offered me invaluable advice and support in sharing thoughts in Statistics, discussing experiences in research, and expressing the attitude in chasing academic achievements on our pathes to the PhD.

I have also had the privilege to talk and discussion with a number of talented individual in Department of Statistics at Virginia Tech. Thank you all the professors for your inspiring courses and guidance. Thank you all the reference authors for their significant work which are the shoulders of giants I could stand on. Thank you all my friends for your generous collaboration and discussion on statistical issues that widened my vision. I also would like to thank all staff of Department of Statistics for the selfless support to the students.

Lastly, I would like to thank my beloved family, for their great and never-ending support. None of this would be possible without the love of my family.

Contents

List of Figures

List of Tables

Chapter 1

Motivation and Outline of the Dissertation

We provide a Frequentist-Bayesian hybrid test statistic in this dissertation for two testing problems. One is to test the significance in difference between non-parametric functions and the other is to test allowing any departure of predictors of high dimensional X from constant. The implementation is also given in construction of the proposal test statistics for both problems. We propose a Adaptive Bayes Sum Test (ABST) for the first problem and "Hybrid Omnibus Test" (HOT) for the second problem.

The motivation of the first problem is the following. Most methods (Faraway, 1997; Behseta & Kass, 2005) that test for the significant differences between non-parametric functions have two step procedures. They first require the estimation of a non-parametric function and then access a Hotelling T test or a Bayes factor related to another model. For example Faraway (1997) proposed smoothing on the functional data first and then using traditional analysis of variance. Behseta & Kass (2005) also proposed an equality test for two functions but smoothing here is also required using the Bayesian adaptive regression splines and testing of the Bayes factor. Their algorithm is developed using a reversible-jump MCMC, which is computationally expensive. Another possible technique is to use a smoothed principal component analysis (Besse & Ramsay, 1986; Rice &

Silverman, 1991) and to project data on the first few important principal directions. The question, then, is how many principal directions should be chosen.

Our goal is to develop a single step procedure which directly tests curved data as well as fits a non-parametric function. For this type of test procedure, Fan & Lin (1998) proposed the Adaptive Neyman Test. The techniques by Fan & Lin were based on orthogonal transforms, such as the Fourier transform and the wavelet transforms to compress signals. This required that data be sampled at an equal-spaced design point. For irregularly spaced designs, either binning or interpolation methods can be used to preprocess the data (e.g., see Cai & Brown 1998; Hall & Turlach 1997 in a different context). Once the data are transformed with Fourier or wavelet techniques, useful coefficients in the transformed domain are then adaptively chosen to yield a test statistic. It remains to be seen how efficient these simple and crude binning and interpolation methods are. Their approaches required knowledge of the tuning parameter in addition to equal-spaced design point. In Chapter 2 of this dissertation, we propose a testing procedure without such restrictions on the tuning parameter selection and equal-spaced design.

The motivation of the second problem is the following. Numerous statistical methods have been developed for analyzing high-dimensional data (Efron et al., 2004; Fan & Li, 2001; Tibshirani, 1996; Zou & Hastie, 2005). Such methods are mainly focused on variable selection based on various penalized regression. One of them is Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Tibshirani (1996). LASSO was based on L1 norm penalty on the estimator in the estimating algorithm, which proves efficient in reducing insignificant predictors when the number of variables is moderately larger than the number of observations. However the estimators based on LASSO are biased estimator. Fan & Li (2001) developed a special penalty function that was singular at the origin to produce sparse coefficient estimators. Their penalty function was called Smoothly Clipped Absolute Deviation (SCAD). The estimators based on SCAD are unbiased estimators and have sparsity and continuity properties. Zou & Hastie (2005) proposed an elastic net penalty (ENET), which is a linear combination of both L1-norm and L2-norm with convex constraint, thus having the characteristics of both the LASSO and ridge regression. They are mainly interested in variable selection, not testing. However, for the purpose of testing, the statistical approach is very limited for high-dimensional data. Verzelen & Villers (2010) developed a method for high-dimensional goodness-of-fit test. However this method is constrained to linear models with Gaussian error assumption and a known likelihood. Unlike existing approaches, our approach is developed in the semi-parametric framework where no likelihood function is available. In Chapter 3 of this dissertation, we propose test statistic, called “Hybrid Omnibus Test”, for high-dimensional data, and as high-dimensional predictor presents , we model with single index frame.

Therefore, in Chapter 2, we introduce a flexible non-parametric testing approach, “Adaptive Bayes Sum Test”, for equality of two non-parametric functions and compare it with the adaptive Neyman Adaptive test. In Chapter 3, we propose three HOT approaches, for allowing any departure of high-dimensional predictors from constant in the generalized semi-parametric model framework: HOT, constraint HOT, and Bayesian HOT based on maximum posterior estimator. We compare them with empirical likelihood ratio test and Bayesian inference based on Bayes factor. We conclude with Chapter 4, and in table of contents.

Chapter 2

Adaptive Bayes Sum Test for the Equality of Two Non-parametric Functions

2.1 Introduction

In neurophysiology, imaging, engineering, and other related fields, massive data sets are common. These data sometimes have nonlinear curves, depending on coordinate-based effect and non-normal random error. The main question of interest is to test whether there is any statistically significant difference between curved datasets.

By extending Hart's Bayes sum test statistic for our test on the equality of two curves, we propose an Adaptive Bayes Sum Test, and we have further developed a resampling procedure for this adaptive test. Our approach can manage the testing of the significance between two nonlinear curves by taking into account spatial dependence and reducing the effect of non-white noise. We compare our approach with an alternative approach on the adaptive Neyman test in terms of type I error and power.

2.2 Test Statistic for Equality of Two Curves

In this section, we describe two test statistics: one is our Adaptive Bayes Sum Test in section 2.2.1, and the other is the adaptive Neyman test in section 2.2.2.

2.2.1 Adaptive Bayes Sum Test Statistic

Let $X_i(s_1,s_2)$ and $Y_i(s_1,s_2)$ be the ith measurement values at spatial location (s_1,s_2) for two groups, respectively, $i=1,\ldots,N$, $s_1=1,\ldots,n_1$, $s_2=1,\ldots,n_2$. In our motivational example, $N=100$, $n_1=16$, and $n_2=32$. We assume that these observed values are random samples from the following model:

$$\begin{aligned} X_i(s_1,s_2) &= f_1(s_1,s_2)+\varepsilon_i(s_1,s_2), \\ Y_i(s_1,s_2) &= f_2(s_1,s_2)+\varepsilon_i^{'}(s_1,s_2), \end{aligned}$$

where $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ are unknown and nonlinear functions for two groups and the random error process $\varepsilon_i(s_1,s_2)$ and $\varepsilon_i^{'}(s_1,s_2)$ are stationary with a mean of 0.

Our main question of interest is to test $H_0 : f_1(s_1,s_2) = f_2(s_1,s_2)$ vs $H_0 : f_1(s_1,s_2) \neq f_2(s_1,s_2)$. We define $r(s_1,s_2) \equiv f_1(s_1,s_2) - f(s_1,s_2)$. We assume that $r(s_1,s_2)$ can be characterized by the Fourier coefficients $\phi_1,\phi_2,\ldots$ (Appendix A) and that the null hypothesis is equivalent to $H_0 : \phi_1 = \phi_2 = \cdots = 0$.

Let $\hat{\phi}_1,\ldots,\hat{\phi}_n$ be sample Fourier coefficients estimators of $\phi_1,\ldots,\phi_n$, where $n=n_1n_2$. The observed data are sample Fourier coefficients $\hat{\phi}_1$,...,$\hat{\phi}_n$ that estimate $\phi_1,\ldots,\phi_n$, respectively, where $\hat{\phi}_1,\ldots,\hat{\phi}_n$ are independent and $\hat{\phi}_j \sim N(\phi_j,\sigma^2/n)$, $j=1,\ldots,n$. The scalar parameter σ is allowed to be unknown. Our focus is on non-parametric tests of H_0, i.e., tests that are consistent against virtually any alternative to the null as $n\to\infty$.

Hart (2009) proposed omnibus test, called the "Bayes sum test", which is subsequently used in Frequentist fashion using a Laplace approximation to obtain the posterior probability of the null hypothesis. Hart (2009) showed that his test has powerful properties that are highly competitive with those of a popular class of non-parametric, or omnibus, lack-of fit tests. His omnibus test statistic is weighted sums of exponentiated squared and normalized Fourier coefficients, where the weights depend on prior probabilities. Exponentiation of the Fourier components leads to tests that can be exceptionally powerful compared to high-frequency alternatives. This method is the "hybrid Bayes-Frequentist", which combines Bayesian and Frequentist thinking (Good, 1957; Bayarri & Berger, 2004; Aerts et al., 2004; Chang & Chow, 2005). Testing whether or not a regression function has a prescribed linear form is one application of Hart's test.

By incorporating Bayes sum test into our testing problem, i.e., testing for the equality of two non-parametric functions, we propose the following test statistic:

$$T_{AB} = \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} k^{-2} l^{-2} \exp\left(\frac{n\hat{\phi}_{kl}^2}{2\hat{\sigma}^2}\right), \tag{2.1}$$

where $\hat{\phi}_{ij}$ are the 2D Discrete Fourier transformation coefficients and $\hat{\sigma}^2 = \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \hat{\phi}_{ij}^2$. We refer to this test statistic as the "Adaptive Bayes Sum Test statistic". In Appendix A, we briefly describe how we derive this test statistic (2.1) using the results of Hart (2009).

2.2.2 Adaptive Neyman Test Statistic

The alternative test is the Adaptive Neyman Test (Fan & Lin, 1998), $H_0 : f_1(s_1, s_2) = f_2(s_1, s_2)$ vs $H_1 : f_1(s_1, s_2) \neq f_2(s_1, s_2)$. Let $\{X_i^*(s_1, s_2)\}_{i=1}^N$ and $\{Y_i^*(s_1, s_2)\}_{i=1}^N$ be the discrete fourier transforms of $\{X_i(s_1, s_2)\}_{i=1}^N$ and $\{Y_i(s_1, s_2)\}_{i=1}^N$.

We define

$$D^*(s_1,s_2) = N^{-1}\sum_{i=1}^{N} X_i^*(s_1,s_2) - N^{-1}\sum_{i=1}^{N} Y_i^*(s_1,s_2),$$

and obtain its standardized form

$$Z(s_1,s_2) = \frac{D^*(s_1,s_2)}{\{\hat{\sigma}_1^2(s_1,s_2)/N + \hat{\sigma}_2^2(s_1,s_2)/N\}},$$

where $\hat{\sigma}_1^2(s_1,s_2)$ and $\hat{\sigma}_2^2(s_1,s_2)$ are the sample variance of $\{X_i^*(s_1,s_2)\}_{i=1}^N$ and $\{Y_j^*(s_1,s_2)\}_{i=1}^N$.

Let $Z^* = \Gamma Z$ is the orthonormal transformation, and for some constant c_S tending to infinity with $c_S < n$, the adaptive Neyman test statistic under stationary errors is

$$T_{AN}^* = \max_{1 \le m \le c_S, m = m_1 m_2} (2m)^{-1/2} \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \{|Z^*(k_1,k_2)|^2 - 1\}.$$

We define the standardized adaptive Neyman test statistic:

$$T_{AN} = \sqrt{2\log\log c_S} T_{AN}^* - \{2\log\log c_S + .5\log\log\log c_S - .5\log 4\pi\}.$$

Note that this test T_{AN}^* is equivalent to T_{AN} rejecting H_0. The asymptotic distribution under H_0 is given by

$$P(T_{AN} < x) \to \exp\{-\exp(-x)\} \qquad as \qquad n \to \infty$$

which Darling & Erdös (1956) discussed in a corollary. It has been shown by Fan (1996) that the convergence is very slow.

2.3 Resampling Based Test

In practice, the n_1 and n_2 are relatively small, i.e. in our case, $n_1 = 16$ and $n_2 = 32$, thus, we use resampling technique to obtain empirical p-value. Before explaining these procedures, first we define some notation as follows:

$$\begin{aligned}
\mathbf{x}(s_1,s_2) &= \Big(x_1(s_1,s_2),x_2(s_1,s_2),\ldots,x_N(s_1,s_2)\Big) \\
\mathbf{y}(s_1,s_2) &= \Big(y_1(s_1,s_2),y_2(s_1,s_2),\ldots,y_N(s_1,s_2)\Big) \\
\mathbf{x}_j &= \Big(x_j(1,1),x_j(1,2),\ldots,x_j(1,n_2),\ldots,x_j(n_1,1),x_j(1,2),\ldots,x_j(n_1,n_2)\Big) \\
\mathbf{y}_j &= \Big(y_j(1,1),y_j(1,2),\ldots,y_j(1,n_2),\ldots,y_j(n_1,1),y_j(1,2),\ldots,y_j(n_1,n_2)\Big).
\end{aligned}$$

That is, $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$ are the collection of all observed values of x and y at location (s_1, s_2), while $\mathbf{x}_j$ and $\mathbf{y}_j$ are the collection of all observed values of x and y at the jth measurement point. We then consider the following four resampling procedures.

2.3.1 Resampling by Exchanging Two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$

- Step 1: Randomly select two measurement points, j and j'.
- Step 2: Exchange two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$.
- Step 3: Calculate the test statistic using the Adaptive Bayes Sum Test statistic and the adaptive Neyman test statistic.
- Step 4: Repeat Steps 1-3, T times (i.e., $T = 10,000$).
- Step 5: The empirical p value is obtained from $p - value = \#(T_{perm} \geq T_{obs})/T$, where T_{perm} is the test statistic value from the permutated data and T_{obs} is the test statistic value from the

observed data.

This procedure is performed by keeping the spatial location but ignoring dependence on measurements.

2.3.2 Resampling by Exchanging Two $x_j(s_1,s_2)$ and $y_j(s_1,s_2)$

This procedure is similar to the resampling procedure above by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$, except for steps 1-2

- Step 1: Randomly select jth measurement.
- Step 2: Exchange two $x_j(s_1,s_2)$ and $y_j(s_1,s_2)$.

This procedure is performed by keeping both the geographical locations and the order of measurements.

2.3.3 Resampling by Exchanging Two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$

This procedure is similar to the procedure above by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$ except for steps 1-2

- Step 1: Randomly select the spatial location (s_1,s_2).
- Step 2: Exchange $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$ of two dataset at the same the spatial location (s_1,s_2).

This resampling procedure is performed by keeping variations among spatial locations but ignoring variations among measurements.

2.3.4 Resampling by Exchanging Two $\mathbf{x}_j$ and $\mathbf{y}_j$

This procedure is similar to the resampling procedure above by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$ except for steps 1-2

- Step 1: Randomly select the jth measurement point.
- Step 2: Randomly exchange two $\mathbf{x}_j$ and $\mathbf{y}_j$.

This resampling procedure is performed by keeping both variations among spatial locations and variations among measurements.

2.4 A Simulation Study

We conduct simulations to compare the performance of following three methods:

- ANTs: adaptive Neyman test with resampling by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$;
- ANTsv: adaptive Neyman test with resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$;
- ABSTs: adaptive Bayesian sum test with resampling by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$.

2.4.1 Type I Error and Power

As with our motivated data, we consider $j = 1,\ldots,100$, $s_1 = 1,\ldots,16$, $s_1 = 1,\ldots,32$, and $s = 1,\ldots,512$. X_j and Y_j are simulated from the following six cases, with each case involving 1000 simulations. The first three cases are for the univariate functions and the last three cases are for the bivariate functions:

- Case 1: $f(x) = \sqrt{2\frac{\lambda}{512}}\cos(m\pi_2)$ and $\varepsilon \sim N(0,1)$, where $x = (s-0.5)/512$, $s = 1,\ldots,512$, $m = 1$, and $\lambda = 16$

- Case 2: the same setting as case 1 except that $m = 7$ and $\lambda = 2$

- Case 3: the same setting as case 1 except that $m = 1$ and $\lambda = 1$

- Case 4:

$$\begin{aligned} f(x_1,x_2) &= \frac{\lambda}{16}\cos(m\pi x_1) + U\exp\left(\frac{x_1x_2}{x_1+x_2}\right) + \varepsilon, \\ U &\sim N(0,2), \\ \varepsilon &\sim N(0,1), \end{aligned}$$

  where $m = 7$, $\lambda = 2$, $x_1 = (s_1-0.5)/16$, $x_2 = (s_2-0.5)/32$.

- Case 5: the same setting as case 4 except that $U \sim \text{Gamma}(2,3)$

- Case 6: the same setting as case 4 except that $U \sim \text{Extreme}(3,1,0)$

We calculate the type I error rate and power of the tests. The type I error rate, the probability of rejecting a null hypothesis when it is true, is calculated, using the two data sets which were generated from the same case, thus in this scenario a test is preferable when having a higher null hypothesis rejection rate. The estimated Type I error are summarized in Table 2.1. For cases 1-3, all results from tests are close to nominal values 0.05, which is the significant level. Thus we have 0.05 chance to reject a null hypothesis when it is true. However, for cases 4-6, ABSTs has a smaller Type I error, while ANTs and ANTsv are close to nominal value.

Table 2.1: Type I error for three methods; ANTs=adaptive Neyman test with resampling by exchange two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$; ANTsv: adaptive Neyman test with resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$; ABSTs: adaptive Bayesian sum test with resampling by exchange two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$. 1000 simulations.

Two datasets from	ANTs	ANTsv	ABSTs
Case 1	0.046	0.046	0.046
Case 2	0.034	0.036	0.034
Case 3	0.053	0.052	0.052

Power, the probability of rejecting null hypothesis when the alternative is true, is also obtained using two data sets that were generated from two different cases. The powers are summarized in Table 2.2. The power of ABSTs is 1, indicating this test has great power for detecting small differences between models. However, ANTs and ANTsv are not relatively sensitive in cases 1-3.

Table 2.2: Power for three methods; ANTs=adaptive Neyman test with resampling by exchange two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$; ANTsv: adaptive Neyman test with resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$; ABSTs: adaptive Bayesian sum test with resampling by exchange two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$. 1000 simulations.

Two datasets from	ANTs	ANTsv	ABSTs
Cases 1-2	0.001	0	1
Cases 2-3	0.040	0.039	1
Cases 1-3	0.009	0.010	1
Cases 4-5	1	1	1
Cases 4-6	1	0.900	1
Cases 5-6	1	0.874	1

2.4.2 Capability of Detecting the Difference between Two Functions

We also conduct a simulation study to understand how small of a difference between two functions could be detected by our method in a special case. In general it is difficult to derive the bound of this difference because the non-parametric estimation of two unknown nonlinear curves with smoothing tools, which leads to difficulty for both determining the degree of freedom and asymptotic distribution.

As in Section 2.4.1, we use the similar setting: $j = 1,\ldots,100$, $s_1 = 1,\ldots,16$, $s_1 = 1,\ldots,32$, and $s = 1,\ldots,512$. Our X_j and Y_j are simulated from the following Case,

- Case 7

  $$f(x_1,x_2) = m\cdot[\sin\{\pi(x_1-A)/(B-A)\}][\sin\{\pi(x_2-A)/(B-A)\}] + m\cdot\varepsilon,$$

  where $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $B = \sqrt{3}/2 + 1.645/\sqrt{12}$, $x_1 = (s_1-0.5)/16$, $x_2 = (s_2-0.5)/32$, and $\varepsilon \sim N(0,0.25)$.

We vary m across the interval [1.00,1.10] in increments of 0.01. One data set is generated from Case 7 with $m = 1$ and the other data set is generated from Case 7 with different m values. We simulate this 1000 times to obtain power, the result of which is summarized in Table 2.3. When a difference of m is smaller than 1.01, all three methods are not able to detect such small difference. When m is larger than 1.03, the power of ABSTs is about 1, indicating that this test has great power for detecting such differences between models. However, ANTs and ANTsv have a power of less than 0.3. This may be because our Adaptive Bayes Sum Test statistic assigns different weight (which is π_{jk} in Appendix B) to the basis of the non-parametric filters. Indeed, by making use of wavelets filters, our Adaptive Bayes Sum Test can filter out the components of interest. For example, in our case, low-frequency components are more important and hence must be given

Table 2.3: Power for three methods; ANTs=adaptive Neyman test with resampling by exchange two $x_j(s_1, s_2)$ and $y_{j'}(s_1, s_2)$; ANTsv: adaptive Neyman test with resampling by exchanging two $\mathbf{x}(s_1, s_2)$ and $\mathbf{y}(s_1, s_2)$; ABSTs: adaptive Bayesian sum test with resampling by exchange two $x_j(s_1, s_2)$ and $y_{j'}(s_1, s_2)$. 1000 simulations.

Two datasets from Case 7 with	ANTs	ANTsv	ABSTs
m=1 vs 1.01	0.055	0.056	0.068
m=1 vs 1.02	0.065	0.067	0.406
m=1 vs 1.03	0.094	0.091	0.974
m=1 vs 1.04	0.152	0.142	1
m=1 vs 1.05	0.156	0.148	1
m=1 vs 1.06	0.17	0.167	1
m=1 vs 1.07	0.205	0.2	1
m=1 vs 1.08	0.225	0.231	1
m=1 vs 1.09	0.236	0.237	1
m=1 vs 1.10	0.294	0.285	1

a higher weight. In contrast, the adaptive Neyman test treats every data point uniformly, and is unable to filter out more important components.

2.5 Application

2.5.1 Engineering

Our application is as follows. During the manufacturing of electronic chips, small imperfections in the manufacturing process introduce random variations in the chip performance. These imperfections generally degrade uniform chip performance. However, they also have an important benefit. The pattern for every electronic chip manufactured is unique, and hence they can be used to distinguish one chip from another. Being able to distinguish one electronic chip from another chip is similar to distinguishing persons by their biometrics. Indeed, the biometric of a chip has important applications in supply chain tracking, detection of chip counterfeiting, and authentication of electronic products in general (Sadeghi & Naccache, 2010). In this application, we measured a collection of digital chips used in embedded computers. Each chip is characterized by measuring its performance at 512 different spatial locations in each chip. The 512 locations are chosen according to a rectangular 16 by 32 grid that spans the entire chip surface. Each location is characterized as an integer representing the local performance of the chip; the higher the integer, the better the performance. A detailed description of the measurement setup, and the associated database, is summarized in Maiti et al. (2010).

The data set that we analyzed consists of measurements for 10 different chips. Each chip data set is a collection of 100 measurements of each of the 512 locations characterized per chip. We found that these data have nonlinear curves as shown in Figures 2.1-2.2. Figure 2.1 is obtained from chip 1 and Figure 2.2 is an estimated curve using chip 1. Other chips also have nonlinear curved data.

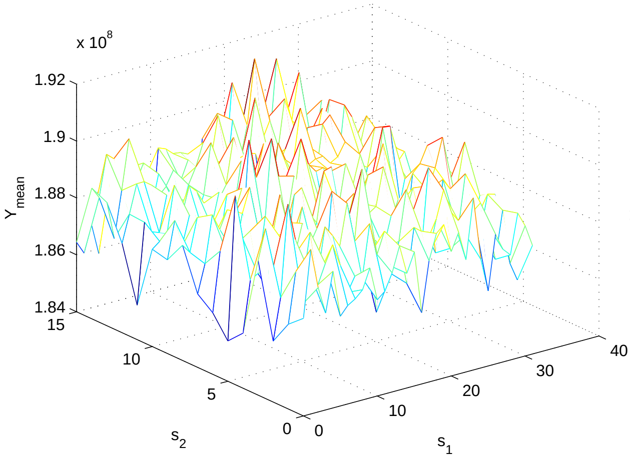

Figure 2.1: Raw data obtained from chip1

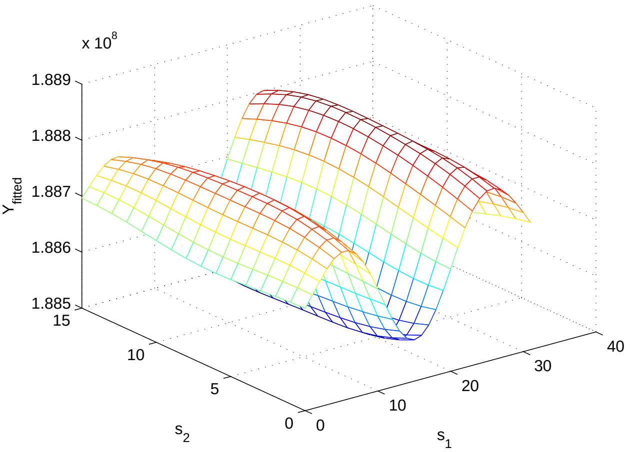

Figure 2.2: Fitted curve obtained from chip1

The objective is to distinguish each chip from all others. We apply our four resampling approaches to these data sets.

- ANTs: Adaptive Neyman test by resampling by exchanging two $x_j(s_1, s_2)$ and $y_{j'}(s_1, s_2)$

- ANTsm: Adaptive Neyman test by resampling by exchanging two $x_j(s_1, s_2)$ and $y_j(s_1, s_2)$
- ANTsv: Adaptive Neyman test by resampling by exchanging two $\mathbf{x}(s_1, s_2)$ and $\mathbf{y}(s_1, s_2)$
- ANTmv: Adaptive Neyman test by resampling by exchanging two $\mathbf{x}_j$ and $\mathbf{y}_j$
- ABSTs: Adaptive Bayes Sum Test by resampling by exchanging two $x_j(s_1, s_2)$ and $y_{j'}(s_1, s_2)$
- ABSTsm: Adaptive Bayes Sum Test by resampling by exchanging two $x_j(s_1, s_2)$ and $y_j(s_1, s_2)$
- ABSTsv: Adaptive Bayes Sum Test by resampling by exchanging two $\mathbf{x}(s_1, s_2)$ and $\mathbf{y}(s_1, s_2)$
- ABSTmv: Adaptive Bayes Sum Test by resampling by exchanging $\mathbf{x}_j$ and $\mathbf{y}_j$

Tables 2.4-2.7 are obtained from ANTs, ANTsm, ANTsv, and ANTmv, respectively. Tables 2.8-2.11 are obtained from ABSTs, ABSTsm, ABSTsv, and ABSTmv, respectively. Tables show both the empirical pvalue and the test statistic.

Table 2.4: The empirical pvalue and the test statistic; $T_{AN;0.95}$, obtained from Adaptive Neyman test by resampling by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AN;0.95}/T_{AN;obs}$, where $T_{AN;0.95}$ and $T_{AN;obs}$ are 95 percentile value of adaptive Neyman test statistic value based on 10,000 permutations and observed adaptive Neyman test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	**0.1721**	0.0111	0.0000	0.0010	**0.0791**	0.0000	**0.0987**	0.0001
chip2	-4.248/-4.343		0.0000	0.0000	0.0054	0.0010	0.0000	0.0000	0.0234	0.0000
chip3	-4.349/-4.343	-4.272/-4.344		0.0051	0.0000	0.0172	0.0059	0.0000	**0.4575**	0.0000
chip4	-4.337/-4.344	-4.295/-4.344	-4.332/-4.344		0.0000	0.0002	0.0498	**0.7773**	0.0044	**0.3122**
chip5	-4.301/-4.344	-4.334/-4.344	-4.305/-4.343	-4.290/-4.343		0.0008	0.0000	0.0000	0.0014	0.0000
chip6	-4.326/-4.343	-4.322/-4.343	-4.339/-4.344	-4.316/-4.344	-4.325/-4.343		0.0000	0.0000	**0.7660**	0.0000
chip7	-4.346/-4.344	-4.211/-4.343	-4.333/-4.344	-4.343/-4.343	-4.282/-4.344	-4.306/-4.343		0.0000	0.0028	0.0340
chip8	-4.264/-4.343	-4.093/-4.343	-4.245/-4.344	-4.353/-4.343	-4.226/-4.344	-4.203/-4.343	-4.305/-4.344		0.0000	0.0087
chip9	-4.346/-4.343	-4.340/-4.343	-4.352/-4.344	-4.332/-4.344	-4.326/-4.344	-4.353/-4.343	-4.330/-4.343	-4.272/-4.344		0.0000
chip10	-4.314/-4.344	-4.103/-4.343	-4.287/-4.343	-4.351/-4.343	-4.243/-4.344	-4.252/-4.343	-4.342/-4.344	-4.336/-4.344	-4.294/-4.343	

Table 2.5: The empirical pvalue and the test statistic, $T_{AN;0.95}$, obtained from Adaptive Neyman test by resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AN;0.95}/T_{AN;obs}$, where $T_{AN;0.95}$ and $T_{AN;obs}$ are 95 percentile value of adaptive Neyman test statistic value based on 10,000 permutations and observed adaptive Neyman test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	-4.248/-4.354		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	-4.349/-4.354	-4.272/-4.354		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip4	-4.337/-4.354	-4.295/-4.354	-4.332/-4.354		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip5	-4.301/-4.354	-4.334/-4.354	-4.305/-4.354	-4.290/-4.354		0.0000	0.0000	0.0000	0.0000	0.0000
chip6	-4.326/-4.354	-4.322/-4.354	-4.339/-4.354	-4.316/-4.354	-4.325/-4.354		0.0000	0.0000	0.0003	0.0000
chip7	-4.346/-4.354	-4.211/-4.354	-4.333/-4.354	-4.343/-4.354	-4.282/-4.354	-4.306/-4.354		0.0000	0.0000	0.0000
chip8	-4.264/-4.354	-4.093/-4.354	-4.245/-4.354	-4.353/-4.354	-4.226/-4.354	-4.203/-4.354	-4.305/-4.354		0.0000	0.0000
chip9	-4.346/-4.354	-4.340/-4.354	-4.352/-4.354	-4.332/-4.354	-4.326/-4.354	-4.353/-4.354	-4.330/-4.353	-4.272/-4.354		0.0000
chip10	-4.314/-4.354	-4.103/-4.354	-4.287/-4.354	-4.351/-4.354	-4.243/-4.354	-4.252/-4.354	-4.342/-4.354	-4.336/-4.354	-4.294/-4.354	

Table 2.6: The empirical pvalue and the test statistic, $T_{AN;0.95}$, obtained from Adaptive Neyman test by resampling by exchanging two $x_j(s_1,s_2)$ and $y_j(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AN;0.95}/T_{AN;obs}$, where $T_{AN;0.95}$ and $T_{AN;obs}$ are 95 percentile value of adaptive Neyman test statistic value based on 10,000 permutations and observed adaptive Neyman test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5
chip1		0.0000	**0.1358**	0.0113	0.0000
chip2	-4.248/-4.343		0.0000	0.0000	0.0058
chip3	-4.349/-4.345	-4.272/-4.344		0.0033	0.0001
chip4	-4.337/-4.344	-4.295/-4.344	-4.332/-4.344		0.0000
chip5	-4.301/-4.344	-4.334/-4.344	-4.305/-4.343	-4.290/-4.343	
chip6	-4.326/-4.344	-4.322/-4.344	-4.339/-4.344	-4.316/-4.343	-4.325/-4.344
chip7	-4.346/-4.345	-4.211/-4.344	-4.333/-4.344	-4.343/-4.344	-4.282/-4.343
chip8	-4.264/-4.343	-4.093/-4.344	-4.245/-4.344	-4.353/-4.352	-4.226/-4.343
chip9	-4.346/-4.344	-4.340/-4.345	-4.352/-4.346	-4.332/-4.344	-4.326/-4.344
chip10	-4.314/-4.344	-4.103/-4.343	-4.287/-4.344	-4.351/-4.347	-4.243/-4.344

	chip6	chip7	chip8	chip9	chip10
chip1	0.0023	**0.0595**	0.0000	**0.0847**	0.0001
chip2	0.0016	0.0000	0.0000	0.0180	0.0000
chip3	0.0170	0.0061	0.0000	0.4094	0.0000
chip4	0.0011	0.0429	**0.5155**	0.0042	**0.2251**
chip5	0.0020	0.0000	0.0000	0.0020	0.0000
chip6		0.0001	0.0000	**0.7451**	0.0000
chip7	-4.306/-4.344		0.0000	0.0029	0.0319
chip8	-4.203/-4.343	-4.305/-4.343		0.0000	0.0080
chip9	-4.353/-4.346	-4.330/-4.344	-4.272/-4.344		0.0000
chip10	-4.252/-4.343	-4.342/-4.344	-4.336/-4.344	-4.294/-4.343	

Table 2.7: The empirical pvalue and the test statistic, $T_{AN;0.95}$, obtained from Adaptive Neyman test by resampling by exchanging two $\mathbf{x}_j$ and $\mathbf{y}_j$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AN;0.95}/T_{AN;obs}$, where $T_{AN;0.95}$ and $T_{AN;obs}$ are 95 percentile value of adaptive Neyman test statistic value based on 10,000 permutations and observed adaptive Neyman test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	-4.248/-4.351		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	-4.349/-4.351	-4.272/-4.351		0.0000	0.0000	0.0000	0.0000	0.0000	**0.9183**	0.0000
chip4	-4.337/-4.351	-4.295/-4.351	-4.332/-4.351		0.0000	0.0000	0.0000	**0.9803**	0.0000	**0.1646**
chip5	-4.301/-4.351	-4.334/-4.351	-4.305/-4.351	-4.290/-4.351		0.0000	0.0000	0.0000	0.0000	0.0000
chip6	-4.326/-4.351	-4.322/-4.351	-4.339/-4.351	-4.316/-4.351	-4.325/-4.351		0.0000	0.0000	**0.9763**	0.0000
chip7	-4.346/-4.351	-4.211/-4.351	-4.333/-4.351	-4.343/-4.351	-4.282/-4.351	-4.306/-4.351		0.0000	0.0000	0.0000
chip8	-4.264/-4.351	-4.093/-4.351	-4.245/-4.351	-4.353/-4.350	-4.226/-4.351	-4.203/-4.351	-4.305/-4.351		0.0000	0.0000
chip9	-4.346/-4.351	-4.340/-4.351	-4.352/-4.351	-4.332/-4.351	-4.326/-4.351	-4.353/-4.348	-4.330/-4.351	-4.272/-4.351		0.0000
chip10	-4.314/-4.351	-4.103/-4.351	-4.287/-4.351	-4.351/-4.351	-4.243/-4.351	-4.252/-4.351	-4.342/-4.351	-4.336/-4.351	-4.294/-4.351	

Table 2.8: The empirical pvalue and the test statistic, $T_{AB:0.95}$, obtained from Adaptive Bayes sum test by resampling by exchanging two $x_j(s_1,s_2)$ and $y_{j'}(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AB;0.95}/T_{AB;obs}$, where $T_{AB;0.95}$ and $T_{AB;obs}$ are 95 percentile value of Adaptive Bayes sum test statistic value based on 10,000 permutations and observed Adaptive Bayes sum test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	4.130e+108/2.571		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	2.367e+063/2.572	1.758e+106/2.571		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip4	8.616e+109/2.571	6.354e+110/2.571	4.852e+109/2.571		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip5	2.219e+110/2.571	2.058e+109/2.571	2.744e+110/2.571	7.408e+110/2.571		0.0000	0.0000	0.0000	0.0000	0.0000
chip6	1.177e+102/2.571	1.356e+102/2.571	2.584e+089/2.572	4.005e+110/2.571	1.085e+110/2.571		0.0000	0.0000	0.0000	0.0000
chip7	3.938e+092/2.571	1.095e+110/2.571	8.664e+099/2.572	1.204e+109/2.571	4.366e+110/2.571	3.637e+107/2.572		0.0000	0.0000	0.0000
chip8	1.633e+110/2.572	7.682e+110/2.571	1.165e+110/2.571	5.107e+061/2.571	8.597e+110/2.571	4.533e+110/2.571	4.443e+109/2.571		0.0000	0.0000
chip9	1.508e+100/2.571	4.684e+104/2.572	4.996e+060/2.572	2.365e+110/2.571	1.185e+110/2.571	2.332e+024/2.572	1.512e+108/2.571	3.740e+110/2.571		0.0000
chip10	1.014e+107/2.571	4.340e+110/2.571	3.679e+108/2.571	1.273e+103/2.571	8.169e+110/2.571	6.292e+109/2.571	1.766e+101/2.571	8.057e+106/2.571	8.667e+109/2.571	

Table 2.9: The empirical pvalue and the test statistic, $T_{AB;0.95}$, obtained from Adaptive Bayes sum test by resampling by exchanging two $x_j(s_1,s_2)$ and $y_j(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AB;0.95}/T_{AB;obs}$, where $T_{AB;0.95}$ and $T_{AB;obs}$ are 95 percentile value of Adaptive Bayes sum test statistic value based on 10,000 permutations and observed Adaptive Bayes sum test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	4.130e+108/2.586		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	2.367e+063/2.585	1.758e+106/2.584		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip4	8.616e+109/2.584	6.354e+110/2.585	4.852e+109/2.583		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip5	2.219e+110/2.584	2.058e+109/2.584	2.744e+110/2.584	7.408e+110/2.586		0.0000	0.0000	0.0000	0.0000	0.0000
chip6	1.177e+102/2.585	1.356e+102/2.586	2.584e+089/2.585	4.005e+110/2.585	1.085e+110/2.584		0.0000	0.0000	0.0000	0.0000
chip7	3.938e+092/2.586	1.095e+110/2.585	8.664e+099/2.585	1.204e+109/2.584	4.366e+110/2.584	3.637e+107/2.585		0.0000	0.0000	0.0000
chip8	1.633e+110/2.585	7.682e+110/2.584	1.165e+110/2.584	5.107e+061/2.585	8.597e+110/2.590	4.533e+110/2.585	4.443e+109/2.585		0.0000	0.0000
chip9	1.508e+100/2.586	4.684e+104/2.584	4.996e+060/2.585	2.365e+110/2.586	1.185e+110/2.584	2.332e+024/2.585	1.512e+108/2.586	3.740e+110/2.585		0.0000
chip10	1.014e+107/2.584	4.340e+110/2.585	3.679e+108/2.584	1.273e+103/2.586	8.169e+110/2.583	6.292e+109/2.586	1.766e+101/2.584	8.057e+106/2.585	8.667e+109/2.585	

Table 2.10: The empirical pvalue and the test statistic, $T_{AB;0.95}$, obtained from Adaptive Bayes sum test by resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AB;0.95}/T_{AB;obs}$, where $T_{AB;0.95}$ and $T_{AB;obs}$ are 95 percentile value of Adaptive Bayes sum test statistic value based on 10,000 permutations and observed Adaptive Bayes sum test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5
chip1		0.0000	0.0000	0.0000	0.0000
chip2	4.130e+108/10.372		0.0000	0.0000	0.0000
chip3	2.367e+063/10.686	1.758e+106/10.617		0.0000	0.0000
chip4	8.616e+109/10.352	6.354e+110/10.377	4.852e+109/10.517		0.0000
chip5	2.219e+110/10.296	2.058e+109/10.281	2.744e+110/10.573	7.408e+110/10.215	
chip6	1.177e+102/10.254	1.356e+102/10.553	2.584e+089/10.517	4.005e+110/10.125	1.085e+110/10.691
chip7	3.938e+092/10.419	1.095e+110/10.297	8.664e+099/10.704	1.204e+109/10.243	4.366e+110/10.378
chip8	1.633e+110/10.178	7.682e+110/10.299	1.165e+110/10.235	5.107e+061/10.754	8.597e+110/10.347
chip9	1.508e+100/10.070	4.684e+104/10.455	4.996e+060/10.807	2.365e+110/10.274	1.185e+110/10.408
chip10	1.014e+107/10.323	4.340e+110/10.570	3.679e+108/10.269	1.273e+103/10.574	8.169e+110/10.412

	chip6	chip7	chip8	chip9	chip10
chip1	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	0.0000	0.0000	0.0000	0.0000	0.0000
chip4	0.0000	0.0000	0.0000	0.0000	0.0000
chip5	0.0000	0.0000	0.0000	0.0000	0.0000
chip6		0.0000	0.0000	0.0000	0.0000
chip7	3.637e+107/10.227		0.0000	0.0000	0.0000
chip8	4.533e+110/10.331	4.443e+109/10.440		0.0000	0.0000
chip9	2.332e+024/10.585	1.512e+108/10.132	3.740e+110/10.479		0.0000
chip10	6.292e+109/10.165	1.766e+101/10.280	8.057e+106/10.304	8.667e+109/10.353	

Table 2.11: The empirical pvalue and the test statistic, $T_{AB;0.95}$, obtained from Adaptive Bayes sum test by resampling by exchanging $\mathbf{x}_j$ and $\mathbf{y}_j$; The upper triangular indicates p-value and the lower triangular indicates the test statistic, $T_{AB;0.95}/T_{AB;obs}$, where $T_{AB;0.95}$ and $T_{AB;obs}$ are 95 percentile value of Adaptive Bayes sum test statistic value based on 10,000 permutations and observed Adaptive Bayes sum test statistic value, respectively.

	chip1	chip2	chip3	chip4	chip5	chip6	chip7	chip8	chip9	chip10
chip1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip2	4.130e+108/21796.115		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip3	2.367e+063/334.130	1.758e+106/17710.905		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip4	8.616e+109/22196.034	6.354e+110/26523.205	4.852e+109/23277.880		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
chip5	2.219e+110/24492.204	2.058e+109/21942.350	2.744e+110/22764.339	7.408e+110/27255.356		0.0000	0.0000	0.0000	0.0000	0.0000
chip6	1.177e+102/11592.106	1.356e+102/11427.941	2.584e+089/3662.282	4.005e+110/25574.395	1.085e+110/23669.096		0.0000	0.0000	0.0000	0.0000
chip7	3.938e+092/5037.090	1.095e+110/24469.787	8.664e+099/9599.773	1.204e+109/19762.237	4.366e+110/26531.487	3.637e+107/19838.913		0.0000	0.0000	0.0000
chip8	1.633e+110/24172.891	7.682e+110/26907.399	1.165e+110/25224.285	5.107e+061/226.207	8.597e+110/25725.026	4.533e+110/26710.958	4.443e+109/23299.348		0.0000	0.0000
chip9	1.508e+100/9606.989	4.684e+104/14861.033	4.996e+060/258.294	2.365e+110/25725.289	1.185e+110/23987.296	2.332e+024/10.500	1.512e+108/19936.054	3.740e+110/26124.360		0.0000
chip10	1.014e+107/18720.100	4.340e+110/26258.852	3.679e+108/21275.078	1.273e+103/12527.224	8.169e+110/26646.525	6.292e+109/24078.374	1.766e+101/10167.936	8.057e+106/18647.454	8.667e+109/24444.931	

The upper triangular indicates empirical p-value and the lower triangular indicates the test statistic, $T_{AN;0.95}/T_{AB;obs}$ or $T_{AB;0.95}/T_{AB;obs}$, where $T_{AN;0.95}$ and $T_{AN;obs}$ are 95th percentile value of the adaptive Neyman test statistic value based on 10,000 permutations and observed the adaptive Neyman test statistic value, respectively; $T_{AB;0.95}$ and $T_{AB;0.95}$ are similarly defined using the Adaptive Bayes Sum Test statistic value. We found that all chips are significantly different using only ANTsv. On the other hand, using other methods (ANTs, ANTsm, ANTmv), some chips cannot be distinguished. Thus we consider ANTsv the best method with the simulation and application result, with the unstructured resampler.

However, the results based on the adaptive Bayes sum show that all chips are significantly different. We have the same results for ABSRs, ANSTsm, ABSTsv, and ABSTmv. Figures 2.3-2.6 display both the empirical probability density and distribution of test statistic for ABSTs, ABSTsm, ABSTsv, and ABSTmv, respectively.

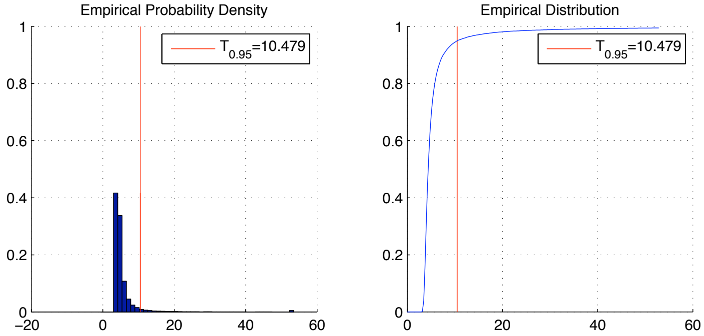

Figure 2.3: Empirical probability density and distribution of ABSTs; ABSTs=Adaptive Bayes sum test by resampling by exchanging two $x_j(s_1, s_2)$ and $y_{j'}(s_1, s_2)$

As we observed in simulation, the Adaptive Bayes Sum Test is more powerful in detecting both small and huge differences, so we conclude that all chips are significantly different.

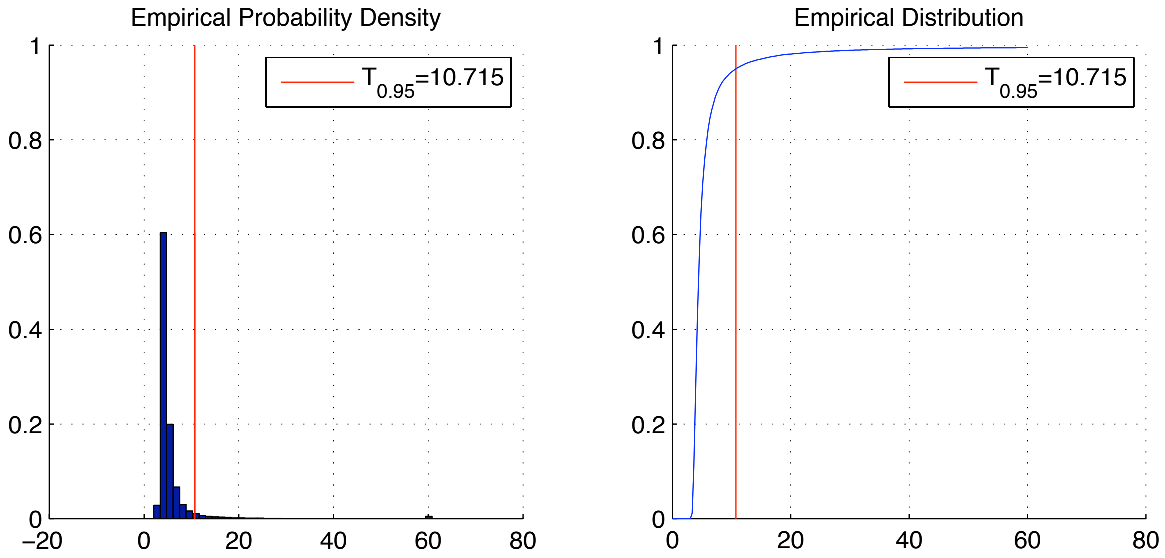

Figure 2.4: Empirical probability density and distribution of ABSTsm; ABSTsm=Adaptive Bayes sum test by resampling by exchanging two $x_j(s_1,s_2)$ and $y_j(s_1,s_2)$

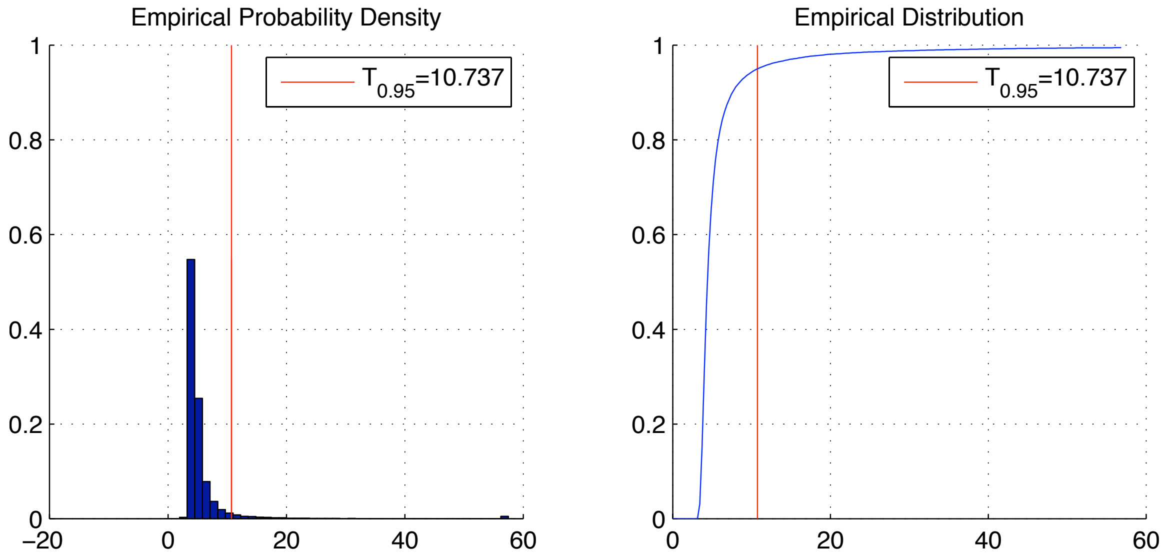

Figure 2.5: Empirical probability density and distribution of ABSTsv; ABSTsv=Adaptive Bayes sum test by resampling by exchanging two $\mathbf{x}(s_1,s_2)$ and $\mathbf{y}(s_1,s_2)$

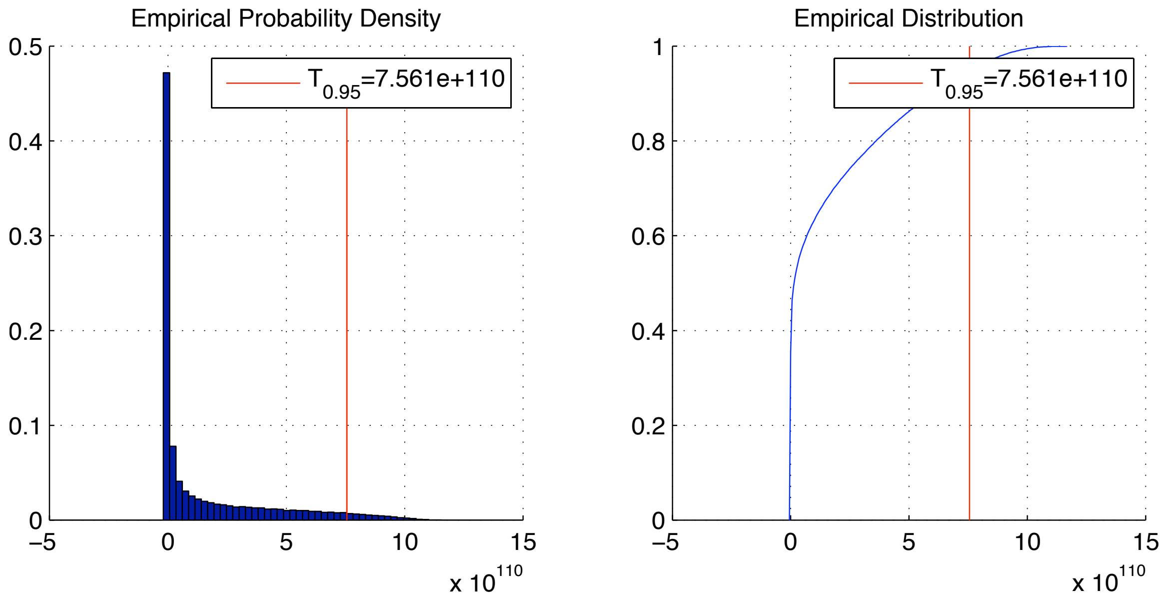

Figure 2.6: Empirical probability density and distribution of ABSTmv; ABSTmv=Adaptive Bayes sum test by resampling by exchanging $\mathbf{x}_j$ and $\mathbf{y}_j$

2.5.2 Precipitation

We can also apply our Adaptive Bayes Sum Test to a precipitation study. Our application is as follows. We obtain the precipitation data and location (latitude and longitude) data through the National Weather Service (NWS) River Forecast Centers (RFCs) from about 500,000 of precipitation site. The data are in XMRG format and projected in the Hydrologic Rainfall Analysis Project (HRAP) grid coordinate system, a polar stereographic projection true at 60^oN / 105^oW, visualized in Figure 2.7. The month precipitation data are collected at 287×581 grid with latitude and longitude rounded to tenth.

Precipitation is a major component of the water cycles which describes the continuous movement of water on, above and below the surface of the Earth, as well as the chaos system which studies the behavior of dynamical systems - climate. We can use our Adaptive Bayes Sum Testing approach to evaluate the stability of climate system. We compare the pattern of precipitation over the past five years from May 2008 through May 2013. Figure 2.7 shows the visualization of the precipitation

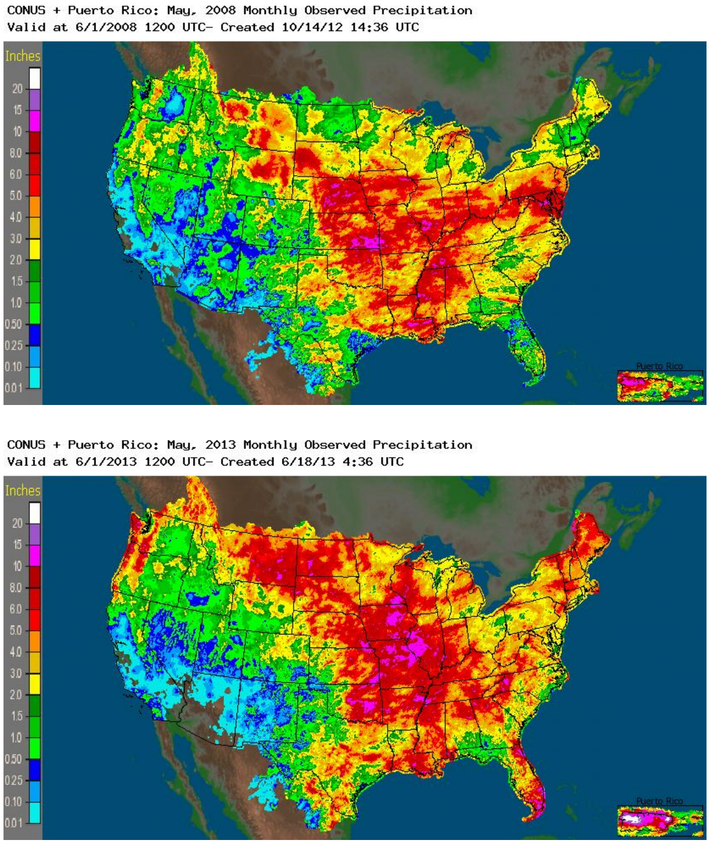

Figure 2.7: Visualization of Precipitation of US, May 2008 and May 2013, Observed (inch)

in the US in 2008 and 2013.

Using our Adaptive Bayes Sum Test, we assert for the equality of two non-parametric functions of precipitations in 2008 and 2013. The test results are summarized in Table 2.12. The empirical pvalues based on 10,000 permutations are close zero, indicating that precipitation changes over times and does not appear to be stable.

Table 2.12: The empirical pvalue and the test statistic, $T_{AB;0.95}$, obtained from Adaptive Bayes sum test by resampling by exchanging $\mathbf{x}_j$ and $\mathbf{y}_j$ using month precipitation from May 2008 through May 2013; $T_{AB;0.95}$ is a 95 percentile value of Adaptive Bayes sum test statistic value based on 10,000 permutations; $T_{AB:obs}$, a observed Adaptive Bayes sum test statistic, were larger than 10^{32}.

Two years	$T_{AB;0.95}$	Empirical pvalue
07-08	11.5467	0.0000
08-09	10.9670	0.0000
09-10	11.4563	0.0000
10-11	12.1339	0.0000
11-12	11.3332	0.0000
12-13	11.9893	0.0000

2.6 Discussion

In this chapter, we proposed an Adaptive Bayes Sum Test statistic for the equality of two curves. Our approach can be used for testing significant difference between two nonlinear curves by taking into account spatial dependence and reducing the effect of non-white noise. We compare our approach with the adaptive Neyman test. Simulation study suggests that both tests are close to nominal level. Our Adaptive Bayes Sum Test is more powerful in detecting both small and large difference, while adaptive Neyman test is not. Indeed, by making use of wavelets filters, the

Bayes Sum Test can filter out the components of interest. For example, in our case, low-frequency components are more important and hence must be given a higher weight. In contrast, the adaptive Neyman test treats every data point uniformly.

Further research is still needed to examine the theoretical properties in detail. Deriving theoretical distributions of resampling procedures will be useful to reduce computation burden as well as to study the theoretical bound of the difference between two functions. Although we found that our method can detect well small difference between two functions in special case, it is difficult to derive the bound of this difference in general case. It will also important to develop multiple comparison method using our approach which may be a challenging problem because of unknown dependence structures among data.

Chapter 3

Hybrid Omnibus Test in Generalized Semi-parametric Single Index Model with Low/High-Dimensional Covariate

3.1 Introduction

The variable selection and testing problem play important role in many research areas such as genomics, data mining, image analysis, text mining, and other areas with high-dimensional data.

Numerous statistical methods have been developed for analyzing high-dimensional data (Efron et al., 2004; Fan & Li, 2001; Tibshirani, 1996; Zou & Hastie, 2005). Such methods are mainly focused on variable selection and are based on various penalized regression models. One of them is Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Tibshirani (1996). LASSO was based on L1 norm penalty on the estimator in the estimating algorithm, which proves efficient in reducing insignificant predictors when the number of variables is moderately larger than the number of observations. However the estimators based on LASSO are biased estimator. Fan & Li (2001) developed a special penalty function that was singular at the origin to produce sparse coefficient estimators. Their penalty function was called smoothly clipped absolute deviation (SCAD). The estimators based on SCAD are unbiased estimators and have sparsity and continuity properties. Zou & Hastie (2005) proposed an elastic net penalty (ENET), which is a linear combination of both L1-norm and L2-norm with convex constraint, thus having the characteristics of both the lasso and ridge regression. They are mainly interested in variable selection, not testing.

For the purpose of testing, those statistical methods are very limited for high-dimensional data. Verzelen & Villers (2010) developed a method for testing high-dimensional goodness-of-fit. However this method was developed assuming a linear model with Gaussian error. Our approach is developed in the semi-parametric framework where likelihood function is no longer necessary. We propose a test statistic, called "Hybrid Omnibus Test" for high-dimensional data with single index model.

Carroll et al. (1997) proposed a framework for estimating generalized partially linear single-index models; it allows for unknown single index function, in low-dimensional case. Carroll et al. (1997) focused on estimating, not testing. Although the single-index model can be considered to be one of the most convenient way in handling the "curse of dimensionality", it is still difficult to handle high-dimensional data. Upon launching the omnibus test for the single index, we stabilize the procedure with a hybrid Frequentist and Bayesian method. The moethod involves a redesigned efficient score without the requirement for the explicit likelihood. The idea for this type of test (Hart, 2009; Ma et al., 2011) was used for traditional measurement error model in low-dimension case.

In this chapter, we propose Freqentist-Bayesian hybrid score-type test for high dimensional case. We refer to our test as "Hybrid Omnibus Test" (HOT). Our HOT development starts from a generalized partial linear single index model where the dependent variable is modeled with link being a function of a generalized partially linear single index frame of the independent variables. The goal of our study is to develop the HOT with the following features: (a) our HOT is based on an estimating equation that is not necessarily an explicit likelihood function; (b) it obtains high computational efficiency relative to the likelihood-ratio-style tests for generalized partial single index models; (c) it achieves a reasonable true positive rate (TPR) and false positive rate (FPR); (d) it is robust to different estimating algorithms; (e) it applies when either $p > n$ or $p >> n$, where n is the sample size and p is the number of parameters. To the best of our knowledge, no current test statistics have all these features.

An example to illustrate these features is when Y is $n \times 1$ binary response, X is a $n \times p$ design matrix for predictor, Z is a $n \times q$ low-dimensional predictor matrix, logistic regression is used. Consider that we want to test whether the effect of the predictor X is constant. Let $H(\cdot)$ be the logistic function and $f(\cdot)$ be a known parametric function of Z. Thus for example, the null model would be

$$\Pr(Y = 1|X,Z) = H\{\kappa_0 + f(Z,\boldsymbol{\gamma})\} \tag{3.1}$$

whereas, for a local test, an alternative model allowing for non-parametric/polynomical departures from constant κ_0 might be

$$\Pr(Y = 1|X,Z) = H\{\kappa_0 + h(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}, \tag{3.2}$$

where $h(X\boldsymbol{\alpha})$ is a linear/polynomical function with parameters $\boldsymbol{\alpha}$. For an omnibus test, an alternative allowing any departures of X from constant κ_0 might be

$$\Pr(Y = 1|X,Z) = H\{\kappa_0 + g(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} \tag{3.3}$$

for the unspecified global function $g(\cdot)$. Models of type (3.3) are referred to as single-index models (Stoker, 1986; Härdle & Stoker, 1989; Ichimura, 1993).

Model (3.1) and (3.2) are easy to make inference using either sufficient scores or likelihood ratio. However, model (3.3) is not trivial because $\boldsymbol{\alpha}$ is unknown and $g(\cdot)$ is unknown function. We propose score type tests, fitting the model only under the null hypothesis to obtain estimates of $\boldsymbol{\gamma}$, and then constructing a test statistic based on the estimating equations (Tsiatis & Ma, 2004; Ma & Carroll, 2006; Ma et al., 2011). Our test is established avoiding parameter estimation for $g(\cdot)$ but is needed to find any feasible solution of estimating equation of $\boldsymbol{\alpha}$. Hence our test statistic is characterized only under the reduced model. Therefore, it is in this way that our approach is similar to scores tests. However, there is a crucial difference from the ordinary score test from likelihood theory: the structure of the general functional semi-parametric model does not involve estimating the non-parametric components directly; no profiles semi-parametric likelihood is available on which to base a score test. Our approach is developed in the semi-parametric model framework where likelihood function is not necessary. We propose an efficient score which does not require to derive likelihood and then develop our Hybrid Omnibus Test using this efficient score.

3.2 Semi-parametric Framework

Let Y be a $n \times 1$ response vector variable which depends on predictor $n \times p$ matrix X through non-parametric function $g(\cdot)$ and predictor $n \times q$ matrix Z through parametric function $f(\cdot)$ with unknown parameter $\boldsymbol{\gamma}$. We consider either $n > p$ or $n < p$ but assume that $n > q$.

Let $p_{Y|X,Z}$ be represent the model for Y given (X,Z). Then the null model is

$$p_{Y|X,Z}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z,\boldsymbol{\gamma})\}$$

for a local test, supposing $h(\cdot) = \sum_{j=1}^{J} h_j(\cdot)\beta_j$ consists of J bases functions $h_j(\cdot)$, where $h_j(\cdot)$ can be linear/polynomical function depending on testing, an alternative model allowing for linear/polynomical departure from constant κ_0

$$p_{Y|X,Z}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + h_j(X\boldsymbol{\alpha})\beta_j + f(Z,\boldsymbol{\gamma})\}.$$

For an omnibus test, an alternative allowing any departure of X from constant κ_0 would be

$$p_{Y|X,Z}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + g(X\boldsymbol{\alpha},\boldsymbol{\beta}) + f(Z,\boldsymbol{\gamma})\}.$$

for an unspecified function $g(\cdot)$.

For identifiability, $||\boldsymbol{\alpha}|| = 1$ is usually assumed. The benefit of single index models is to make high-dimensional predictor convenient to be handled properly using $X\boldsymbol{\alpha}$. For $||\boldsymbol{\alpha}|| = 1$, we use the following parametrization. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_p)^T$, $p \geq 3$.

$$\begin{aligned}
\alpha_1 &= \sin(\phi_1);\\
\alpha_2 &= \cos(\phi_1)\sin(\phi_2);\\
&\vdots\\
\alpha_{p-1} &= \cos(\phi_1)...\cos(\phi_{p-2})\sin(\phi_{p-1});\\
\alpha_p &= \cos(\phi_1)...\cos(\phi_{p-1}),
\end{aligned} \tag{3.4}$$

where $-\pi/2 \leq \phi_j \leq \pi/2$, $j = 1,\ldots,p-1$. The advantages of this parametrization is a) and an identifiable model and b) the parameter ϕ_j has a finite range for all j.

For the logistic case, a generalized semi-parametric single index model can be written as

$$p_{Y|X,Z}(Y|X,Z) = \Pr(Y=1|X,Z) = H\{\kappa_0 + g(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}, \tag{3.5}$$

where $\boldsymbol{\alpha}_{1\times p}$ is the single index parameter.

Because we do not specify an unknown function $g(\cdot)$, we do not have an exact likelihood. On the other hand, if one approximates an unknown function $g(\cdot)$ using wavelet basis functions or splines, one can obtain an approximate likelihood. For example, $g(\cdot)$ can be approximated by splines (Yu & Ruppert, 2002) with K knots:

$$g(X\boldsymbol{\alpha}) = \sum_{j=1}^{p} \beta_j (X\boldsymbol{\alpha})^j + \sum_{k=1}^{K} \beta_{k+p}(X\boldsymbol{\alpha} - \kappa_k)_+^p,$$

where κ_k is kth knot, $1 \leq k \leq K$, and K is the total number of knots, but not limited to splines. For example, a cosine basis can be also applicable, $\cos\{k\pi(X\boldsymbol{\alpha} - x_l)/(x_r - x_l)\}$ for $k = 1, \ldots, K$, where x_l and x_r are the estimated left-hand and right-hand limit of the support of the distribution of $X\boldsymbol{\alpha}$. Under this setting, the null hypothesis becomes whether $\boldsymbol{\alpha} = \mathbf{0}$ or $\boldsymbol{\beta} = \mathbf{0}$. However H_0: $\boldsymbol{\beta} = \mathbf{0}$ is more desirable which does not suffer from the identifiable problem as when H_0: $\boldsymbol{\alpha} = \mathbf{0}$, which is contrast with identifiability condition $\|\boldsymbol{\alpha}\| = 1$.

However, in this chapter, we does not require to estimate $g(\cdot)$ so that the likelihood is not available. Under this framework, we propose a test procedure based on our Hybrid Omnibus Test (HOT).

3.3 Hybrid Omnibus Test (HOT)

Explicit likelihoods may neither be available nor necessary because the actual form of $g(\cdot)$ is unknown. Thus, we may not be able to derive an explicit likelihood. Under this framework we then consider a “score-type” test, which requires neither additional likelihood computation nor full parameter utilization for the full model with the unspecified function $g(\cdot)$.

3.3.1 Score-type Test

Consider a semi-parametric logistic regression models,

$$\mathrm{pr}(y|X,Z) = H\{\kappa_0 + \sum_{j=1}^{J} \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}, \tag{3.6}$$

Suppose that we select J bases $h_j(\cdot)$, $j = 1\ldots,J$ and express a linear form $g(\cdot) = \sum_j^J \beta_j h_j(\cdot)$. Using a likelihood function, we can then obtain a score function for $\boldsymbol{\beta} = (\beta_1,...,\beta_J)$,

$$\boldsymbol{h}(X\boldsymbol{\alpha})^T \left\{ \boldsymbol{y} - \frac{\exp(Z\boldsymbol{\gamma})}{1+\exp(Z\boldsymbol{\gamma})} \right\},$$

where $\boldsymbol{h}(\cdot) = [h_1(\cdot),\ldots,h_J(\cdot)]$.

For general model, we can write the model as the follows,

$$E(y|X,Z) = \mathscr{H}\{G(X;\alpha,\beta,\gamma) + Z\gamma\},$$

with link function $\mathscr{H}(\cdot)$, unknown function $G(\cdot)$ and linear predictor Z with associated coefficient γ. In general, score-type function can be derived from estimating equations rather than strictly likelihood function with similar characteristic to original score function. The estimating equations can be written as

$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\beta}}(x_i,z_i,y_i,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = 0$$
$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\gamma}}(x_i,z_i,y_i,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = 0$$
$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\alpha}}(x_i,z_i,y_i,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = 0,$$

where estimating functions $\Phi_{\boldsymbol{\beta}}(\cdot)$, $\Phi_{\boldsymbol{\gamma}}(\cdot)$, $\Phi_{\boldsymbol{\alpha}}(\cdot)$ have the same dimensions as $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$, respectively. We note that these estimating functions are not necessarily derivatives of some version of any profile likelihood. A profile likelihood hardly exists in our semi-parametric framework.

Under the null hypothesis, the estimating equations are simply

$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\gamma}}(x_i, z_i, y_i, \boldsymbol{\alpha}, 0, \boldsymbol{\gamma}) = 0$$
$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\alpha}}(x_i, z_i, y_i, \boldsymbol{\alpha}, 0, \boldsymbol{\gamma}) = 0.$$

We call the roots of these estimating equations as $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\gamma}}$.

However, we can not use this $\sum_{i=1}^{n} \Phi_{\boldsymbol{\alpha}}(x_i, z_i, y_i, \boldsymbol{\alpha}, 0, \boldsymbol{\gamma}) = 0$ for $\hat{\boldsymbol{\alpha}}$ because of identifiability between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Instead, we use one possible root of any feasible solutions of $\boldsymbol{\alpha}$ from

$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\alpha}}(x_i, z_i, y_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = 0.$$

To find root $\hat{\boldsymbol{\alpha}}$ we need to estimate all parameters under the full model, but we do not use these parameter estimates except $\boldsymbol{\alpha}$. They are only served as auxiliary parameters for obtaining $\hat{\boldsymbol{\alpha}}$ and are not used for testing.

From the form of the score test, we propose estimated score,

$$\hat{U} = n^{-1/2} \sum_{i=1}^{n} \Phi_{\boldsymbol{\beta}}(x_i, z_i, y_i, \hat{\boldsymbol{\alpha}}, 0, \hat{\boldsymbol{\gamma}}).$$

Analyzing $\hat{U}$ is not difficult. We first make the following definitions, with all expectations taken under the null hypothesis:

$$\begin{aligned}
A_1 &= E\{\partial\Phi_{\boldsymbol{\gamma}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T\},\\
A_2 &= E\{\partial\Phi_{\boldsymbol{\beta}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T\},\\
A_3 &= E\{\partial\Phi_{\boldsymbol{\gamma}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})/\partial\boldsymbol{\beta}^T\},\\
A_4 &= E\{\partial\Phi_{\boldsymbol{\beta}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})/\partial\boldsymbol{\beta}^T\},\\
B_{11} &= E\{\Phi_{\boldsymbol{\gamma}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})\Phi_{\boldsymbol{\gamma}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})^T\},\\
B_{22} &= \mathrm{cov}\{\Phi_{\boldsymbol{\beta}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})\},\\
B_{12} &= \mathrm{cov}\{\Phi_{\boldsymbol{\gamma}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma}),\Phi_{\boldsymbol{\beta}}(X,Z,Y,\boldsymbol{\alpha},0,\boldsymbol{\gamma})^T\}\\
V_{\boldsymbol{\beta}} &= A_1^{-1}B_{11}(A_1^{-1})^T,\\
\Sigma_0 &= \mathrm{cov}\{\Phi_{\boldsymbol{\beta}}(\cdot,\boldsymbol{\alpha},\boldsymbol{\gamma},0)-A_2A_1^{-1}\Phi_{\boldsymbol{\gamma}}(\cdot,,\boldsymbol{\alpha},\boldsymbol{\gamma},0)\},
\end{aligned}$$

and further define following matrices

$$\mathscr{A}=\begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}, \quad \mathscr{B}=\begin{pmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{pmatrix}, \quad \mathscr{V}=\mathscr{A}^{-1}\mathscr{B}(\mathscr{A}^{-1})^T=\begin{pmatrix} V_{11} & V_{12} \\ V_{12}^T & V_{22} \end{pmatrix}.$$

All these quantities can be estimated by replacing expectations and covariance matrices by their sample versions. We denote the resulting sample estimate of Σ_0 by $\hat{\Sigma}_0$.

Theorem 1. Under the null hypothesis, $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\to N(0,V_{\boldsymbol{\beta}})$ and $\hat{U}\to N(0,\Sigma_0)$. Hence $T=\hat{U}^T\hat{\Sigma}_0^{-1}\hat{U}$ is asymptotically χ^2 with p_T degree of freedom.

3.3.2 Construction of HOT

In this section we explain how we derive Hybrid Omnibus Test (HOT) without specifying an unknown function $g(x)$. It is based on the idea that a smooth function can be approximated arbitrarily well by a linear combination of sufficiently many basis functions such as splines, wavelet bases, and so on. For any fixed system of basis functions, a departure from the null model in the direction of a given basis function is a problem of local testing. However, because departures along different direction needs to be considered and multiple testing need to be avoided, local tests need to be combined. Hart (2009) proposed a goodness-of fit test that is a hybrid of Bayesian and Frequentist idea. Here we use this idea and develop it to the general semi-parametric framework.

Consider a set of basis functions $h_1(X\boldsymbol{\alpha})$, $h_2(X\boldsymbol{\alpha})$,…, $h_J(X\boldsymbol{\alpha})$, which are arranged from lowest to highest frequency. Let $p_{Y|X,Z}$ be represent the model for Y given (X,Z). We consider J different local tests with the following null hypothesis (H_0) and the jth alternative hypothesis (H_{j1}), $j = 1 \ldots J$:

$$\begin{aligned} H_0: &\quad p_{Y|X,Z}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z,\boldsymbol{\gamma})\} \\ H_{j1}: &\quad p_{Y|X,Z}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}. \end{aligned}$$

This test is equivalent to H_0: $\beta_j = 0$ vs H_{j1}: $\beta_j \neq 0$. For example of the logistic model, $p_{Y|X,Z}(Y|X,Z) = H\{\kappa_0 + f(Z,\boldsymbol{\gamma})\}$ under H_0 and $H\{\kappa_0 + \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}$ under H_{j1}, respectively. We can do a score-type test according to Theorem 1: For each of the $h_j(\cdot)$, the test statistic is then can be written as $T_j^2 = u_j^2/\sigma_j^2$, where u_j and σ_j^2 are the one-dimensional version of $\hat{U}_j$ and $\hat{\Sigma}_{0j}$ respectively, and T_j and Σ_{0j} are defined in Theorem 1.

Let $\hat{\boldsymbol{T}}_{\text{HOT}} = (\hat{T}_1, \ldots, \hat{T}_J)^T$. Using results of Ma et al. (2011), we can show that the joint limit distribution is $n^{-1/2}\hat{\boldsymbol{T}}_{\text{HOT}} \sim N(\mathbf{0}, \Sigma)$ asymptotically, where the (i,j)th element of Σ can be derived as

$$E\left(\frac{\hat{u}_j}{\hat{\sigma}_j}\frac{\hat{u}_k}{\hat{\sigma}_k}\right) \doteq \frac{1}{\sigma_j\sigma_k}[A_{j2}A_1^{-1}B_{11}A_1^{-1T}A_{k2}^T - A_{j2}A_1^{-1}B_{k12} - A_{k2}A_1^{-1}B_{j12} + E\{\Phi_{\beta_j}(\cdot;\kappa_0,0)\Phi_{\beta_k}(\cdot;\kappa_0,0)\}] \quad (3.7)$$

because $E(\hat{u}_j/\hat{\sigma}_j) = 0$ for any j. This also means that the marginal limit distribution is $\hat{u}_j/\hat{\sigma}_j \to N(0,1)$ in distribution. Our score-type function is

$$\hat{U} = n^{-1/2}\sum_{i=1}^{n}\Phi_{\boldsymbol{\beta}}(x_i, z_i, y_i, \hat{\boldsymbol{\alpha}}, 0, \hat{\boldsymbol{\gamma}}).$$

We then propose Hybrid Omnibus Test using local tests. We combine the local test statistic to build an omnibus test like the F-test, even though it is difficult to obtain a closed form distribution or asymptotic distribution. Following the approach taken to obtain a χ^2 from the sum of square normal, we design our HOT statistic which can be expressed as

$$\hat{\mathscr{T}}_{HOT} = \sum_{j=1}^{J}\omega_j \exp\left(\frac{\hat{T}_j^2}{2}\right),$$

where ω_j is a weight resulting from the prior (π_j) of the jth local test statistic, The purpose of this weight term is that low frequency components contain more global features than high frequency components contain. Assuming that each of the indices j is proportional to the local test statistic associated to the corresponding frequency component, it is reasonable to specify the prior of the corresponding local test statistic as follows,

$$\pi_j = \frac{1}{1+(j)^c},\ j = 1,\ldots,J,$$

where $c > 1$ for the local test statistics.

Our weight ω_j is $\pi_j/(1-\pi_j)$. With this weight, our HOT statistic can be rewritten as

$$\hat{\mathscr{T}}_{HOT} = \sum_{j=1}^{J} \frac{\pi_j}{1-\pi_j} \exp\left(\frac{\hat{T}_j^2}{2}\right).$$

Note that Hart (2009) showed that this test statistic was proportional to the Bayes Factor.

With the selection of $c = 2$, we have $\omega_j = 1/j^2$. Our HOT statistic, $\hat{\mathscr{T}}_{HOT}$,

$$\hat{\mathscr{T}}_{HOT} = \sum_{j=1}^{J} \frac{1}{j^2} \exp\left(\frac{\hat{T}_j^2}{2}\right). \tag{3.8}$$

The power of the omnibus test against a local test can be obtained from the result that $\hat{\boldsymbol{T}}_{\mathrm{HOT}}$ has an asymptotically multivariate normal distribution. This result implicitly determines the asymptotic power of our omnibus test against root n local alternatives. In practice, this power can be approximated by generating samples from $N(0,\Sigma)$ and comparing them with an observed value. This procedure is explained in Section 3.5.1.

3.4 HOT with Constraint Estimating Equation

Solving estimating equations for coefficients involve numerical problems such as convergence issue or matrix inverting issue for small sample size because of the curse of dimensionality. Hence in this section, we propose two remedy approaches to solve these problems by introducing constraint estimation equations. One is based on constraint estimation equation on $\boldsymbol{\beta}$ and the other is based on constraint estimation equation on $\boldsymbol{\alpha}$. They are described in Section 3.4.1 and section 3.4.2, respectively.

3.4.1 HOT with Constraints on Non-parametric Basis Coefficients

We solve estimating equation subject to constraints on $\boldsymbol{\beta}$. The constraint estimating equation can be written as

$$\sum_{i=1}^{n} \Phi_{\boldsymbol{\beta}}(x_i, z_i, y_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = 0, \quad \text{subject to} \quad \|\boldsymbol{\beta}\|^2 < \Lambda. \tag{3.9}$$

We note that we do not need to estimate $\boldsymbol{\beta}$. However, it affects the estimates of $\boldsymbol{\alpha}_{\mathbf{root}}$, thus the stability of $\boldsymbol{\beta}$ is still necessary.

In practice, when likelihood is available, HOT can be performed using the following estimating equation obtained from the penalized likelihood under the logistic regression

$$\Phi_{\boldsymbol{\beta}}(X, Z, Y, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}},$$

where

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_i(X\boldsymbol{\alpha}) + f(Z, \boldsymbol{\gamma})\} - \mathbf{1}^T \log[\mathbf{1} + \exp\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_i(X\boldsymbol{\alpha}) + f(Z, \boldsymbol{\gamma})\}] - \lambda \|\boldsymbol{\beta}\|^2.$$

Equivalently, the penalized likelihood can be treated as a mixed effect model with the penalty as a solvable parameter for non-parametric basis:

$$p_{Y|X,Z}(Y|X, Z) = p_{Y|X,Z}\{Y, \kappa_0 + h_j(X\boldsymbol{\alpha})u_j + f(Z, \boldsymbol{\gamma})\}, \tag{3.10}$$

where

$$U = (u_1, ..., u_K)^T \sim N(0, \sigma_u^2 I).$$

Therefore, the estimating equation can be obtained via penalized quasi-likelihood (PQL) which is

$$\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} - \mathbf{1}^T\log[\mathbf{1}+\exp\{\kappa_0 + \sum_{j=1}^{p+K}u_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}] - \frac{1}{2}U^T G^{-1}U,$$

where $G = \sigma_u^2 I$.

3.4.2 HOT with Shrinkage Constraints on Single Index Coefficients

Although any one feasible solution of $\boldsymbol{\alpha}_{\text{root}}$ is suitable, we found that convergence problems often exist with multiple solutions because of the curse of dimension. The algorithm may jump from one solution to another which cause multiple solutions and instability. The identification constraint $\|\boldsymbol{\alpha}\|^2 = 1$ has a good performance in simulations. But it is necessary to check if we can improve with stronger penalty for faster convergence or better stability. Hence we consider the L_1 shrinkage constraints on single index coefficients. In this case, the constraint estimating equation can be obtained from the following penalized likelihood

$$\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} - \mathbf{1}^T\log[\mathbf{1}+\exp\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}] - \lambda\|\boldsymbol{\alpha}\|,$$

subject to $\|\boldsymbol{\alpha}\|^2 = 1$.

3.5 HOT with Bayesian Posterior Estimation

HOT can be also implemented with fully Bayesian framework. We can assign non-informative priors $\pi(\boldsymbol{\alpha},\boldsymbol{\beta}) \propto 1, \pi(\boldsymbol{\gamma}) \propto 1$. We use Metropolis-Hastings (MH) algorithm to obtain empirical posterior distribution $\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)$. Then the Maximum Posterior Estimator (MAP) is obtained by solving the following estimating equation

$$\Phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{\partial \pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)}{\partial(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}.$$

Since $\pi(\boldsymbol{\alpha},\boldsymbol{\beta}) \propto 1, \pi(\boldsymbol{\gamma}) \propto 1$, the posterior distribution is proportional to the likelihood,

$$\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z) \propto L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)$$

therefore, the estimating equation is

$$\Phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{\partial \pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)}{\partial(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} = C \times \frac{\partial L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)}{\partial(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}.$$

where C is a constant due to the non-informative flat prior $\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) \propto 1$.

For the log-transformation of the posterior, it turns out that

$$\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z) \propto \ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z) + C$$

and

$$\Phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{\partial \log(\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z))}{\partial(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} = \frac{\partial \ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)}{\partial(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})},$$

which has the same form to the HOT with MLE, respectively.

Therefore the procedure to construct HOT is exactly the same as that described in Section 3.3.2, because the local test statistics are invariant to C. Hence if equation (3.9) is used with a non-informative flat prior and fixed penalty, the penalized MAP is equivalent to penalized maximum likelihood estimator. We can obtain the acceptance rate of Metropolis-Hastings using PQL in equation (3.10) when likelihood is available.

3.5.1 Algorithm for HOT

With the semi-parametric framework, our HOT algorithm proceeds in the following steps:

Step H0: Perform Fitting steps;

- Step F1: Select non-parametric basis for the approximation function $\boldsymbol{h}(\cdot) = \{h_1(\cdot), ..., h_J(\cdot)\}$ and approximation function $\tilde{g}(X\boldsymbol{\alpha}) = \boldsymbol{h}(X\boldsymbol{\alpha})\boldsymbol{\beta} = \sum_{j=1}^{J} \beta_j h_j(X\boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_J)^T$.
- Step F2: Select estimating equations $\Phi_{\boldsymbol{\alpha}}$,$\Phi_{\boldsymbol{\beta}}$, and $\Phi_{\boldsymbol{\gamma}}$ corresponding to $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$:
- Step F3: Select solving algorithm for the estimating equations using Frequentist and Bayesian approaches:
  - For Frequentist approach, Interior Trust Region (Coleman & Li, 1996) algorithm is selected for solving the estimating equations with non-linear constraint; The term “Trust Region” is used in mathematical optimization, which denotes the subset of the region of the objective function to be optimized and the objective function is approximated using a model function; If we found an adequate model of the objective function within the Trust Region, then we expand the region; otherwise, if the adequate model approximation is poor then the region is contracted; This method is selected because it does not require a quadratic programming subproblem under inequality constraints.
  - For Bayesian approach, Metropolis-Hastings sampler is implemented to obtain a sample from posterior distribution, and corresponding summary, such as MAP.
- Step F4: With selected estimating equations of $\Phi_{\boldsymbol{\alpha}}$, $\Phi_{\boldsymbol{\beta}}$ and $\Phi_{\boldsymbol{\gamma}}$, the root, $\hat{\boldsymbol{\alpha}}_{\text{root}}$ is obtained under the model $\text{pr}(Y|X,Z) = H\{\kappa_0 + \sum_{i=1}^{J} \beta_j h_j(X\boldsymbol{\alpha}) + f(Z, \boldsymbol{\gamma})\}$. Dur-

ing this step, we do estimate parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ under the full model and denote them as $\hat{\boldsymbol{\beta}}_{aux}$ and $\hat{\boldsymbol{\gamma}}_{aux}$. We use them to obtain a root $\hat{\boldsymbol{\alpha}}_{\text{root}}$ in order to avoid the identifiable problem. Both $\hat{\boldsymbol{\beta}}_{aux}$ and $\hat{\boldsymbol{\gamma}}_{aux}$ are only served as auxiliary parameter estimates for $\hat{\boldsymbol{\alpha}}_{\text{root}}$ and are not involved in HOT statistic;

Step F5: With $\Phi_{\boldsymbol{\gamma}}$, the root $\hat{\boldsymbol{\gamma}}_{\text{root}}$ is obtained, under the reduced model $\text{pr}(Y|Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z,\boldsymbol{\gamma})\}$;

Step H1: Perform Testing steps:

Step T1: Based on the roots $\hat{\boldsymbol{\alpha}}_{\text{root}}$ and $\hat{\boldsymbol{\gamma}}_{\text{root}}$, we define $\hat{\boldsymbol{T}}_{\text{HOT}} = (\hat{T}_1, ..., \hat{T}_J)^T$, where $\hat{T}_j$ is obtained from the following loop

do $j = 1 : J$

$\{\hat{T}_j = \hat{u}_j^2/\hat{\sigma}_j^2$ with jth local test$\}$;

Step T2: Let $\hat{\Sigma}$ be the covariance estimator of $\hat{\boldsymbol{T}}_{\text{HOT}}$. The (j,k) element of $\hat{\Sigma}$ can be derived from formula (3.7) with all terms replaced by its estimator. Then $\hat{\mathscr{T}}_{\text{HOT}} = \sum_{j=1}^{J} j^{-2} \exp(\hat{T}_j^2/2)$;

Step T3: Obtain local test statistic $\boldsymbol{T}_b$ under H_0 and then calculate HOT test statistic $\mathscr{T}_{b,HOT}$ by following loop

do $b = 1 : B$

$\{$Generate $\boldsymbol{T}_b \sim N(\mathbf{0}_{J\times 1}, \hat{\Sigma})$, and $\mathscr{T}_{b,HOT} = \sum_{j=1}^{J} j^{-2} \exp(T_{bj}^2/2)\}$;

Step T4: The p-value is announced as $p = B^{-1}\sum_{b=1}^{B} I\{\mathscr{T}_{b,HOT} > \hat{\mathscr{T}}_{\text{HOT}}\}$.

3.6 Empirical Likelihood Ratio Test (ELRT)

Alternative test can be likelihood ratio based test. Since the function $g(\cdot)$ is unknown, we can approximate it using spline basis or other basis with easier estimating process. Spline is one of possible choices to proceed such a test. However, under non-parametric circumstances, especially in single index model, the degree freedom for the likelihood remains in darkness essentially for the non-linear frame which prevents us from deriving a convincing asymptotic characteristic. Due to the difficulty in theory with degree of freedom, simulation based test is necessary with the alternative choice to obtain empirical distribution.

First we estimate model under H_0 :

$$\mathrm{pr}(Y|Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z, \boldsymbol{\gamma})\}$$

and the model using splines under H_1 :

$$\mathrm{p}(Y|X,Z) = p_{Y|X,Z}\{Y, \kappa_0 + \sum_{j=1}^{J} \beta_j h_j(X\boldsymbol{\alpha}) + f(Z, \boldsymbol{\gamma})\}$$

and then obtain the likelihood ratio T_{obs} using the data.

Next we generate Y under the null hypothesis. Using these generated data, we estimate models under H_0 and H_1 and then obtain the likelihood ratio $T_{\mathrm{emprical},b}$, where $T_{\mathrm{emprical},b} = \ell(\hat{\boldsymbol{\gamma}}_{0b}) - \ell(\hat{\boldsymbol{\alpha}}_{1b}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\gamma}}_{1b})$ is the likelihood ratio based on the bth generated data, $\ell(\hat{\boldsymbol{\gamma}}_{0b})$ is the loglikelihood under H_0, and $\ell(\hat{\boldsymbol{\alpha}}_{1b}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\gamma}}_{1b})$ is the loglikelihood under H_1. Then we can obtain the empirical distribution of T_{emprical} where the null hypothesis is true. We can expect that the T_{obs} will be most likely to be larger that the empirical likelihood ratio test (ELRT) when X is important variables. Hence we can construct such empirical likelihood ratio test by comparing T_{obs} with $T_{\mathrm{empirical}}$. We then calculate the empirical power and Type I error rate by simulation based approach.

3.6.1 Algorithm for ELRT

The algorithm for ELRT is summarized as follows:

Step E1 Using observed data, estimate $\hat{\boldsymbol{\gamma}}_{0obs}$ under H_0 and $(\hat{\boldsymbol{\alpha}}_{1obs}, \hat{\boldsymbol{\beta}}_{1obs}, \hat{\boldsymbol{\gamma}}_{1obs})$ under H_1;

Step E2 Calculate $T_{obs} = \ell(\hat{\boldsymbol{\gamma}}_{0\text{obs}}) - \ell(\hat{\boldsymbol{\alpha}}_{1\text{obs}}, \hat{\boldsymbol{\beta}}_{1\text{obs}}, \hat{\boldsymbol{\gamma}}_{1\text{obs}})$, where $\ell(\hat{\boldsymbol{\gamma}}_{0\text{obs}})$ and $\ell(\hat{\boldsymbol{\alpha}}_{1\text{obs}}, \hat{\boldsymbol{\beta}}_{1\text{obs}}, \hat{\boldsymbol{\gamma}}_{1\text{obs}})$ are loglikelihood under H_0 and H_1, respectively;

Step E3 Perform the following loop to obtain empirical likelihood ratio test statistic under H_0

do $b = 1 : B$

{Generate Y_{ib} for $i = 1, ...n$, based on $\Pr(y|Z, \hat{\boldsymbol{\gamma}}_{0b})$, calculate $T_{\text{emprical},b} = \ell(\hat{\boldsymbol{\gamma}}_{0b}) - \ell(\hat{\boldsymbol{\alpha}}_{1b}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\gamma}}_{1b})$ };

Step E4 The p-value is announced as $p = \sum_{b=1}^{B} I(T_{\text{emprical},b} > T_{\text{obs}})$.

3.7 Approximate Bayes Factor (ABF)

Another alternative approach is to use Bayes factor under fully Bayesian framework. The difficulty for ELRT is the computational burden because we have to obtain empirical distribution of the likelihood ratio test using resampling procedure and iterative estimating algorithm. To avoid these computational burden, we can use approximated Bayes factor which provides less computation. Bayes factor also requires to calculate integrals using samples from the posterior distribution. Numerical computation for integral is still needed but the computation burden is reduced because we substituted resampling procedure and iterative estimating algorithm. When likelihood is available, we can sample from the posterior distribution using Metropolis-Hastings algorithm.

Given a model selection problem in which we have to choose between two models on the basis of observed data **D**, the plausibility of the two different models, and in our context, the model under null-hypothesis and the model under alternative hypothesis, which are $\mathscr{M}_{\text{H}_0}$ and $\mathscr{M}_{\text{H}_1}$ parameterized by model parameter vectors θ_{H_0} and θ_{H_1}, is assessed by the Bayes factor B given by

$$B = \frac{\Pr(\mathbf{D}|\mathscr{M}_{\mathrm{H}_1})}{\Pr(\mathbf{D}|\mathscr{M}_{\mathrm{H}_0})} = \frac{\int \Pr(\theta_{\mathrm{H}_1}|\mathscr{M}_{\mathrm{H}_1}) \Pr(\mathscr{M}|\theta_{\mathrm{H}_1}, \mathscr{M}_{\mathrm{H}_1}) d\theta_{\mathrm{H}_1}}{\int \Pr(\theta_{\mathrm{H}_0}|\mathscr{M}_{\mathrm{H}_0}) \Pr(\mathscr{M}|\theta_{\mathrm{H}_0}, \mathscr{M}_{\mathrm{H}_0}) d\theta_{\mathrm{H}_0}},$$

where $\Pr(\theta_{\mathrm{H}_l}|\mathscr{M}_{\mathrm{H}_l})$ is the prior distribution and $\Pr(\mathscr{M}|\theta_{\mathrm{H}_l}, \mathscr{M}_{\mathrm{H}_l})$ is the likelihood function, $l = 0, 1$.

In general Bayes factor calculation is also tricky because we cannot obtain a explicit distribution for posterior distribution. However if we use samples from posterior distribution using both Metropolis-Hastings algorithm and $\pi(\alpha, \beta, \gamma|X, Z) \propto L(\alpha, \beta, \gamma|X, Z)$, we can approximate Bayes factor using the algorithm described in Section 3.7.1. We implemented volume tessellation algorithm (VTA) proposed by Weinberg (2012) in order to estimate approximate Bayes Factor. We further explain how to calculate this ABF in Appendix C.

3.7.1 Algorithm for ABF

With the semi-parametric framework, the ABF algorithm with flat non-informative prior proceeds the following steps:

Step 00: Perform posterior sampler steps;

- Step F1: Select non-parametric basis for approximation function $\boldsymbol{h}(\cdot) = \{h_1(\cdot), ..., h_J(\cdot)\}$ and approximation function $\tilde{g}(X\boldsymbol{\alpha}) = \boldsymbol{h}(X\boldsymbol{\alpha})\boldsymbol{\beta} = \sum_{j=1}^{J} \beta_j h_j(X\boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_J)^T$;
- Step F2: With selected basis, obtain likelihoods under H_0 and H_1;
- Step F3: Initial α, β, γ correspondingly under H_1 and γ under H_0;
- Step F4: We obtain the posterior sample of (α, β, γ) from the following Metropolis-Hastings samplers with denoting I as the total number of Metropolis-Hastings

iterations and i as the ith Metropolis-Hastings iteration;

```
While    $i < I$
• Sample new $\boldsymbol{\alpha}_{new}$ from proposal distribution $\mathscr{G}_1(\boldsymbol{\alpha}_{i-1} \rightarrow \boldsymbol{\alpha}_{\text{new}})$;
    – Initial flag$_{\boldsymbol{\alpha}} = 0$;
      While flag$_{\boldsymbol{\alpha}} = 0$;
    – Set
          $\text{AR}_1 = \frac{L(\boldsymbol{\alpha}_{\text{new}}, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} | Y, X, Z)\mathscr{G}_1(\boldsymbol{\alpha}_{i-1} \rightarrow \boldsymbol{\alpha}_{\text{new}})}{L(\boldsymbol{\alpha}_{i-1}, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} | Y, X, Z)\mathscr{G}_1(\boldsymbol{\alpha}_{new} \rightarrow \boldsymbol{\alpha}_{\text{i-1}})}$
    – Sample $U_{i1} \sim U(0,1)$;
    – If $U_{i1} < \text{AR}_1$, then $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{new}$, flag$_{\boldsymbol{\alpha}} = 1$;
• Sample $(\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$ from proposed distribution $\mathscr{G}_2(\boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} \rightarrow \boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$;
    – Initial flag$_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 0$;
      While flag$_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 0$;
    – Set
          $\text{AR}_2 = \frac{L(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_{\text{new}}, \boldsymbol{\gamma}_{\text{new}} | Y, X, Z)\mathscr{G}_2(\boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} \rightarrow \boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})}{L(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} | Y, X, Z)\mathscr{G}_2(\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new} \rightarrow \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1})}$
    – Sample $U_{i2} \sim U(0,1)$;
    – If $U_{i2} < \text{AR}_2$, $(\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i) = (\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$, flag$_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 1$;
• Set $i = i + 1$;
```

Step F5: With similar steps, we also obtain posterior distribution sample of γ under the reduced model $\text{pr}(Y|Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z, \boldsymbol{\gamma})\}$;

Step 11: After we obtain samples from posterior distribution, we calculate ABF using VTA which is described in Appendix C.

3.8 Simulation

We conduct several simulation studies to understand the performance of our Hybrid Omnibus Test (HOT). We apply our HOT approach using both non-constraint and constraint estimating equations as well as using Bayesian maximum posterior estimation. We denote them as “HOT”, “CHOT”, “BHOT”. We then compare them with the alternatives, empirical likelihood ratio test (ELRT) and Bayesian inference based on approximated Bayes factor (ABF), in terms of false positive rate (FPR) and true positive rate (TPR). We consider two settings for nonlinear models and two cases in each setting. One is low-dimensional case which has large sample size (n) relative to the number of parameter (p) and the other is high-dimensional case which has relatively small sample size to the number of parameter. We apply our approaches for both low and high dimensional cases, while ELRT is only applicable to low dimensional case because it is very unstable in estimating parameters.

3.8.1 Simulation Setting

We consider two settings: one is Sine-bump function and the other is Polynomial function. The sine-bump function has significant nonlinearity. We would expect to see that our HOT approach dominates the other approaches. The polynomial function was chosen to investigate the loss of efficiency when compared to the correctly parameterized likelihood approach. For each setting, we simulated data for low dimensional case and high dimensional case.

3.8.1.1 Setting 1: Sine Bump Function

Binary predictor Z takes 0 and 1 with half 0’s and half 1’s of the sample size (n), continuous predictors X are generated from normal and uniform distribution for low and high dimensional cases which are the followings:

- Case 1: Low dimension case

  We set $n = 500$ and $p = 3$. Each of X generates from N(0, 1). The true parameters are set as $\boldsymbol{\alpha} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $B = \sqrt{3}/2 + 1.645/\sqrt{12}$, and $\gamma = 0.3$;

- Case 2: High dimension case

  We set $n = 100$ and $p = (100, 150, 200, 250, 300, 500, 800, 1000)$. Each of X generates from U(0, 1). We consider $\boldsymbol{\alpha} = (0, ..., 0, 1, ..., 1)$ with half 0's and half 1's of p and then standardize. We set $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $B = \sqrt{3}/2 + 1.645/\sqrt{12}$, and $\gamma = 0.3$.

Under this setting, binary response Y is generated using the sine-bump function,

$$\text{pr}(Y_i = 1) = H[m \cdot \sin\{\pi(X_i\boldsymbol{\alpha} - A)/(B - A)\} + \gamma Z_i],\ i = 1 \ldots, n$$

where m is an amplifying multiplier; we vary it in range [3,8] with 0.5 increment.

We generate 100 simulated data set.

3.8.1.2 Setting 2: Polynomial Function

We also conduct simulation to compare our HOT with a traditional score test using polynomial model under high dimensional case.

We consider $n = 100$ and p=(100,150,200,250,300,500,800,1000). X was generated from $U(0,1)$. We set $\gamma = 0.3$ and $\boldsymbol{\alpha} = (\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2})$ which represents half 0's and half 1's of p. We then standardize it. The binary response variable is generated from the following polynomial function,

$$\text{pr}(Y_i = 1) = H[m\{-0.9619 + X\alpha + (X\alpha)^2 + (X\alpha)^3 + (X\alpha)^4\} + \gamma Z_i],\ i = 1, \ldots, n$$

where m is an amplifying multiplier; we vary it in range [4.25,6] with 0.5 increment.

We generate 100 simulated data set.

3.8.2 Test Hypothesis and Methods for Testing

Under setting 1 and Case 1, we apply our HOT, CHOT, and BHOT and then compare them with ELRT and ABF. Both HOT and ELRT need to choose bases functions. Under Setting 1 and Case 2, we apply our three approaches and then compare them with ABF. As we mention before, ELRT is very unstable and computational intractable in Case 2 so that we do not provide convergent results.

3.8.2.1 HOT, CHOT, and BHOT

For HOT methods, a set of basis is considered as $h_j(X\boldsymbol{\alpha}) = (X\boldsymbol{\alpha})^j$ for $j = 1, \ldots, p$ and $h_j(X\boldsymbol{\alpha}) = (X\boldsymbol{\alpha} - \kappa_{j-p})_+^p$ for $j = (p+1), \ldots, (p+K)$, where κ_k is kth knot, $1 \le k \le K$, and K is the total number of knots. The smoother is chosen with quadratic order $p = 2$ and knots $K = 20$.

Let $p_{Y|X,Z}$ represent the logistics model for binary response Y given (X,Z). We consider $J = (1 + p + K)$ different local tests with the following null hypothesis and the jth alternative hypothesis, $j = 1 \ldots p + K$:

$$\begin{aligned} H_0: &\quad \Pr(Y|X,Z) = H\{\kappa_0 + f(Z,\boldsymbol{\gamma})\}; \\ H_{j1}: &\quad \Pr(Y|X,Z) = H\{\kappa_0 + \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}. \end{aligned}$$

We now proceed logistic-specified HOT algorithm by following Steps H0-H5;

Step H1: Perform non-parametric fitting steps;

Step F1: Select non-parametric basis for approximation function:

Define

$$\begin{aligned} h_j(X\boldsymbol{\alpha}) &= (X\boldsymbol{\alpha})^j \qquad \text{if} \qquad 1 \le j \le p, \\ h_j(X\boldsymbol{\alpha}) &= (X\boldsymbol{\alpha} - \kappa_{j-p})_+^p \qquad \text{if} \qquad (p+1) \le j \le p+K; \end{aligned}$$

Step F2: Select estimating equations which are the following:

For β and $\boldsymbol{\gamma}$, we have

$$\begin{aligned} \Phi_\beta(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma}) &= \frac{\partial \ell(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma})}{\partial \beta}; \\ \Phi_{\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma}) &= \frac{\partial \ell(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}, \end{aligned}$$

where Φ_β and $\Phi_{\boldsymbol{\gamma}}$ are score functions, which depend on the choice of formula (3.12), (3.13), (3.14), corresponding to HOT, CHOT and BHOT;

For $\boldsymbol{\alpha}$,

$$\hat{\boldsymbol{\alpha}} = \text{argmax}_{\boldsymbol{\alpha}} \ell(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma}) \quad \text{subject to} \quad \|\boldsymbol{\alpha}\|^2 = 1, \tag{3.11}$$

with no explicit expression;

Step F3: Select Interior Trust Region solving algorithm for the estimating equations;

Step F4: Solve equation (3.11) with $\Phi_\beta(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma})$ and $\Phi_{\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma})$ for $\hat{\boldsymbol{\alpha}}_{\text{root}}$;

Solve equation $\Phi_{\boldsymbol{\gamma}}(Y,X,Z,\boldsymbol{\alpha},0,\boldsymbol{\gamma})$ for $\hat{\boldsymbol{\gamma}}_{\text{root}}$;

Step H2: Perform non-parametric fitting steps;

Step T1: Based on the roots $\hat{\boldsymbol{\alpha}}_{\text{root}}$ and $\hat{\boldsymbol{\gamma}}_{\text{root}}$, we define $\hat{\boldsymbol{T}}_{\text{HOT}} = (\hat{T}_1, ..., \hat{T}_J)^T$, where $\hat{T}_j$ is obtained from the following loop

```
        do    $j = 1 : J$
            $\{\hat{T}_j = \hat{u}_j^2/\hat{\sigma}_j^2$ with $j$th local test$\}$;
Step T2: Let $\hat{\Sigma}$ be the covariance estimator of $\hat{\boldsymbol{T}}_{\text{HOT}}$. The $(j,k)$ element of $\hat{\Sigma}$ can be
         derived from formula (3.7) with all terms replaced by its estimator. Then
         $\hat{\mathscr{T}}_{\text{HOT}} = \sum_{j=1}^{J} j^{-2}\exp(\hat{T}_j^2/2)$;
Step T3: Obtain local test statistic $\boldsymbol{T}_b$ under $H_0$ and then calculate HOT test statistic
         $\mathscr{T}_{b,HOT}$ by following loop
        do    $b = 1 : B$
            $\{$Generate $\boldsymbol{T}_b \sim N(\mathbf{0}_{J\times 1}, \hat{\Sigma})$, and $\mathscr{T}_{b,HOT} = \sum_{j=1}^{J} j^{-2}\exp(T_{bj}^2/2)\}$;
Step T4: The p-value is announced as $p = B^{-1}\sum_{b=1}^{B} I\{\mathscr{T}_{b,HOT} > \hat{\mathscr{T}}_{\text{HOT}}\}$.
```

Our HOT, CHOT and BHOT perform very similar procedures except for Step F2 which is explained in the following:

- HOT: it is following the Step H1-H2 using Φ with $\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, where

$$\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} - \mathbf{1}^T\log[\mathbf{1} + \exp\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}]; \tag{3.12}$$

- CHOT: it is the same as HOT except for using Φ with $\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, where

$$\ell(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} - \mathbf{1}^T\log[\mathbf{1} + \exp\{\kappa_0 + \sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}] - \lambda\|\boldsymbol{\beta}\|^2; \tag{3.13}$$

  ; we determine λ which has the maximum value of difference between TPR and FPR; the best λ was chosen at 30 which is shown in Figure 3.1;

- BHOT: it is the same as HOT except for using Φ with $\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)$, where

$$\begin{aligned}\log(\pi(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|Y,X,Z)) =& y^T\{\kappa_0+\sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha})+f(Z,\boldsymbol{\gamma})\}-\mathbf{1}^T\log[\mathbf{1}+\exp\{\kappa_0+\sum_{j=1}^{p+K}\beta_j h_i(X\boldsymbol{\alpha})\\ &+f(Z,\boldsymbol{\gamma})\}]+n\pi(\boldsymbol{\alpha},\boldsymbol{\beta})+n\pi(\boldsymbol{\gamma});\end{aligned} \tag{3.14}$$

We choose the proposal distributions for Metropolis-Hastings to ensure the detailed balance,

$$\begin{aligned}\mathscr{G}_{\boldsymbol{\phi}}(\boldsymbol{\phi}_{i-1}\to\boldsymbol{\phi}_{\text{new}}) &= \mathscr{G}_{\boldsymbol{\phi}}(\boldsymbol{\phi}_{\text{new}}\to\boldsymbol{\phi}_{i-1});\\ \mathscr{G}_2(\boldsymbol{\beta}_{i-1},\boldsymbol{\gamma}_{i-1}\to\boldsymbol{\beta}_{new},\boldsymbol{\gamma}_{new}) &= \mathscr{G}_2(\boldsymbol{\beta}_{new},\boldsymbol{\gamma}_{new}\to\boldsymbol{\beta}_{i-1},\boldsymbol{\gamma}_{i-1}),\end{aligned}$$

and use the following proposal distributions:

  - Proposal distributions for $\boldsymbol{\alpha}$, $\mathscr{G}_1(\boldsymbol{\alpha}_{i-1}\to\boldsymbol{\alpha}_{\text{new}})$:

    It is selected via the reparameterization formula (3.4) using the proposal distribution of $\boldsymbol{\phi}$, $\mathscr{G}_{\boldsymbol{\phi}}(\boldsymbol{\phi}_{i-1}\to\boldsymbol{\phi}_{new})$ where

$$\begin{aligned}&\boldsymbol{\phi}_{\text{temp}}\sim N(\boldsymbol{\phi}_{i-1},0.3I),\\ &\phi_{\text{new},j}=\phi_{\text{temp},j}, && \text{if}\quad 0\le\phi_{\text{temp},j}<2\pi\\ &\phi_{\text{new},j}=\phi_{\text{temp},j}-2\pi\times\lfloor\phi_{\text{temp},j}/2\pi\rfloor, && \text{otherwise}\\ &\qquad j=1,2,...,p-1,\end{aligned} \tag{3.15}$$

    where $\lfloor\phi\rfloor$ represents the nearest integer value less than ϕ and then $\boldsymbol{\alpha}_{\text{new}}$ is obtained via formula (3.4);

  - Proposal distribution for $(\boldsymbol{\beta},\boldsymbol{\gamma})$, $\mathscr{G}_2(\boldsymbol{\beta}_{i-1},\boldsymbol{\gamma}_{i-1}\to\boldsymbol{\beta}_{new},\boldsymbol{\gamma}_{new})$:

It is multivariate normal distribution,

$$(\boldsymbol{\beta}_{\text{new}}, \boldsymbol{\gamma}_{\text{new}}) \sim N\Big((\boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1}), 12I\Big). \tag{3.16}$$

3.8.2.2 ELRT

We then compare ours with ELRT. For ELRT, null and alternative hypothesis are the following,

$$\begin{aligned} H_0: &\quad \Pr(Y|X,Z) = H\{\kappa_0 + f(Z,\boldsymbol{\gamma})\}; \\ H_1: &\quad \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}. \end{aligned}$$

and we proceed the following steps E1-E4 to obtain empirical pvalues.

Step E1 Estimate $\hat{\boldsymbol{\gamma}}_0$ under H_0 and $(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\gamma}}_1)$ under H_1, respectively;

Step E2 Calculate $T_{obs} = \ell(0, \hat{\boldsymbol{\gamma}}_{0b}, 0) - \ell(\hat{\boldsymbol{\alpha}}_{1b}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\gamma}}_{1b})$, where

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = y^T\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} - \mathbf{1}^T \log[\mathbf{1} + \exp\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}];$$

Step E3 Run the loop, do $b = 1 : B$

{Generate Y_{ib} for $i = 1, \ldots n$, based on $\Pr(y|Z, \hat{\boldsymbol{\gamma}}_{0b})$ and calculate $T_{\text{emprical},b} = \ell(\hat{\boldsymbol{\gamma}}_{0b}) - \ell(\hat{\boldsymbol{\alpha}}_{1b}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\gamma}}_{1b})$ };

Step E4 The p-value is announced as $p = \sum_{b=1}^{B} I(T_{\text{emprical},b} > T_{obs})$.

3.8.2.3 ABF

We also compare ours with ABF. For ABF, null and alternative hypothesis models are the following,

$$\begin{aligned} \mathscr{M}_{H_0}: &\quad \Pr(Y|X,Z;\boldsymbol{\gamma}) = H\{\kappa_0 + f(Z,\boldsymbol{\gamma}_{\mathscr{M}_{H_0}})\}\pi(\boldsymbol{\gamma}_{\mathscr{M}_{H_0}}); \\ \mathscr{M}_{H_1}: &\quad \Pr(Y|X,Z;\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = H\{\kappa_0 + \sum_{j=1}^{p+K}\beta_{j,\mathscr{M}_{H_1}} h_j(X\boldsymbol{\alpha}_{\mathscr{M}_{H_1}}) + f(Z,\boldsymbol{\gamma}_{\mathscr{M}_{H_1}})\}\pi(\boldsymbol{\alpha}_{\mathscr{M}_{H_1}},\boldsymbol{\beta}_{\mathscr{M}_{H_1}},\boldsymbol{\gamma}_{\mathscr{M}_{H_1}}). \end{aligned}$$

With the semi-parametric framework, the ABF algorithm with flat non-informative prior proceeds the following steps:

Step 00: Perform the model fitting steps;

Step F1: Select non-parametric basis for approximation function $\boldsymbol{h}(\cdot) = \{h_1(\cdot),...,h_J(\cdot)\}$ and approximation function $\tilde{g}(X\boldsymbol{\alpha}) = \boldsymbol{h}(X\boldsymbol{\alpha})\boldsymbol{\beta} = \sum_{j=1}^{J}\beta_j h_j(X\boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1,...,\beta_J)^T$;

Step F2: With selected basis, obtain likelihoods under H_0 and H_1;

Step F3: Initial α,β,γ correspondingly under H_1 and γ under H_0;

Step F4: We obtain the posterior sample of (α,β,γ) from the following Metropolis-Hastings sampler with the total number of Metropolis-Hastings iteration I and the ith Metropolis-Hastings iteration;

While $\quad i < I$

- Sample new $\boldsymbol{\alpha}_{new}$ from proposal distribution $\mathscr{G}_1(\boldsymbol{\alpha}_{i-1} \rightarrow \boldsymbol{\alpha}_{\text{new}})$ explained in equation (3.15);
  - Initial $\text{flag}_{\boldsymbol{\alpha}} = 0$;

    While $\text{flag}_{\boldsymbol{\alpha}} = 0$;

- Set

$$\mathrm{AR}_1 = \frac{L(\boldsymbol{\alpha}_{\mathrm{new}}, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1}|Y,X,Z)\mathscr{G}_1(\boldsymbol{\alpha}_{i-1} \rightarrow \boldsymbol{\alpha}_{\mathrm{new}})}{L(\boldsymbol{\alpha}_{i-1}, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1}|Y,X,Z)\mathscr{G}_1(\boldsymbol{\alpha}_{new} \rightarrow \boldsymbol{\alpha}_{\text{i-1}})},$$

  where

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = &\exp(y^T\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_j(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\} \\ &- \mathbf{1}^T \log[\mathbf{1} + \exp\{\kappa_0 + \sum_{j=1}^{p+K} \beta_j h_i(X\boldsymbol{\alpha}) + f(Z,\boldsymbol{\gamma})\}]); \end{aligned} \tag{3.17}$$

- Sample $U_{i1} \sim U(0,1)$;
- If $U_{i1} < \mathrm{AR}_1$, then $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{new}$, $\mathrm{flag}_{\boldsymbol{\alpha}} = 1$;

• Sample new $(\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$ from proposed distribution $\mathscr{G}_2(\boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} \rightarrow \boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$ explained in equation (3.16);

- Initial $\mathrm{flag}_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 0$;

  While $\mathrm{flag}_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 0$;
- Set

$$\mathrm{AR}_2 = \frac{L(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_{\mathrm{new}}, \boldsymbol{\gamma}_{\mathrm{new}}|Y,X,Z)\mathscr{G}_2(\boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1} \rightarrow \boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})}{L(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1}|Y,X,Z)\mathscr{G}_2(\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new} \rightarrow \boldsymbol{\beta}_{i-1}, \boldsymbol{\gamma}_{i-1})},$$

- Sample $U_{i2} \sim U(0,1)$;
- If $U_{i2} < \mathrm{AR}_2$, then $(\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i) = (\boldsymbol{\beta}_{new}, \boldsymbol{\gamma}_{new})$, $\mathrm{flag}_{\boldsymbol{\beta},\boldsymbol{\gamma}} = 1$;

• Set $i = i+1$;

Step F5: With similar steps, we also obtain posterior distribution sample of γ under the reduced model $\mathrm{pr}(Y|Z) = p_{Y|X,Z}\{Y, \kappa_0 + f(Z,\boldsymbol{\gamma})\}$;

Step 11: After we obtain samples from posterior distribution, we calculate ABF using VTA which is described in Appendix C.

3.8.3 Simulation Result

We obtain the average values of TPR and FPR. We vary *m* in range [3,8] with 0.5 increment. The simulations results are summarized in Table 3.1-3.2 and Figure 3.2-3.3 for low-dimensional case and Table 3.3-3.12 and Figure 3.4-3.7 for high-dimensional case.

Under the Case 1 of Setting 1 where low-dimensional case and sign bump function are used, as Table 3.1 is shown, average FPRs of ELRT and HOTS under Case 1 of Setting 1 are comparable. ELRT's FPRs are between 0.04 and 0.10. HOT's FPRs are between 0.04 and 0.1. BHOT's are between 0.01 and 0.014. CHOT's FPRs are between 0 and 0.05. ABF's FPRs are between 0 and 0.06. CHOT provides the smallest FPR values. The average value of FPR for each methods are displayed in Figure 3.2. For average TPR which is shown in Table 3.2, as expected, average TPRs of ELRT increases as *m* increases in Case 1 of Setting 1. HOT, CHOT, and BHOT's TPRs are also increasing as *m* increases. The average TPRs of CHOT is between 0.99 and 1. It is larger than that of ELRT and ABF. BHOT's TPRs are between 0.36 and 0.81. Therefore, HOT and CHOT perform better than ELRT, BHOT, and ABF. We display the average value of TPR for each methods in Figure 3.3.

Under the Case 2 of Setting 1 where high-dimensional case and sign bump function are used, as Table 3.3 is shown, average FPRs of HOT under Case 2 of Setting 1 is between 0 and 0.10. We display the average value of FPR for HOT in Figure 3.4. The average TPRs of HOT also increases as *m* increases. The average values of TPRs of HOT is between 0.69 and 0.98. They are also summarized in Table 3.4 and in Figure 3.5. Therefore, HOT performes well in high-dimensional case. As Table 3.7 is shown, the average FPRs of CHOT under Case 2 of Setting 2 is between 0.00 and 0.08 The average values of TPRs of CHOT is betwee 0.89 and 1 which is shown in Table 3.8. Therefore, CHOT also performes well in high-dimensional case. The average value of FPR for CHOT is displayed in Figure 3.6. As Table 3.9 is shown, average FPRs of BHOT is between 0.00 and 0.30. For average TPR which is shown in Table 3.10, the average values of TPRs of BHOT is between 0.91 and 1 under Case 2 of Setting 1. Therefore, BHOT performs well in high-dimensional case.

Under the Case 2 of Setting 2 where high-dimensional case and polynomial function are used, the average FPRs of CHOT is between 0.01 and 0.03 which is shown in Table 3.5 . For average TPR which is shown in Table 3.6, the average values of TPRs of HOT is 1. Therefore, CHOT performs well in high-dimensional case. We display the average value of FPR for CHOT in Figure 3.7.

Under Case 2 and Setting 1 where high-dimensional case and sign bump function are used, as Table 3.11 is shown, average FPRs based on Bayes factor is 0. For average TPR which is shown in Table 3.12, the average values of TPRs based on Bayes factor is between 0.21 and 0.44. Therefore, ABF does not provide good performance in high-dimensional case.

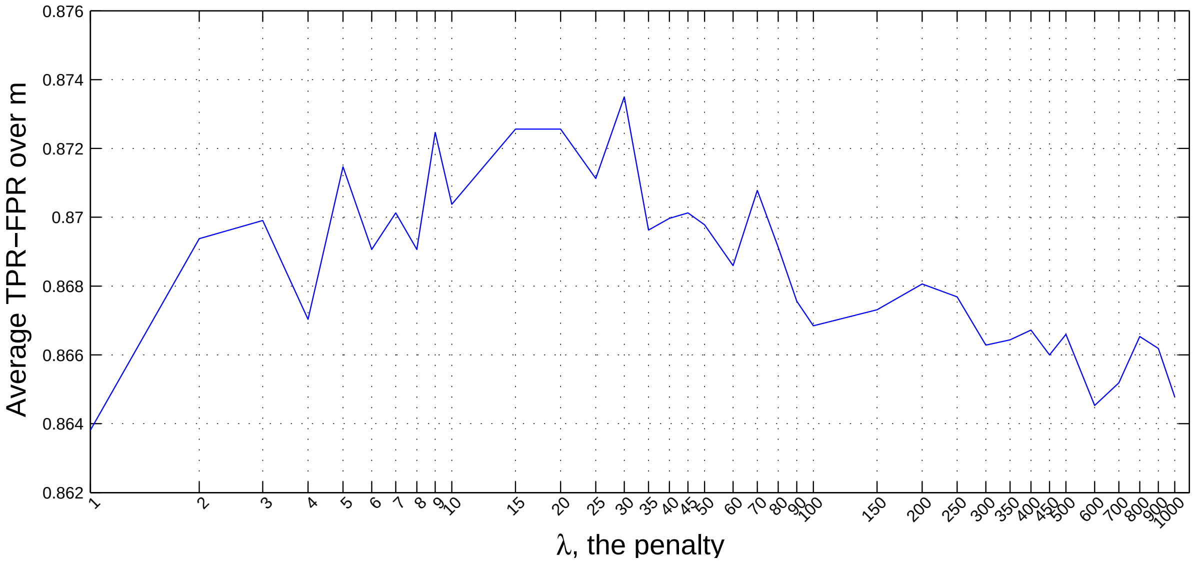

Figure 3.1: Determination of the best λ which maximize the difference between average value of true positive rate and average value of false positive rate for Constraint Hybrid Omnibus Test (CHOT) under low dimensional case;

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B - A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR and $\boldsymbol{\alpha} = \{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\}$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$,$\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ =the inverse logistic link; The predictors are generated from the following setting $X \sim \text{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The test hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$. The results are based on the sample size of $n = 500$ and 1000 simulated data sets.

Table 3.1: The average value of false positive rate obtained from ELRT, HOT, CHOT, BHOT and ABF in simulation study under low dimensional case; *m*=the amplifying multiplier; ELRT=Empirical Likelihood Ratio Test; HOT=Hybrid Omnibus Test; CHOT=HOT with constraint estimating equation; BHOT= Bayesian MPE Hybrid Omnibus Test; ABF=Approximate Bayes Factor.

m	3.00	3.50	4.00	4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00
ELRT	0.05	0.04	0.05	0.06	0.05	0.07	0.08	0.09	0.09	0.09	0.10
HOT	0.04	0.04	0.09	0.10	0.10	0.04	0.01	0.01	0.01	0.01	0.01
CHOT	0.02	0.05	0.00	0.05	0.01	0.02	0.00	0.00	0.00	0.01	0.00
BHOT	0.05	0.05	0.10	0.14	0.06	0.09	0.02	0.05	0.02	0.01	0.01
ABF	0.00	0.00	0.01	0.00	0.01	0.04	0.03	0.04	0.06	0.05	0.04

Table 3.2: The average value of true positive rate from ELRT, HOT, CHOT, BHOT and ABF in simulation study under low dimensional case; *m*=the the amplifying multiplier; ELRT=Empirical Likelihood Ratio Test; HOT=Hybrid Omnibus Test; CHOT= HOT with constraint estimating equation; BHOT= Bayesian MPE Hybrid Omnibus Test; ABF=Approximate Bayes Factor.

m	3.00	3.50	4.00	4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00
ELRT	0.87	0.85	0.87	0.84	0.87	0.88	0.91	0.85	0.88	0.92	0.87
HOT	0.67	0.85	0.94	0.97	0.98	1.00	0.99	1.00	1.00	1.00	1.00
CHOT	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BHOT	0.36	0.49	0.56	0.59	0.60	0.72	0.68	0.73	0.76	0.72	0.81
ABF	0.75	0.88	0.90	0.99	0.99	1.00	1.00	1.00	0.99	1.00	1.00

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B - A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha} = \{0,0,0\}$ for FPR, $\boldsymbol{\alpha} = \{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\}$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}, \gamma = 0.3$, $m =$ the amplifying multiplier, $H(\cdot) =$ the inverse logistic link; The predictors are generated from the following setting: $X \sim N(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 :$ $\Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; These results are based on sample size n=500 and 1000 simulated data sets. Length of MCMC chain=5000 after 2500 burn-in.

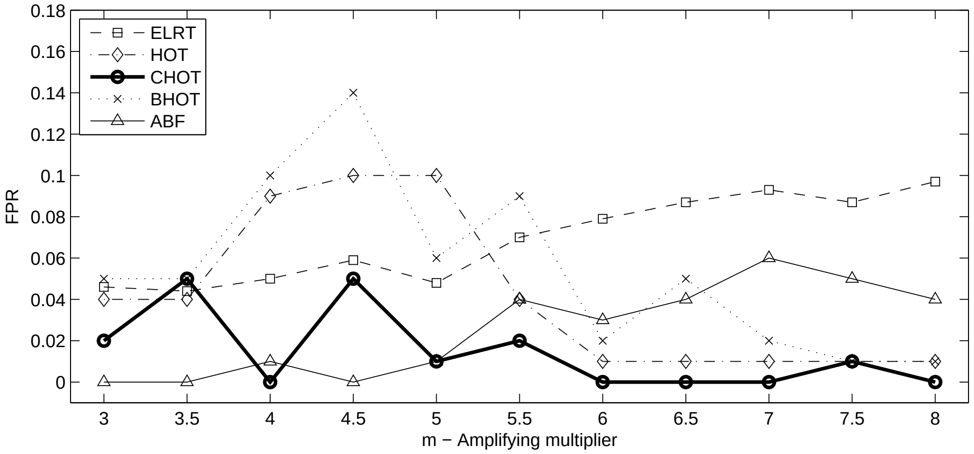

Figure 3.2: The values of False positive rate vs the amplifying multiplier (m) in simulation study under low-dimensional case; ELRT=Empirical Likelihood Ratio Test; HOT=Hybrid Omnibus Test; CHOT=HOT with constraint estimating equation; BHOT= Bayesian MPE Hybrid Omnibus Test; ABF=Approximate Bayes Factor

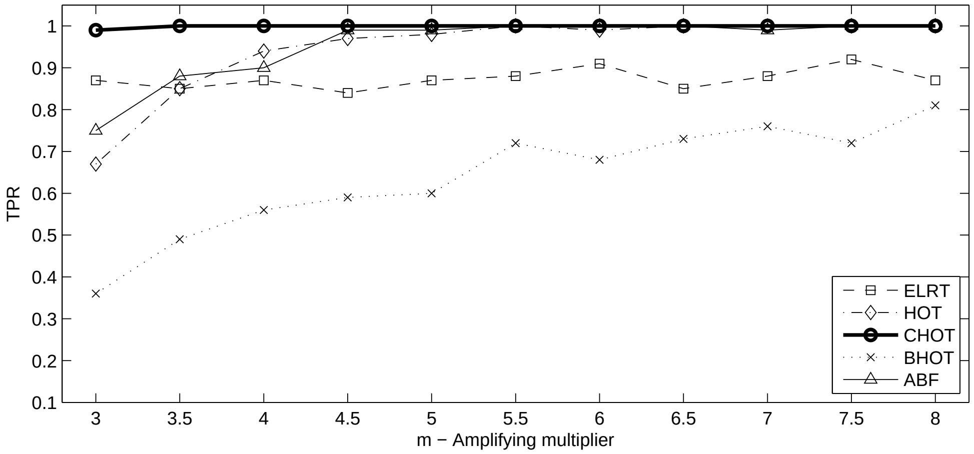

Figure 3.3: The values of true positive rate vs the amplifying multiplier (m) in simulation study under low-dimensional case; ELRT=Empirical Likelihood Ratio Test; HOT=Hybrid Omnibus Test; CHOT=HOT with constraint estimating equation; BHOT= Bayesian MPE Hybrid Omnibus Test; ABF=Approximate Bayes Factor

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha} = \{0,0,0\}$ for FPR, $\boldsymbol{\alpha} = \{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\}$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$,$\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ =the inverse logistic link; The predictors are generated from the following setting: $X \sim N(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; The results are based on sample size n=500 and 1000 simulated data sets

Table 3.3: The average values of False positive rate vs the amplifying multiplier (m) in simulation study under high-dimensional case with HOT; HOT=Hybrid Omnibus Test

m \ dim(p)	100	150	200	250	300	500	800	1000
4.25	0.06	0.10	0.08	0.06	0.03	0.05	0.04	0.00
4.50	0.03	0.05	0.03	0.04	0.00	0.02	0.02	0.00
4.75	0.04	0.06	0.03	0.02	0.02	0.02	0.00	0.02
5.00	0.03	0.01	0.00	0.00	0.01	0.01	0.00	0.00
5.25	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.00
5.50	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
5.75	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
6.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.4: The average values of true positive rate vs the amplifying multiplier (m) in simulation study under high-dimensional case with HOT; HOT=Hybrid Omnibus Test

m \ dim(p)	100	150	200	250	300	500	800	1000
4.25	0.98	0.98	0.98	0.98	0.98	0.95	0.95	0.98
4.50	0.98	1.00	0.98	0.96	0.93	0.96	0.97	0.95
4.75	0.99	1.00	0.99	0.96	0.94	0.97	0.96	0.98
5.00	0.97	0.96	0.97	0.96	0.93	0.89	0.97	0.96
5.25	0.96	0.89	0.89	0.91	0.89	0.88	0.91	0.94
5.50	0.94	0.90	0.84	0.88	0.80	0.76	0.96	0.94
5.75	0.90	0.84	0.90	0.77	0.79	0.75	0.94	0.88
6.00	0.82	0.90	0.80	0.72	0.71	0.69	0.81	0.92

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR, $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$,$\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The predictors are generated from the following setting: $X \sim \mathrm{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; These results are based on sample size n=100 and 1000 simulated data sets.

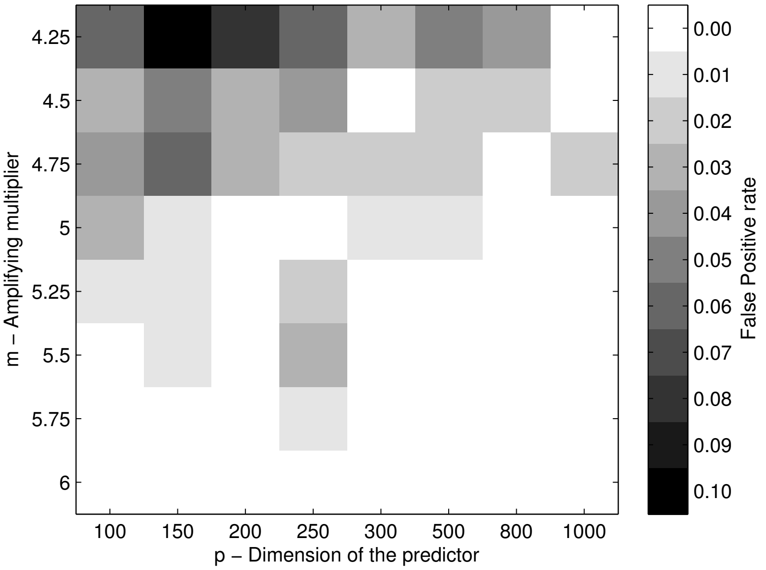

Figure 3.4: Grayscale plot between false positive rate vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under high-dimensional case; HOT=Hybrid Omnibus Test

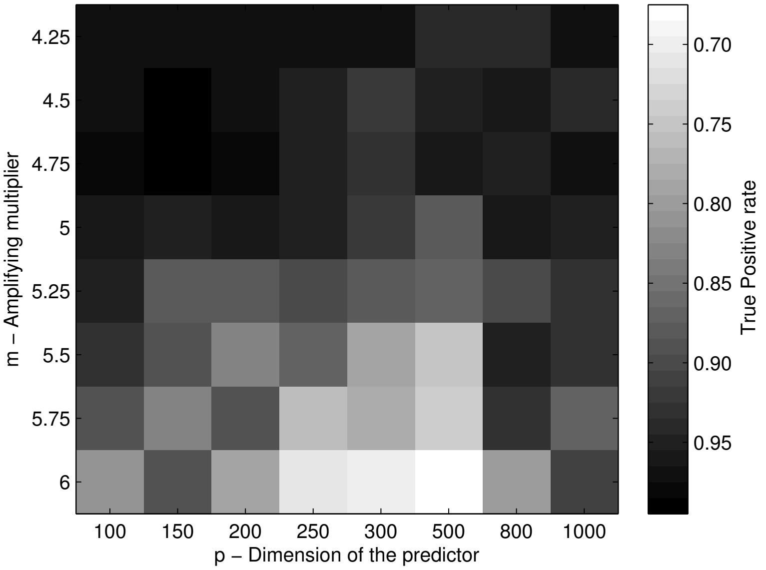

Figure 3.5: Grayscale plot between true positive rate vs the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under high-dimensional case; HOT=Hybrid Omnibus Test; The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR, $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The data are generated from the following setting: $X \sim \mathrm{N}(0,1)$, $Z = 0$ if odd observation, $Z = 1$ if even observation; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; The results are based on sample size n=100 and 1000 simulated sets.

Table 3.5: The average value of false positive rate using CHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under polynomial setting and high-dimension case; CHOT= Constraint Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02
4.50	0.02	0.02	0.02	0.01	0.03	0.02	0.02	0.02
4.75	0.03	0.03	0.02	0.01	0.03	0.02	0.03	0.02
5.00	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02
5.25	0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.03
5.50	0.03	0.02	0.03	0.03	0.03	0.02	0.02	0.03
5.75	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.03
6.00	0.02	0.03	0.03	0.03	0.03	0.03	0.02	0.01

Table 3.6: The average value of true positive rate using CHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under polynomial setting and high-dimension case; CHOT= Constraint Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The data were generated from: $\mathrm{pr}(Y_i = 1) = H[m\{-0.9619 + X\alpha + (X\alpha)^2 + (X\alpha)^3 + (X\alpha)^4\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR; $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot) =$ the inverse logistic link; The predictors are generated from the following setting: $X \sim \mathrm{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \mathrm{Pr}(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \mathrm{Pr}(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; These results are based on sample size n=100 and 1000 simulated data sets.

Table 3.7: The average value of False positive rate using CHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under sin-bump setting and high-dimension case; CHOT= Constraint Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.08
4.50	0.00	0.01	0.05	0.00	0.01	0.03	0.01	0.03
4.75	0.00	0.02	0.00	0.03	0.02	0.00	0.00	0.04
5.00	0.00	0.00	0.02	0.04	0.01	0.04	0.00	0.02
5.25	0.00	0.02	0.01	0.02	0.01	0.01	0.00	0.01
5.50	0.00	0.01	0.00	0.01	0.02	0.00	0.01	0.01
5.75	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01
6.00	0.00	0.02	0.00	0.00	0.01	0.02	0.00	0.00

Table 3.8: The average value of true positive rate using CHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under sine-bump setting and high-dimension case; CHOT= Constraint Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	1.00	1.00	1.00	1.00	0.99	0.99	0.98	1.00
4.50	1.00	1.00	1.00	1.00	0.97	1.00	0.96	0.99
4.75	1.00	1.00	1.00	1.00	0.98	1.00	0.96	1.00
5.00	0.99	1.00	1.00	1.00	1.00	1.00	0.95	0.98
5.25	1.00	1.00	1.00	1.00	0.98	1.00	0.96	1.00
5.50	1.00	1.00	1.00	1.00	0.99	1.00	0.92	0.99
5.75	1.00	1.00	1.00	1.00	0.99	0.99	0.89	0.99
6.00	1.00	1.00	1.00	1.00	0.98	1.00	0.94	0.99

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR, $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The predictors are generated from the following setting: $X \sim \text{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2} (X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; These results are based on sample size n=100 and 100 simulated data sets.

Table 3.9: The average value of false positive rate using BHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under sine-bump setting and high-dimension case; BHOT= Bayesian MPE Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.23	0.20	0.24	0.17	0.21	0.26	0.25	0.22
4.50	0.17	0.16	0.30	0.25	0.20	0.28	0.16	0.21
4.75	0.22	0.18	0.18	0.26	0.15	0.26	0.17	0.13
5.00	0.16	0.19	0.13	0.20	0.23	0.25	0.24	0.16
5.25	0.18	0.16	0.17	0.16	0.13	0.16	0.15	0.12
5.50	0.24	0.16	0.13	0.18	0.16	0.21	0.16	0.14
5.75	0.08	0.11	0.06	0.15	0.20	0.17	0.11	0.09
6.00	0.11	0.09	0.07	0.09	0.08	0.14	0.08	0.14

Table 3.10: The average value of true positive rate using BHOT vs the the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under sine-bump setting and high-dimension case; BHOT= Bayesian MPE Hybrid Omnibus Test

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.93	0.98	0.96	0.92	0.95	0.94	0.98	0.97
4.50	0.98	0.94	1.00	0.98	0.97	0.94	1.00	0.92
4.75	0.99	0.93	0.97	0.99	0.98	0.98	0.99	0.97
5.00	0.96	0.92	0.98	0.95	0.96	0.96	0.96	0.93
5.25	0.99	0.95	0.95	0.97	0.98	1.00	0.96	0.97
5.50	0.95	0.97	1.00	0.98	0.97	0.97	0.97	0.98
5.75	0.99	0.91	0.98	0.99	0.95	0.96	0.95	0.97
6.00	0.99	0.94	0.96	0.94	0.97	1.00	0.99	0.98

The data were generated from: $\Pr(Y_i=1)=H[m\cdot\sin\{\pi(\boldsymbol{\alpha}^T X_i-A)/(B-A)\}+\gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1}=\mathbf{0}_{p,1}$ for FPR, $\boldsymbol{\alpha}_{p\times 1}=\{\mathbf{0}_{1,p/2},\mathbf{1}_{1,p/2}\}^T$ for TPR; $A=\frac{\sqrt{3}}{2}-\frac{1.645}{\sqrt{12}}$, $B=\frac{\sqrt{3}}{2}+\frac{1.645}{\sqrt{12}}$,$\gamma=0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The predictors are generated from the following setting: $X\sim$ Unif$(0,1)$, $Z=0$ if observation is odd, $Z=1$ if observation is even; The testing hypothesis is $H_0:\Pr(Y|Z)=H\{\kappa_0+\gamma Z_i\}$ vs $H_1:\Pr(Y|X,Z)=H\{\kappa_0+\sum_{k=1}^{2}\beta_k(X\boldsymbol{\alpha})^k+\sum_{k=1}^{20}\beta_{k+2}(X\boldsymbol{\alpha}-\kappa_k)_+^2+\gamma Z\}$. These results are based on sample size n=100 and 100 simulated data sets. Length of MCMC chain=5000 after 2500 burn-in.

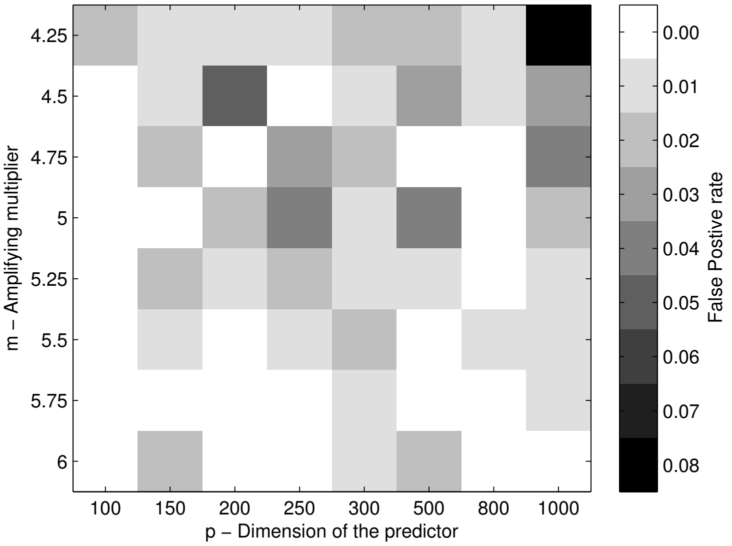

Figure 3.6: Grayscale plot between false positive rate of CHOT vs the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under high-dimensional case; CHOT=Constraint Hybrid Omnibus Test

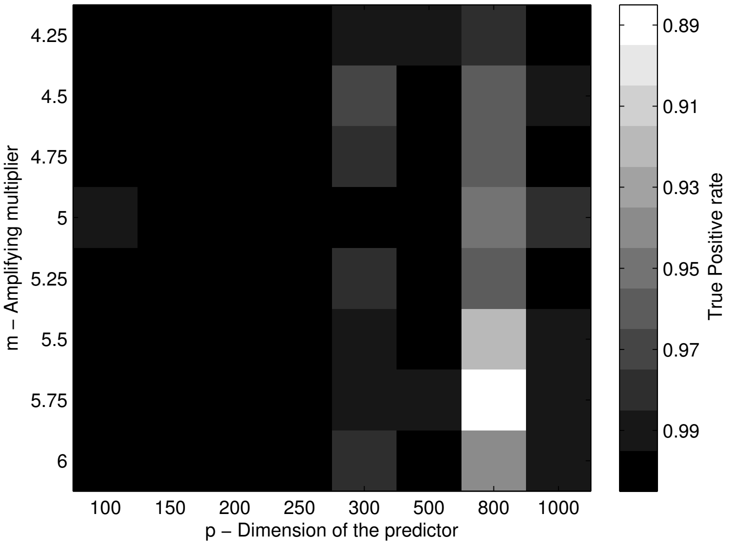

Figure 3.7: Grayscale plot between true positive rate of CHOT vs the amplifying multiplier (m) and the dimension of predictor (p) in simulation study under high-dimensional case; CHOT=Constraint Hybrid Omnibus Test

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B - A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR, $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\gamma = 0.3$, $m =$the amplifying multiplier, $H(\cdot) =$ the inverse logistic link; The predictors are generated from the following setting: $X \sim \mathrm{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$. The results are based on sample size n=100 and 100 simulated data sets.

Table 3.11: The average value of false positive rate (Approximate Bayes Factor (ABF) > 10) using ABF vs the amplifying multiplier (m) and the dimension of predictor (p) in simulation under high dimension case

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.12: The average value of true positive rate (Approximate Bayes Factor (ABF) > 10) using ABF vs the amplifying multiplier (m) and the dimension of predictor (p) in simulation under high dimension case

dim(p) m	100	150	200	250	300	500	800	1000
4.25	0.27	0.36	0.27	0.37	0.26	0.41	0.38	0.21
4.50	0.35	0.39	0.26	0.36	0.34	0.44	0.34	0.31
4.75	0.34	0.42	0.28	0.27	0.27	0.28	0.29	0.37
5.00	0.40	0.33	0.37	0.30	0.33	0.23	0.38	0.37
5.25	0.35	0.32	0.35	0.30	0.31	0.39	0.44	0.34
5.50	0.40	0.30	0.32	0.29	0.36	0.38	0.37	0.25
5.75	0.34	0.33	0.33	0.28	0.34	0.39	0.21	0.35
6.00	0.32	0.35	0.43	0.36	0.39	0.33	0.31	0.21

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B - A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR; $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\gamma = 0.3$, $m =$the amplifying multiplier, $H(\cdot) =$ the inverse logistic link; The predictors are generated from the following setting: $X \sim \text{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; These results are based on sample size n=100, 100 simulated data sets. Length of MCMC chain=5000 after 2500 burn-in.

3.8.4 Simulation Summary

For low-dimensional case, HOT and CHOT perform as very close to ELRT with comparable FTP rate. HOT has both FPR rate and TPR increase as the amplifier m increases, as expected, which is result from the weight component j^{-2} since the effect is slight to be decomposing with a good linear combination of basis. When m increases, as shown in high-dimension case, HOT becomes stable. For both low and high dimensional cases, CHOT performs the best. For the performance of ELRT, which is based on maximum likelihood estimator, can be constructed with the best unbiased estimator when $n > p$, and the test statistic may achieve the asymptotic χ^2 distribution as $n > p$ along with low-dimensional p. For high-dimensional case, ELRT requires devastating computational time and is also very unstable in estimating $\hat{\alpha}$ whose unbiased estimator is difficult to obtain. On the contrast, HOT, CHOT, and BHOT perform well in terms of FPR and TPR in high-dimensional case. It provides reasonable FTP and large TPR. FPR is in the range [0,0.1] which depends m and p and TPR is in the range [0.7, 1] which depends on m and p.

We also compare our HOT approaches with testing based on Bayes factor. The inference based on Bayes factor does not provide good performance in high dimensional case. For polynomial setting, traditional score test does not work because it provides zero FPR and zero TPR when $p \geq 100$. In this case X is ill-conditioned since the ratio of the number of column to the number of row is high. As a result, the score test statistic is expected to be very small because the inverse of Fisher information matrix will be extremely large. Bayesian frame is also fragile when dimension increases. One obvious problem is that the convergence problem. When the dimension of parameter is very high, MCMC samplers such as Metropolis-Hastings and slice-sampler are not reliable since they are easily stuck in particular status. Another problem is that the MPE is likely to be biased because of autocorrelated MCMC sample.

For each method, we obtained the overall means of FPR and TPR for both low and high dimensional cases. The results under low and high-dimensional cases are summarized in Table 3.13. When we consider low-dimensional case, CHOT outperformed because CHOT provides the smallest average value FPR and largest average value of TPR. ELRT also has good performance because it use known likelihood function. The performance of HOT and BHOT are similar. However when we consider high dimensional case, the ELRT is very unstable so that we are not able to obtain FPR and TPR. For all other four methods except for ELRT, we are able to have them. HOT, CHOT, and BHOT outperform. Our three HOT approaches are comparable. The BHOT's FPR and TPR have larger variance than other two. The boxplots of all FPR values and TPR values of each method are displayed in Figures 3.8 and 3.9, respectively. The FPR values of BHOT is larger than those of HOT and CHOT. The TPR values of HOT is smaller than those of CHOT and BHOT. Overall in both low and high dimensional cases, CHOT is the best performer.

Table 3.13: The average values of false positive rate (FPR) and true positive rate (TPR) for five methods in simulation study: ELRT=Empirical Likelihood Ratio Test; HOT=Hybrid Omnibus Test; CHOT=Contraint HOT; BHOT=Bayesian HOT; ABF=Approximated Bayes Factor.

Method	Low Dim		High Dim	
	Average FPR	Average TPR	Average FPR	Average TPR
ELRT	0.0691	0.8736	N/A	N/A
HOT	0.0418	0.9455	0.0152	0.9116
CHOT	0.0145	0.9991	0.0120	0.9898
BHOT	0.0545	0.6382	0.1765	0.9649
ABF	0.0255	0.9545	0.0000	0.3327

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{1}_p$ for low dim; $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for high dim; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The predictors are generated from the following setting: $X \sim \mathrm{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$. Length of MCMC chain=5000 after 2500 burn-in.

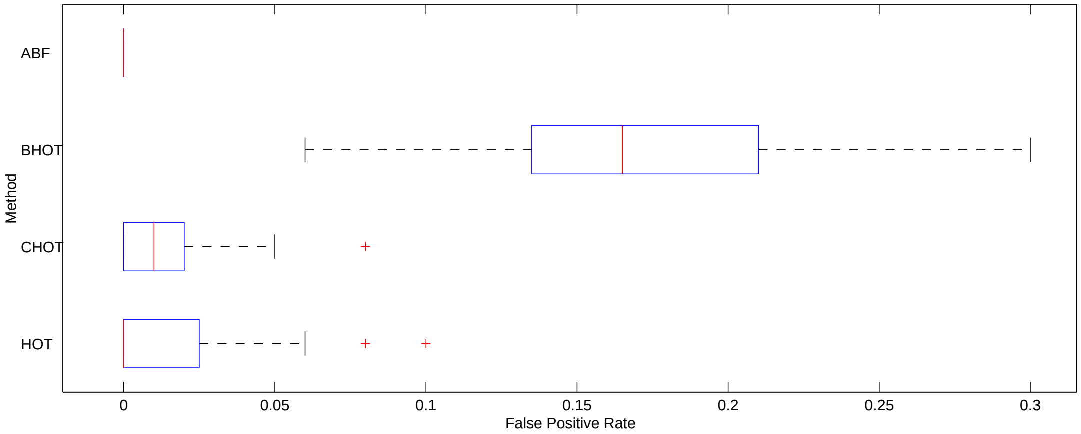

Figure 3.8: Summary of false positive rate values in simulation study under high-dimension case; HOT=Hybrid Omnibus Test; CHOT=Constraint HOT; BHOT=Bayesian HOT; ABF=Approximated Bayes Factor.

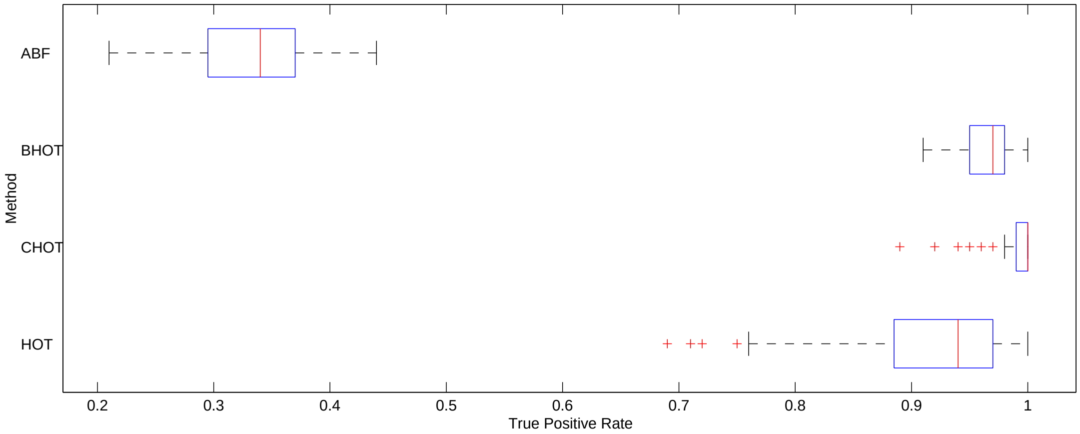

Figure 3.9: Summary of true positive rate values in simulation study under high-dimension case; HOT=Hybrid Omnibus Test; CHOT=Constraint HOT; BHOT=Bayesian HOT; ABF=Approximated Bayes Factor.

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B - A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{0}_{p,1}$ for FPR; $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for TPR; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$,$\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot) =$ the inverse logistic link; The predictors are generated from the following setting: $X \sim \text{Unif}(0, 1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even; The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X, Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2} (X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$; The results are based on sample size n=100 and 100 simulated data sets. Length of MCMC chain=5000 after 2500 burn-in.

3.8.5 Sensitivity Study on Initial Value of α

As we describe in the section 3.8.2, we solve the estimating equation for $\boldsymbol{\alpha}_{\text{root}}$. Since the estimating equation usually does not have a analytical solution, it can be solved by numerical method based on iteration. We specify the initial value of $\boldsymbol{\alpha}_{\text{root},0}$ for iteration. Hence, in this section, we would like to investigate how much initial value of $\boldsymbol{\alpha}_{\text{root}}$ affects our testing results.

We consider the same setting as setting 2 with $m \in [5,8]$ and $p \geq 40$. True values of $\boldsymbol{\alpha}_{\text{root}}$ is $(0,\ldots,0,1,\ldots,1)^T$. We compare FPR and TPR using three different initial values:

- $\boldsymbol{\alpha}_{\text{root},0} = (1,\ldots,1)/\sqrt{p}$
- $\boldsymbol{\alpha}_{\text{root},0} = (\mathbf{1}_{\lfloor p/2 \rfloor}, \mathbf{0}_{\lceil p/2 \rceil})/\sqrt{\lfloor p/2 \rfloor}$, where $\lceil p/2 \rceil$ represents the nearest integer greater than $p/2$.
- $\boldsymbol{\alpha}_{\text{root},0,\boldsymbol{Rs}}$ which represents that randomly assigning the initial with half 1's and half 0's. Technically, take a random shuffle $\boldsymbol{Rs}$ from the subscript $\{1,...,p\}$ and obtain

  $$\boldsymbol{Rs} \;=\; (Rs_1, Rs_2, ..., Rs_p).$$

  So the random initial value is set as

  $$\alpha_{\text{root},0,Rs_1} = 1/\sqrt{\lfloor p/2 \rfloor}, ..., \alpha_{\text{root},0,Rs_{\lfloor p/2 \rfloor}} = 1/\sqrt{\lfloor p/2 \rfloor}, \alpha_{\text{root},0,Rs_{\lceil p/2 \rceil}} = 0, ..., \alpha_{\text{root},0,Rs_p} = 0.$$

Average values of FPR and TPR obtained using each initial value are summarized in Table 3.14. The first initial values are reasonable. The second initial values are the same as the true values. Average values of FPR and TPR using the first and third initial values are all comparable to those using the second ones. Hence this result suggests that if reasonable initial value is chosen, they do not affect testing results.

Table 3.14: Averages values of True positive rate (TPR) and false positive rate (FPR) using different initial value of α

Initial values	average FPR	average TPR
$\boldsymbol{\alpha}_{\text{root},0} = (1,1,\ldots,1)\sqrt{p}$	0.0100	0.9861
$\boldsymbol{\alpha}_{\text{root},0} = (\mathbf{1}_{\lfloor p/2 \rfloor}, \mathbf{0}_{\lceil p/2 \rceil})/\sqrt{\lfloor p/2 \rfloor}$	0.0834	0.9562
$\boldsymbol{\alpha}_{\text{root},0,\boldsymbol{Rs}}$	0.0166	0.9911

The data were generated from: $\Pr(Y_i = 1) = H[m \cdot \sin\{\pi(\boldsymbol{\alpha}^T X_i - A)/(B-A)\} + \gamma Z_i]$, with the following settings: $\boldsymbol{\alpha}_{p\times 1} = \mathbf{1}_p$ for low dim; $\boldsymbol{\alpha}_{p\times 1} = \{\mathbf{0}_{1,p/2}, \mathbf{1}_{1,p/2}\}^T$ for high dim; $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $B = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$,$\gamma = 0.3$, m =the amplifying multiplier, $H(\cdot)$ = the inverse logistic link; The predictors are generated from the following setting: $X \sim \text{Unif}(0,1)$, $Z = 0$ if observation is odd, $Z = 1$ if observation is even.The testing hypothesis is $H_0 : \Pr(Y|Z) = H\{\kappa_0 + \gamma Z_i\}$ vs $H_1 : \Pr(Y|X,Z) = H\{\kappa_0 + \sum_{k=1}^{2} \beta_k (X\boldsymbol{\alpha})^k + \sum_{k=1}^{20} \beta_{k+2}(X\boldsymbol{\alpha} - \kappa_k)_+^2 + \gamma Z\}$.

3.9 Application

We applied our HOT approach to a microarray expression data set on type II diabetes (Kim et al., 2012; Mootha et al., 2003), where a microarray expression data from 22,283 genes were measured in 17 male patients with normal glucose tolerance and 18 male samples with type II diabetes mellitus. We studied a total of 278 pathways consisting of 128 KEGG pathways and 149 curated pathways. The 149 curated pathways were constructed from known biological experiments by Mootha and colleagues. This data also contains body mass index (BMI) of patients. Certainly, this relationship among disease status and pathways is not linear and has no function being established.

3.9.1 Significant Pathways Obtained from HOT

In our analysis, let Y be the binary response representing normal and type II diabetes mellitus, X be the $n \times p$ gene expression levels within each pathway, where n is 35, i.e. the number of subjects, p is the number of genes in a specific pathway which varied from 4 to 200 across these pathways, and

Z is the clinical predictors such as BMI. Our goal is to identify pathways that distinguish between normal and type II diabetes mellitus after adjusting for the linear BMI effect. We use a set of spline bases functions and perform HOT. We identify pathways and summarize top 20 pathways in Table 3.15. And the p-values of all pathways are summarized in Figure 3.10.

Table 3.15: Top 20 Significant Pathways obtained from Hybrid Omnibus Test (HOT)

Pathway ID	Pathway Name	P-val
50	c29_U133_probes	0.005
31	c12_U133_probes	0.007
44	c23_U133_probes	0.007
62	c8_U133_probes	0.007
91	GLUCO_HG-U133A_probes	0.007
155	MAP00380_Tryptophan_metabolism	0.007
27	c0_U133_probes	0.008
30	c11_U133_probes	0.009
37	c18_U133_probes	0.009
38	c19_U133_probes	0.009
45	c24_U133_probes	0.009
48	c27_U133_probes	0.009
51	c3_U133_probes	0.009
35	c16_U133_probes	0.01
36	c17_U133_probes	0.01
39	c2_U133_probes	0.01
71	Complement and coagulation cascades	0.01
154	MAP00380_Tryptophan_metabolism	0.011
213	MAPK signaling pathway	0.013
34	c15_U133_probes	0.015

The five of them are also found as important pathways to identify normal and type II diabetes (Pang et al., 2006). One of them is MAPK signaling pathway. It is one of members of the MAPK family that are activated by a variety of environmental stresses and inflammatory cytokineses. As with other MAPK cascades, the membrane-proximal component is a MAPKKK, typically a MEKK

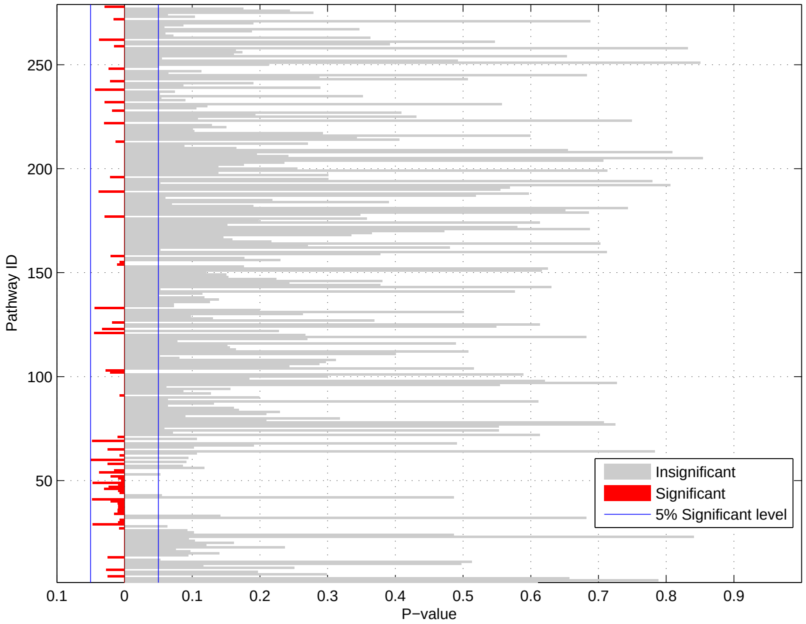

Figure 3.10: P-values of 278 pathways in type II diabetes data obtained from Hybrid Omnibus Test (HOT).

or a mixed lineage kinase (MLK). The MAPKKK phosphorylate and activates MKK3/6, the p38 MAPK kineses. MKK3/6 can also be activated directly by ASK1, which is stimulated by apoptosis stimuli. p38 MAPK is involved in regulation of HSP27, MAPKAPK-2 (MK2), MAPKAPK-3 (MK3), and several transcription factors including ATF-2, Stat1, the Max/ Myc complex, MEF-2, Elk-1, and indirectly CREB via activation of MSK1.

We also found that pathway 36, c17_U133_probes which was found by Random forest (Pang et al., 2006) and Bayesian approach (Kim et al., 2012). Genes MEF2C, NR4A1, SOX1, TPS1 are known to be related to glucose (Voisine et al., 2004; Zhang et al., 2006). Gene CAP1 is related to human insulin signaling (Dahlquist et al., 2002). Genes MAP2K6, ARF6, and SGK which are known to be related to human insulin signaling (Dahlquist et al., 2002). Gene ARF6 has been reported to play a role in the activation of protein kinase and phospholipase under high glucose condition that was hypothesized to be an important intracellular event linked to diabetic nephropathy (Padival et al., 2004). The SGK haplotype was significantly higher in individuals with type II diabetes than in healthy volunteers in the Romanian population (Schwab et al., 2008; Boini et al., 2006) reported in journal Diabetes that salk intake decreases SGK-dependent glucose uptake and thus SGK plays a role in glucose intolerance in mice. We also found other pathways which were not detected. We need to be further validated biologically. The above findings can help scientists identify potential biomarkers, drug targets, and generate further biological hypothesis for testing.

3.9.2 Significant Pathways Obtained from CHOT

We also apply our CHOT to our type II diabetes data. We compare two results based on HOT and CHOT. The empirical p-values are shown in Table 3.16.

Table 3.16: Significant pathways using our HOT approaches: HOT=Hybrid Omnibus Test; CHOT=Contraint HOT

Pathway ID	Pathway Name	P-value (HOT)	P-value (CHOT)
29	c10_U133_probes(user defined)	0.047	0.01
35	c16_U133_probes(user defined)	0.01	0.012
37	c18_U133_probes(user defined)	0.009	0.008
40	c20_U133_probes(user defined)	0.02	0.024
46	c25_U133_probes(user defined)	0.03	0.033
47	c26_U133_probes(user defined)	0.023	0.047
55	c33_U133_probes(user defined)	0.015	0.028
62	c8_U133_probes(user defined)	0.007	0.011
102	GO_0005739_HG-U133A_probes(user defined)	0.021	0.04
103	Histidine metabolism	0.028	0.042
154	MAP00380_Tryptophan_metabolism(user defined)	0.011	0.009
155	MAP00380_Tryptophan_metabolism (user defined)	0.007	0.014
213	MAPK signaling pathway	0.013	0.048

We compare how many pathways are significant in both HOT and CHOT. Total 227 pathways are same results using both methods. The total 13 out of 227 are significant using both methods. The total 214 out of 227 are insignificant using both methods. However, 51 pathways does not have same results using both methods. The 43 out of 51 are significant using HOT and 8 out of 51 are significant using CHOT. These results are also summarized in Table 3.17.

Table 3.17: Matched and mismatached results between HOT and CHOT; HOT=Hybrid Omnibus Test; CHOT=Constraint HOT

Matched 227		Mismatched 51	
Both significance	Both insignificance	HOT only significance	CHOT only significance
13	214	43	8

3.10 Discussion

Numerous statistical methods have been developed for analyzing high-dimension data. Their methods are mainly focused on variable selection, do not apply to high-dimensional data and require a known likelihood function. In this chapter we propose Hybrid Omnibus Test (HOT) for high-dimensional data. Our approach is developed using a semi-parametric framework where a likelihood function is not available. We propose an efficient score, which serves as a local test statistic associating with estimating equations, to avoid likelihood derivations when unavailable. With this score, we further develop our HOT within a Bayesian setting which assigns different weights to non-parametric bases in order to emphasize important components.

We compare our three approaches denoted as “HOT”, “CHOT”, and “BHOT” with a empirical likelihood ratio test and Bayes factor using simulation study. Our simulation results suggest that our CHOT approach outperforms in Low/High-Dimensional cases. Our HOT and BHOT also perform well in high dimensional cases. They are comparable to CHOT. ELRT performs well for low dimension case as expected, our approaches was comparable to ELRT. We was not able to obtain ELRT because of $p > n$. When the dimension p is larger than n, we hardly establish a estimating algorithm which provides unbiased estimators. In addition, the algorithm involving resampling and re-estimating requires devastating computational cost. But for the high-dimension case, our proposed HOT approaches perform well in terms of FPR and TPR. Tesing based on Approximate Bayes factor does not provide good performance in high dimensional.

Our Hybrid Omnibus Testing approaches have the following several advantages: (1) they do not require to have likelihood function; (2) they are applicable even in high-dimension case $p >> n$; (3) they do not reply on specified estimating equations; (4) they are flexible to build by choosing different bases functions for example spline, fourier, wavelet, triangle and so on; (5) they do not depends on estimating algorithm; (6) they have high computational efficiency. To the best of knowledge, our approach is the novel one to provide these advantage.

We note that we have analyzed each pathway separately in our analysis. It is known that pathways are not independent of each other because of shared genes and interactions among pathways, which makes it difficult to adjust p-value for resampling based inference, while our HOT approaches do not consider multiple comparisons. A proper analysis of the data should take this aspect of the multiple testing problem into account. Developing such multiple comparison method for pathway based analysis will be an interesting and challenging problem because of the complex dependence structure among pathways.

Last but not least, although some of pathways are identified to distinguish diabetes. They need to be further validated biologically.

Chapter 4

Conclusion/Future Work

In this dissertation, we developed the Frequentist-Bayesian Hybrid Tests in semi-parametric and nonparametric models test statistic described in Chapter 2 and Chapter 3. One is to test the significant differences between non-parametric functions and the other is to test allowing any departure of predictors of high dimensional X from constant.

In Chapter 2, we proposed an Adaptive Bayes Sum Test statistic for the equality of two curves. Our approach can be used for testing of significance between two nonlinear curves by taking into account spatial dependence and reducing the effect of non-white noise. We compare our approach with the adaptive Neyman test. Simulation study suggests that both tests are close to nominal level. Our Adaptive Bayes Sum Test is more powerful in detecting both small and large difference, while adaptive Neyman test is not. Indeed, by making use of wavelets filters, the Bayes sum test can filter out the components of interest. For example, in our case, low-frequency components are more important and hence must be given a higher weight. In contrast, the adaptive Neyman test treats every data point uniformly, and is unable to filter out more important components.

In Chapter 3, we proposed the efficient score, which serves as a local test statistic associating with estimating equation, to avoid likelihood derivation when unavailable. With this score, we further develop our Hybrid Omnibus Test (HOT) with Bayesian setting which assigns different weight to non-parametric basis in order to emphasizing important components (basis). We propose three HOT approaches, dentated as "HOT", "CHOT", and "BHOT", which represent HOT without constraints, HOT with constraint estimating equation, and HOT using Bayesian MPE, respectively. We compare our three HOT approaches with a empirical likelihood ratio test and Bayesian inference based on Bayes factor using simulation study. Our simulation results suggest that our CHOT approach outperforms in terms of FPR and TPR under Low/High-Dimensional cases. Our HOT and BHOT also perform well in high dimensional cases. They are also comparable to CHOT. ELRT performs well for low dimension case as expected, our approaches was comparable to ELRT. We was not able to obtain ELRT because of $p > n$. When the dimension p is larger than n, we hardly establish a estimating algorithm which provides unbiased estimators. In addition, the algorithm involving resampling and re-estimating requires devastating computational cost. But for the high-dimension case, our proposed HOT approaches perform well in terms of FPR and TPR. Testing based on Approximate Bayes factor does not provide good performance in high dimensional case.

Our HOT approaches have several advantages. First, they do not require a likelihood function;. Second, they are applicable even in high-dimension case $p >> n$. Third they do not reply on specified estimating equations. Fourth they are flexible to build by choosing different bases functions for example spline, fourier, wavelet, triangle and so on. Fifth they do not depend on estimating algorithms. Sixth, they have high computational efficiency. To the best of our knowledge, our approach is the first one has these advantage.

We proposed a hybrid test statistic still has many remaining issues to address. Many interesting problems require further research works.

Further research is still needed to examine the theoretical properties in detail: Deriving theoretical distributions of resampling procedures will be useful to reduce computation burden as well as to study the theoretical bound of the difference between two functions. Although we found that our method can detect well small difference between two functions in special case, it is difficult to derive the bound of this difference in general case. It will also important to develop multiple comparison method using our approach which may be a challenging problem because of unknown dependence structures among data.

For the second testing problem, we can further focus on the following directions: In our simulation we used polynomial function and sine-bump function. However, we would like to examine more simulations based on different classes of nonlinear function and also study the performance of our approach in higher dimensional case so that we can understand the limitation of our proposed test statistic. We would like to consider different identification constraints for single index coefficients rather than $\|\boldsymbol{\alpha}\|^2 = 1$ to understand flexibility of our approach on identification constraints. In our approach, we used spline basis. However there are other basis functions we can used such as kernel smoother and plate smoother by Franke (1982) to study the sensitivity of choice of basis function. We also would extend our approach for mixed model frame work. In this case, we can develop our approach using estimating equations suggested by Zeger et al. (1988) for mixed model frameworks. We may improve the computational efficiency or achieve better convergence by substituting current optimization algorithm, for such as Sequential Quadratic Programming by Boggs & Tolle (1995), and Bayesian sampler for slice sampler proposed by Neal (2003) along with proposal distributions. For BHOT, we use non-informative prior. However, BHOT may perform better if we incorporate informative prior into the method. However, it may be a challenging problem because of non-identifiability between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ so that it is difficult to find proper prior or conjugate prior for $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta})$. It will be also an interesting and challenging problem to develop multiple comparison method using our approach because pathways are not independent of each other because of shared genes and interactions among pathways, which makes it difficult to adjust p-value; A proper analysis of the data should take this aspect of the multiple testing problem into account.

Bibliography

Aerts, M., Claeskens, G., & Hart, J. D. (2004). Bayesian-motivated tests of function fit and their asymptotic frequentist properties. *The Annals of Statistics*, *32*(6), 2580–2615.

Bayarri, M. J. & Berger, J. O. (2004). The interplay of bayesian and frequentist analysis. *Statistical Science*, *19*(1), 58–80.

Behseta, S. & Kass, R. E. (2005). Testing equality of two functions using bars. *Statistics in medicine*, *24*(22), 3523–3534.

Besse, P. & Ramsay, J. (1986). Principal components analysis of sampled functions. *Psychometrika*, *51*(2), 285–311.

Boggs, P. T. & Tolle, J. W. (1995). Sequential quadratic programming. *Acta numerica*, *4*, 1–51.

Boini, K. M., Hennige, A. M., Huang, D. Y., Friedrich, B., Palmada, M., Boehmer, C., Grahammer, F., Artunc, F., Ullrich, S., Avram, D., et al. (2006). Serum-and glucocorticoid-inducible kinase 1 mediates salt sensitivity of glucose tolerance. *Diabetes*, *55*(7), 2059–2066.

Cai, T. T. & Brown, L. D. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, *26*(5), 1783–1799.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Chang, M. & Chow, S.-C. (2005). A hybrid bayesian adaptive design for dose response trials. *Journal of Biopharmaceutical Statistics*, *15*(4), 677–691.

Coleman, T. F. & Li, Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, *6*(2), 418–445.

Dahlquist, K. D., Salomonis, N., Vranizan, K., Lawlor, S. C., & Conklin, B. R. (2002). Genmapp, a new tool for viewing and analyzing microarray data on biological pathways. *Nature genetics*, *31*(1), 19–20.

Darling, D. & Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J*, *23*(1), 143–155.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association*, *91*(434), 674–688.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fan, J. & Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, *93*(443), 1007–1021.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, *39*(3), 254–261.

Franke, R. (1982). Smooth interpolation of scattered data by local thin plate splines. *Computers & Mathematics with Applications*, *8*(4), 273–281.

Good, I. (1957). Saddle-point methods for the multinomial distribution. *The Annals of Mathematical Statistics*, *28*(4), 861–881.

Hall, P. & Turlach, B. A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *The Annals of Statistics*, *25*(5), 1912–1925.

Härdle, W. & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, *84*(408), 986–995.

Hart, J. D. (2009). Frequentist-bayes lack-of-fit tests based on laplace approximations. *Journal of Statistical Theory and Practice*, *3*(3), 681–704.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, *58*(1), 71–120.

Kim, I., Pang, H., & Zhao, H. (2012). Bayesian semiparametric regression models for evaluating pathway effects on continuous and binary clinical outcomes. *Statistics in Medicine*, *31*(15), 1633–1651.

Ma, Y. & Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association*, *101*(476), 1465–1474.

Ma, Y., Hart, J. D., Janicki, R., & Carroll, R. J. (2011). Local and omnibus goodness-of-fit tests in classical measurement error models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 81–98.

Maiti, A., Casarona, J., McHale, L., & Schaumont, P. (2010). A large scale characterization of ro-puf. In *Hardware-Oriented Security and Trust (HOST), 2010 IEEE International Symposium on*, (pp. 94–99). IEEE.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov,

J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., & Groop, L. C. (2003). Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*, *34*(3), 267–273.

Neal, R. M. (2003). Slice sampling. *Annals of statistics*, 705–741.

Padival, A. K., Hawkins, K. S., & Huang, C. (2004). High glucose-induced membrane translocation of pkc βi is associated with arf6 in glomerular mesangial cells. *Molecular and cellular biochemistry*, *258*(1-2), 129–135.

Pang, H., Lin, A., Holford, M., Enerson, B. E., Lu, B., Lawton, M. P., Floyd, E., & Zhao, H. (2006). Pathway analysis using random forests classification and regression. *Bioinformatics*, *22*(16), 2028–2036.

Rice, J. A. & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 233–243.

Sadeghi, A.-R. & Naccache, D. (2010). *Towards hardware-intrinsic security*. Springer.

Schwab, M., Lupescu, A., Mota, M., Mota, E., Frey, A., Simon, P., Mertens, P. R., Floege, J., Luft, F., Asante-Poku, S., et al. (2008). Association of sgk1 gene polymorphisms with type 2 diabetes. *Cellular Physiology and Biochemistry*, *21*(1-3), 151–160.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society*, 1461–1481.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tsiatis, A. A. & Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, *91*(4), 835–848.

Verzelen, N. & Villers, F. (2010). Goodness-of-fit tests for high-dimensional gaussian linear models. *The Annals of Statistics*, *38*(2), 704–752.

Voisine, P., Ruel, M., Khan, T. A., Bianchi, C., Xu, S.-H., Kohane, I., Libermann, T. A., Otu, H., Saltiel, A. R., & Sellke, F. W. (2004). Differences in gene expression profiles of diabetic and nondiabetic patients undergoing cardiopulmonary bypass and cardioplegic arrest. *Circulation*, *110*(11 suppl 1), II–280.

Wang, Q. (2011). *Frequentist-Bayes goodness of fit tests*. PhD thesis, Texas A&M University.

Weinberg, M. D. (2012). Computing the bayes factor from a markov chain monte carlo simulation of the posterior distribution. *Bayesian Analysis*, *7*(3), 737–770.

Yu, Y. & Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*(460).

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1049–1060.

Zhang, D., Zhou, Z., Li, L., Weng, J., Huang, G., Jing, P., Zhang, C., Peng, J., & Xiu, L. (2006). Islet autoimmunity and genetic mutations in chinese subjects initially thought to have type 1b diabetes. *Diabetic medicine*, *23*(1), 67–71.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.

Appendix A

Wavelet Transformation

A function $\phi \in L^2(\mathbb{R})$ can be used to define an orthonormal system, for the Hilbert space $L^2(\mathbb{R})$ of square integrable functions is an ”orthonormal wavelet”.

The Hilbert basis is constructed as the family of functions $\{\phi_{jk} : j,k \in \mathbb{Z}\}$ by means of Dyadic transformations and dilations of ϕ,

$$\phi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k)$$

for integers $j,k \in \mathbb{Z}$.

This family is defined as an orthonormal system if it is orthonormal under the standard inner product

$$\langle f,g \rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx$$

on $L^2(\mathbb{R})$.

$$\langle \phi_{jk}, \phi_{lm} \rangle = \delta_{jl}\delta_{km}$$

where δ_{jl} is the Kronecker delt.

Completeness is satisfied if every function $h \in L^2(\mathbb{R})$ may be expanded in the basis as

$$h(x) = \sum_{j,k=-\infty}^{\infty} c_{jk}\phi_{jk}(x)$$

with convergence of the series. Such a representation of a function "f" can be referred as a "wavelet basis".

Appendix B

Adaptive Bayes Sum Test Statistic

In this section, we briefly describe how we derive our test statistic using the results of Hart (2009).

Let $\phi_{j,k}$, $k = 0, 1, \ldots, 2^J - 1$ denote the true wavelet coefficients at resolution level j, $j = 0, 1, \ldots, J - 1$. By assigning the same prior probability of each of $\phi_{j,0}, \ldots, \phi_{j,2^j-1}$,

$$\pi_j = \frac{1}{(1 + 2^{j\alpha})}, \ \alpha > 1$$

the test statistic B_n takes the form

$$B_n = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} 2^{-j\alpha} \exp(\frac{n\hat{\phi}_{jk}^2}{2\sigma^2}).$$

If one uses a trigonometric basis, the sample Fourier coefficients will be a function of two frequency indices, say j and k, and may be written $\hat{\phi}_{jk}$, $j = 1, \ldots, \sqrt{n}$, $k = 1, \ldots, \sqrt{n}$, where for convenience we assume that $\sqrt{n}$ is an integer. The statistic B_n has the form

$$B_n = \sum_{j=1}^{\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{\pi_{jk}}{1 - \pi_{jk}} \exp(\frac{n\hat{\phi}_{jk}^2}{2\sigma^2}).$$

Assuming that each of the indices j and k is proportional to frequency, it is reasonable to take as vague probabilities

$$\pi_{jk} = \frac{1}{1+(jk)^c},\ j = 1,\ldots,\sqrt{n}, k = 1,\ldots,\sqrt{n},$$

for some $c > 1$.

Therefore, with the selection of the prior $\pi_{jk} = \frac{1}{1+(jk)^2}$, $c = 2$, and $j = 1,\ldots,n_1$, $k = 1,\ldots,n_2$, our test statistic T_{AB} (1) can be established.

We note that the number of frequencies is not considered important when it is greater than 15. We choose $p = 2$ to achieve better power in balancing the emphasis on lower frequencies, which is shown by Wang (Wang, 2011).

Appendix C

Computation of Bayes Factor Using MCMC of the Posterior Distribution

Consider a model selection problem on the basis of observed data $\mathbf{D}$ and the plausibility of the model $\mathscr{M}$ parameterized by model parameter vectors $\theta \in \Omega$, where Ω is the sample space. Let $f(\theta)$ be a density function and Ω_s be the trimmed subdomain of Ω, that is $\Omega_s \subset \Omega$. Define $Y_0 = \inf\{Y(D|\theta) : \theta \in \Omega_s\}$ and $Y_1 = \sup\{Y(D|\theta) : \theta \in \Omega_s\}$. Define the following notations

$$
\begin{aligned}
&Z \equiv P(\mathbf{D}|\mathscr{M}) = \int d\theta P(\theta|\mathscr{M})P(\mathbf{D}|\theta,\mathscr{M}),\\
&M(y) \equiv \int_{f(\theta)>y} d\theta = \int_{P(\theta|\mathbf{D})>y} d\theta\\
&I = \int_0^{\sup\{f(\theta):\theta\in\Omega\}} M(y)dy,\\
&J \equiv \int_{\Omega_s} d\theta P(\theta),\\
&Y \equiv 1/P(D|\theta),\\
&K \equiv \int_{\Omega_s} \frac{d\theta P(\theta|D)}{P(D|\theta)} = \int_0^{Y_1} M(Y)dY == \int_{Y_0}^{Y_1} M(Y)dY + M(Y_0)Y_0,
\end{aligned}
$$

Then $Z = J/K$ and for the Bayes Factor B, with respect to models $\mathscr{M}_1$ and $\mathscr{M}_2$,

$$B = \frac{Z_{\mathscr{M}_1}}{Z_{\mathscr{M}_2}},$$

where $Z_{\mathscr{M}_1}$ is Z under $\mathscr{M}_1$ and $Z_{\mathscr{M}_2}$ is Z under $\mathscr{M}_2$, respectively.

Hence we need to calculate K and J under each model. We calculate them using Weinberg (2012). We explain how to calculate them in the following:

- The numeric algorithm to calculate K is to use Riemann sums: We begin with the grid defined by the ordered set $\{Y_i\}$ induced by the MCMC sample. The Monte Carlo expression of $M(y)$ is

  $$M_i^{[l]} \equiv \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{Y_j > Y_i\}}, \quad \text{or} \quad M_i^{[u]} \equiv \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{\{Y_j \geq Y_i\}}, \quad \text{or} \quad M_i \equiv \frac{M_i^{[l]} + M_i^{[u]}}{2},$$

  where the indicator function includes a contribution from the index j only if $\{Y_j \geq Y_i\}$ or $\{Y_j \geq Y_i\}$ for the lower and upper form, respectively. Then similarly, we evaluate K from $\{M_i\}$ by upper and lower Riemann sums:

  $$K^{[l]} \equiv \sum_{j=1}^{N} (h_j M_{j-1} + M(Y_0)Y_0 + \delta_j^{[l]}) \quad \text{or} \quad K^{[u]} \equiv \sum_{j=1}^{N} (h_j M_{j-1} + M(Y_0)Y_0 + \delta_j^{[u]})$$

  with $K \equiv (K^{[l]} + K^{[u]})/2$, $h_j \equiv Y_j - Y_{j-1}$, and

  $$\delta_j^{[l]} \equiv \frac{h_j^2}{2} M'(Y_{j-1}) \approx \frac{h_j^2}{2}\left(\frac{M_j - M_{j-2}}{Y_j - Y_{j-2}}\right) \quad \text{or} \quad \delta_j^{[u]} \equiv \frac{h_j^2}{2} M'(Y_j) \approx \frac{h_j^2}{2}\left(\frac{M_{j+1} - M_{j-1}}{Y_{j+1} - Y_{j-1}}\right).$$

- The numeric algorithm to calculate J: We partition Ω_s into exclusive subsets $\{\omega_s\}$ containing a fixed number of leaves c and let $\upsilon(\omega_s)$ be the volume enclosing each subset ω_s as determined by the partition. Let $f(\theta_i)$ denote $P(\theta_i)$, respectively for $\theta_i \in \omega_s$. Then we can evaluate J as a Riemann sum, multiplying $\upsilon(\omega_s)$ by some representative value of the prior probability, $f_*(\omega_s)$, for each element of $\{\omega_s\}$ (such as a p-quantile or mean value) and then sum the contributions for all $\{\omega_s\}$:

$$J \approx \sum_{\omega_s \in \Omega_s} \upsilon(\omega_s) f_*(\omega_s).$$