Dynamic Conditional Density Models for Multivariate Longitudinal Data

by Yuan Gu

B.S. in Engineering, July 2009, University of Shanghai for Science and Technology
M.S. in Biostatistics, August 2015, University of Pittsburgh

A Dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial satisfaction of the requirements
for the degree of Doctor of Philosophy

May 17, 2020

Dissertation directed by

Colin O. Wu
Mathematic Statistician, National Institute of Health, NHLBI

The Columbian College of Arts and Sciences of The George Washington University certifies that Yuan Gu has passed the Final Examination for the degree of Doctor of Philosophy as of February 21, 2020. This is the final and approved form of the dissertation.

Dynamic Conditional Density Models for Multivariate Longitudinal Data

Yuan Gu

Dissertation Research Committee:

Colin O. Wu, Mathematic Statistician, National Institute of Health, NHLBI, Dissertation Director

Hua Liang, Professor of Statistics, Dissertation Co-Director

Angelo Elmi, Associate Professor of Biostatistics, Committee Member

Xin Tian, Mathematic Statistician, National Institute of Health, NHLBI, Committee Member

© Copyright 2020 by Yuan Gu
All rights reserved

Acknowledgments

With many thanks to my mentors Dr. Colin O. Wu and Dr. Hua Liang for their help and guidance during this research and work for years. To Dr. Xin Tian, for her support and encouragement throughout the dissertation process and my PHD life. I also appreciate Dr. Eileen E. Navarro-Almario and Dr. Frank Pucino's long-term help and trust. Furthermore, to all my committee members, Dr. Angelo Elim, Dr. Tatiyana V. Apanasovich, Dr. Subrata Kundu for providing great insights and suggestions, without these, the research would not have been possible. Finally, to my beloved parents, who provide me the psychological and financial support during my whole life.

Abstract

Dynamic Conditional Density Models for Multivariate Longitudinal Data

Longitudinal studies allow to evaluate and make statistical inferences of the relationships between risk factors and the development of disease or death over some period of time. However, when the outcome variables of interest have more than two dimensions, the density estimation and the computation are much more complicated and time-consuming. Here, we propose a simple and easily applicable but more flexible method to estimate the dynamic conditional density models of a two-dimensional outcome variable based on other risk factors for multivariate longitudinal data.

Previous studies were all based on either conditional mean or conditional distribution modeling schemes. To address the limitations of all previous work, our improved approach extends the univariate time-varying models to bivariate time-varying models by incorporating both conditional mean and conditional distribution modeling structures. Our study is motivated by the attempt of assessing joint densities of clinical variables in two real word studies, the Framingham Heart Study (FHS) and the National Growth and Health Study (NGHS).

Based on the previous work of the nonparametric time-varying modeling under appropriate statistical hypotheses, we extend the models by incorporating dynamic joint density distributions of a two-dimensional response variable. Unlike the other direct estimation methods for the multivariate longitudinal data, we construct dynamic models through the conditional modeling approach. Our conditional approach provides us very useful insights in both the FHS and NGHS data: by only making relaxed assumptions: assuming both X|Y and Y|X are conditional normally distributed, given values of age and other health status covariates, e.g. the cholesterol (Chol) or the diastolic blood pressure (DBP), the exact dynamic joint density functions of a two-dimensional response variable, e.g. the bivariate cardiovascular risk factor: (SBP, BMI) over time can be easily generated. For example,

given the mean values of cholesterol, we can estimate the joint density function of SBP and BMI for subjects at age 60. Hence, the exact cumulative densities of SBP < 250 and BMI < 30 at different ages also can be calculated.

The thesis is organized as follows. In Chapter 1, we introduce the motivation of our study; review methods commonly used in analyzing multivariate longitudinal data; the available smoothing methods to study time-varying models: the smoothing framework: two-step smoothing for building the conditional models and the conditional specified modeling for cross-sectional data. In Chapter 2, we look at the special cases of the joint models with exponential and normal conditionals and review their theoretical derivations. In Chapter 3, we propose our approach of estimating dynamic conditional specified joint density models, following similar framework in previous chapters. In Chapter 4, we apply our methods on the Framingham Heart Study and the National Growth and Health Study. In Chapter 5, we explore the feasibility of our approach using simulated data, under three different simulation scenarios. In Chapter 6 we will present the consistency properties of all our estimators at each step. And finally, we will make and discussion about our research and the future work.

Table of Contents

Acknowledgments . . . . . **iv**

Abstract . . . . . **v**

List of Figures . . . . . **ix**

List of Tables . . . . . **xii**

1 Introduction **1**

1.1 Objective of This Longitudinal Study 1
1.2 Motivation Data Sets and Justification 2
1.2.1 Framingham Heart Study (FHS) 2
1.2.2 National Growth and Health Study (NGHS) 3
1.2.3 Summary of current estimation methods 4
1.3 Conditional Specified Approach Model 5
1.3.1 Difference between Conditional Normal and Joint Normal 5
1.3.2 Conditional in Specified Families 7
1.4 Structured Nonparametric Time-varying Coefficient Models 9
1.4.1 Structured Univariate Nonparametric Time-varying Models 10
1.5 Structured Multivariate Nonparametric Models 14

2 Conditional Specified Bivariate Joint Model **17**

2.1 Compatibility and Uniqueness 18
2.1.1 Uniqueness 19
2.1.2 Markov Chain Setting and Uniqueness 21
2.2 Joint Models with Normal Conditionals 21
2.2.1 Noncentered Joint Models with Normal Conditionals 21
2.2.2 Centered Joint Models with Normal Conditionals 22

3 Conditional Specified Time-varying Joint Density Estimation **24**

3.1 Data Structure 24
3.2 Nonparametric Time-varying Conditional Models 25
3.3 Estimation Methods 27
3.3.1 Two-step Time-varying Estimation Method for Conditional Models 27
3.3.2 Estimation of variances for Joint Density of LogSBP and LogBMI 31
3.3.3 Bootstrap Confidence Intervals 35

4 Application **38**

4.1 Application in the Framingham Heart Study (FHS) 38
4.1.1 Data Description 38
4.1.2 Two-step Time-varying estimation for univariate models of LogSBP and LogBMI . 40
4.1.3 Dynamic MME for Joint Density of LogSBP and LogBMI 44
4.1.4 Bootstrap Pointwise Confidence Intervals 52

4.2 Application in the National Growth and Health Study (NGHS) 54
4.2.1 NGHS Data 54
4.2.2 Time Point Binning for NGHS 54
4.2.3 Data Description 54
4.2.4 Two-step time-varying coefficients for conditional models of LogSBP and LogBMI 56
4.2.5 Dynamic MME of variances for Joint Density of LogSBP and LogBMI . . . 56
4.2.6 Discussion 64

5 Simulation **65**

5.1 Simulation Objective 65
5.2 Simulation Setting 65
5.3 Gibbs Sampling 66
5.4 Simulation Scenarios 67
5.4.1 Simulation Scenario 1: Dynamic Bivariate Normal Distribution 67
5.4.2 Simulation Scenario 2: Dynamic Conditional Normal Distributions 75
5.4.3 Simulation Scenario 3: Dynamic Conditional Normal Distributions with dynamic correlations 85
5.5 Results and Discussion 93

6 Asymptotic Property **95**

6.1 Consistency of the two-step time-varying estimators of univariate models . . . 95
6.2 Consistency of MME for joint models 97
6.2.1 Consistency of raw MMEs over time 98
6.2.2 Consistency of dynamic MMEs over time 104

7 Summary and Future Work **105**

Bibliography **108**

A Appendix **112**

A.1 Asymptotic Assumptions for Two-step Time-varying Estimators 112
A.1.1 Notations 113
A.2 Asymptotic properties of the raw estimators in the two-step time-varying estimation 113
A.3 Asymptotic properties of the smoothing estimators in the two-step time-varying estimation 114
A.4 Distribution Theory for Joint Models 114
A.4.1 Assumptions of Error Terms in Condition Models 114
A.4.2 Moment Generation Functions 115
A.4.3 Parameter k and δ 116

List of Figures

1.1 Contour plots of joint distribution of X and Y. 6

1.2 A comparison of three popular kernels for local smoothing. Each is symmetric and has been calibrated to integrate to 1. The Tri-Cube kernel is compact and has two continuous derivatives at the boundary of its support, while the Epanechnikov kernel has none. The Gaussian kernel is continuously differentiable and has infinite support. 11

3.1 functions of LTCV and Bandwidth selection using Grid Search 30

3.2 Generating the Dynamic MME and the Bootstrap Pointwise Confidence Intervals 37

4.1 FHS: Histogram plots of SBP (left), logSBP (right) 39

4.2 FHS: Histogram plots of BMI (left), logBMI (right) 39

4.3 FHS: The coefficients over time for first model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)Chol(t)$. Left: black: $\hat{\alpha}_1(t)$, red: $\tilde{\alpha}_1(t)$. Right: black: $\hat{\beta}_0(t)$, red: $\tilde{\beta}_0(t)$. 41

4.4 FHS: The coefficients over time for first model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)Chol(t)$. Black: $\hat{\beta}_1(t)$, Red: $\tilde{\beta}_1(t)$. 42

4.5 FHS: The coefficients over time for first model:$logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)Chol(t)$. Left: black: $\hat{\alpha}_2(t)$, red: $\tilde{\alpha}_2(t)$. Right: black: $\hat{\gamma}_0(t)$, red: $\tilde{\gamma}_0(t)$. . 43

4.6 FHS: The coefficients over time for first model:$logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)Chol(t)$. Black: $\hat{\gamma}_1(t)$, Red: $\tilde{\gamma}_1(t)$. 44

4.7 The 3D density distributions and the contour plots of bivariate normal distributions given different parameters 46

4.8 FHS: Estimators of the variances of logSBP (left), logBMI (right) at different time point. Blue lines are estimators of the variances of time. Pink and Green curves are 95% bootstrap pointwise confidence intervals. Black dots are original variances of logSBP and logBMI over time directly calculated from data. . . 48

4.9 FHS: Contour plots at t=37 and 53. The corresponding estimated c(t) are 0.075 and 0.206. 49

4.10 FHS: Contour plots at t= 67 and 73. The corresponding estimated c(t) are 0.682 and 0.709. 50

4.11 FHS: Estimated CDF plot of logSBP $< log(150)(SBP < 250)$ and logBMI $< log(30)(BMI < 30)$ based on Mean(Chol) at age 37, 53, 67 and 73 51

4.12 NGHS: Histogram plots of SBP (left), logSBP (right) 55

4.13 NGHS: Histogram plots of BMI (left), logBMI (right) 55

4.14 NGHS: The estimated coefficients over time in model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)DBP(t)$. left: blue $\hat{\alpha}_1(t)$. red: $\tilde{\alpha}_1(t)$ right: blue: $\hat{\beta}_0(t)$. red: $\tilde{\beta}_0(t)$ 57

4.15 NGHS: The estimated coefficients blue: $\hat{\beta}_1(t)$ and red: $\tilde{\beta}_1(t)$ over time in model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)DBP(t)$. 57

4.16 NGHS: The estimated coefficients over time in model: $logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)DBP(t)$. left: blue: $\hat{\alpha}_2(t)$ red: $\hat{\alpha}_2(t)$. right: blue: $\hat{\gamma}_0(t)$ red : $\tilde{\gamma}_0(t)$. 58

4.17 NGHS: The estimated coefficients blue: $\hat{\gamma}_1(t)$ red: blue: $\tilde{\gamma}_1(t)$ over time in model: $logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)DBP(t)$. 58
4.18 NGHS: Estimators of the variances of logSBP (left), logBMI (right) conditioning on mean values of DBP at different time points. Blue lines are estimators of the variances of time. Purple and Green curves are 95% pointwise bootstrap confidence intervals. Black dots are original variances of logSBP and logBMI calculated directly from the data. 60
4.19 NGHS: Contour plots at age 9.1, 11, 15 and 19. The corresponding estimated c values are 0.048, 0.055, 0.197 and 0.77. 61
4.20 NGHS: Estimated CDF plot of logSBP $< log(150)$ ($SBP < 150$) and logBMI $< log(30)$ ($BMI < 30$) based on Mean(Chol) at age 9.1, 11, 15 and 19 . . . 62

5.1 Gibbs Sampling algorithm 67
5.2 Scenario 1: Simulated mean functions of time for X, Y and Z. 69
5.3 Scenario 1: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Red curves: estimated time-varying coefficient curves; Blue dots: raw coefficients at different time points 71
5.4 Scenario 1: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$. Red curves: estimated time-varying coefficient curves; Blue dots: raw coefficients at different time points 71
5.5 Scenario 1: Estimated C(t) over time and 95% Bootstrap pointwise confidence intervals based on 800 bootstrap samples. Black: estimated C(t); Purple and Orange: CIs 72
5.6 Scenario 1: Dynamic MME of variances for X (a) and Y (b) over time and 95% Bootstrap pointwise CIs conditioning on the mean value of Z. Black curve: estimated curve; Red dots: true value; Purple and Orange: CIs 74
5.7 Scenario 2: Simulated mean of time for X, Y and Z 76
5.8 Scenario 2: Simulated histograms and density contours for x and y at time point t= 1 77
5.9 Scenario 2: Simulated histograms and density contours for x and y at time point t= 25 78
5.10 Scenario 2: Simulated histograms and density contours for x and y at time point t= 50 78
5.11 Scenario 2: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points. 79
5.12 Scenario 2: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$.Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points. 80
5.13 Scenario 2: Estimated dependent parameter c(t). Black line: estimated c(t), Green line: true c(t), Red and Purple lines: 95% Bootstrap pointwise confidence intervals of c(t) 81

5.14 Scenario 2: Dynamic MME of variances over time and bootstrap point wise confidence intervals for the X and Y conditioning on the mean value of Z. Red dots: true variances; Black: estimated MME; Purple and Orange: 95% Bootstrap pointwise CIs 82
5.15 Scenario 2: Simulated density contours for x and y at time point t= 1, 20, 40 and 50 84
5.16 Scenario 3: Density contours for simulated x and y at time point t= 1 and t= 25 87
5.17 Scenario 3: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points. 88
5.18 Scenario 3: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$.Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points. 88
5.19 Scenario 3: Estimated dependent parameter c(t). Black curve: estimated c(t), Green curve: true c(t), Red and Purple curves: 95% Bootstrap pointwise confidence intervals 89
5.20 Scenario 3: Dynamic MME of variances over time and bootstrap point wise confidence intervals for the X and Y conditioning on the mean value of Z. Black: estimated MME; Purple and Orange: 95% Bootstrap pointwise confidence intervals 90
5.21 Scenario 3: Simulated density contours for x and y at time point t= 1, 20, 40 and 50 92

List of Tables

4.1 FHS: Estimated c(t) and variances of logSBP(t) and logBMI(t) at age 37, 53, 67 and 73 48
4.2 FHS: Estimated of CDF when $SBP < 150$ and $BMI < 30$ based on Mean(Chol) at age 37, 53, 67 and 73 in FHS. 51
4.3 FHS: 95% Boostrap pointwise CI of $\hat{\sigma}^2_{logSBP}(t)$ and $\hat{\sigma}^2_{logSBP}(t)$ at age 37, 53, 67 and 73 based on Mean(Chol). 52
4.4 NGHS: Estimated of CDF when $SBP < 150$ and $BMI < 30$ based on Mean(DBP) at age 9.1, 11, 15 and 19 63
4.5 NGHS: 95% Boostrap pointwise CI of MME parameters for Joint model based on Mean(DBP) at age 9.1, 11, 15 and 19 63

5.1 Scenario 1: Estimated $\hat{c}(t)$ versus True $c(t)$ at 5 time points 72
5.2 Scenario 1: First column: time points; Second column: true variance of x at t; Third column: MME of $\sigma^2_x(t)$; Forth column: 95% Bootstrap pointwise CIs based on 800 bootstrap samples; Fifth column: absolute error of $\hat{\sigma}^2_x(t)$. . . . 73
5.3 Scenario 1: First column: time points; Second column: true variance of x at t; Third column: MME of $\sigma^2_y(t)$; Forth column: 95% Bootstrap pointwise CIs based on 800 bootstrap samples; Fifth column: absolute error of $\hat{\sigma}^2_y(t)$. . . . 73
5.4 Scenario 2: Estimated $\hat{c}(t)$ versus True c(t) at different time points. First column: time points; Second column: true c(t); Third column: estiamted c(t) 80
5.5 Scenario 2: Estimated $\hat{\sigma}^2_x(t)$, 95% Bootstrap pointwise CIs based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}^2_x(t)$ 83
5.6 Scenario 2: Estimated $\hat{\sigma}^2_y(t)$, 95% Bootstrap pointwise CIs based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}^2_y(t)$ 83
5.7 Scenario 3: Estimated $\hat{c}(t)$ versus True c(t) at different time points 89
5.8 Scenario 3: Estimated $\hat{\sigma}^2_x(t)$, 95% Bootstrap pointwise confidence intervals based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}^2_x(t)$ 91
5.9 Scenario 3: Estimated $\hat{\sigma}^2_y(t)$, 95% Bootstrap pointwise confidence intervals based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}^2_y(t)$ 91

Chapter 1: Introduction

In biomedical studies, researchers are often interested in evaluating the relationships between risk factors and the development of some disease or death over a period of time — can be years or even decades. This kind of study, observational or experimental, using the data gathered for subjects with repeated observations over a period of time, is called the longitudinal study.

1.1 Objective of This Longitudinal Study

For modeling observational longitudinal data, the conditional mean based regression models are most commonly used for evaluating the dynamic behaviors of risk factors or treatment effects on the disease development over time.

The class of univariate mean based models, the linear regression models at different time points, aims to estimate the conditional expectation, which is notated as the conditional mean of Y given X: $E(X(t)|Y(t))$ [7] This class of model inference crucially relies on assumptions, for example, we assume the outcome variables follow in some symmetric distribution or normal distributions. When the assumption on outcome variables are violated, or when the modeling inference is less informative for the scientific objectives, the other regression modeling schemes based on conditional distributions will be more capable to investigate the dynamic information of response variables give a set of risk factors, for example, the quantile regression [47, 31, 10]. On the other hand, despite of a lot of modeling approaches have been studied, when the outcome variables of interest have more than two dimensions, the density estimation are getting more complicated and the computation are more time-consuming.

In this thesis study, we propose a simple and easily applicable but also very flexible method to incorporate both conditional mean and conditional distribution modeling struc-

tures to estimate the dynamic joint density models of a two-dimensional outcome response variable based on some other risk factors for multivariate longitudinal data.

1.2 Motivation Data Sets and Justification

In this thesis study, we focus on a class of specified dynamic joint density estimation method, based on univariate conditional time-varying models. This approach is motivated by getting some practical sense of how joint risk factors evolve by clinical visiting times. We use two real-life examples, the Framingham Heart Study (FHS) and the National Growth and Health Study (NGHS), both are long-term, ongoing cardiovascular prospective cohort studies.

We attempt to model the dynamic covariate effects on the joint dynamic density distributed two-dimensional variables, which are, joint variables of Systolic Blood Pressure (SBP) and BMI in FHS, conditioning on cholesterol over ages or the joint variables of SBP and BMI in NHGS, conditioning on DBP over ages.

1.2.1 Framingham Heart Study (FHS)

In 1948, the Framingham Heart Study – under the direction of the National Heart, Lung, and Blood Institute; NHLBI – embarked on an ambitious project in health research. The objective of the Framingham Heart Study was to identify the common factors or characteristics that contribute to cardiovascular disease (CVD) by following its development over a long period of time in a large group of participants who had not yet developed over symptoms of CVD or suffered a heart attack or stroke. This study was one of the first long-term prospective studies and was planned for 20 years initially. It continues today with more than 50 years of follow up on the original cohort recruited in 1948.

Over the years, a total of 32 exams were performed every two years, collecting extensive longitudinal data on clinical information, careful monitoring of the Framingham Study population has led to the identification of the major CVD risk factors – such as high blood pressure, high blood cholesterol, obesity, diabetes, and physical inactivity – as well as a

great deal of valuable information on the effects of related factors such as blood triglyceride and HDL cholesterol levels, age, gender, and psychosocial factors, and lifestyle habits such as smoking and alcohol consumptions, cognitive, physical and daily living activities, medical history including the occurrence of any thyroid disease, respiratory problems, CVD procedures, cancer, diabetes, prostate disease, gout, among others. The incidences and longitudinal progression of any heart diseases were also documented, including coronary heart disease (CHD) such as angina pectoris, coronary insu ciency, myocardial infartion, atrial brillation, syncope, cerebrovascular disease, peripheral arterial disease, and so on.

On the FHS, researchers have studied the longitudinal effects of SBP using multilevel modeling [9] , multivariate analysis by showing descriptively the average annual incidence of CHD over 14 years according to the level of SBP and DBP [19], multivarite variance-components analysis of longitudinal blood pressure measurements [33], semiparametric partial linear mean modeling structure to find the effect of age and cholesterol level on blood pressure [44], etc. Those were all mean based methods gave information of what factors were significant for SBP from an univariate conditional model or multivariate models of SBP at different age intervals by using variance components.

1.2.2 National Growth and Health Study (NGHS)

The National Growth and Health Study is also a typical example of longitudinal observational study. The research goal was to evaluate the racial disparity in obesity on CVD risk factors from pre-adolescence through maturation between African-American and Caucasian girls. This multiple clinical centers collaborating study was initiated between 1987 to 1988 and renewed twice through 2000, until the subjects reached ages of 19 to 20. A total of 1213 African-American and 1166 Caucasian girls, aged at 9 or 10, were recruited [13].

In each visit, the anthropometric data, bioelectrical impedance, blood pressure, serum lipids, lipoproteins and apolipoprotein, medical history, dietary intake and eating patterns, physical activity levels and patterns, health attitudes and beliefs, body satisfaction and family

influence, and psychosocial measures were collected on each participant.

On the NGHS, a couple of research groups [38] used mixed effect models to evaluate the effects of race and other baseline covariates on the incidence of blood pressure, including SBP and DBP. To study the covariate effects on the time-varying conditional distributions of various cardiovascular risk factors with the NGHS data, other researchers [49, 48] introduced the time-varying transformation model. Another nonparametric estimation method based on the conditional CDFs and the "tracking ability" of SBP given age, race and height percentile [47] was also developed to estimated the time-trends of SBP.

Among all the above methods, the mixed effects modeling approach has strong assumptions of parametric distributions. Also although among those papers, nonparametric conditional density estimations was used, the dependent variables were all univariate variables.

1.2.3 Summary of current estimation methods

Besides the above studies on the FHS and NGHS data, during the last a few decades, a variety of joint modeling schemes of multivariate longitudinal data have been proposed.

One approach is specifying directly the joint density by marginal modeling, for example, using the copula model [36], which is defined by an m-dimensional cumulative distribution function $C(u_1, ...u_m)$ with uniform marginals. For the m outcomes, the $Y_1, ...Y_m$ have univariate distribution functions $F_1(y_1), ...F_m(y_m)$, respectively. However, although the copula model is mathematical elegant, the parameter estimation is not that evident and the application in longitudinal contexts is often not straightforward due to the unbalanced data structures. Another limitation is that in most settings of multivariate modeling of distributions using copulas, marginal distributions are usually arbitrarily, and many of those methods using multivariate copulas cannot fully exploit the accurate margin information over time.

Another direct estimation approach is the conditional modeling, since the joint den-

sity equal to the marginal distributions multiply the conditional distributions: $f(y_1, y_2) = f(y_1|y_2)f(y_2) = f(y_2|y_1)f(y_1)$. This approach can reduce the modeling task to the specification of a model for each outcome separately. However, this approach needs to specify the variance-covariance structures of marginal models, although different parsimonious structures were used to reduce the number of massive parameters of the variance-covariance structures, it has limitations usually because of that the marginal densities $f(y_1), f(y_2)$ may not be straightforward, hence the modeling are not realistic in applications [43].

Based on those limitations of previous work on the nonparametric time-varying models under appropriate statistical hypotheses [48], we want to make a unified dynamic modeling scheme of both conditional mean and conditional distribution and also extend those univariate conditional models to joint distributions of a two-dimensional response variable given a set of risk factors over time.

According to [4], provided that the families of conditional distributions are compatible, and a related Markov process is indecomposable, then these families of conditional distributions will uniquely determine the joint distribution. In this longitudinal context of our thesis study, we propose an approach to model dynamic bivariate distributions $f(X(t), Y(t))$ with conditional models $f(X(t)|Y(t))$ and $f(Y(t)|X(t))$ which are associated with some parametric families of distributions.

1.3 Conditional Specified Approach Model

1.3.1 Difference between Conditional Normal and Joint Normal

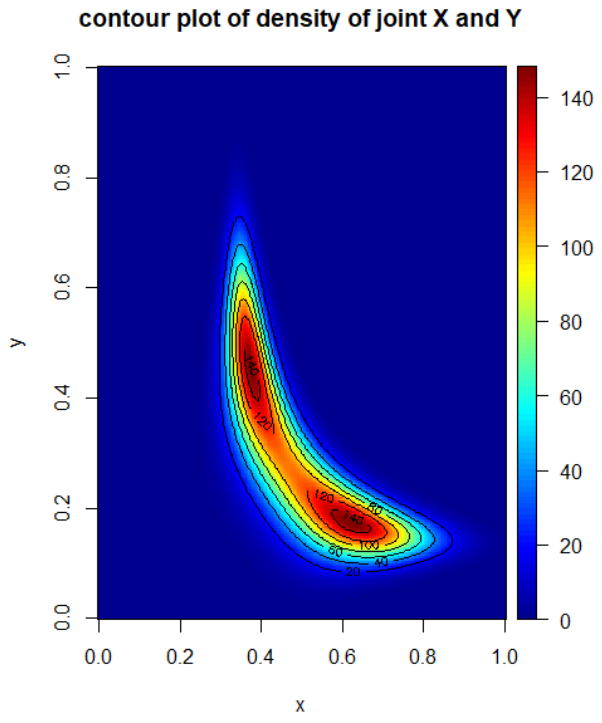
If two variables are joint normally distributed, then either one variable is normal. However, the reverse does not always hold. Let's see an example presenting a distribution that is clearly different from joint normal but conditional normal.

Suppose the conditional densities $X|Y$ and $Y|X$ are both normally distributed and their

distributions are of the following forms:

$$X|Y \sim \mathcal{N}(\frac{5Y+3}{Y^2+1}, \frac{1}{Y^2+1}) ,$$
$$Y|X \sim \mathcal{N}(\frac{5X+3}{X^2+1}, \frac{1}{X^2+1}) ,$$

Plotting probability contours helps us to visualize bivariate distributions. Each one contour line in the contour plot represents constant probability within a certain probability of the distribution lies. The contour plot of joint normal distributions should have all elliptical contours. Also if the contours are not elliptical, then the joint distribution will not be normal. The following contour plot is of joint distribution of X and Y when $X|Y$ and $Y|X$ are conditional normal.



(a)

Figure 1.1: Contour plots of joint distribution of X and Y.

The joint distribution of X and Y is:

$$f(x,y) \propto exp\{-\frac{1}{2}[x^2y^2+x^2+y^2-6x-6y]\}$$

Clearly, the contours in above Figure 1.1 are not elliptical, demonstrating that X and Y are not joint normally distributed. Another quick proof by contradiction is that: assume $X|Y$ and $Y|X$ are both normal distributed, if X and Y are joint normal, then the bivariate density of $U = X - Y$ and $V = X + Y$ are also joint normal, hence the joint distribution of U and V will only have one modal. However, only having the conditional normal $X|Y$ and $Y|X$ might encounter a bimodal distribution of $X - Y$ and $X + Y$. So, we can see that if two variables are conditional normal, their joint distribution might not be bivariate normal. If we start from the assumption of conditional normal and the marginal density: we assume $X|Y$ and Y are normal, then it is easily to get the joint density of X and Y are also normal. However, in most cases, especially in the real studies, the marginal density of Y are impossible to derive.

1.3.2 Conditional in Specified Families

Before embarking on this study, we may need to face the above difficulties and be a little taken aback by the mathematical consequences of only assuming two conditional densities $X|Y$ and $Y|X$ are following in some quite general parametric families, for instance, the exponential family.

Let X, Y be a bivariate density whose conditional densities satisfy:

$$f(x|y) = f_1(x; \underline{\theta}(y))$$
$$f(y|x) = f_2(y; \underline{\tau}(x))$$

where f_1 and f_2 represent two distribution families, they can be the same or different, $S(X)$ and $S(Y)$ are the support sets for X and Y. θ and τ are the parameter spaces for the two distribution families. $\forall x \in S(X)$, $\forall y \in S(Y)$, $\underline{\theta} \in \theta$ and $\underline{\tau} \in \tau$, if the above two equations

are hold, and since we also assume X and Y are following the marginal distributions $f_X(x)$ and $f_Y(y)$ and the joint density exist. Then we must have :

$$f_Y(y)f_1(x;\underline{\theta}(y)) = f_X(x)f_2(y;\underline{\tau}(x))$$

Whether we can solve the equations corresponds to the joint densities which depend on the nature of the functions f_1 and f_2. In a lot of cases, the solutions cannot be achieved. For understanding more, we refer to the articles by [2] for details on theoretical aspect of this question. But in our thesis study, we only consider the eligible cases that solutions of those equations can be solved.

Exponential family conditionals

Assume the conditional distributions are:

$$f(x|y) = f_1(x;\underline{\theta}(y)) = r_1(x)\beta_1(\underline{\theta}(y))exp(\sum_{i=1}^{l_1}\theta_i(y)q_{1i}(x))$$
$$f(y|x) = f_2(y;\underline{\tau}(x)) = r_2(y)\beta_2(\underline{\tau}(x))exp(\sum_{j=1}^{l_2}\tau_j(x)q_{2i}(y))$$

Let $f(x,y)$ be a bivariate density whose conditional densities satisfy the above two equations, it follows that $f(x,y)$ is of the form [2]:

$$f(x,y) = r_1(x)r_2(y)exp(\underline{q}^{(1)}(x)'M\underline{q}^{(2)}(y))$$

where

$$\underline{q}^{(1)}(x) = (q_{10}(x), q_{11}(x), q_{12}(x), ...q_{1l_1}(x)),$$
$$\underline{q}^{(2)}(y) = (q_{20}(y), q_{21}(y), q_{22}(y), ...q_{2l_2}(y))$$

here, $q_{10}(x) = q_{20}(y) \equiv 1$, and the M is a matrix of constants parameters of appropriate dimensions $(i.e., (l_1+1)\times(l_2+1))$ subject to the requirement that :

$$\int_{D_1}\int_{D_2} f(x,y)d\mu_1(x)d\mu_2 y = 1$$

For convenience we can partition the matrix M as follows:

$$M = \left[\begin{array}{c|c} m_{00} & M_{11} \\ \hline M_{12} & \tilde{M} \end{array}\right]$$

where $M_{11} = (m_{01},...m_{0l_2})'$, and $M_{12} = (m_{10},...m_{l_10})$. And note that the case of independence is included, it corresponds to choice $\tilde{M} \equiv 0$.

1.4 Structured Nonparametric Time-varying Coefficient Models

Parametric models are regarding to assumptions of underlying population distributions or relationships between response and predictors. Despite the simplicity in model specification with a finite number of parameters and convenient in model interpretation, the modeling inference may not be correct or can be less efficient if we construct the model based on parametric schemes, especially when the data does not follow a specific distribution.

In many clinical studies, we do not make use of random sampling and the model assumptions are often relaxed. And also in some models, the domain of smoothing is different from human visualization. Hence, the nonparametric analysis will be used to avoid those limitations [16]. The nonparametric statistical model will require weaker assumptions such us the functional smoothness, differentiability, or symmetric of the population than parametric models but it also gives functional relationships between observed variables, unobservable and the outcome variables.

However, since the distributions of those variables may be unknown with unspecified parameters and will be derived by some nonlinear smoothing technique, such as kernel

smoothing, local polynomial smoothing [16], smoothing splines [28], basis expansion smoothing [27] and so on. The nonparametric models may become very complex with an increasing amount of data. Among all those nonparametric models, due to the "curse of dimensionality", completely unstructured nonparametric models are often impractical in the sense that the resulting covariate effects are difficult to interpret and the estimation and inference procedures are numerically unstable [16, 48] . For improving the interpretability and the stability of the numerical estimation and inference procedures, the structured nonparametric models approach should be considered [48].

And when the interests are focusing on response variables through time points $\{(Y_{ij}, t_{ij} : i = 1, ...n; j = 1, ..., n_i)\}$, we need the time-varying data analysis. Hence, the nonparametric time-varying data analysis emerged. Let's look at the structured univariate and multivariate nonparmetric time-varying models.

1.4.1 Structured Univariate Nonparametric Time-varying Models

A simple form of the structured univariate nonparametric time-varying models is as following:

$$Y_{ij} = X_{ij}^T \beta(t_{ij}) + \varepsilon_i(t_{ij}), \tag{1.1}$$

where $\beta(t) = (\beta_0(t), ..., \beta_K(t))^T$ is a random vector which needs to be estimated by some smoothing functions at any time point t_i, $X_{ij} = (X_{ij}^{(0)}, ..., X_{ij}^{(K)})'$ is the observed explanatory covariate vector, t_{ij} is the j^{th} measurement time of the i^{th} subject, $i = 1, ...n$ and Y_{ij} is the outcome dependent vector. ε_i are considered to be identically distributed and the $\varepsilon_i(t_{ij})$ has zero mean and finite variance for the i^{th} subject.

The model permits the coefficients evolves over time as smooth functions. The coefficient function can be obtained by different smoothing methods, which will be discussed in the following chapters. Note that at each fixed time point t_i, this model structure assumes a linear relationship between the dependent variable Y and the set of K number of covariate variables X.

Existing Smoothing Methods

Huge number of literatures have been devoted to the nonparametric smoothing methods, including kernel smoothing, local polynomial regression, basis approximations, spline smoothing, and two-step smoothing methods [16, 48, 41].

(1) Kernel Smoothing

Kernel smoothing is a weighted average smoothing of neighboring observed data determined by a fixed bandwidth h. It achieves flexibility in local estimation and is mathematically tractable. Hence in the setting of kernel smoothing, one is the form of the weights, which depends on the form of kernel $k(x)$, need to be determined.

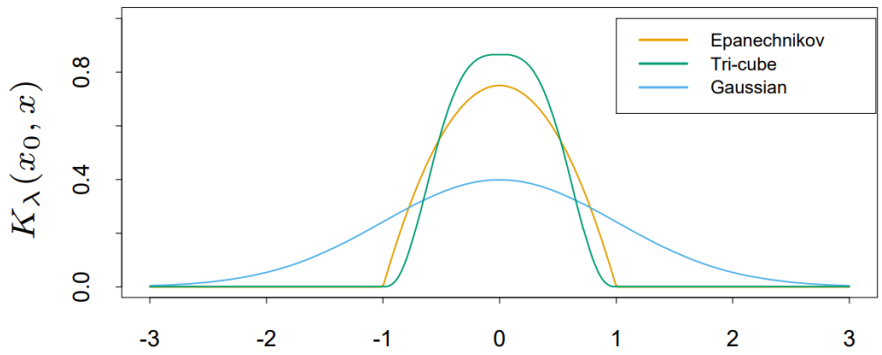


Figure 1.2: A comparison of three popular kernels for local smoothing. Each is symmetric and has been calibrated to integrate to 1. The Tri-Cube kernel is compact and has two continuous derivatives at the boundary of its support, while the Epanechnikov kernel has none. The Gaussian kernel is continuously differentiable and has infinite support.

Here is the comparison of three different kernels for local smoothing. They are all symmetric and integrates to 1. For example, the Gaussian kernel function of t is:

$$K(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} exp^{-\frac{t^2}{2}}, |t| \leqslant 1 \\ 0, |t| > 1 \end{cases} \tag{1.2}$$

For a given scale parameter, the bandwidth h, the weight function is defined:

$$W_{hi}(t_0) = \frac{K(\frac{t_j - t_0}{h})}{\sum_{i=1}^{n} K(\frac{t_j - t_0}{h})}, \tag{1.3}$$

Notice: $\sum_{i=1}^{n} W_{hi}(t_0) = 1$. And the kernel smoother is defined for any t_0 as:

$$\hat{f}(t_0) = \frac{\sum_{i=1}^{n} K(\frac{t_j - t_0}{h}) y_i}{\sum_{i=1}^{n} K(\frac{t_j - t_0}{h})}, \tag{1.4}$$

From above equations, the other one parameter determine the weights of local neighborhoods of t are further determined by this nearest neighbor bandwidth h, which is a nonnegative number. The bandwidth h can vary, adapting to information in the data on the roughness of the function. A larger h indicates the more neighborhoods observations selected and the variance is reduced. The smaller h indicates the less neighborhoods observations selected, hence the more wiggly the smoothing function but the less bias. The bandwidth is usually optimized by least-squares cross-validation [24].

(2) Two-step Time-varying Estimation Method

The two-step smoothing method was proposed by [17]. The name is self-explained, it is using two steps to do the coefficients smoothing. Due to the lack of direct observations of β_t, if there is a way to construct some interim observations $\tilde{\beta}_t$ which keeps track of the shapes and degrees of smoothness of coefficients β_t over time, then it will be feasible to make final smoothing estimators of β_t.

Time Point Binning

In reality studies, most of time, at each distinct time points, the sample has different number of independent observations, especially in medial studies, we always met problems of losing patients' observations during the study time period. At some design time points, the number of observations may be very small, which makes the estimation more difficult.

Before making the two-step estimation, we will need to encounter and solve this problem.

To reduce the effects of minor observation errors, time point binning will be used to group a number of more or less continuous values into a smaller number of "time point bins", which means, make the original design time points fall into a given small interval, a time-point bin [50]. For example, if we have float number of ages as design time points, we might want to arrange the float ages into a smaller number of age intervals (e.g. grouping every five years together).

Suppose we have a sample containing n independent subjects. For each subject, there are n_i number of repeated observations. We denote T_j be the set of subjects whose observations are available at time t_j. So the longitudinal outcome variables and $P \times 1$ covairates for the i_{th} observations are $Y_i(t_{ij})$ and $X_i(t_{ij}) = (X_{i1}(t_{ij}), ..., X_{ip}(t_{ij}))^T$. We will drop the observations either $Y_i(t_{ij})$ or $X_i(t_{ij})$ is not observed. Then we will make the time point binning before heading to the following two steps of two-step smoothing.

First step

The first step is obtaining the "raw estimates" from a preliminary parametric estimation procedure. Let's look at the conditional model:

$$Y_1(t)|Y_2(t) = \beta_0(t) + \beta_1(t)Y_2(t) + \beta_2(t)X_1(t) + ... + \beta_p(t)X_p(t) + e_1(t) \qquad (1.5)$$

Suppose we have enough observations n_j at time point $t_j, j = 1, 2, ...J$, so that the coefficients can be estimated by the maximum likelihood estimators $\tilde{\beta}_p(t_j)s$, also the least square estimators in linear regression, which represents the time trends of the covariate effects. Hence, we get the raw estimators of the coefficients $\beta_0(t), \beta_1(t), ..., \beta_p(t)$ for the conditional regression models. We can write $Y_1(t) = \tilde{\beta}_0(t) + \tilde{\beta}_1(t)Y_2(t) + \tilde{\beta}_2(t)X_1(t) + ... + \tilde{\beta}_p(t)X_p(t) + e_1(t)$. This step is very computational simple.

Second step

In the second step, we use nonparametric methods to estimate those raw estimators $\tilde{\beta}_p(t_j)s$ at different time point t_j. One of the advantages for the two-step estimation method is that the conditional density of errors has less dependence on independent variables than the conditional density of dependent variable itself. And typically parametric estimates converge at a $n^{-1/2}$ rate, nonparametric estimates converge at a rate slower than $n^{-1/2}$, here the second step estimation has a much lower convergence rate [10] than the first step estimator. Besides that, it reduces the variability of the curve estimates at any time point by borrowing the information obtained from the raw estimates at the design time points which are adjunct to t.

Hence, the two-step time-varying estimation method has advantages over the one-step method. And since the two-step smoothing methods only depend on the existing estimation methods, they are computational simple, and their bandwidths can be easily selected by modifying the cross-validation procedures with the classical cross-sectional i.i.d. data [48].

1.5 Structured Multivariate Nonparametric Models

The univariate distribution estimation is most commonly used to analysis outcomes separately. However, the univariate models ignore the association between different outcomes, this may induce bias in the inferences, also due to the large amounts of data in clinical studies and the limited power of one dimensional modeling, it cannot be dissociated from most of the problems found in health studies. For example, after lung transplant, we need to evaluate postoperative normalized forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) [34]. When we want to look at what clinical factors have significant effects on patients with two outcome variables, then a bivariate model must be considered.

Multivariate Nonparametric Smoothing Methods

A multivariate statistical model is a model in which the multiple response variables are modeled jointly. In multivariate modeling practice, many scientists today have already been aware of the power of multivariate statistical algorithms, especially in medical studies, we observe more than one outcomes. However, when the dimension of data is getting higher, or when the data contains complex structures, the computation can be more complicated and the multivariate density estimation are getting progressively more difficult. There are different approaches for the multivariate nonparametric modeling, including the direct smoothing and the indirect approaches.

Direct multivariate nonparametric smoothing

In particular, the direct smoothing approaches, for example, the multivariate kernel estimation, the copula estimation and the local likelihood estimation, can be generalized to higher dimensions. [45] extended univariate regression estimation to the bivariate case, however, their approach has strong limitations, namely constraining to grid kernel support. As discussed in previous text, the copula estimation also has drawbacks, for example, using multivariate copulas cannot fully exploit the accurate margin information.

Also, when the dimensions are increasing, the more likely to have the almost surely empty "local" neighborhoods, this makes less effective of the direct density estimation. Besides that, the asymptotic arguments require massive amounts of data also makes it very difficult to understand and interpret the estimators. It seems that when making the analysis of the high dimensional data through lower-dimensional subspaces, it can be identified and understood much easier.

The Conditional Specified Bivariate Joint Model

Although when the dimensions are increasing, the computations are more complexed, the greatest challenge comes from without knowing a parametric form for the multivariate density.

In particular, when the data contains complex structures, the direct specification approach will be much more difficult to achieve and be less efficient than the conditional approach. In contrast of direct multivariate density estimation for the entire set of random variables, constructing models through the conditional modeling approach, which means specifies a conditional distribution for each response variable given a set of known covariates in the lower dimensional space, then construct the multivariate model according to some mathematical structures, has some advantages.

Our interest throughout the dissertation is on a specific subclass models, the dynamic conditional specification of joint densities for longitudinal data. Based on the previous work on the nonparametric time-varying models and under appropriate statistical hypotheses, we want to extend those models to incorporate the joint distributions of a two-dimensional response variable [48]. In the biomedical studies, these dynamic joint density estimators lead to useful statistical indications to quantitatively track the changes of health status. Hence, we also want to find out if the specified dynamic joint density conditioned on other covariates works on real world medical data.

Chapter 2: Conditional Specified Bivariate Joint Model

Dealing with multi-dimensional datasets which have typically more than two attributes is much more difficult hence more problems will be envisioned. For example, when we have a dependent variable of three dimensions, in addition to checking the prerequisite conditions of a two dimensional response, we need to consider and check if additional requirements are sufficient to satisfy a distribution of the three dimensional response exists. In our dissertation study, we will only focus on the modeling of bivariate response variables, which already presents a plethora of difficulties. To make multivariate estimation extensions, the ideas are already visible in the two dimensions case, so it is helpful for us to start from the simplified bivariate modeling case.

Let (Y_1, Y_2) be a two-dimensional random variable on R^2 and $f_{Y_1,Y_2}(y_1, y_2)$ is a density function with respect to some measure on R^2. For making sure it has a joint density, the following must be required:

(1), $f_{Y_1,Y_2}(y_1, y_2) >= 0$

(2), the integration or summation on R^2 is 1.

An alternative way is to specify the joint characteristic function by the joint Laplace transform:

$$\Phi_{Y_1,Y_2}(t,s) = E(exp[-(tY_1 + sY_2)]), t, s > 0 \tag{2.1}$$

which uniquely determine $F_{Y_1,Y_2}(y_1, y_2)$ by all degree of moments and mixed moments of Y_1 and Y_2, for example, using the joint characteristic function, the joint mean residual life function or the joint hazard functions will all uniquely determine the joint distribution [2].

The other exotic characterization methods exist but they often rely on some specific form of $F_{Y_1,Y_2}(y_1, y_2)$. If we only have the marginal distributions of the two variables, the joint distribution cannot be determined. However, if we using the conditional specification, or incorporate the conditional with the marginal specification, then the joint distribution

can be determined [21]. For going to much detail of this topic, we need to look at what necessary conditions should be satisfied of an existing and unique joint density.

2.1 Compatibility and Uniqueness

Although the joint density distributions can be sometimes characterized completely according to the conditional distributions, one disadvantage is that although specified conditional models $Y_1|Y_2$ and, $Y_2|Y_1$ exist, the consistency must be checked on the given conditional information, because the joint models that compatible with all given conditional models might not exist.

Hence, we need to find out under which conditions there will at least exist one bivariate distribution of (Y_1, Y_2) with respect to some product measure given the conditional distributions of $Y_1|Y_2$ and $Y_2|Y_1$. If the incompatibility exist, there might be some serious issues on the statistical inference. So, first, the compatibility must be checked.

2.1.0.1 Compatibility

The study of the compatibility is usually done through the study of conditionally specified densities. Extensive previous studies have been worked on the compatibility by [5, 25, 2].

Consider the conditional specification for a set of n random variables, $Y = \{Y_1, Y_2, ...Y_n\}$ such that for any $Y_i \in \Omega_i$, $i = 1, ..., n,$, where Ω_i is the support set for Y_i, a set of conditional distributions $\{f_i(y_i|Y_{-i}) : i = 1, ...n\}$ is said to be compatible if there exists at least one joint distribution for Y that is absolutely continuous with respect to some product measure on product Ω and satisfies all the given conditionals. A general condition for compatible conditionals in the bivariate case was derived by Arnold [4].

Consider a set of variables (Y_1, Y_2) such that $Y_1 \in \Omega_1$ and $Y_2 \in \Omega_2$. A joint density or mass function $f(y_1, y_2)$ with $f_1(y_1|y_2)$ and $f_2(y_2|y_1)$ as its conditionals exists if and only if : (1):

$$\{(y_1, y_2) : f_1(y_1|y_2) > 0\} = \{(y_1, y_2) : f_2(y_2|y_1) > 0\} \equiv \Omega, \tag{2.2}$$

(2): there exist functions $u_1(.)$ and $u_2(.)$ such that for all $(Y_1, Y_2) \in \Omega$,

$$\frac{f_1(y_1|y_2)}{f_2(y_2|y_1)} = u_1(y_1)u_2(y_2) \tag{2.3}$$

such that

$$\int_{\Omega_1} u_1(y_1)dv_1(y_1) < \infty \tag{2.4}$$

Notice that the integrability of u_1 implies the integrability of u_2 over Ω_2.
Clearly, since the factor of $f_1(y_1|y_2)$ and $f_2(y_2|y_1)$ can be multiplied by functions u_1 and u_2, then $f_1(y_1|y_2) \propto u_1(y_1)$ and $f_2(y_2|y_1) \propto u_2(y_2)$.

According to the Tonelli's theorem, the condition: $\int_{\Omega_1} u_1(y_1)dv_1(y_1) < \infty$ equals to that $\int_{\Omega_2} \frac{1}{u_2(y_2)}dv_2(y_2) < \infty$. These also implied the marginal densities must be integrable and integrate to 1.

2.1.1 Uniqueness

Once we find that the two conditional distributions are compatible, we may reasonably ask if the associated joint density is unique. Hence, we need to look at the second condition required for a unique bivariate distribution — uniqueness. The uniqueness can be proved by recasting the problem in a Markov Chain setting. Roughly speaking, instructed by the Bayes' theorem, we can treat the conditional probability distributions as the Markov Kernels which are the building blocks interconnected inside of a stochastic system [39, 4]. In our case, the system is the joint probability distributions.

2.1.1.1 Markov Chain

Markov Chain is original proposed by a Russian Mathematician Andrey Markov in 1913. It was defined as a stochastic process describing a sequence of possible events $X_1, X_2, ...X_m$ in an arbitrary probability state space and the conditional distribution of X_{m+1} given the past events: $X_1, X_2, ...X_m$ will only depends on X_m. A Markov chain has stationary transition

probabilities if the conditional distribution of X_{m+1} given X_m does not depend on m. According to [37, 35], we will restrict the Markov Chain by referring it to a specific chain with stationary transition probabilities.

State Space and Markov Kernel For a general state space B, the transition probabilities are specified by defining a Markov kernel, also called a stochastic kernel or a regular conditional probability distribution, by [8]:

$$P(x,S) = \int_S f(y|x)d(y),$$

where S is a measurable set in B and $x \in B$,

Transition Probability Hobert and Casella [26] defined the invariance measure: $\pi(.)$ on the measurable space (A, B) using function $g(x_1,...x_m) : A \to [0,\infty)$ as follows:

$$\pi(X) = \int_A g(x_1,...x_m)\mu(d(x_1,...x_m))$$

and proposed the following theorem:

The measure $\pi(.)$ is a sigma-finite invariant measure for Φ, which is a Markov Chain of conditional distributions, so, for any measurable set $S \in B$:

$$\pi(S) = \int_A \pi(dX)P(X,S)$$

In the above theorem, $X = (x_1,...x_m)$, $m \geqslant 0$ is a Markov Chain living on a continuous state space A with the transition probability density $P(S)$ which is a Markov Kernel, hence we note it as $P(X,S)$. And the probability density function $\pi(S)$ is a stationary distribution for $x_1,...x_m$ on A with stationary transition probabilities. To understand the transition probability and use it to recast the problem of uniqueness, let us to define the Markov chain setting.

2.1.2 Markov Chain Setting and Uniqueness

Now suppose (X,Y) is absolutely continuous with respect to $\mu_1 \times \mu_2$ with support sets $S(X)$ and $S(Y)$, consider a Markov chain with the state space $S(X)$ and transition kernel.

In our study, suppose $m = 2$ and we define the marginal density $\tau(x)$ for x and the transition probability is the following transition Markov kernel:

$$P(X,S(y)) = Transiton\,Kernel(x|z) = \int_{S(y)} f_1(x|y) f_2(y|z) du_2(y).$$

[8, 23, 25, 3] pointed out the necessary and sufficient conditions for uniqueness — or "coherency" of the joint density. If the transition probability is stationary, the conditional densities $f_1(x|y)$ and $f_2(y|x)$ are compatible with the marginal density τ for x and by :

$$joint = conditional \times marginal$$

For such a chain $X_1, X_2, ...X_m$ with the τ which is a stationary distribution (as the π in 2.1.1.1), will be unique if and only if the chain is indecomposable.

2.2 Joint Models with Normal Conditionals

2.2.1 Noncentered Joint Models with Normal Conditionals

Among the distributions of exponential families, the normal conditionals are the most common case. If we assume all the joint distributions have all conditionals in the normal family and the mean and variance are unknown. According to the formula and [2], let $l_1 = 2, l_2 = 2$, and we have $r_1(t) = r_2(t) = 1$ and we may parametrize in a form that :

$$\underline{q}^{(1)}(t) = \underline{q}^{(1)}(t) = \begin{bmatrix} 1 \\ t \\ t^2 \end{bmatrix}$$

yielding the bivariate density of the form:

$$f(x,y) = exp\{\underline{q}^{(1)'}(x)M\underline{q}^{(2)}(y)\} = exp\left\{(1,x,x^2)M\begin{bmatrix}1\\ y\\ y^2\end{bmatrix}\right\}$$

$$= exp\{\underline{q}^{(1)'}(x)M\underline{q}^{(2)}(y)\} = exp\left\{(1,x,x^2)\begin{bmatrix}m_{00} & m_{01} & m_{02}\\ m_{10} & m_{11} & m_{12}\\ m_{20} & m_{21} & m_{22}\end{bmatrix}\begin{bmatrix}1\\ y\\ y^2\end{bmatrix}\right\}$$

The normal conditionals will satisfy one of the following two sets of conditions:

$$(i), m_{22} = m_{12} = m_{21} = 0, m_{20} < 0, m_{20} < 0, m_{11}^2 < 4m_{02}m_{20}$$

$$(ii), m_{22} < 0, 4m_{22}m_{02} > m_{12}^2, 4m_{22}m_{02} > m_{21}^2$$

2.2.2 Centered Joint Models with Normal Conditionals

According to [6, 2], the most general nonstandard density functions which have normal conditional distributions with zero means was identified, the centered joint density function with different parameters is as the following:

$$f_{Y_1,Y_2}(y_1,y_2) = k(c)\frac{1}{2\pi\sigma_1\sigma_2}exp\{-\frac{1}{2}[(\frac{y_1}{\sigma_1})^2 + (\frac{y_2}{\sigma_2})^2 + c(\frac{y_1}{\sigma_1})^2(\frac{y_2}{\sigma_2})^2]\} \tag{2.5}$$

where, $\sigma_1, \sigma_2 > 0, c \geq 0$ and $k(c)$ is normalizing constant, c is the dependent parameter. And we have :

$$Y_1|Y_2 = y_2 \sim N(0, \sigma_1^2(y_2))$$
$$Y_2|Y_1 = y_1 \sim N(0, \sigma_2^2(y_1))$$

where $\sigma_1^2(y_2), \sigma_2^2(y_1)$ are unknown functions and $\sigma_1^2(y_2) > 0, \sigma_2^2(y_1) > 0$ for all y_1, y_2. Then proved by [12]:

$$\sigma_1^2(y_2) = \frac{\sigma_1^2}{(1 + c(y_2/\sigma_2)^2)}$$
$$\sigma_2^2(y_1) = \frac{\sigma_2^2}{(1 + c(y_1/\sigma_1)^2)}$$

By integrating the joint density function with respect to the two variables, using the normal density function, we can also obtain the marginal densities,

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} k(c) \frac{1}{(2\pi)\sigma_1\sigma_2} exp\{-\frac{1}{2}[(\frac{y_1}{\sigma_1})^2 + (\frac{y_2}{\sigma_2})^2 + c(\frac{y_1}{\sigma_1})^2(\frac{y_2}{\sigma_2})^2]\}dy_2$$

let $\frac{y_2}{\sigma_2} = U, U \in (-\infty, \infty)$, hence : $d\frac{y_2}{\sigma_2} = dU = \frac{1}{\sigma_2}dy_2$

$$\begin{aligned}
&\int_{-\infty}^{\infty} k(c) \frac{1}{(2\pi)\sigma_1} exp\{-\frac{1}{2}[(\frac{y_1}{\sigma_1})^2 + U^2 + c((\frac{y_1}{\sigma_1})^2 U^2]\}dU \\
=&\frac{k(c)}{(2\pi)\sigma_1} exp\{-\frac{1}{2}(y_1/\sigma_1)^2\} \int_{-\infty}^{\infty} exp\{-\frac{1}{2}[U^2 + c_t((\frac{y_1}{\sigma_1})^2 U^2]\}dU \\
=&C_1 \int_{-\infty}^{\infty} exp\{-\frac{1}{2}[1 + c(\frac{y_1}{\sigma_1})^2]U^2\}dU \\
=&C_1 \int_{-\infty}^{\infty} exp\{-\frac{1}{2}\frac{U^2}{(1 + c((\frac{y_1}{\sigma_1})^2))^{-1}}\}dU \\
=&C_1(2\pi)^{-0.5}(1 + c(\frac{y_1}{\sigma_1})^2)^{-0.5} \\
=&k(c)\frac{1}{\sigma_1\sqrt{2\pi}}[1 + c(\frac{y_1}{\sigma_1})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{y_1}{\sigma_1})^2]
\end{aligned}$$

Integrating on y_1, we have the marginal density function of y_2:

$$f_{Y_2}(y_2) = k(c)\frac{1}{\sigma_2\sqrt{2\pi}}[1 + c(\frac{y_2}{\sigma_2})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{y_2}{\sigma_2})^2]$$

Hence, from the two marginal functions, only when $c = 0$, then both the marginal densities are normals.

Chapter 3: Conditional Specified Time-varying Joint Density Estimation

As discussed in chapter 1 and 2, the varying-coefficient univariate conditional mean models are most commonly used in all kinds of studies. However, the univariate models have not taken account of the association relationships between different response variables which are ultimately needed to be considered simultaneously in a lot of realistic studies. The existing methods for multivariate extensions have the following drawbacks:

(1) no unified dynamic modeling scheme of both conditional mean and conditional distribution;

(2) the direct estimation methods cannot fully exploit the accurate marginal information over time.

(3) dynamic variance-covariance structures need to be prespecified which might not be accurate.

3.1 Data Structure

Let's assume we have a longitudinal data of n independent subjects, for the i^{th} subject, we have n_i number of time points $t_{ij}, j = 1,2,...n_i$ repeated measurements. The dataset can be denoted by $\{(Y_{1i}(t_{ij}), Y_{2i}(t_{ij}), X_i(t_{ij}))\}$ and $X_i(t_{ij}) = (X_{i1}(t_{ij}), X_{i2}(t_{ij}), ...X_{ip}(t_{ij}))$ is a vector of length p, where $1 \leqslant p \leqslant P$, $X_{i1}(t_{ij})$ can be continuous or discrete. Assume we have two conditional variables $Y_1(t)|Y_2(t)$ and $Y_2(t)|Y_1(t)$ are both normally distributed, the conditional mean models with estimated time-varying coefficients are:

$$Y_1(t)|Y_2(t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)Y_2(t) + \hat{\beta}_2(t)X_1(t) + ... + \hat{\beta}_p(t)X_p(t) + \varepsilon_1(t) \tag{3.1}$$

$$Y_2(t)|Y_1(t) = \hat{\gamma}_0(t) + \hat{\gamma}_1(t)Y_1(t) + \hat{\gamma}_2(t)X_1(t) + ... + \hat{\gamma}_p(t)X_p(t) + \varepsilon_2(t) \tag{3.2}$$

At each time point t, each model has the covariate vector of length $p+1$, where $1 \leqslant p \leqslant P$, $X_i(t_{ij})$ are common covariates of both models and $Y_2(t)$, $Y_1(t)$ are the additional covariates corresponding to the first Equation (3.1) and second model Equation (3.2). The estimated coefficients $\hat{\beta}(t) = (\hat{\beta}_0(t), \hat{\beta}_1(t), ...\hat{\beta}_p(t))$ and $\hat{\gamma}(t) = (\hat{\gamma}_0(t), \hat{\gamma}_1(t), ...\hat{\gamma}_p(t))$ are unknown smooth functions of t. Among those functions, $\hat{\beta}_0(t)$ and $\hat{\gamma}_0(t)$ are baseline functions, the other coefficients functions represent the covariates effects on the outcome variables over time. For example, in our application study on FHS, to investigate the joint dynamic density distribution of $log(SBP(t))$ and $log(BMI(t))$ given some fixed value of $Chol(t)$, we will set: $Y_1(t) = log(SBP(t)), Y_2(t) = log(BMI(t))$, and $X_1(t) = Chol(t)$, hence $p = 2$ and $X_1(t) = Chol(t)$ is the common covariate of both conditional models.

The $\varepsilon_1(t)$ and $\varepsilon_2(t)$ represent the mean zero stochastic process at time t, we assume that $(\varepsilon_1(t), \varepsilon_2(t))$ has a standard bivariate normal distribution and they are independent of the covariate vectors.

3.2 Nonparametric Time-varying Conditional Models

According to the conditional specification paradigm in chapter 2, we need to check if the nonparametric time-varying conditional models have the properties of both compatibility and uniqueness. Only when the two requirements are fulfilled, then the joint distribution of $Y_1(t), Y_2(t)$ at any time t are uniquely existing.

Compatibility

At any time $t \in T$, $\{(y_1, y_2) : f_1(y_1|y_2) > 0\} = \{(y_1, y_2) : f_2(y_2|y_1) > 0\} \equiv \Omega$, take the function at both sides of our conditioning models, we have:

$$\frac{f(Y_1(t)|Y_2(t))}{f(Y_1(t)|Y_2(t))} = \frac{f(\beta_0(t)+\beta_1(t)Y_2(t)+\beta_2(t)X_1(t)+...+\beta_p(t)X_p(t)+\varepsilon_1(t))}{f(\gamma_0(t)+\gamma_1(t)Y_1(t)+\gamma_2(t)X_1(t)+...+\gamma_p(t)X_p(t)+\varepsilon_2(t))} = u_1(y_1(t))u_2(y_2(t))$$

where

$$u_1(y_1) = \frac{1}{f(\gamma_0(t) + \gamma_1(t)Y_1(t) + \gamma_2(t)X_1(t) + ... + \gamma_p(t)X_p(t) + \varepsilon_2(t))}$$
$$u_2(y_2) = f(\beta_0(t) + \beta_1(t)Y_2(t) + \beta_2(t)X_1(t) + ... + \beta_p(t)X_p(t) + \varepsilon_1(t))$$

We have the standard bivariate normal distribution between $\varepsilon_1(t), \varepsilon_2(t)$ which are also independent with the covariates in the assumption, so that we can factorized the function by u_1 which is a function of y_1 and u_2 which is a function of y_2. We have both u_1 and u_2 which are finite functions, according to chapter 2, the compatibility condition is satisfied.

Uniqueness

According to Chapter 2, if the stochastic kernel, the transition kernel is indecomposable, which makes the marginal densities unique, hence the joint distribution will be unique. Since they have the common supports fields $\Omega = S(Y_1) \times S(Y_2)$. In this case, then we have the kernel greater than 0 for every x and z, which means that the Markov chain with the transition kernel is indecomposable. As discussed in previous content, the uniqueness is also satisfied.

3.3 Estimation Methods

3.3.1 Two-step Time-varying Estimation Method for Conditional Models

3.3.1.1 Time Point Binning and Data Transformation

In this step, to reduce the effects of minor observation errors and when we have small number of data observations at some time point, we use the time point binning approach. If a subject's observation time points t_j^* stay within a small neighborhood of t_j, they are grouped into one of the design time points t_j, all will be treated as t_j.

It is often easier to meet the assumptions if the conditional variables is normally distributed. However, variables in the analysis might be left-skewed, with a long left tail, or right-skewed, with a long right tail. Remind that in 3.1, we assume $Y_1(t)|Y_2(t)$ and $Y_2(t)|Y_1(t)$ are conditional normal. So, making data transformation such as log, square root, will make it fit the assumptions better [14] .

3.3.1.2 First Step: Raw Estimators of the Coefficients for Univariate Conditional Models

As introduced in chapter 1, the first step of the two-step time-varying estimation method for conditional models is to obtaining the "pseudo observations" or "raw estimators" prepared for the second step. For getting the raw estimators, we need to get the maximum likelihood estimators (MLEs), which are the same as the least-square estimators at a fixed time point $t \in T$ in a linear model.

Let's look at the first conditional model Equation (3.1). Suppose we have independent and enough observations n_j at $t_j, j = 1,2,...J$, so that the coefficients can be estimated by the least square criteria:

$$J(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ Y_{1_{ij}} - [\beta_0(t_{ij}) + Y_{2_{ij}}\beta_1(t_{ij}) + \sum_{l=1}^{p} X_{ijl}\beta_l(t_{ij})] \right\}^2$$

and the estimators $\tilde{\beta}_p(t_j)s$ represent the time trends of the covariate effects. We can write $Y_1(t) = \tilde{\beta}_0(t) + \tilde{\beta}_1(t)Y_2(t) + \tilde{\beta}_2(t)X_1(t) + ... + \tilde{\beta}_p(t)X_p(t) + e_1(t)$. Also $Y_2(t) = \tilde{\gamma}_0(t) + \tilde{\gamma}_1(t)Y_1(t) + \tilde{\gamma}_2(t)X_1(t) + ... + \tilde{\gamma}_p(t)X_p(t) + e_2(t)$. This step is computational simple.

3.3.1.3 Second Step: Dynamic Estimators of the MLE

Secondly, we will use nonparametric methods to estimate those raw estimators $\tilde{\beta}_p(t_j)s$ at any time $t_j \in T$. Here, we use local kernel smoothing estimators.

The Kernel Smoothing with Bandwith Selection Using Leave-One-Time-Point-Out Cross-Validation (LTCV) We define at any time design points $t_j \in T$, by treating the raw estimates $\tilde{\beta}_p(t_j)$ as the pseudo-observations for $\beta_p(t_j)$. The kernel estimation methods applied on $\tilde{\beta}_p(t_j)s$ are the same as on $\tilde{\gamma}_p(t_j)s$, so we will only describe how we apply this method on $\tilde{\beta}_p(t_j)s$ in details.

A kernel smoothing estimator of $\hat{\beta}_p(t_j)$ can be expressed as a weighted linear estimator of the form $\sum_{j=1}^{J} w(t_j,t)\tilde{\beta}_p(t_j), t \in T = (t_1,t_2,...t_J)$, where $w(t_j,t)$ is the smoothing weight function as we discussed in 1.4.1, which gives more weight if the time point t_j is close to t. Kernel density estimation also depends on the bandwidth h. It turns out that the choice of kernel has much less influence on the estimation result than does the bandwidth. Given the local neighborhood observations of a particular observation $\tilde{\beta}_p(t_j)s$, the bandwidth h_j, a nonnegative number, which is defined as the largest distance from $\tilde{\beta}_p(t_j)s$ to its local neighbor observations.

The cross-validation (CV) is a method to assess how accurately our estimators performs [46] . The basic idea is to split data into the training and the testing sets. The training set is used to construct estimators and the testing set is used to evaluate the performance of the prediction.

A simplest form of cross-validation is the leave-one-time-point-out cross-validation (LTCV), which means the testing set will be the observations of only one time point. When

we have J number of time points, LTCV is making splitting on data J times, each time leaving one time point out and use the training set to train the estimators and then predict the value for that we left out.

For optimizing the time-varying coefficient estimation, we need to define a evaluation metric, for example, the most commonly used one, the mean squared error (MSE). To optimize the bandwith h, we use the grid search to select the bandwidth h which returned the minimum MSE. For example, we can set $h = (h_1, h_2, ...h_l)^T$ be a sequence of scalar vector of length L and $\tilde{\beta}^{(-j)}(t_j, h_l)$ be an estimator computed using the kernel smoothing method based on all time points but with the j^{th} time point measurements excluded.

Let's define the LTCV score function of bandwidth h :

$$LTCV_l(h_l) = \sum_{j=1}^{J} [\hat{\beta}_l(t_j) - \tilde{\beta}_l{}^{(-j)}(t_j, h_l)]^2$$

The bandwidth h_{LTCV} is the minimizer of the cross-validation score $LTCV_l(h_l)$ which can be uniquely minimized. The following is the psuedo code Figure 3.1 for functions of calculating the score and optimizing the span h.

Algorithm 1 LTCV function

procedure LTCV(vector y, bandwidth h)
    input time vectors x of length J
    initialize null vector for the prediction errors
    for each integer i in $1 : J$ **do**
        $xi = x[-i]$
        $yi = y[-i]$
        $temp = kernelsmoothing(yi \sim xi, bandwidth = h, kernal = "normal")$
        $errors[i] = (y[i] - predict(temp))^2$
    end for
    Return mean squared errors
end procedure

Algorithm 2 Bandwidth selection using Grid Search

input the vector of raw coefficient estimators
initialize $h = (h_1, h_2, ...h_l)^T$
initialize a null cross validation error vector
for each integer i in length(h) **do**
    cross validation error[i] = LTCV function(raw coefficient estimators, h)
end for
$h_{LTCV} = h$ of the index of minimum cross validation error

Figure 3.1: functions of LTCV and Bandwidth selection using Grid Search

Kernel Smoothing Estimators of the Raw Estimators Based on the raw estimators of the coefficients, using the optimal bandwidth selected by the LTCV function, we can get the kernel estimator $\hat{\beta}(t)$ of $\tilde{\beta}(t)$.

3.3.1.4 Centering Procedure

At any design point t_j, the kernel smoothing estimators of coefficients at different time points will be used to proceed to the centering procedure. Since we assume $(\varepsilon_1(t), \varepsilon_2(t))$

follows a standard bivariate normal distribution, and the expectation of either $\varepsilon_1(t)$ or $\varepsilon_2(t)$ equals to 0.

Take the expectation on both sides of the equations, we have: $E(Y_1(t)) = E(\hat{\beta}_0(t) + \hat{\beta}_1(t)Y_2(t) + \hat{\beta}_2(t)X_1(t) + ... + \hat{\beta}_p(t)X_p(t))$ and $E(Y_2(t)) = E(\hat{\gamma}_0(t) + \hat{\gamma}_1(t)Y_1(t) + \hat{\gamma}_2(t)X_1(t) + ... + \hat{\gamma}_p(t)X_p(t))$. Subtracting the expectation from $Y_1(t), Y_2(t)$ will make the centering steps.

3.3.2 Estimation of variances for Joint Density of LogSBP and LogBMI

After making the estimation on univariate conditional models and the centering steps, we proceed to the estimation of parameters for joint density functions.

3.3.2.1 Raw MME for Joint Density Distributions over time

Method of moments is a popularized estimation technique. It can be routinely employed to estimate the population parameters which characteristic the distributions. By setting expression of population moments equal to that of the sample parameters, the solution will be the estimates of population parameters.

The method of moments was introduced by Pafnuty Chebyshev in the proof of the central limit theorem in 1887, so it is based solely on the law of large numbers. As we will see in the later simulation studies, using this method to estimate two-dimensional parameters, a sample size of hundreds of thousands is required when using the MCMC approach to make the sample parameters converge to the distributional parameters. As we know, the moments are the expected values of a sequence of variables, the most well-known moments are mean and variances, which are the first two orders of central moments. For convenience, the moment generating function (MGF) is an alternative way to encodes all the moments: the n^{th} moment about 0 is the n^{th} derivative of the MGF, evaluated at 0.

From 3.3.1.4, we can obtain the centered longitudinal data, and we have, $Y_1(t)|Y_2(t) = y_2(t) \sim N(0, \sigma_1^2(y_2(t)))$, $Y_2(t)|Y_1(t) = y_1(t) \sim N(0, \sigma_2^2(y_1(t)))$ and $(Y_1(t), Y_2(t))$ follows centered joint conditional normal distribution: then by the MGF, the moments can be easily

get. Since only the moments of even order are different from zero, we shall start from the second order moment, calculating the MGF of $(Y_1^2(t), Y_2^2(t))$.

At time point t, the density form of centered joint models $f_{Y_1(t),Y_2(t)}(y_1(t), y_2(t))$ with normal conditionals is:

$$k(c(t))\frac{1}{2\pi\sigma_1(t)\sigma_2(t)}exp\{-\frac{1}{2}[(\frac{y_1(t)}{\sigma_1(t)})^2+(\frac{y_2(t)}{\sigma_2(t)})^2+c(t)(\frac{y_1(t)}{\sigma_1(t)})^2(\frac{y_2(t)}{\sigma_2(t)})^2]\}, \qquad (3.3)$$

where,

$$\sigma_1^2(y_2(t)) = \frac{\sigma_1^2(t)}{1+c(t)(\frac{y_2(t)}{\sigma_2(t)})^2}$$
$$\sigma_2^2(y_1(t)) = \frac{\sigma_2^2(t)}{1+c(t)(\frac{y_1(t)}{\sigma_1(t)})^2}$$

we can calculating the MGF function using the equation:

$$M_{Y_1^2(t),Y_2^2(t)}(s,w) = E(e^{sY_1^2(t)+wY_2^2(t)})$$

Note that the normalizing constant $k(c) = \delta$ is as the fourth parameter which depends on c. The joint moments of time point t is calculated by the following steps:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{k(c(t))}{(2\pi)\sigma_1(t)\sigma_2(t)}exp\{-\frac{1}{2}[(\frac{y_2(t)}{\sigma_2(t)})^2+(\frac{y_1(t)}{\sigma_1(t)})^2+c(t)(\frac{y_2(t)}{\sigma_2(t)})^2(\frac{y_1(t)}{\sigma_1(t)})^2-2sy_1^2(t)-2wy_2^2(t)]\}dy_1(t)dy_2(t)$$

$$= k(c(t))\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)\sigma_1(t)\sigma_2(t)}exp\{-\frac{1}{2}[(\frac{1}{\sigma_1^2(t)}-2s)y_1^2(t)+(\frac{1}{\sigma_2^2(t)}-2w)y_2^2(t)+c(t)(\frac{y_1(t)}{\sigma_1(t)})^2(\frac{y_1(t)}{\sigma_1(t)})^2]\}dy_1(t)dy_2(t)$$

Where $s < \frac{1}{2\sigma_1^2(t)}, w < \frac{1}{2\sigma_2^2(t)}$ in order to guarantee the convergence of the integrand.

$$
\begin{aligned}
&(\frac{1}{\sigma_1^2(t)} - 2s)y_1^2(t) = m^2 \\
&(\frac{1}{\sigma_2^2(t)} - 2w)y_2^2(t) = k^2, \\
&\frac{y_1^2(t)}{\sigma_1^2(t)} = \frac{m^2}{1-2s\sigma_1^2(t)}, \\
&\frac{y_2^2(t)}{\sigma_2^2(t)} = \frac{k^2}{1-2w\sigma_2^2(t)} \\
&dy_2(t) = \frac{\sigma_2^2(t)}{(1-2w\sigma_2^2(t))^{\frac{1}{2}}}dk \\
&dy_1(t) = \frac{\sigma_1^2(t)}{(1-2s\sigma_1^2(t))^{\frac{1}{2}}}dm
\end{aligned}
$$

then, the above moment functions will be

$$
\begin{aligned}
&k(c(t))\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{(2\pi)\sigma_1(t)\sigma_2(t)}exp\{-\frac{1}{2}[m^2+k^2+c(t)\frac{k^2}{1-2w\sigma_2^2(t)}\frac{m^2}{1-2s\sigma_1^2(t)}]\}dy_1(t)dy_2(t) \\
&= k(c(t))\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi(1-2w\sigma_2^2(t))^{\frac{1}{2}}(1-2s\sigma_1^2(t))^{\frac{1}{2}}}exp\{-\frac{1}{2}[m^2+k^2+c(t)\frac{k^2}{1-2w\sigma_2^2(t)}\frac{m^2}{1-2s\sigma_1^2(t)}]\}dkdm \\
&= \frac{k(c(t))(1-2w\sigma_2^2(t))^{-\frac{1}{2}}(1-2s\sigma_1^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2w\sigma_2^2(t))^{-1}(1-2s\sigma_1^2(t))^{-1}]}
\end{aligned}
$$

Let $\delta(c_t) = \frac{dlogk(c_t)}{dc_t} = \frac{k^{'}(c_t)}{k(c_t)}$ and combine with:

$M_{Y_1^2(t),Y_2^2(t)}(s,w) = \frac{k(c(t))(1-2w\sigma_2^2(t))^{-\frac{1}{2}}(1-2s\sigma_1^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2w\sigma_2^2(t))^{-1}(1-2s\sigma_1^2(t))^{-1}]}$,

Let $w = 0$, $E(e^{sy_1^2(t)}) = \frac{k(c(t))(1-2s\sigma_1^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2s\sigma_1^2(t))^{-1}]}$. Take the derivative with respect to s when $s = 0$,

we can get

$$
\begin{aligned}
&E(y_1^2(t)) = \frac{dE(e^{sy_1^2(t)})}{ds}|_{s=0} \\
&= \frac{\sigma_1^2(t)(1-2s\sigma_1^2(t))^{-\frac{3}{2}}k(c(t))k[c(t)(1-2s\sigma_1^2(t))^{-1}]-2c(t)\sigma_1^2(t)(1-2s\sigma_1^2(t))^{-\frac{5}{2}}k(c(t))k'[c(t)(1-2s\sigma_1^2(t))^{-1}]}{(k[c(t)(1-2s\sigma_1^2(t))^{-1}])^2}|_{s=0} \\
&= \frac{\sigma_1^2(t)(k(c(t)))^2 - 2c(t)\sigma_1^2(t)(k(c)(t)^2\delta(c(t))}{(k(c(t)))^2} \\
&= \sigma_1^2(t)[1-2c(t)\delta(c(t))]
\end{aligned}
$$

Using the same method, we can easily get $E(y_2^2(t)) = \sigma_2^2(t)[1-2c(t)\delta(c(t))]$. More calculation is showed in the appendix, and the results are as following:

$$
\begin{aligned}
&\overline{Y_1^2(t)} = \frac{1}{n}\sum_{i=1}^{n} Y_{1i}^2(t), \\
&\overline{Y_2^2(t)} = \frac{1}{n}\sum_{i=1}^{n} Y_{2i}^2(t), \\
&S^2_{Y_1^2(t)} = \frac{1}{n}\sum_{i=1}^{n} (Y_{1i}^2(t) - \overline{Y_1^2(t)})^2, \\
&S^2_{Y_2^2(t)} = \frac{1}{n}\sum_{i=1}^{n} (Y_{2i}^2(t) - \overline{Y_2^2(t)})^2 \\
&S^2_{Y_1^2(t)Y_2^2(t)} = \frac{1}{n}\sum_{i=1}^{n} (Y_{1i}^2(t) - \overline{Y_1^2(t)})(Y_{2i(t)}^2(t) - \overline{Y_2^2(t)})
\end{aligned}
$$

then by the strong law of large numbers, we may conclude that as $n \to \infty$:

$$
\begin{aligned}
&\overline{Y_1^2(t)} \xrightarrow{\text{a.s.}} \sigma_1^2(t)(1-2c(t)\delta(t), \\
&\overline{Y_2^2(t)} \xrightarrow{\text{a.s.}} \sigma_2^2(t)(1-2c(t)\delta(t)) \\
&T = \frac{S_{Y_1^2(t)}S_{Y_2^2(t)}}{\overline{Y_1^2(t)Y_2^2(t)}} \xrightarrow{\text{a.s.}} \frac{1+2\delta(t)-4c^2(t)\delta^2(t)}{(1-2c(t)\delta(t))^2} \\
&R = \frac{S_{Y_1^2(t)Y_2^2(t)}}{S_{Y_1^2(t)}S_{Y_2^2(t)}} \xrightarrow{\text{a.s.}} \frac{1-2\delta(t)-4c(t)\delta(t)+4c^2(t)\delta^2(t)}{-1-2\delta(t)+4c^2(t)\delta^2(t)}
\end{aligned}
$$

By equating these statistics to their almost surely limits and solving, the following

strongly consistent estimators are obtained:

$$
\begin{aligned}
\hat{c(t)} &= \frac{1}{4}[1+RT-2T-\frac{1-T^2}{1+RT}], \\
\hat{\delta}(t) &= \frac{2(1+RT)}{(1-R)^2T^2}, \\
\hat{\sigma_1^2}(t) &= \frac{\overline{Y_1^2(t)}}{1-2\hat{c}(t)\hat{\sigma}(t)}, \\
\hat{\sigma_2^2}(t) &= \frac{\overline{Y_2^2(t)}}{1-2\hat{c}(t)\hat{\sigma}(t)}
\end{aligned}
$$

and, $\hat{k} = exp(\hat{c}(t)\hat{\sigma}(t))$ will be the estimator of the normalizing constant. The details will be discussed in the later chapters and the appendix.

3.3.2.2 Dynamic MME Estimators for Joint Distributions

Secondly, we will use nonparametric methods to estimate those raw MME estimators $\hat{\sigma_1^2}(t)$ and $\hat{\sigma_2^2}(t)$ at any time $t_j \in T$. The gain from the second step estimation method is that it reduces the variability of the curve estimates at any $t_j \in T$ by borrowing the information obtained from the raw estimates at the design time points which are adjunct to t_j. Here, we use local kernel density estimators.

3.3.3 Bootstrap Confidence Intervals

Since different asymptotic distributions of the smoothing estimators may be obtained depending on the longitudinal designs and whether and how fast the number of observations at each time point converge to infinity, statistical inferences based on the asymptotic approximations may not be an appropriate option in practice.

In out study, the observations of each subject are assumed to be independent, a natural bootstrap sampling scheme, introduced by [15] is a general intuitive method just about sampling statistic that keeps the application away from some technical tedium. Bootstrap is using resampling technique to draw repeated measurements of each subjects with replace-

ment from the original data set, then it creates a large number of "phantom samples" known as bootstrap samples. The estimations is then conducted on each of the bootstrap samples [42].

The subject resampling bootstrap procedure for an approximate $(1-\alpha)$ pointwise percentile interval has the following steps.

Step 1: We randomly select k number of bootstrap subjects $\{(Y_{ij}^*, t_{ij}^* : i = 1, ..., k; j = 1, ...k_i\}$ with replacement, then randomly select the longitudinal bootstrap samples total B times, B is usually several hundreds to several thousands.

Step 2: After resampling, based on the each one bootstrap sample, repeat all the procedures to estimate the dynamic MME estimators of variances for joint distributions as discussed in 3.3.1 to 3.3.3.

Step 3: As in 3.3.3, after the LTCV step, we can get B numbers of sets of estimators at different time points for all B numbers of bootstrap samples. The estimators of the b^{th} sample are $\tilde{\sigma_1^2}^{i,b}(t, LSCV(h)), b = 1, ...B$.

Step 4: At each time point t, we have B numbers of estimators, suppose we set $\alpha = 0.05$ then the $[100 \times (1-0.05)]\% = 95\%$ bootstrap percentile pointwise confidence intervals for $\tilde{\sigma_1^2}(t)$ can be calculated at each time point:

$$(L_{0.025}^b(t, LSCV(h_{opt})), U_{0.975}^b(t, LSCV(h_{opt})))$$

The following is the psuedo code in Figure 3.2 for functions of generating the dynamic MME estimators and the Bootstrap pointwise confidence intervals.

Algorithm 3 Generating Dynamic MME

Function Smoothed estimators (data)

initial coefficient and time vector

for each t **do**

    Two-step estimation of the coefficients for Univariate time-varying conditional mean models

    Centering procedure, get the centered conditional normal data

    Generate MME raw estimators

end for

Call the bandwidth selection function on raw MME estimators

Return the predicted smoothed MME estimator functions of time $\tilde{\sigma_1^2}(t)$ and $\tilde{\sigma_2^2}(t)$

Algorithm 4 Bootstrap Inference

Function Bootstrap (longitudinal data)

IDlist = all unique subjects

n= length(IDlist)

Initialize the number of bootstrap sample B

for each integer i in $1 : B$ **do**

    Draw a bootstrap sample of the same IDlist with size M with replacement from longitudinal data

end for

Call Function Smoothed estimators on each bootstrap data sample

Return $(L^b_{0.025}(t, LSCV(h_{opt})), U^b_{0.975}(t, LSCV(h_{opt})))$ — the 2.5% and 97.5% Bootstrap percentile pointwise confidence intervals at time point t.

Figure 3.2: Generating the Dynamic MME and the Bootstrap Pointwise Confidence Intervals

Chapter 4: Application

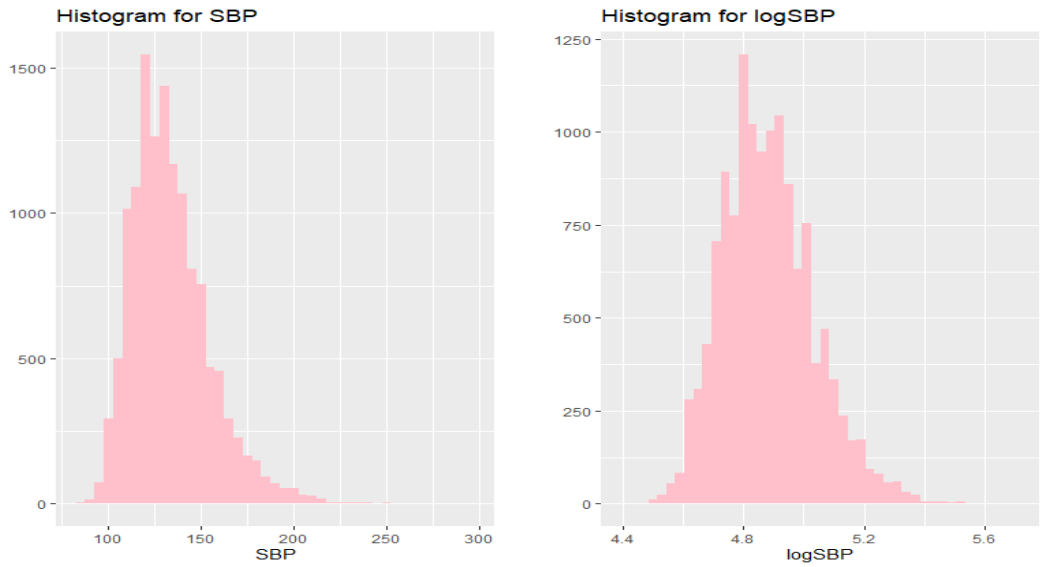
4.1 Application in the Framingham Heart Study (FHS)

The Framingham heart study data set is provided for the first 40 years of follow up in the Original Cohort, it contains 1213 subjects from a systematic sample of the households in the town of Framingham, Massachusetts. The repeated measurements on each subject were performed every two years, each subject has 1 to 20 times repeated measurements. It was estimated that about total 25473 observations were in the age range $29-101$ and thus were eligible for study, the ages were all binned at integer numbers.

4.1.1 Data Description

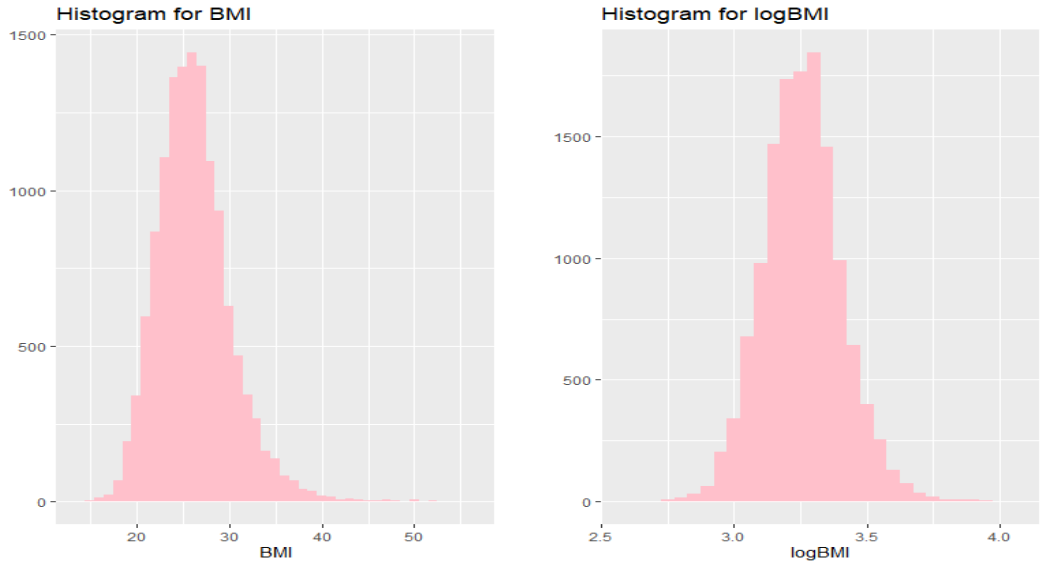
Although the data has already bean binned, the availability of the observations from different age (also the time t variable) groups, < 33 or > 80 years old, are still very small. We only select those age groups which have more than 80 observations, the selected age range are from $33-80$. We use subject ID, systolic blood pressure (SBP), Body Mass Index (BMI), total cholesterol (Chol) and Age variables, which is the time point variable, in our analysis. In this longitudinal data, each row represents one observation at one time point on the three variables SBP, BMI and Chol. Any row contains unobserved values will be dropped. Our final analysis data set has a dimension of 13182×5, containing 1209 subjects.

We assume that SBP given BMI and BMI given SBP are conditional normally distributed. Both the histograms of SBP|BMI (Figure 4.1 (a)) and BMI|SBP (Figure 4.2 (a)) are right-skewed with long right tails. The histograms of log transformation of SBP (logSBP) and BMI (logBMI) are also showing in (Figure 4.1 (b) and (Figure 4.2 (b))).



(a) (b)

Figure 4.1: FHS: Histogram plots of SBP (left), logSBP (right)



(a) (b)

Figure 4.2: FHS: Histogram plots of BMI (left), logBMI (right)

We observed that after the log transformation, both logSBP and logBMI are approximately normally distributed. Among all 1209 subjects from age at 33 to 80, the range of SBP is (83, 295), hence the corresponding log transformation logSBP is in (4.42, 5.68).

The range of BMI is (13.64, 56.13) and logBMI is in (2.61, 4.03). And the range of Chol is (87, 483). By looking at the conditional plots of logSBP|logBMI and logBMI|logSBP, although due to the limited observations of one variable conditioning on a fixed value of the other variable at a given age, we will assume the populations of both logSBP|logBMI and logBMI|logSBP are following normal distributions.

In previous chapters, we have discussed the two-step time-varying coefficients estimation method for conditional models. We will skip the time point binning step since the ages (t variable) in the FHS data has already been binned at integer numbers. In the FHS, we will assume the the joint density $f(logSBP(t), logBMI(t))$ is:

$$\frac{k(c(t))}{2\pi\sigma_1(t)\sigma_2(t)} exp\{-\frac{1}{2}[(\frac{logSBP(t)}{\sigma_1(t)})^2+(\frac{logBMI(t)}{\sigma_2(t)})^2+c(t)(\frac{logSBP(t)}{\sigma_1(t)})^2(\frac{logBMI(t)}{\sigma_2(t)})^2]\}$$

In the above centered normal conditional distribution, c(t) and variances over time points are unknown parameters. The k(c(t)) is a function c(t), which can be calculated by the c(t) values.

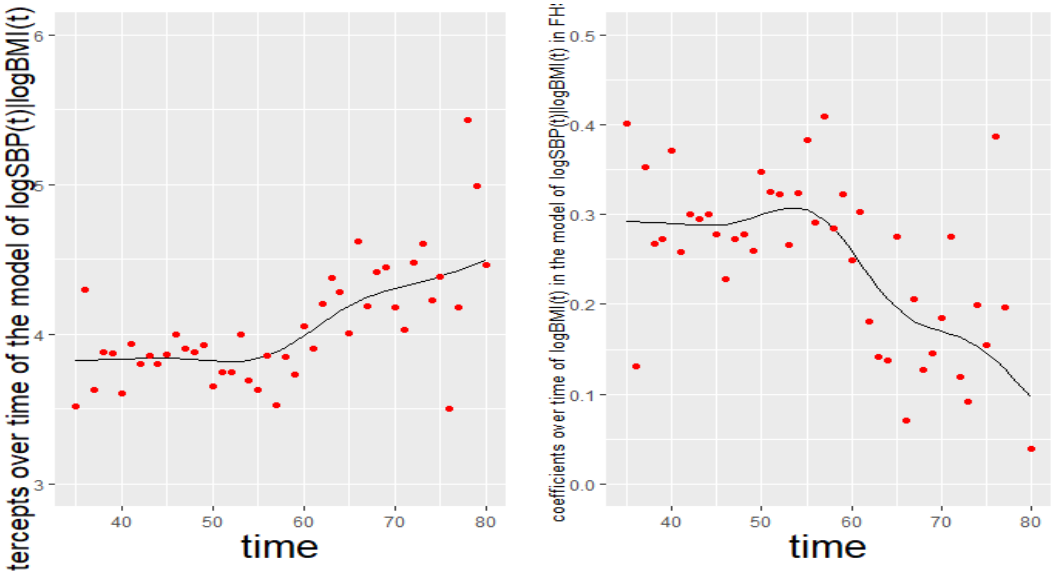
4.1.2 Two-step Time-varying estimation for univariate models of LogSBP and LogBMI

The two-step estimation is following the steps in 3.3.1.2 and 3.3.1.3. Our univariate conditional models of $logSBP|logBMI$ and $logBMI|logSBP$ are:

$$logSBP(t) = \alpha_1(t) + \beta_0(t) logBMI(t) + \beta_1(t) Chol(t) \tag{4.1}$$

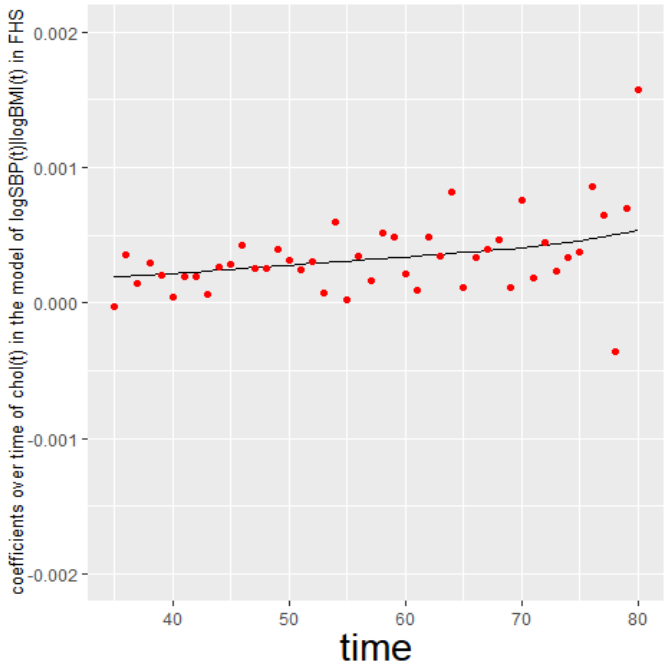
$$logBMI(t) = \alpha_2(t) + \gamma_0(t) logSBP(t) + \gamma_1(t) Chol(t) \tag{4.2}$$

The MLEs of the coefficients: $\tilde{\alpha}_1(t)$, $\tilde{\alpha}_2(t)$ represent the baseline values of logSBP and logBMI, $\tilde{\beta}_0(t), \tilde{\beta}_1(t), \tilde{\gamma}_0(t)$, and $\tilde{\gamma}_1(t)$ represent the time trends of the covariate effects. The kernel smoothing of raw MLE coefficients are: $\hat{\alpha}_1(t), \hat{\alpha}_2(t)$, and $\hat{\beta}_0(t), \hat{\beta}_1(t), \hat{\gamma}_0(t), \hat{\gamma}_1(t)$.



(a) (b)

Figure 4.3: FHS: The coefficients over time for first model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)Chol(t)$. Left: black: $\hat{\alpha}_1(t)$, red: $\tilde{\alpha}_1(t)$. Right: black: $\hat{\beta}_0(t)$, red: $\tilde{\beta}_0(t)$.



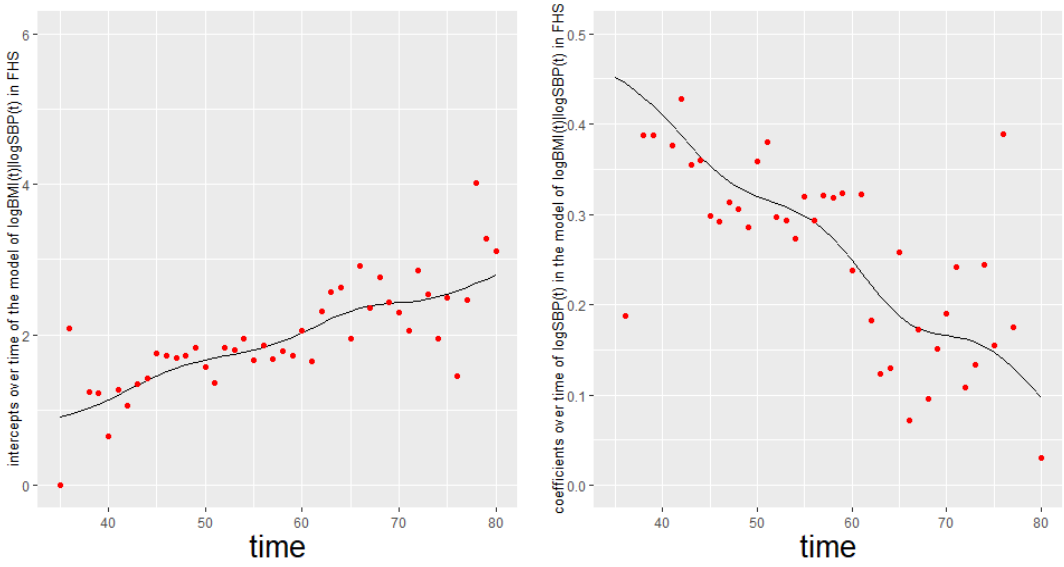
(a)

Figure 4.4: FHS: The coefficients over time for first model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)Chol(t)$. Black: $\hat{\beta}_1(t)$, Red: $\tilde{\beta}_1(t)$.

The scatter points in Figure 4.3 (a), are the MLE of coefficients in model Equation (4.1), represent the baseline values of logSBP at different time points. The black curves, are kernel estimations of the time-varying coefficients. Before age 55, $\hat{\alpha}_1(t)$ are nearly horizontally distributed around 3.8. This means that the baseline logSBP are around 3.8 on average before age 55. Figure 4.3 (b) shows the estimated coefficients $\hat{\beta}_0(t)$ of logBMI are all positive, but decrease after age 55, meaning that when age is increasing and Chol is fixed, the positive effects on logSBP caused by logBMI is decreasing. At age 40, $\hat{\beta}_0(t)$ is nearly 0.3, means that when Chol is holding constantly, if logBMI increased by one unit, the logSBP will increase by 0.3 unit. But at age 70, $\hat{\beta}_0(t)$ is near 0.17, demonstrating that if logBMI increases by one unit (and Chol does not differ), logSBP will only increase by 0.17 units, on average.

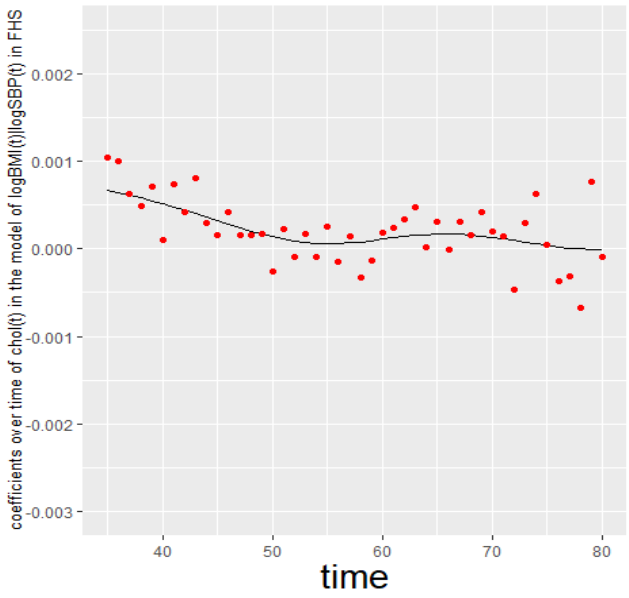
In Figure 4.4, all $\hat{\beta}_1(t)$ are very close to 0, indicating that Chol has very small positive

effect on logSBP and Chol can be regarded as independent of logSBP over time. The following plots are presenting the time-varying coefficients of model Equation (4.2).



(a) (b)

Figure 4.5: FHS: The coefficients over time for first model:$logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)Chol(t)$. Left: black: $\hat{\alpha}_2(t)$, red: $\tilde{\alpha}_2(t)$. Right: black: $\hat{\gamma}_0(t)$, red: $\tilde{\gamma}_0(t)$.



(a)

Figure 4.6: FHS: The coefficients over time for first model:$logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)Chol(t)$. Black: $\hat{\gamma}_1(t)$, Red: $\tilde{\gamma}_1(t)$.

Above figures, Figure 4.5 and Figure 4.6 are based on the model Equation (4.2). The subplot Figure 4.5 (a) is showing an increasing baseline trend of logBMI(t) over time. In Figure 4.5 (b), all coefficients $\hat{\gamma}_0(t)$ of logSBP are all positive but decreasing over time, meaning that elder people with one unit increasing of logSBP have less positive effect on their logBMI compare with younger people when their Cholesterol level are the same. In Figure 4.6, all coefficients $\hat{\gamma}_1(t)$ are very close to 0, indicating the Chol variable has very rare effect on logBMI at different ages when their logSBP level is holding as a constant.

4.1.3 Dynamic MME for Joint Density of LogSBP and LogBMI

At the first step, we generate raw MME of variances as the input of the second smoothing step. And at the second step, we apply the local kernel smoothing to get the dynamic MME of variances.

4.1.3.1 Raw MME of variances for Joint Density of LogSBP and LogBMI

For making the computing easier, we apply the centering step on logSBP and logBMI first.

Centering Procedures:

At any design point t_j, all variables are following the normal distributions in the conditional linear models, and the error terms have zero expectations. Take the expectation on both sides of the equations Equation (4.1) and Equation (4.2), we have: $E(logSBP(t)) = E(\hat{\alpha}_1(t) + \hat{\beta}_0(t)logBMI(t) + \hat{\beta}_1(t)Chol(t))$ and $E(logBMI(t)) = E(\hat{\alpha}_2(t) + \hat{\gamma}_0(t)logSBP(t) + \hat{\gamma}_1(t)BMI(t))$. The centered variables can be calculated by subtracting the expectation from $logSBP(t)$ and $logBMI(t)$.

Raw MME of variances:

According to 3.3.1.4, the consistent estimators for the joint density can be easily obtained by calculating $\overline{Y_1^2(t)}, \overline{Y_2^2(t)}, S^2_{Y_1^2(t)}, S^2_{Y_2^2(t)}, S^2_{Y_1^2(t)Y_2^2(t)}$. Remember that when the parameter c(t) equals to 0, the two variables $Y_1(t), Y_2(t)$ follow a bivariate normal distribution.

Contour Plots of Bivariate Normal Distributions:

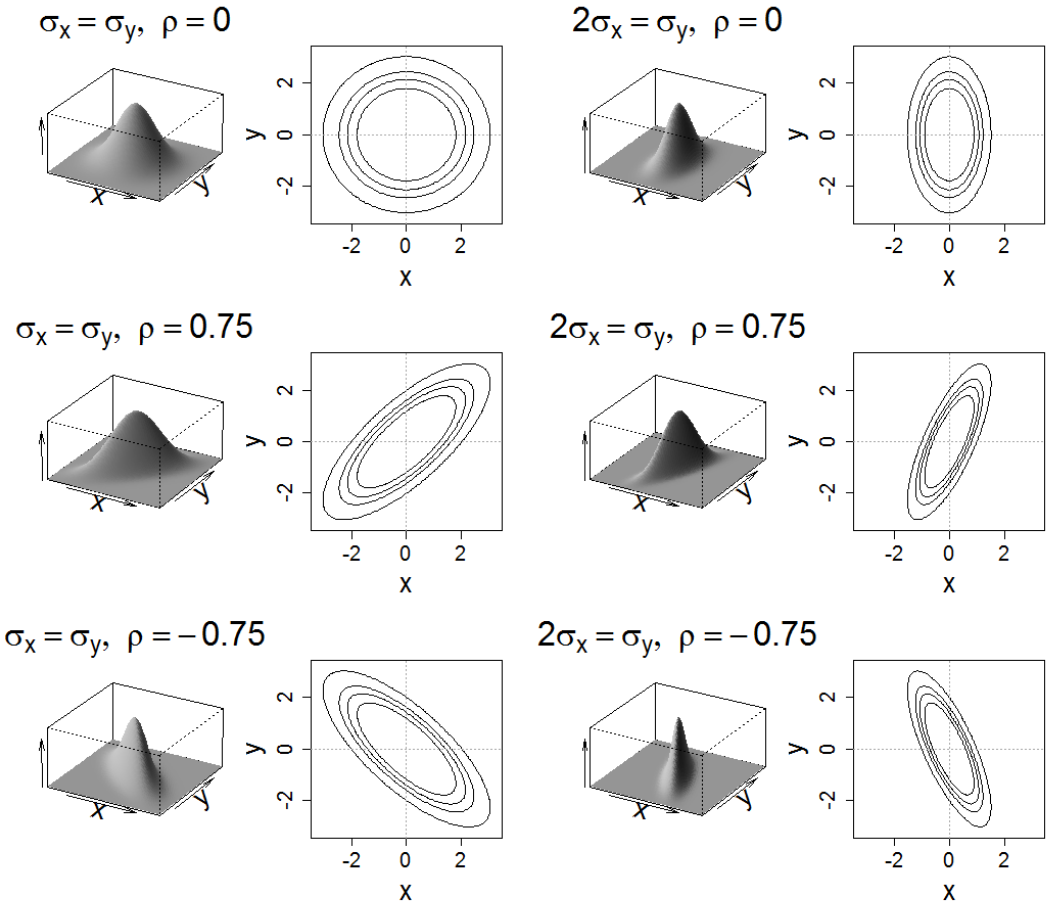


Figure 4.7: The 3D density distributions and the contour plots of bivariate normal distributions given different parameters

The contour plots of bivariate normal distributions at different parameter settings are listed in Section 4.1.4 . Note that all density contour plots for bivariate normal distributions should be ellipsoid or circular. In Section 4.1.4, we assume two variables X and Y are following a bivariate normal distributions. If X and Y are independent with each other and they have equal variances, their contours should be circles, as shown in the subplot (the upper left panel). When X and Y have correlations and different variances, their contours will appear to be elliptical but not circular (other panels). Comparing with the left panels, all the right panels indicate that the different ratios between variances of X and Y also change

the shape of the contours. For example, in the right bottom pane, when $2\sigma_x = \sigma_y$, the major axis is of twice length of the minor axis in the ellipse.

4.1.3.2 Dynamic MME of variances and Bootstrap Pointwise Confidence Intervals

The dynamic joint density function is given in 4.1.1. To get the full joint density functions, we need to generate K(c(t)) based on the estimated c(t) values.

Generation of K(c(t))

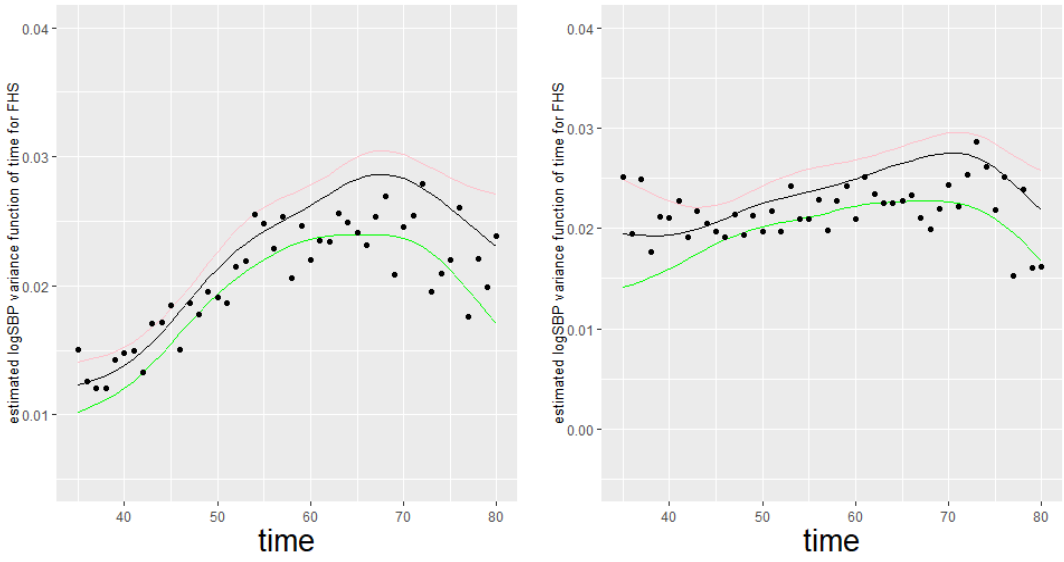
The values of K(c(t)) can be expressed as [3]:

$$k^{-1}(c(t)) = E\left[\frac{1}{\sqrt{(1+c(t)Z^2)}}\right] \tag{4.3}$$

where Z follows a standard normal distribution. The k(c(t)) can be generated by a function of c(t). Given a sequence of c values, for example, an equally spaced sequence from 0 to 1 with length 1000, the k(c) values can be calculated by Equation (4.3) after generating the standard normally distributed samples Z. The k value which is most close to the calculated k(c) given the estimated c(t) will be the estimated k(c(t)).

Dynamic MME of variances for joint distributions of LogSBP and LogBMI

The estimated dynamic variances (black lines) for logSBP and logBMI have similar trend with the original variances (black dots) over time directly calculated from data, showing in Figure 10. Remember that the marginal variances are always slightly larger than the conditional variances except when the two variables are independent with each other. In figure 10 (a), before age 65, the variances of LogSBP are increasing over time, but decreases after age 65. The variances of LogBMI also increase slightly before age 75, then decrease after age 75. The pink and green curves are 95% bootstrap pointwise confidence intervals for variances of logSBP and logBMI over time.



(a) (b)

Figure 4.8: FHS: Estimators of the variances of logSBP (left), logBMI (right) at different time point. Blue lines are estimators of the variances of time. Pink and Green curves are 95% bootstrap pointwise confidence intervals. Black dots are original variances of logSBP and logBMI over time directly calculated from data.

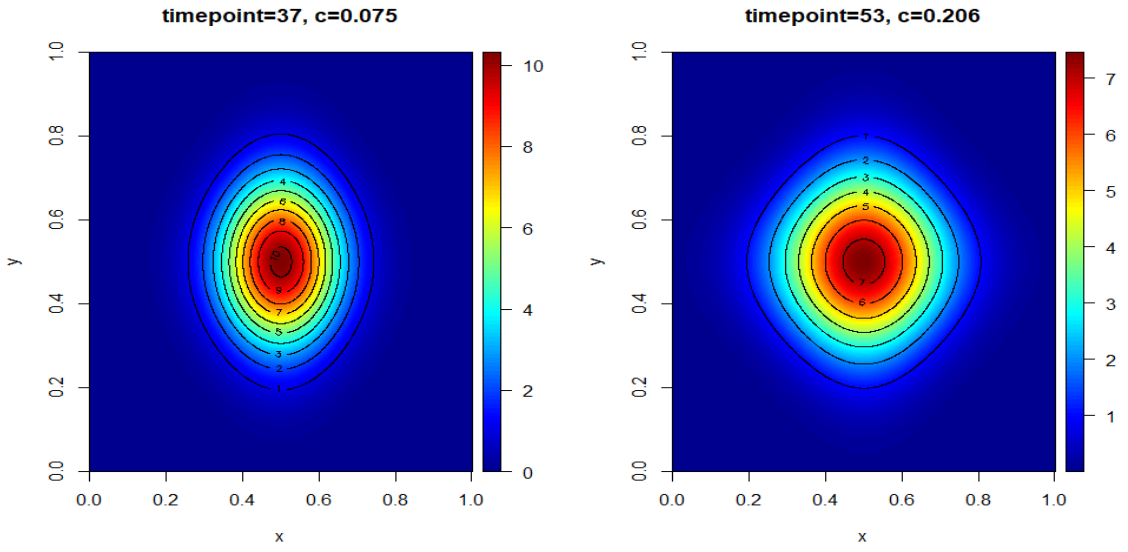
The dynamic MME estimators for variances of logSBP and logBMI at 4 age groups, age 37, 53, 67 and 73, conditioning on the mean value of Chol are listed in Table 4.1.

Age	$\hat{c}(t)$	$\hat{\sigma}^2_{logSBP}(t)$	$\hat{\sigma}^2_{logBMI}(t)$	k(c(t))
37	0.07543551	0.01270199	0.01993034	1.031991
53	0.20662318	0.02334528	0.02257274	1.077314
67	0.68248030	0.03011370	0.02750718	1.194570
73	0.70990396	0.02794060	0.02952759	1.200002

Table 4.1: FHS: Estimated c(t) and variances of logSBP(t) and logBMI(t) at age 37, 53, 67 and 73

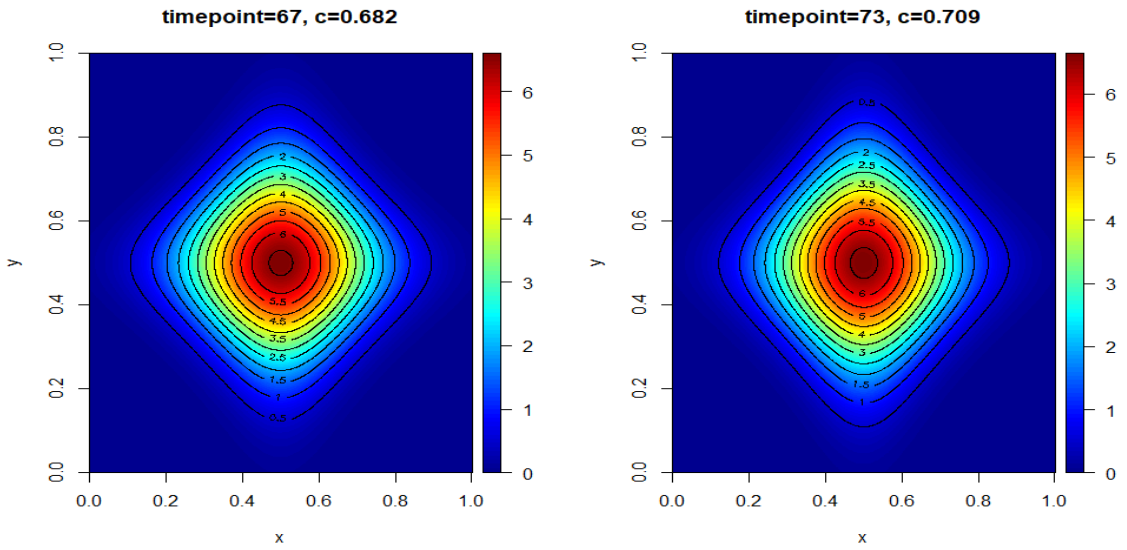
At age 37, the estimated c(t)=0.075, close to 0, indicate that, logSBP and logBMI are linearly independent with each other conditioning on the mean values of Chol, and their

joint distribution is bivariate normal, this is also can be shown in Figure 4.8 (a): at early age 37, the estimated variance $\hat{\sigma}^2_{logSBP}(t)$, $\hat{\sigma}^2_{logBMI}(t)$ line in the middle of the black dots, the variance directly calculated from data. However, the estimated c(t) at other age groups are not close to 0, telling that logSBP and logBMI are not bivariate normal. When age is increasing, the $\hat{c}(t)$ are getting larger, in the Figure 4.8 (a), there is a increasing departure between $\hat{\sigma}^2_{logSBP}(t)$, $\hat{\sigma}^2_{logBMI}(t)$ and the variance directly calculated from data (black dots). This finding is consistent with the information illustrated in the next 4 contour plots in Figure 4.9.



(a) (b)

Figure 4.9: FHS: Contour plots at t=37 and 53. The corresponding estimated c(t) are 0.075 and 0.206.



(a) (b)

Figure 4.10: FHS: Contour plots at t= 67 and 73. The corresponding estimated c(t) are 0.682 and 0.709.

The above contour plots at time points: 37, 53, 67 and 73 are also shown in Figure 4.9 and Figure 4.10. The corresponding estimated c values are 0.075, 0.206, 0.682 and 0.709. From the contour plots, at age of 37 and 53, the c(t)= 0.075 and 0.206, almost equal to zero, the contours are elliptical. At ages of 67 and 73, the contours are of diamond shapes, not elliptical. This indicates that at age 67 and 73, logSBP and logBMI conditioning on Chol are not following joint normal distributions but conditional normal distributions. Before age 67, the contour plots also show an increasing radius (the distance from the centor of the contours to the outermost contour lines). But the radius is smaller at age 73 compare to that of age 67. This presents that the estimated variance is increasing before age 67 but decreases after age 67.

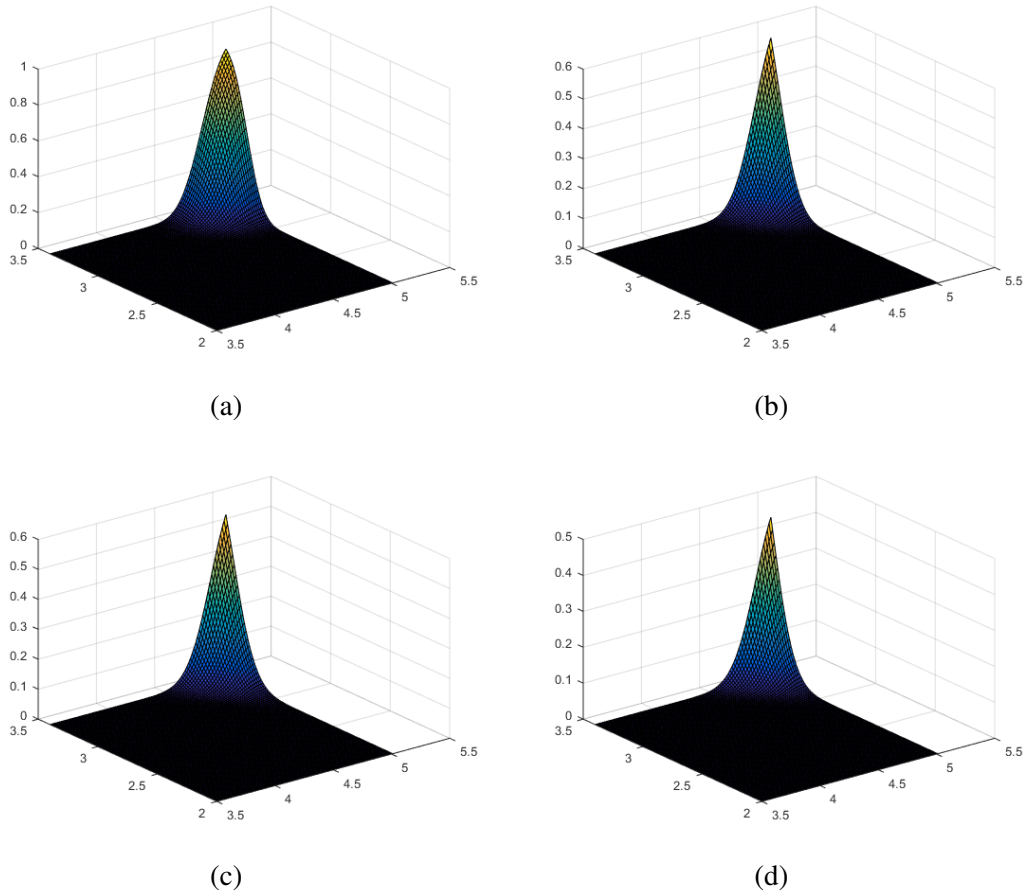


Figure 4.11: FHS: Estimated CDF plot of logSBP $< log(150)(SBP < 250)$ and logBMI $< log(30)(BMI < 30)$ based on Mean(Chol) at age 37, 53, 67 and 73

Figure 4.11 is generated by MATLAB, shows the estimated CDF of $SBP < 150$ and $BMI < 30$, given the mean value of Chol at age 37, 53, 67 and 73. When age is increasing, the cumulative density of $SBP < 150$ and $BMI < 30$ is decreasing, the exact CDF values are listed in Table 4.2.

age	CDF
37	0.8785
53	0.5640
67	0.5423
73	0.4445

Table 4.2: FHS: Estimated of CDF when $SBP < 150$ and $BMI < 30$ based on Mean(Chol) at age 37, 53, 67 and 73 in FHS.

4.1.4 Bootstrap Pointwise Confidence Intervals

The boostrap pointwise confidence intervals of the MME parameters for joint model are calculated by the algorithms and steps in 3.3.3. The 2.5% and 97.5% Boostrap pointwise confidence intervals of MME at age 37, 53, 67 and 73 are shown in (the green and pink curves are the lower and upper bound of 95% bootstrap CIs for logSBP and logBMI) and listed in Table 4.3.

Age	$\hat{\sigma}^2_{logSBP}(t)$	$L_{0.025}$ bound of CI for $\hat{\sigma}^2_{logSBP}(t)$	$U_{0.975}$ bound of CI for $\hat{\sigma}^2_{logSBP}(t)$	$\hat{\sigma}^2_{logBMI}(t)$	$L_{0.025}$ bound of CI for $\hat{\sigma}^2_{logBMI}(t)$	$U_{0.975}$ bound of CI for $\hat{\sigma}^2_{logBMI}(t)$
37	0.0127	0.0123	0.0129	0.0199	0.0193	0.0204
53	0.0233	0.0232	0.0238	0.0225	0.0220	0.0239
67	0.0301	0.0290	0.0309	0.0275	0.0266	0.0279
73	0.0279	0.0274	0.0288	0.0295	0.0288	0.0301

Table 4.3: FHS: 95% Boostrap pointwise CI of $\hat{\sigma}^2_{logSBP}(t)$ and $\hat{\sigma}^2_{logSBP}(t)$ at age 37, 53, 67 and 73 based on Mean(Chol).

4.1.4.1 Discussion

In this section, to study the medical insights of the two dimensional outcome (SBP, BMI) in the effect of Cholesterol at different ages, based on previous method of the nonparametric two-step time-varying modeling, we extend those work by incorporating the dynamic conditional specified bivariate density model. Different with other direct two-dimensional estimation approach, our approach is taken aback by the most general mathematical consequences: estimation based on the most general nonstandard joint densities with normal conditionals.

Our approach provides us very useful insights in the Framingham Heart Study data. Given values of Cholesterol and age, from the two-step MME estimated joint time-varying models of logSBP and logBMI, we find that although both logSBP conditioning on logBMI and logBMI conditioning on logSBP are normally distributed, the joint distributions are not bivariate normal. For example, the contours of joint density of logSBP and logBMI at age 67 and 73 are not ellipse, demonstrating the violation of bivariate normal distributions.

Usually in medical studies, when our interest is on a bivariate outcome, if the distribution is outside of the scope of a joint normal setting, the density estimation are becoming much more complicated. But by our approach, when conditioning on some specific values of Cholesterol and age, we can easily generate the exact joint density function of logSBP and logBMI. Therefore, given Cholesterol values, we can set any cutoff points of SBP and BMI (e.g. some abnormal values) and get their exact CDFs, then compare the risks of high SBP and BMI between different age groups. Also, given a high probability, say 0.8, with the input of our own Cholesterol and ages, we can also check what critical ranges of SBP and BMI are following in the high risk group in population.

4.2 Application in the National Growth and Health Study (NGHS)

4.2.1 NGHS Data

As introduced in previous chapters, the National Growth and Health Study is also a typical longitudinal observational study annually followed up to 10 to 11 years, containing a total of 2378 girls, in which 1213 were African-American and the left 1165 were Caucasian girls. The variables selected in our analysis are: ID, AGE, BMI, SBP and DBP. The rows contain any unobserved values of those selected variables are excluded.

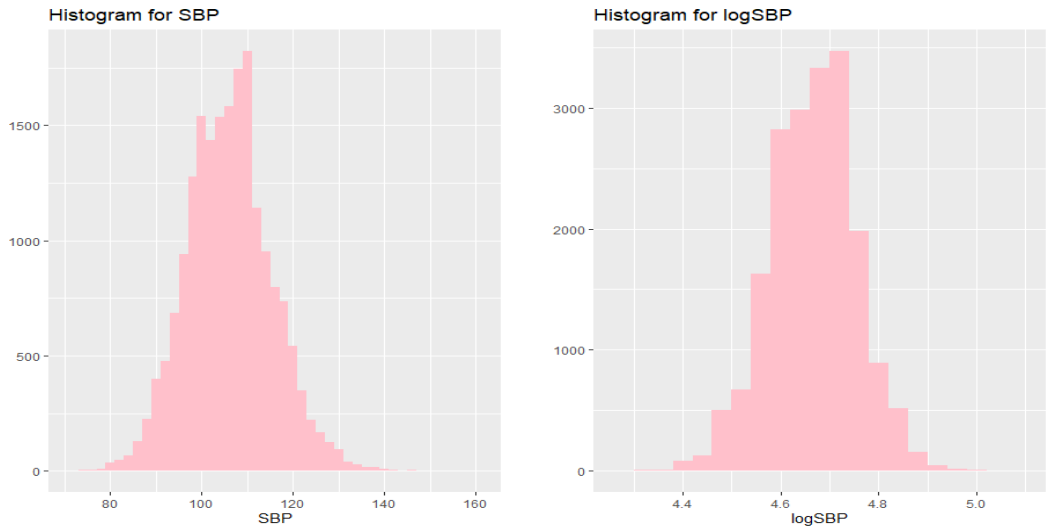
4.2.2 Time Point Binning for NGHS

The decimal number of age variable in NGHS data is 6, for example, one observation is at age of 11.715264. At each exact age, the observed visit times of some subjects might only have 1 to 2. In this study, time point binning is used. To reduce estimation error induced by the minor number of observations, we round the adjacent ages down to the nearest 1 decimal place to increase the number of observations at each time point bin. Those subjects younger than 9.1 will be excluded due to the small number of observations. The resulting age variable, from 9.1 to 19, consists of 100 small time bins, $t = 1, 2, ...100$. The final analysis data has a total of 2375 subjects, including 1164 Caucasian and 1212 African-American girls.

4.2.3 Data Description

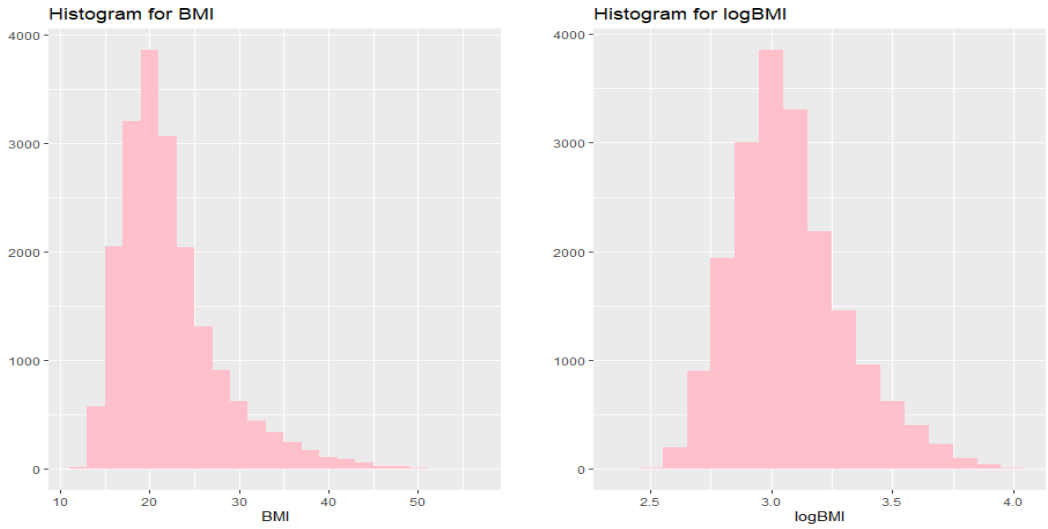
As in the FHS study, we again want to assess the dynamic joint density distribution of SBP and BMI over ages, conditioning on some other covariates, for example, DBP. Again, we compare the histograms of SBP and BMI (left panels in Figure 4.12 and Figure 4.13), with the histograms of log transformations, LogSBP and LogBMI (right panels in Figure 4.12 and Figure 4.13). We observed that after the log transformation, both SBP and BMI are more likely conditional normal distributed. We will assume the populations of the

logSBP conditioning on logBMI and logBMI conditioning on logSBP are following normal distributions. Among those 2375 subjects, the range of SBP is (72, 160) and for logSBP is (4.277, 5.075). The range of BMI is (11.31, 56.05) and for logBMI is (2.426, 4.026). And the range of DBP is (20, 106).



(a) (b)

Figure 4.12: NGHS: Histogram plots of SBP (left), logSBP (right)



(a) (b)

Figure 4.13: NGHS: Histogram plots of BMI (left), logBMI (right)

4.2.4 Two-step time-varying coefficients for conditional models of LogSBP and LogBMI

We assume LogSBP(t) and LogBMI(t) are following the same dynamic joint density distributions in 4.1.1. A slight difference of the univariate conditional models with the model Equation (4.1) and Equation (4.2) is that we will use the covariate DBP instead of Chol. The following are the two univariate conditional models:

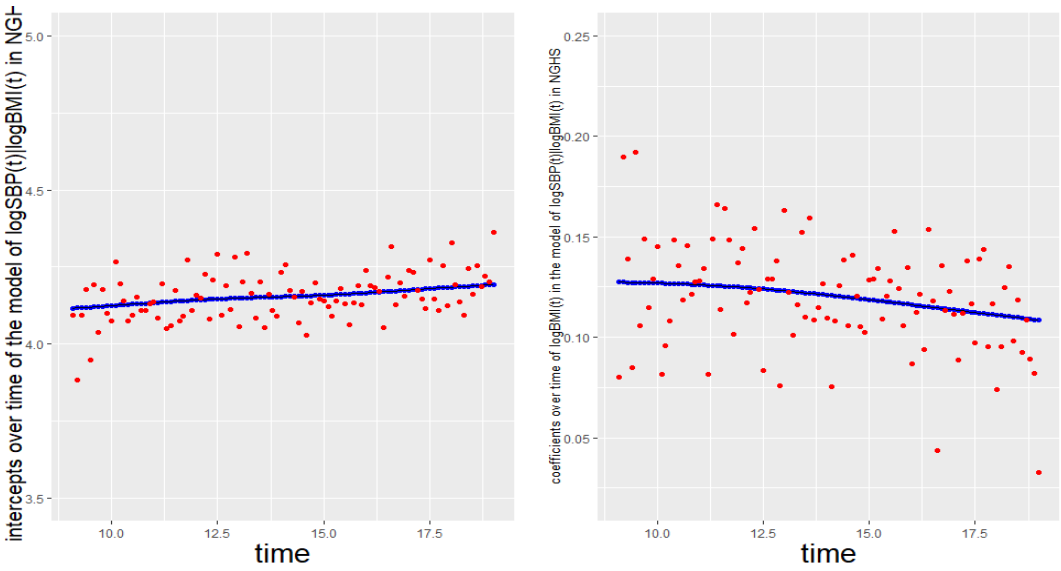
$$logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)DBP(t) \tag{4.4}$$

$$logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)DBP(t) \tag{4.5}$$

We assume LogSBP(t)|LogBMI(t) and LogBMI(t)|LogSBP(t) are conditional normals. The MLE and kernel estimators of coefficients over time are plotted in Figure 4.14 to Figure 4.17. The red dots are MLE estimators and the blue curves are kernel estimators.

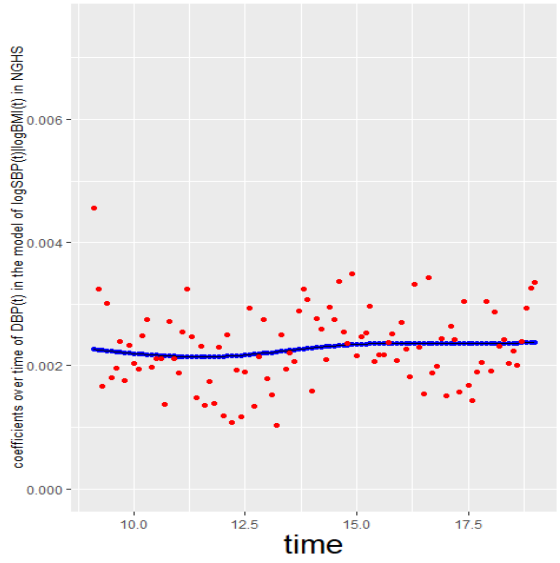
4.2.5 Dynamic MME of variances for Joint Density of LogSBP and LogBMI

Similar in 4.1.2.1, the kernel estimators of the coefficients: $\hat{\alpha}_1(t), \hat{\alpha}_2(t)$ represent the baseline values of logSBP and logBMI, and $\hat{\beta}_0(t), \hat{\beta}_1(t)$, represent the time-varying effects of logBMI and DBP on logSBP. $\hat{\gamma}_0(t), \hat{\gamma}_1(t)$ represent the time-varying effects of logSBP and DBP on logBMI.



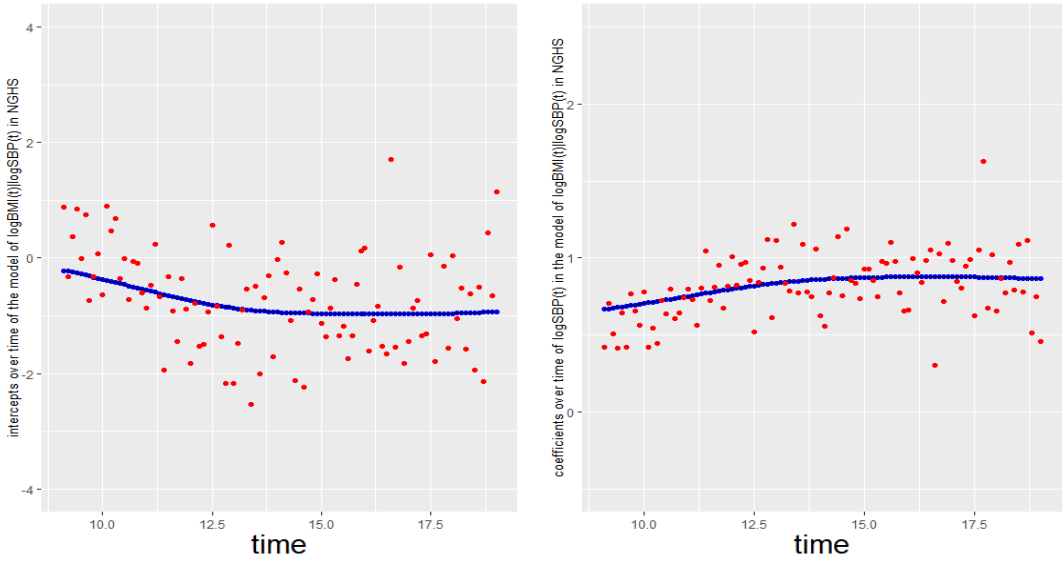
(a) (b)

Figure 4.14: NGHS: The estimated coefficients over time in model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)DBP(t)$. left: blue $\hat{\alpha}_1(t)$. red: $\tilde{\alpha}_1(t)$ right: blue: $\hat{\beta}_0(t)$. red: $\tilde{\beta}_0(t)$



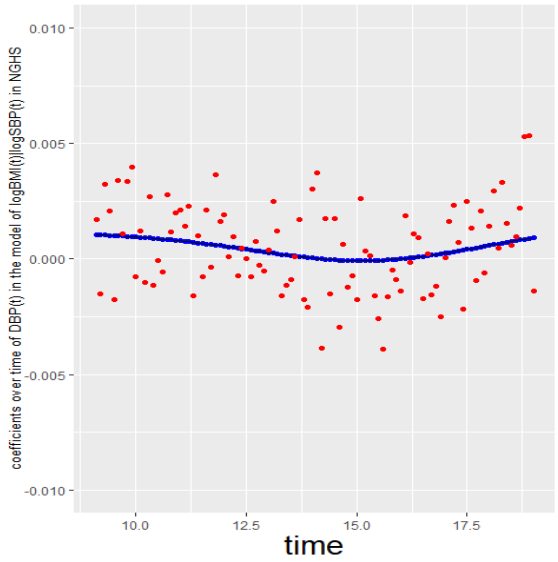
(a)

Figure 4.15: NGHS: The estimated coefficients blue: $\hat{\beta}_1(t)$ and red: $\tilde{\beta}_1(t)$ over time in model: $logSBP(t) = \alpha_1(t) + \beta_0(t)logBMI(t) + \beta_1(t)DBP(t)$.



(a) (b)

Figure 4.16: NGHS: The estimated coefficients over time in model: $logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)DBP(t)$. left: blue: $\hat{\alpha}_2(t)$ red: $\hat{\alpha}_2(t)$. right: blue: $\hat{\gamma}_0(t)$ red : $\tilde{\gamma}_0(t)$.



(a)

Figure 4.17: NGHS: The estimated coefficients blue: $\hat{\gamma}_1(t)$ red: blue: $\tilde{\gamma}_1(t)$ over time in model: $logBMI(t) = \alpha_2(t) + \gamma_0(t)logSBP(t) + \gamma_1(t)DBP(t)$.

The scatter points and the smoothed curves in Figure 4.14 (a), are the MLE $\tilde{\alpha}_1(t)$ and

kernel smoothed estimators $\hat{\alpha}_1(t)$ of coefficients in model Equation (4.4), represent the baseline values of logSBP and logBMI over time respectively. $\hat{\alpha}_1(t)$, the base line of logSBP is increasing by age, mainly ranging from 4.0 to 4.3.

Figure 4.14 (b) shows that the estimators of coefficients of logBMI over time are all positive, decreasing slightly when age increases, indicating that when age is increasing and DBP is not changing, the positive effects of logSBP on logBMI is decreasing. For example, when DBP is given, at age 10, the $\hat{\gamma}_0(t)$ is near 0.13, one unit increasing in logBMI causes 0.13 unit increasing in logSBP, but when at age 17.5, the $\hat{\gamma}_0(t)$ decreased to 0.13, one unit increasing in logBMI will only cause 0.12 unit increasing in logSBP. In Figure 4.15, again, all $\hat{\beta}_1(t)$ are around 0.002, indicating that DBP also has very rare positive effect on logSBP from age 9.1 to 19.

Different from the FHS study, $\hat{\alpha}_2(t)$ in Figure 4.16 (a) are slightly decreasing when age goes up. The $\hat{\gamma}_0(t)$ in the subplot Figure 4.16 (b) shows a slight increasing effect of logBMI on logSBP over time. In the plot Figure 4.17, all the $\hat{\gamma}_1(t)$ are also very close to 0, indicating nearly null effect of DBP on logSBP over time.

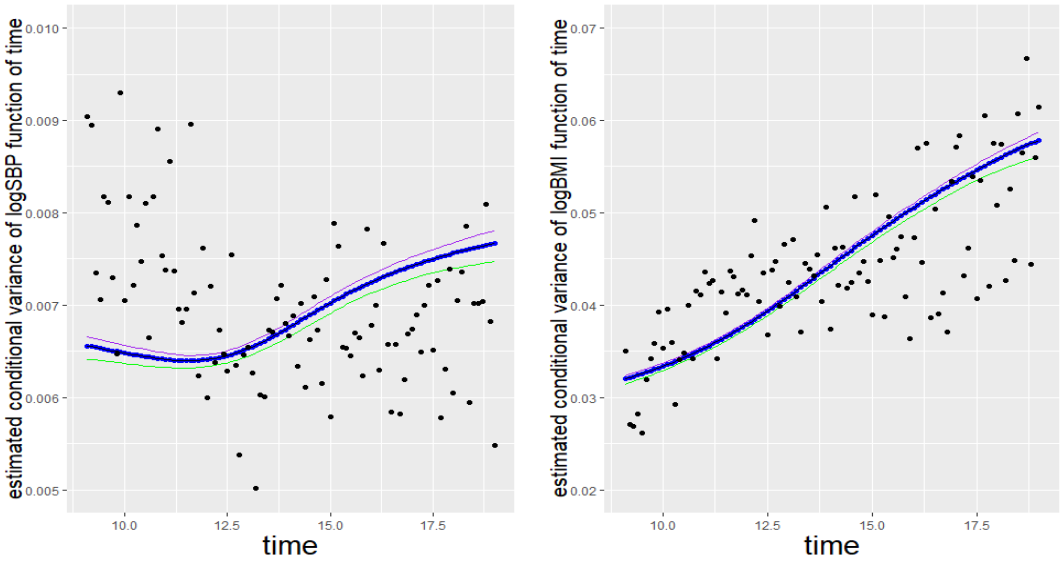
4.2.5.1 Raw MME of variances for Joint Density of LogSBP and LogBMI

The steps are the same as in the FHS, since we assume LogSBP and LogBMI follow the dynamic centered joint density distributions, we make the centering steps as in 4.1.2.2. After taking the expectation on both sides of the equations Equation (4.4) and Equation (4.5), we can calculate the expectations of $logSBP(t)$ and $logBMI(t)$ and subtract the expectation from $logSBP(t)$ and $logBMI(t)$ will make the centering steps. Again, according to 3.3.1.4, the raw consistent estimators over time for the joint density can be easily obtained.

4.2.5.2 Dynamic MME of variances and Bootstrap Point Wise Confidence Intervals for Joint Density of LogSBP and LogBMI

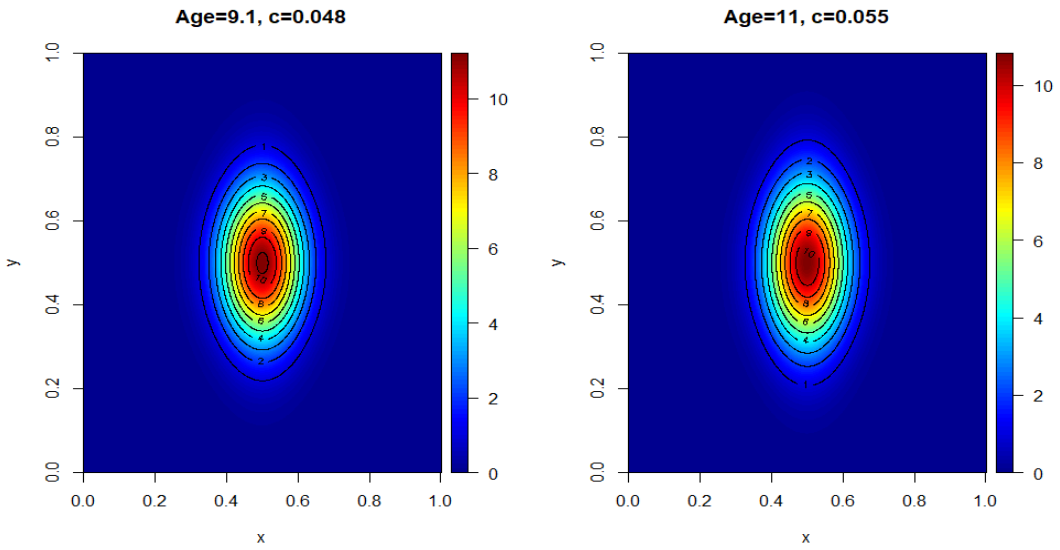
The following figures show the dynamic estimators of logSBP and logBMI over time.

The estimated dynamic variance (blue lines in the left panel in Figure 4.18) of logSBP is decreasing before age 12.5, but increasing after age 12.5. As expected, the logBMI is increasing over time from age 9.1 to 19 (blue lines in the right panel in Figure 4.18). The blue curves and the black dots (original variances directly calculated from the data) have similar trends over time. The purple and green curves are the 95% pointwise bootstrap confidence intervals for variance of logSBP and logBMI.

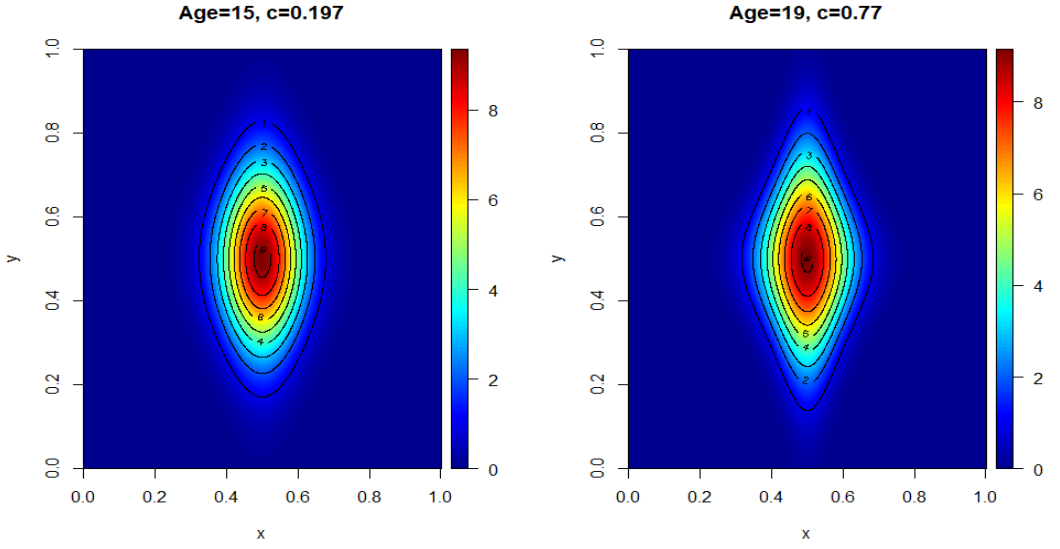


(a) (b)

Figure 4.18: NGHS: Estimators of the variances of logSBP (left), logBMI (right) conditioning on mean values of DBP at different time points. Blue lines are estimators of the variances of time. Purple and Green curves are 95% pointwise bootstrap confidence intervals. Black dots are original variances of logSBP and logBMI calculated directly from the data.



(a) (b)



(c) (d)

Figure 4.19: NGHS: Contour plots at age 9.1, 11, 15 and 19. The corresponding estimated c values are 0.048, 0.055, 0.197 and 0.77.

Figure 4.19 shows that at age 9.1, 11 and 15, the contours are almost elliptical, with estimated c(t) = 0.048, 0.055 and 0.197, respectively. At age 19, the estimated c(t) is 0.77, the density contours are of rhombus shape. Both the c values and contours at different age groups demonstrate that at some ages, the joint distribution of logSBP and logBMI conditioning on the mean values of DBP are not bivariate normal distributions.

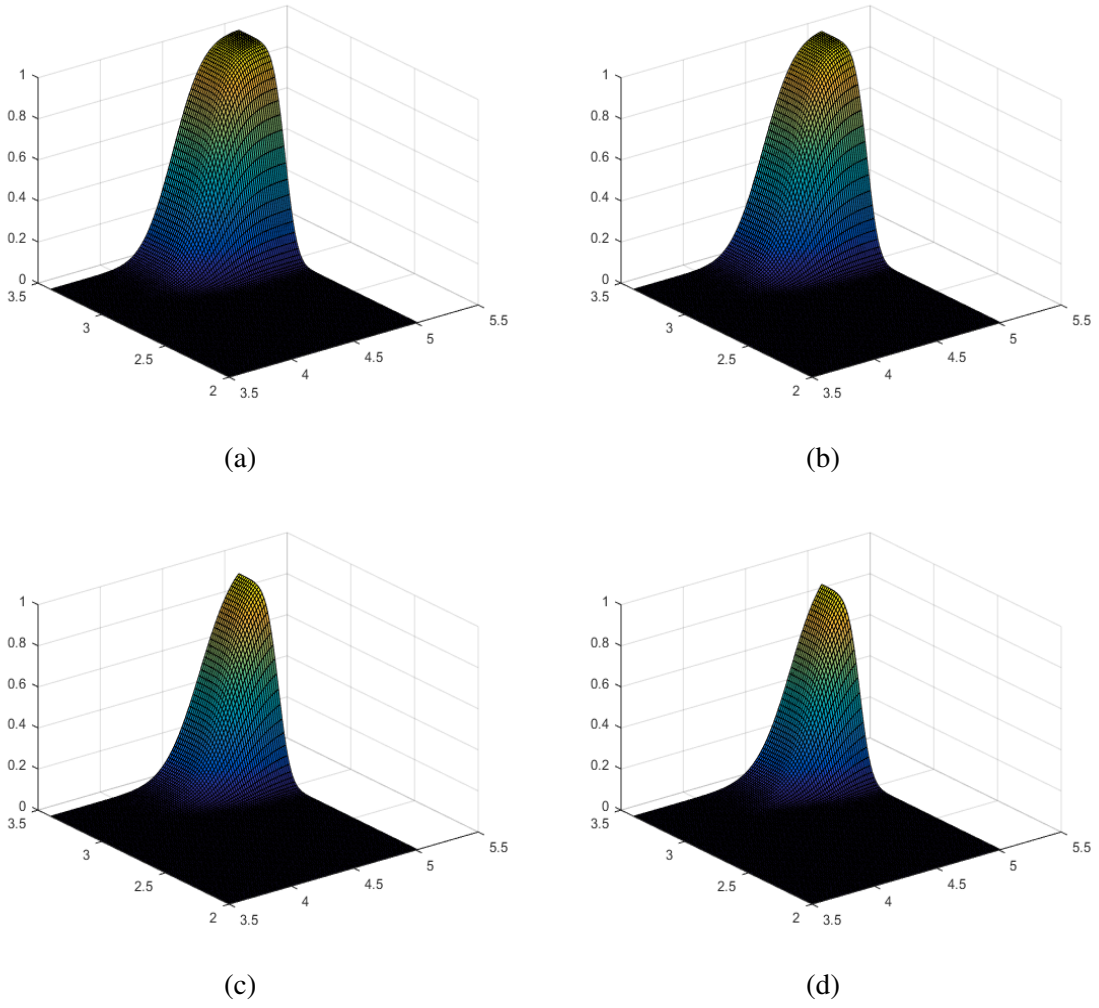


Figure 4.20: NGHS: Estimated CDF plot of logSBP $< log(150)$ $(SBP < 150)$ and logBMI $< log(30)$ $(BMI < 30)$ based on Mean(Chol) at age 9.1, 11, 15 and 19

age	CDF
9.1	0.9961
11	0.9902
15	0.9161
19	0.8640

Table 4.4: NGHS: Estimated of CDF when $SBP < 150$ and $BMI < 30$ based on Mean(DBP) at age 9.1, 11, 15 and 19

age	$\tilde{\sigma}^2_{sbp}(t)$	$L_{0.025}$ bound of CI for $\tilde{\sigma}^2_{sbp}(t)$	$U_{0.975}$ bound of CI for $\tilde{\sigma}^2_{sbp}(t)$	$\tilde{\sigma}^2_{bmi}(t)$	$L_{0.025}$ bound of CI for $\tilde{\sigma}^2_{bmi}(t)$	$U_{0.975}$ bound of CI for $\tilde{\sigma}^2_{bmi}(t)$
9.1	0.0073	0.0070	0.0075	0.0331	0.0321	0.0337
11	0.0067	0.0066	0.0069	0.0360	0.0353	0.0365
15	0.0068	0.0067	0.0070	0.0459	0.0448	0.0465
19	0.0072	0.0069	0.0074	0.0544	0.0525	0.0556

Table 4.5: NGHS: 95% Boostrap pointwise CI of MME parameters for Joint model based on Mean(DBP) at age 9.1, 11, 15 and 19

The CDF of $SBP < 150$ and $BMI < 30$ conditioning on the mean values of DBP at age 9.1, 11, 15 and 19 are calculated and plot above. Notice that log transformation preserves the original trend of the data without log transformation by time. The result is similar as in the FHS, when age is increasing, the cumulative density of $SBP < 150$ and $BMI < 30$ is decreasing, shown in both Figure 4.20 and Table 4.4. The 95% (2.5% lower bound to 97.5% upper bound) Boostrap pointwise confidence intervals of MME at age 9.1, 11, 15 and 19 are also shown in the Table 4.5.

4.2.6 Discussion

In this section, we study the medical insights of the bivariate outcome (SBP, BMI) in the effect of DBP of subjects from 9.1 to 19 years old in the NGHS. At some age groups, logSBP and logBMI are following the bivariate normal distributions given the mean value of DBP, while at some other age groups, they are not following the bivariate normal distributions, indicating from the contour plots.

As in the FHS, we can also calculate the exact joint density function of logSBP and logBMI at different age groups. Therefore, given DBP values and cutoff points of SBP and BMI, e.g. set $SBP < 150$ and $BMI < 30$ or some abnormal thresholds, we can calculate the exact CDFs at different ages. In the NGHS, the CDF of $SBP < 150$ and $BMI < 30$ is decreasing by ages. Conditioning on DBP, the CDF is decreasing from 87.8% to 44.5% for subjects from age 9.1 to 19.

The estimated variance of logSBP decreases before age 12.5 then it goes up after age 12.5. However, the variance estimator of logBMI is increasing from age 9.1 to 19.

In both the application studies of FHS and NGHS, we extend existing research on univariate to bivariate outcomes in the longitudinal studies, by incorporating both conditional mean and conditional density modeling schemes. Our estimation is also based on nonparametric methods, which allow us making relax assumptions comparing with the parametric estimations. We assume the bivariate outcome (logSBP and logBMI) only follow conditional normal distributions instead of bivariate normal distributions at different time points. The estimation methods will be evaluated in the following simulation studies.

Chapter 5: Simulation

5.1 Simulation Objective

The objectives of this simulation study is to investigate and evaluate performance of the proposed estimation procedures. In addition, to access the generalizability of the above approach, we examined different simulation scenarios.

5.2 Simulation Setting

Having described and applied our conditional specified estimation methods, we now proceed to two different simulation scenarios. In both scenarios, we will generate samples according to the following joint density function (**??**) ,

$$k_c(t)\frac{1}{2\pi\sigma_x(t)\sigma_y(t)}exp\{-\frac{1}{2}[(\frac{x(t)}{\sigma_x(t)})^2+(\frac{y(t)}{\sigma_y(t)})^2+c(t)(\frac{x(t)}{\sigma_x(t)})^2(\frac{y(t)}{\sigma_y(t)})^2]\} \tag{5.1}$$

and since the above equation is the most general nonstandard function with centered normal conditional distributions at any time point t, the $X(t)|Y(t)$ and $Y(t)|X(t)$ are following conditional normal:

$$\begin{aligned} X(t)|Y(t)=y(t) &\sim \mathcal{N}(0,\sigma_x^2(y(t))) \\ Y(t)|X(t)=x(t) &\sim \mathcal{N}(0,\sigma_y^2(x(t))) \\ \sigma_x^2(y(t)) &= \frac{\sigma_x^2(t)}{1+c(t)(\frac{y(t)}{\sigma_y(t)})^2} \\ \sigma_y^2(x(t)) &= \frac{\sigma_y^2(t)}{1+c(t)(\frac{x(t)}{\sigma_x(t)})^2} \end{aligned} \tag{5.2}$$

where $c(t)>0$ is the dependent parameter, $\sigma_x^2(y(t))$ and $\sigma_y^2(x(t))$ are conditional variances. Note that $\sigma_x^2(y(t))$ is depending on y(t) and $\sigma_y^2(x(t))$ is depending on x(t). And here $\sigma_x^2(t)$

and $\sigma_y^2(t)$ are marginal variances, $k_c(t)$ depend on $c(t)$, and the parameters needed to be estimated are $c(t), \sigma_x^2(t)$ and $\sigma_y^2(t)$.

5.3 Gibbs Sampling

As mentioned in chapter 2, the uniqueness of joint densities can be proved by recasting the problem in a Markov Chain setting. A Markov Chain can generate random samples based on the most recent samples by the transition probabilities which is indecomposable. A very powerful class of sampling algorithm in a probabilistic space is called the Markov Chain Monte Carlo (MCMC). MCMC can generate a Markov chain such that its stationary distribution coincides with the target posterior distributions. One widely applicable particular MCMC methods is the Gibbs sampling. It was proposed by [22, 20, 1] and fundamentally changed the Bayesian computing in high dimensional space.

Given a multivariate distribution, instead of simulating a n-dimensional sequential data directly using the complex joint density distribution function, the Gibbs Sampling can start by generating univariate conditional variables in turn, that is, this method is alternating drawing samplers of different conditional distributions. When the sample size goes to sufficient large, the generated data will converge to the target joint distribution [11, 32, 40] .

For generating a joint density data note by (x, y) at time point t, we start randomly assigning some initial values of x_0 and y_0. In our study, at time point t, the data will be generated by the following conditional distributions:

$$x_j(t) \sim p(X_j(t) = x_j(t) | Y_j(t) = y_{j-1}(t)) \sim N(0, \sigma_x^2(y_{j-1})(t)) = N(0, \frac{\sigma_x^2(t)}{1 + c(t)(\frac{y(t)}{\sigma_y(t)})^2})$$

$$y_j(t) \sim p(Y_j(t) = y_j(t) | X_j(t) = x_j(t)) \sim N(0, \sigma_y^2(x_j)(t)) = N(0, \frac{\sigma_y^2(t)}{1 + c(t)(\frac{y(t)}{\sigma_x(t)})^2})$$

After iterating the simulation steps sufficient large number of time (say N, in our

simulation, the N is several hundreds of thousands), the first a few number of samples referred to as the burn-in period, often start the chain far away from the target joint distributions. However, those burn-in samples are first "few" samples, relative small compare to the large number N, they will have no effects on the convergence to the target joint distributions [30]. The following pseudo code is describing the Gibbs Sampling algorithm Figure 5.1 .

Algorithm 5 Gibbs Sampling

1: **for** each time point t in $1 : T$ **do**
2: initialize x_0, and y_0 by randomly numbers
3: **for** each integer j in $2 : N$ **do**
4: $x_j \sim rnorm(1, mean = 0, sd = \sigma_x(y_{j-1})(t))$
5: $y_j \sim rnorm(1, mean = 0, sd = \sigma_y(x_j)(t))$
6: **end for**
7: Return as.data.frame(x_j, y_j)
8: **end for**

Figure 5.1: Gibbs Sampling algorithm

5.4 Simulation Scenarios

5.4.1 Simulation Scenario 1: Dynamic Bivariate Normal Distribution

The complexity of correlated data structure in the multivariate longitudinal data suggests that the simulation would be difficult. If we devise an easier simulation scheme in the first scenario, both the simulation strategy and estimation of parameters will be easier. In the second and third scenarios, we will devise more complicate simulation schemes.

In the first scenario, at each time point, we generate data from a multivariate normal distribution, which is a special case of the bivariate conditional normal distributions, and all those variables over time are set to be independent with each other, which means that the $c(t)$ equals to 0.

By setting $c(t) = 0$ in above function Equation (5.1), the joint dynamic density function will be:

$$\frac{1}{2\pi\sigma_x(t)\sigma_y(t)} exp\{-\frac{1}{2}[(\frac{x(t)}{\sigma_x(t)})^2 + (\frac{y(t)}{\sigma_y(t)})^2]\} \tag{5.3}$$

The above function is the bivariate normal density function of X and Y who are independent with each other at each time point t. However, as presented in previous chapters, we want to investigate the dynamic joint distributions based on a third variable Z. So, we generate a dynamic multivariate normal distributions of 3 dimensions, with independent components. And in our simulation study, only the parameters over time of dynamic joint density functions of first two variables will be estimated. The criteria for evaluating these methods will be the absolute errors between the estimated and true parameters. The parameters to be estimated will be: dynamic variance of $X(t)$ and $Y(t)$, and function of dependent parameter over time $c(t)$ conditioning on $mean(Z(t))$. Those parameters will be enough to determine a dynamic centered joint distribution conditioning on $mean(Z(t))$ since the forth parameter $k(c(t))$ only depends on c(t).

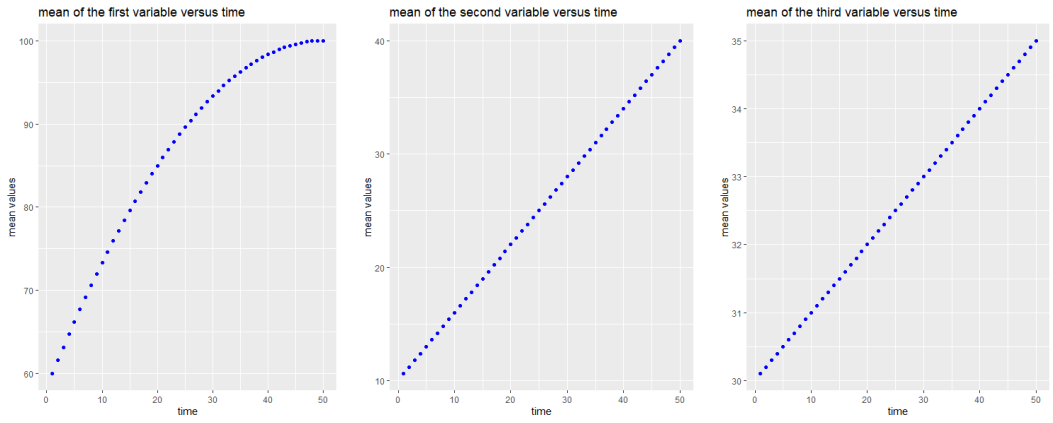
5.4.1.1 Simulation Steps for Scenario 1

To investigate the estimation performance on some nonlinear trends over time and mimic the data structures in the NGHS data, we add some curvature on the mean functions over time. **Step 1 for Scenario 1**: We set three mean functions over time t by the following equations in Equation (5.4), corresponding to three variables, X, Y and Z. For making sure our method can assess complicate distribution structures, we set the mean function of X to be the first half of a parabolic curve by assigning a quadratic function of time t. For the other two

variables Y and Z, the mean trend of time are linear functions of t:

$$\begin{aligned} \mu(x(t)) &= 100 - \frac{(t-50)^2}{60} \\ \mu(y(t)) &= 0.6t + 10 \\ \mu(z(t)) &= 0.1t + 30 \end{aligned} \tag{5.4}$$

where t belongs to a sequence of time points from 1 to 50 with a length of 50. The plots of simulated mean functions of time are showed in Figure 5.2.



(a) (b) (c)

Figure 5.2: Scenario 1: Simulated mean functions of time for X, Y and Z.

Step 2 for Scenario 1: The marginal variance function of time as set by Equation (5.5):

$$\begin{aligned} \sigma_x^2(t) &\sim \mathcal{N}(5, 0.2) \\ \sigma_y^2(t) &= 0.3t \\ \sigma_z^2(t) &= 0.16t + 11 \end{aligned} \tag{5.5}$$

Step 3 for Scenario 1: Thirdly, we simulate multivariate normal distributions of three dimensions at different t, each component of the three dimensions has an univariate normal distribution with different means and variances corresponding to Equation (5.4) and Equation (5.5). The three components, X, Y and Z, are independent with each other. X and Y are

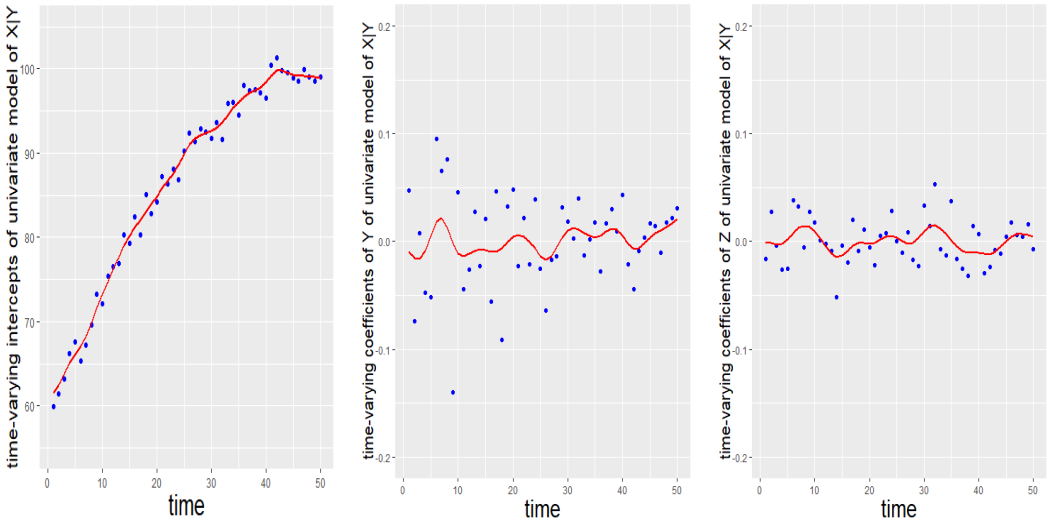
following the distribution in Equation (5.1) .

5.4.1.2 Estimation Results for Scenario 1

The estimation of the parameter functions of time point t is following the steps in 3.3.1, 3.3.2 and 4.3. The whole estimation approach contains two process. First we make the two-step time-varying coefficient estimation, then in the second process, we calculate the raw MME of time according to the formulas in 3.3.2 following by the kernel smoothing on the raw MME to estimate the dynamic parameter functions of time. The optimal bandwidths are selected by the leave-one-out cross-validation.

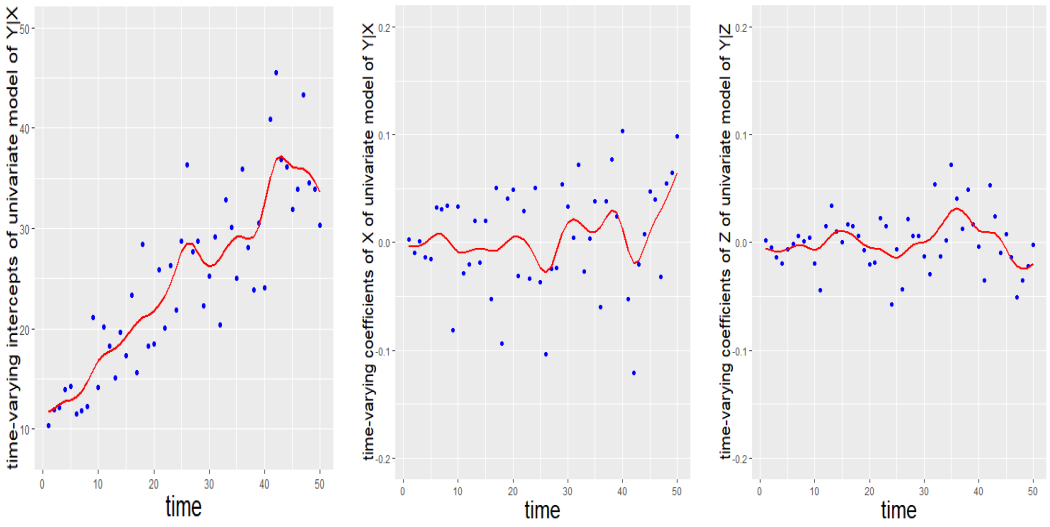
The following Figure 5.3 and Figure 5.4 are showing the time-varying coefficients for conditional univariate variables X (given Y) and Y (given X) based on mean of Z. Note that in each sub plot, the smoothing lines (red curves) are all undersmoothing, which is clearly visualized compared with the MLE coefficients (blue dots) over time. The reason of the undersmoothing is to retain the trend and also to make sure the bias of the raw MLE of coefficients are small. Since that we are estimating the two-dimensional density parameters, a slight change on the two-step coefficient estimators will induce large departure of the estimators for bivariate distributions.

Also, in plots (b) and (c) of Figure 5.3 and Figure 5.4, all those smoothed coefficients are around 0, demonstrating that in the univariate model of X, both Y and Z have nearly no effect on X and in the model of Y, both X and Z also have nearly no effect on Y at all different time points t.



(a) (b) (c)

Figure 5.3: Scenario 1: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Red curves: estimated time-varying coefficient curves; Blue dots: raw coefficients at different time points

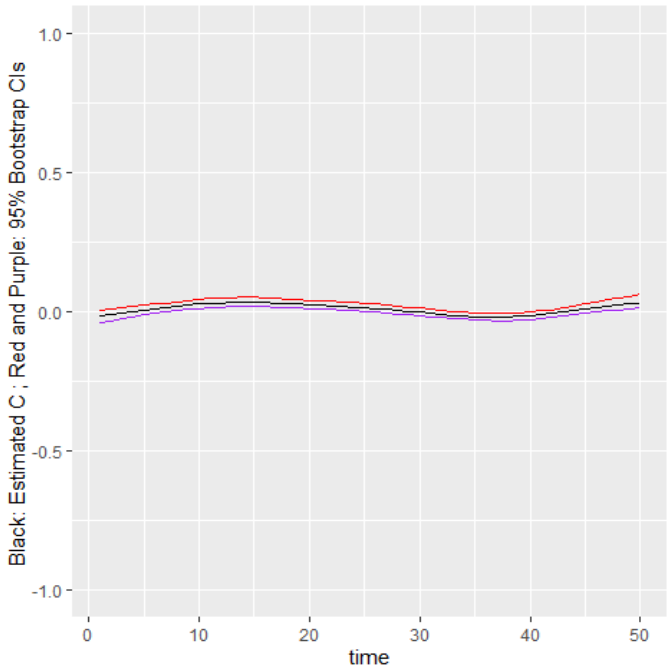


(a) (b) (c)

Figure 5.4: Scenario 1: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$. Red curves: estimated time-varying coefficient curves; Blue dots: raw coefficients at different time points

The following tables and figures are showing all estimated dynamic parameters and their

95% bootstrap pointwise confidence intervals at different time points based on $mean(Z(t))$ under scenario 1.



(a)

Figure 5.5: Scenario 1: Estimated C(t) over time and 95% Bootstrap pointwise confidence intervals based on 800 bootstrap samples. Black: estimated C(t); Purple and Orange: CIs

t	$c(t)$	$\hat{c}(t)$
10	0	0.0269770451
20	0	0.0254296457
30	0	-0.0027417813
40	0	-0.0143375876
50	0	0.0312441357

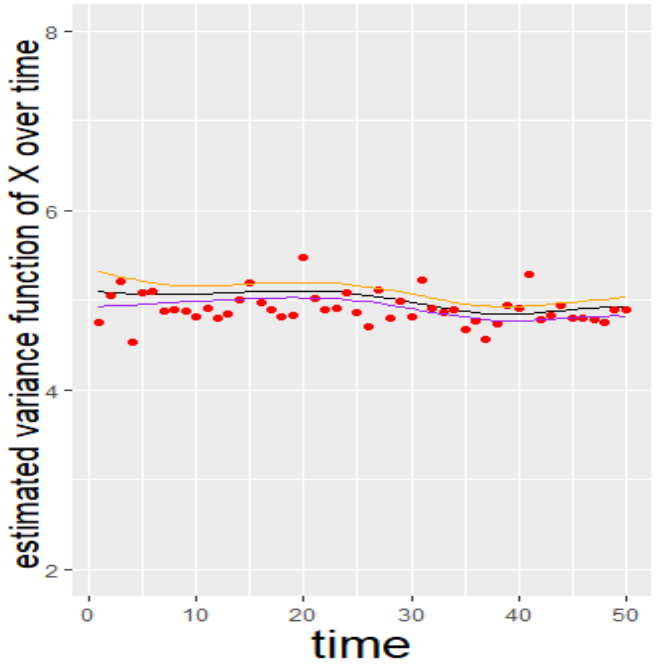
Table 5.1: Scenario 1: Estimated $\hat{c}(t)$ versus True $c(t)$ at 5 time points

t	$\sigma_x^2(t)$	MME of $\hat{\sigma}_x^2(t)$	95% Bootstrap pointwise CIs of $\hat{\sigma}_x^2(t)$	$\lvert\hat{\sigma}_x^2(t) - \sigma_x^2(t)\rvert$
10	4.821992	5.068754	(4.988058,5.159181)	0.246762
20	5.483167	5.107249	(5.029375,5.194994)	0.375918
30	4.812810	4.982054	(4.908211,5.072955)	0.169244
40	4.906820	4.848767	(4.773043,4.930865)	0.058053
50	4.900630	4.931432	(4.823575,5.032509)	0.030802

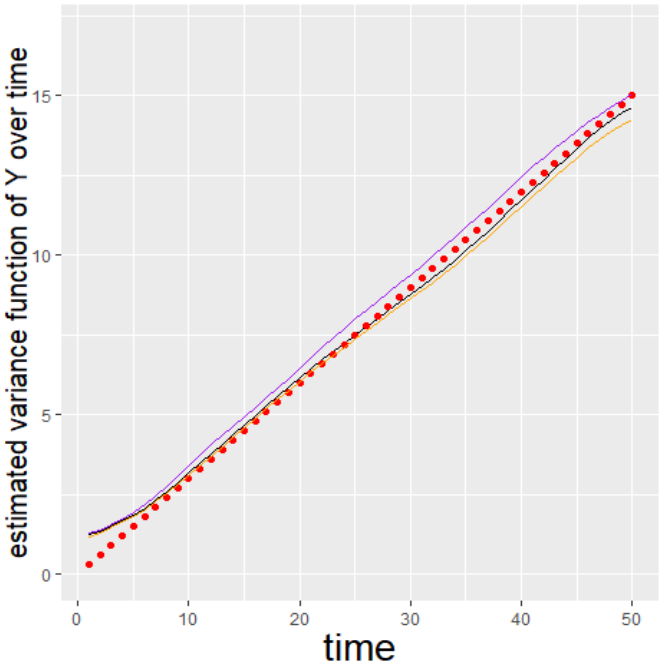
Table 5.2: Scenario 1: First column: time points; Second column: true variance of x at t; Third column: MME of $\sigma_x^2(t)$; Forth column: 95% Bootstrap pointwise CIs based on 800 bootstrap samples; Fifth column: absolute error of $\hat{\sigma}_x^2(t)$.

t	$\sigma_y^2(t)$	MME of $\hat{\sigma}_y^2(t)$	Bootstrap 95% pointwise CI of $\hat{\sigma}_x^2(t)$	$\lvert\hat{\sigma}_y^2(t) - \sigma_y^2(t)\rvert$
10	3.0	3.167188	(3.105294,3.396333)	0.167188
20	6.0	6.148075	(6.042671,6.455805)	0.148075
30	9.0	8.758682	(8.624782,9.362719)	0.241318
40	12.0	11.726439	(11.526711,12.453777)	0.273561
50	15.0	14.646696	(14.244389,15.059168)	0.353304

Table 5.3: Scenario 1: First column: time points; Second column: true variance of x at t; Third column: MME of $\sigma_y^2(t)$; Forth column: 95% Bootstrap pointwise CIs based on 800 bootstrap samples; Fifth column: absolute error of $\hat{\sigma}_y^2(t)$.



(a)



(b)

Figure 5.6: Scenario 1: Dynamic MME of variances for X (a) and Y (b) over time and 95% Bootstrap pointwise CIs conditioning on the mean value of Z. Black curve: estimated curve; Red dots: true value; Purple and Orange: CIs

In Figure 5.5 and Table 5.1 , at all the 5 time points, the estimated MME of $\hat{c}(t)$, are all very close to the true $c(t)$= 0. As interpreted in previous chapters, the 0 value of c(t) means X(t) and Y(t) are independent with each other when giving the mean of Z(t). This result is very consistent with our original simulation setting. All estimated MME variances are very close to the true variances for X(t) and Y(t) over time, showing in Table 5.2 and Table 5.3 . All the absolute errors are listed in the last column in the two tables, no larger than 0.4.

In Figure 5.6, the estimated dynamic MME of variances for X and Y (black curves) and their corresponding bootstrap pointwise 95% confidence intervals (purple and orange curves) are shown. All estimator curves are close to and have the same trends of the original correct variance (red dots), demonstrating that our approach makes good estimation on our simulated data in scenario 1. However, note that at boundary time points, the $\hat{\sigma}_y^2(t)$ have larger bias than at other time points, this might induced by the boundary bias of kernel smoothers.

5.4.2 Simulation Scenario 2: Dynamic Conditional Normal Distributions

In scenario 1, we simulate a sample from normal conditional distributions by generating multivariate normally distributed data with 3 independent components, X(t), Y(t) and Z(t) at different time point t. However, in most real application studies, covariates are usually dependent with each other. Hence, in the scenario 2, different dependent parameters c(t) between pairs of variables X(t) and Y(t) are set to be a linear line over time.

5.4.2.1 Simulation Steps for Scenario 2:

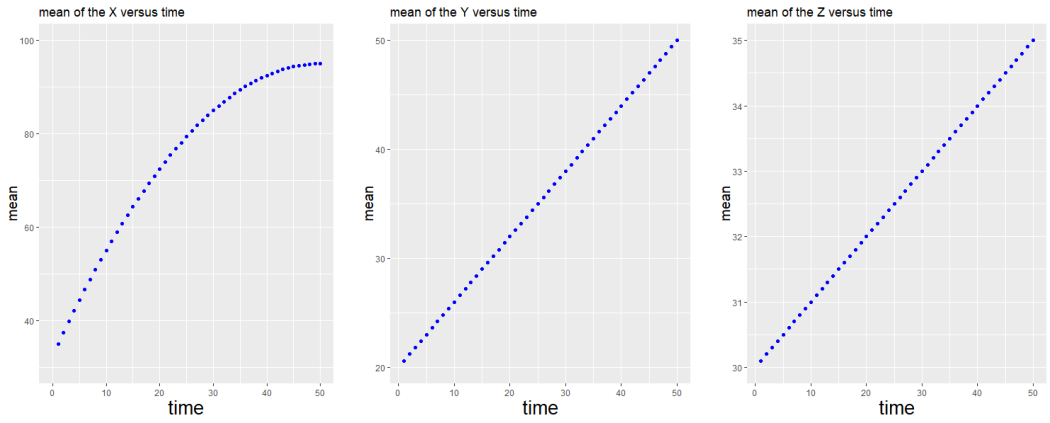
There are 4 steps for this scenario:

Step 1 for Scenario 2: Time points setting, we set the sequence of time: $t = 1, 2, 3, ...50$.

Step 2 for Scenario 2: The 3 mean functions for variables X, Y and Z are similar as in the Scenario 1. The mean function for X over time is set to be a half of a parabolic curve. And the mean function for Y and Z over time is also positive linear.

$$\begin{aligned} \mu(x(t)) &= 95 - (t-50)^2)/40 \\ \mu(y(t)) &= 0.6t + 20 \\ \mu(z(t)) &= 0.1t + 30 \end{aligned} \tag{5.6}$$

The plots of the mean functions over time point t:



(a) (b) (c)

Figure 5.7: Scenario 2: Simulated mean of time for X, Y and Z

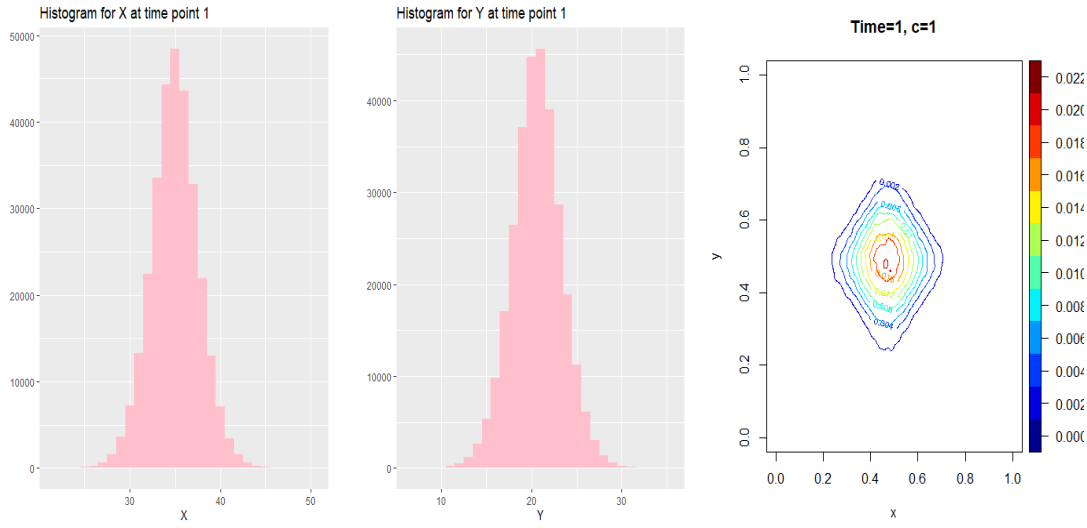
Step 3 for Scenario 2: Generate 3 marginal variance functions over time for X, Y and Z

$$\begin{aligned} \sigma_x^2(t) &\sim \mathcal{N}(10, 0.2) \\ \sigma_y^2(t) &\sim \mathcal{N}(10, 0.2) \\ \sigma_z^2(t) &\sim \mathcal{N}(2, 0.1) \end{aligned} \tag{5.7}$$

Step 4 for Scenario 2: The dependent parameter over time will be $c(t)$: a sequence from 1 to 25, of length = 50, equally spaced.

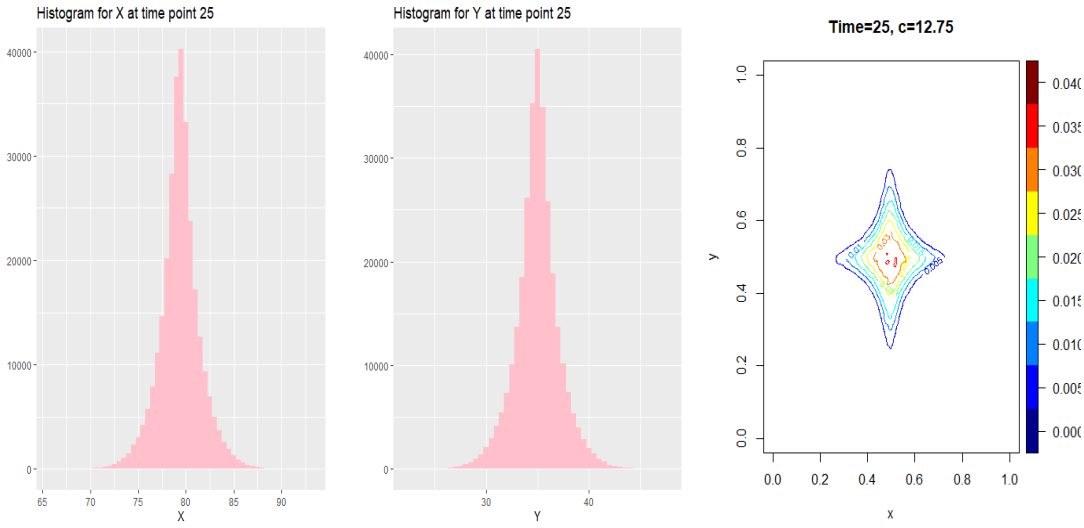
Step 5 for Scenario 2: The Gibbs Sampling needs a very large sample size to make approximation of the target distributions, if the sample size is small, the estimation results will be poor. Here, at each time point t, we generate 300000 samples, so we will have 15000000 samples in total over all 50 time points. Only $X(t)$ and $Y(t)$ are simulated by the Gibbs sampling method, the third variable $Z(t)$ is generated separately from a normal distribution according to the above simulation setting.

The histograms of the simulated conditional $X(t)|Y(t)$ and $Y(t)|X(t)$, at t=1, 25 and 50 are shown below (left two panels in Figure 5.8 , Figure 5.9, Figure 5.10), the density contour plots at the three time points are also shown in the right panels in the three figures. The histograms are all displaying normal distributions and when c(t) value is increasing , the contours are of different shapes. When t=1, the contours are rhombus, and when t = 25 and 50, the contours are cross-shaped.



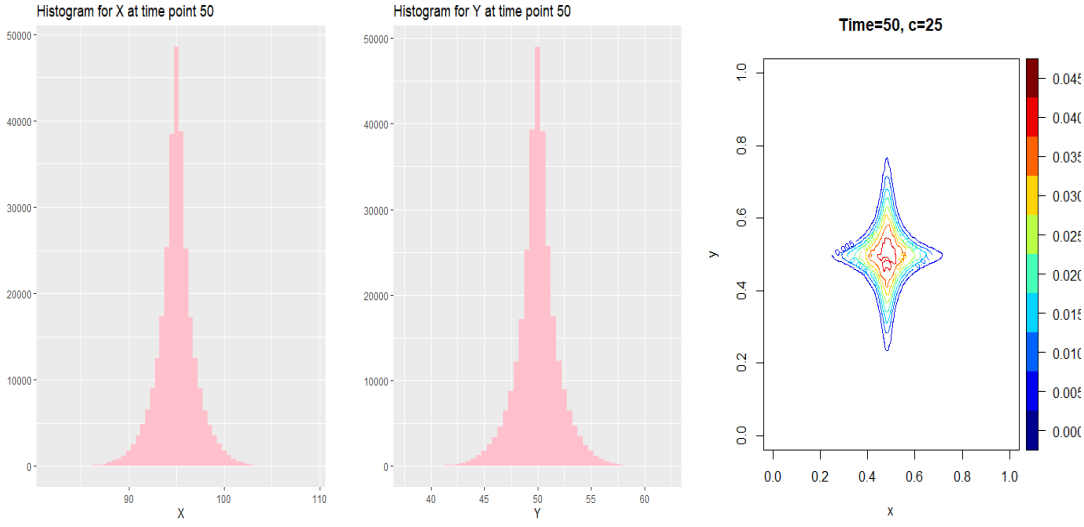
(a) (b) (c)

Figure 5.8: Scenario 2: Simulated histograms and density contours for x and y at time point t= 1



(a) (b) (c)

Figure 5.9: Scenario 2: Simulated histograms and density contours for x and y at time point t= 25



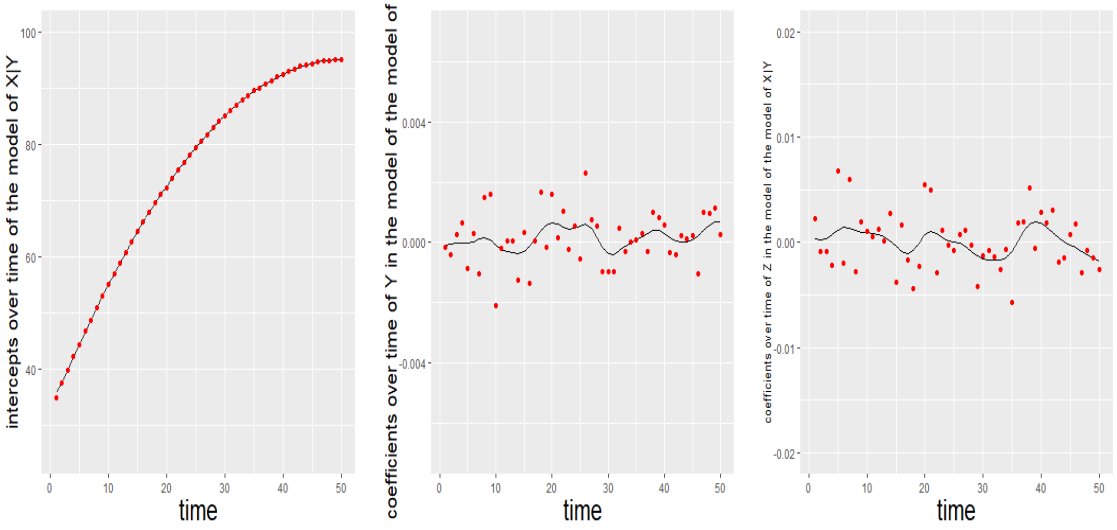
(a) (b) (c)

Figure 5.10: Scenario 2: Simulated histograms and density contours for x and y at time point t= 50

5.4.2.2 Estimation Results for Scenario 2

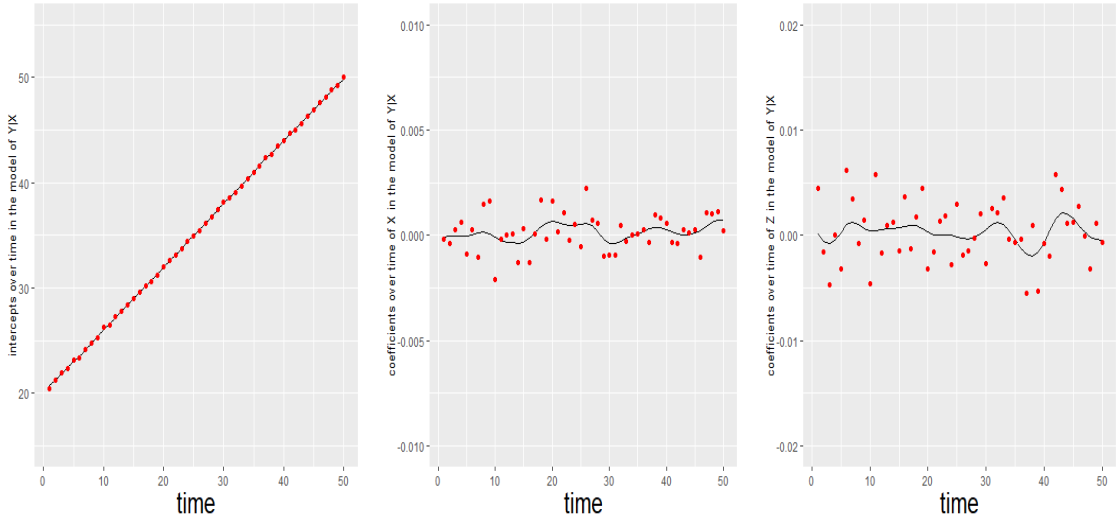
As in Scenario 1, the samples are simulated without any missing data. Again, the estimation steps are the same as we discussed in previous chapters and the Scenario 1.

Figure 5.11 and Figure 5.12 are showing the two-step estimation results on the time-varying coefficients. Again in the second smoothing step, we use undersmoothing on the raw estimators to minimize the bias. In both figures, black curves are estimated kernel smoothers of the time-varying coefficients, and the red dots are the raw estimators from the first step.



(a) (b) (c)

Figure 5.11: Scenario 2: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points.

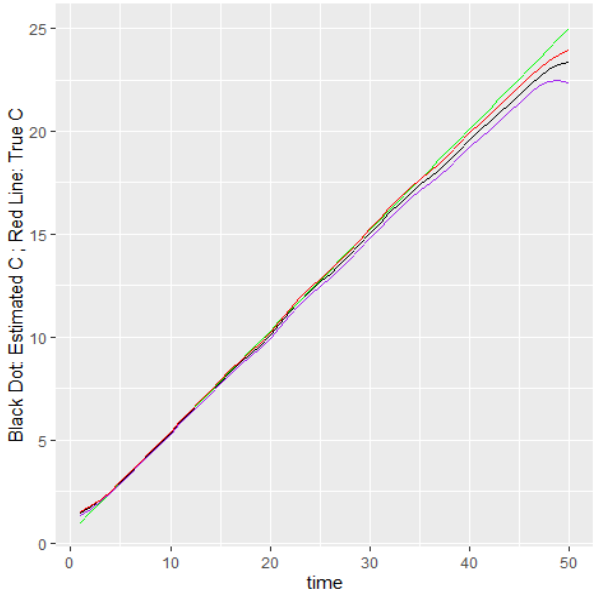


(a) (b) (c)

Figure 5.12: Scenario 2: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$.Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points.

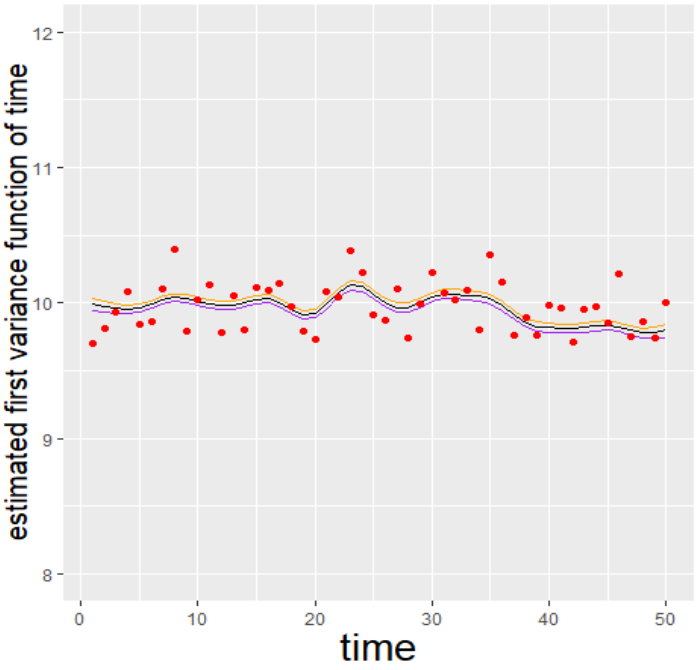
t	$c(t)$	$\hat{c}(t)$	$\hat{k}(c(t))$
1	1	0.8177597	1.221071
10	5.408163	5.3644081	1.703201
20	10.306122	9.9611973	1.978428
30	15.204082	15.3008469	2.217416
40	20.102041	19.4187691	2.370154
50	25.000000	23.6076083	2.507234

Table 5.4: Scenario 2: Estimated $\hat{c}(t)$ versus True c(t) at different time points. First column: time points; Second column: true c(t); Third column: estiamted c(t)

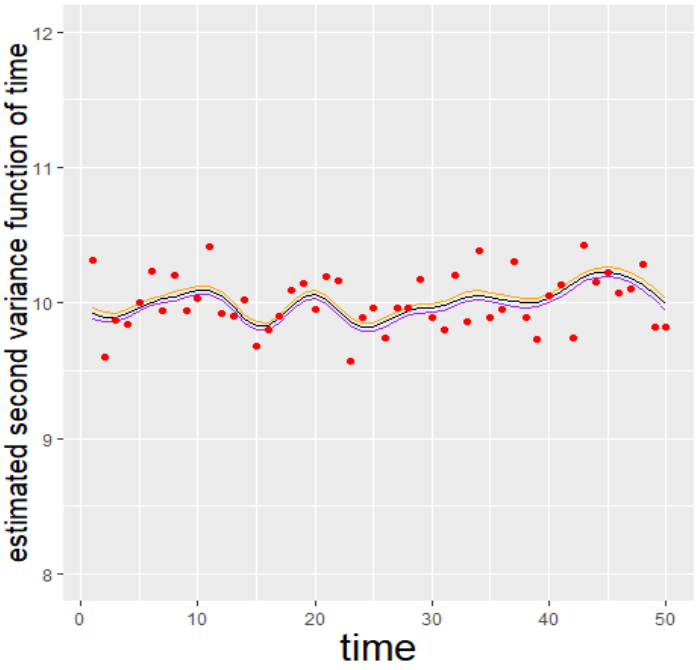


(a)

Figure 5.13: Scenario 2: Estimated dependent parameter c(t). Black line: estimated c(t), Green line: true c(t), Red and Purple lines: 95% Bootstrap pointwise confidence intervals of c(t)



(a)



(b)

Figure 5.14: Scenario 2: Dynamic MME of variances over time and bootstrap point wise confidence intervals for the X and Y conditioning on the mean value of Z. Red dots: true variances; Black: estimated MME; Purple and Orange: 95% Bootstrap pointwise CIs

t	$\sigma_x^2(t)$	MME of $\hat{\sigma}_x^2(t)$	Bootstrap 95% point wise CI of $\hat{\sigma}_x^2(t)$	$\lvert\hat{\sigma}_x^2(t) - \sigma_x^2(t)\rvert$
1	9.703886	9.990932	(9.941761,10.037501)	0.287046
10	10.026225	10.009521	(9.980850,10.040345)	0.016704
20	9.737145	9.925397	(9.890025,9.956479)	0.188252
30	10.222742	10.043596	(10.011285,10.078403)	0.179146
40	9.982104	9.818059	(9.784310,9.851606)	0.164045
50	10.005500	9.801490	(9.742926,9.840006)	0.20401

Table 5.5: Scenario 2: Estimated $\hat{\sigma}_x^2(t)$, 95% Bootstrap pointwise CIs based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}_x^2(t)$

t	$\sigma_y^2(t)$	MME of $\hat{\sigma}_y^2(t)$	Bootstrap 95% point wise CI of $\hat{\sigma}_y^2(t)$	$\lvert\hat{\sigma}_y^2(t) - \sigma_y^2(t)\rvert$
1	10.315434	9.923027	(9.886368,9.966285)	0.392407
10	10.029160	10.093543	(10.067208,10.125990)	0.064383
20	9.949992	10.064083	(10.032241,10.096260)	0.114091
30	9.891622	9.963781	(9.930304,9.997072)	0.072159
40	10.051407	10.032013	(9.999049,10.065833)	0.019394
50	9.818517	9.988919	(9.944418,10.038110)	0.170402

Table 5.6: Scenario 2: Estimated $\hat{\sigma}_y^2(t)$, 95% Bootstrap pointwise CIs based on 500 Bootstrap samples and absolute errors of $\hat{\sigma}_y^2(t)$

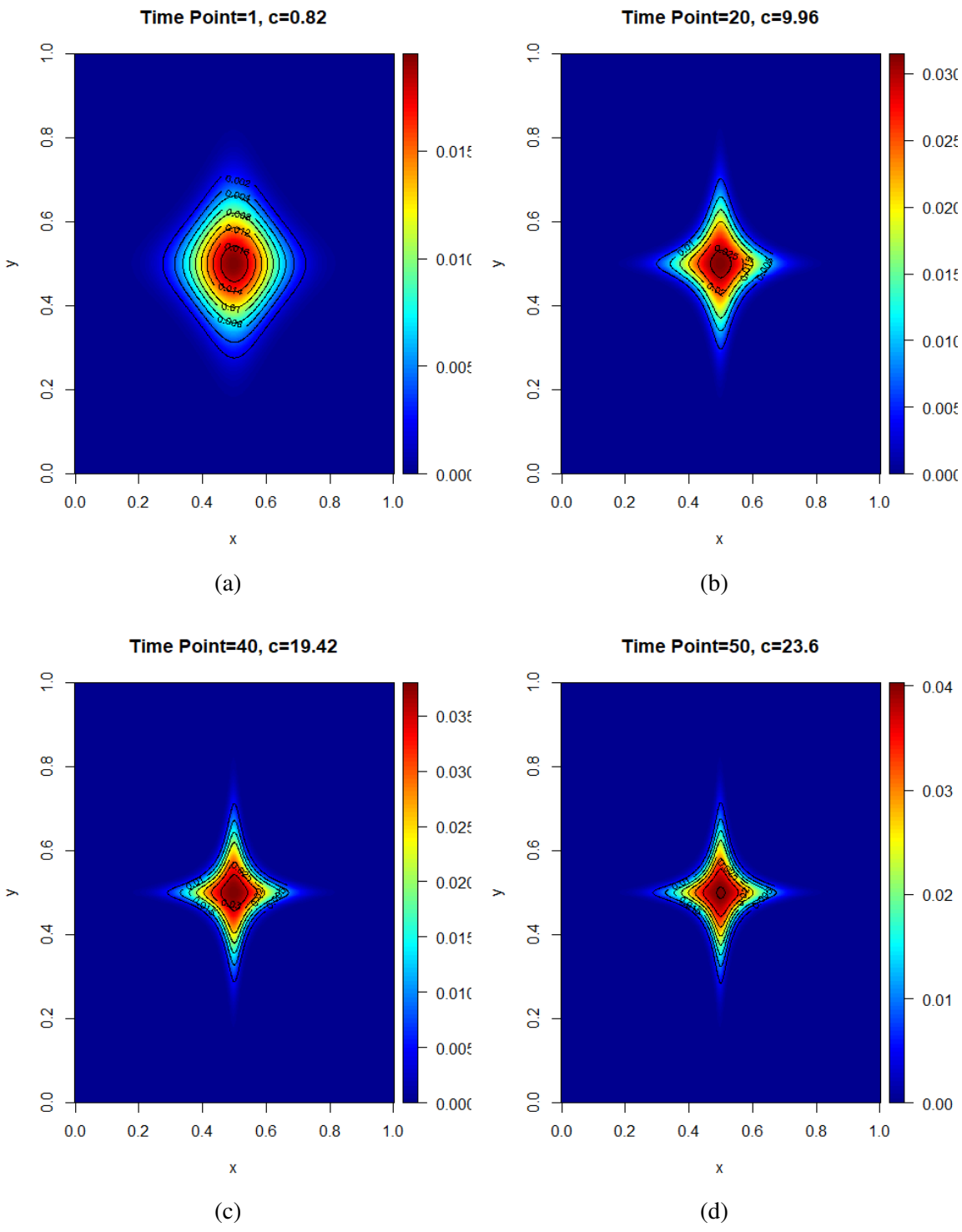


Figure 5.15: Scenario 2: Simulated density contours for x and y at time point t= 1, 20, 40 and 50

At time points: t=1, 10, 20, 30, 40 and 50, the estimated MME of c(t): $\hat{c}(t)$ are listed in

Table 5.4, also a visualized comparison between the $\hat{c}(t)$ and $c(t)$ is shown in Figure 34. All estimated $\hat{c}(t)$ are very close to the true c(t). The corresponding $\hat{k}(c(t))$ are also calculated according to equation (12) in 4.1.3.2.

In Figure 5.14, at all time point t, the estimated dynamic MME of variances for x and y conditioning on the mean value of z (black curves) and their 95% Bootstrap pointwise confidence intervals (purple and orange curves) versus true varainces (red dots) are showing. Also, in Table 5.5 and Table 5.6, the estimated MME variances $\hat{\sigma_x^2}(t)$ and $\hat{\sigma_y^2}(t)$ are compared with the original variance $\sigma_x^2(t)$ and $\sigma_y^2(t)$. The absolute errors $|\hat{\sigma_x^2}(t) - \sigma_x^2(t)|$ and $|\hat{\sigma_y^2}(t) - \sigma_y^2(t)|$ are listed in the right columns, all < 0.4, illustrating the precision of our estimation results.

The contour plots at 4 different time points in Figure 5.15 are showing different shapes. At time point t=1, c=0.82, the density contour is a rhombus. When t is increasing, at t=20, c=0.96; at t=40, c= 19.42 and at t=50, c=23.6, all the density contours are cross-shaped. Since all c values at those time points are not equal to 0, hence, all contour plots are not elliptical, demonstrating that when conditioning on mean(z(t)), X(t) and Y(t) are not bivariate normally distributed.

5.4.3 Simulation Scenario 3: Dynamic Conditional Normal Distributions with dynamic correlations

Similar with the scenario 2, the dependent parameters c(t) between pairs of variables (X(t), Y(t)) at different time points are not set to be 0, but in this scenario, the dependent parameters c(t) are generated by a polynomial curve over time. The varying correlations over time between the third variable Z(t) and X(t) are also considered. The following are the simulation steps:

5.4.3.1 Simulation Steps for Scenario 3:

There are 4 steps for this scenario:

Step 1 for Scenario 3: Time points setting: we set the sequence of equally spaced time points $t = 1, 2, ...50$.

Step 2 for Scenario 3: The mean functions for variables X(t), Y(t) are set the same as in the Scenario 2. The third variable Z(t) are generated by assigning correlations over time after the variable X(t) has been simulated.

$$\begin{aligned} \mu(x(t)) &= 95 - (t-50)^2/40 \\ \mu(y(t)) &= 0.6t + 20 \end{aligned} \tag{5.8}$$

Step 3 for Scenario 3: Again, the 2 marginal variance functions over time for X(t) and Y(t) are set as the same as in scenario 2.

$$\begin{aligned} \sigma_x^2(t) &\sim \mathcal{N}(10, 0.2) \\ \sigma_y^2(t) &\sim \mathcal{N}(10, 0.2) \end{aligned} \tag{5.9}$$

Step 4 for Scenario 3: The dependent parameter over time $c(t)$ is following the equation:

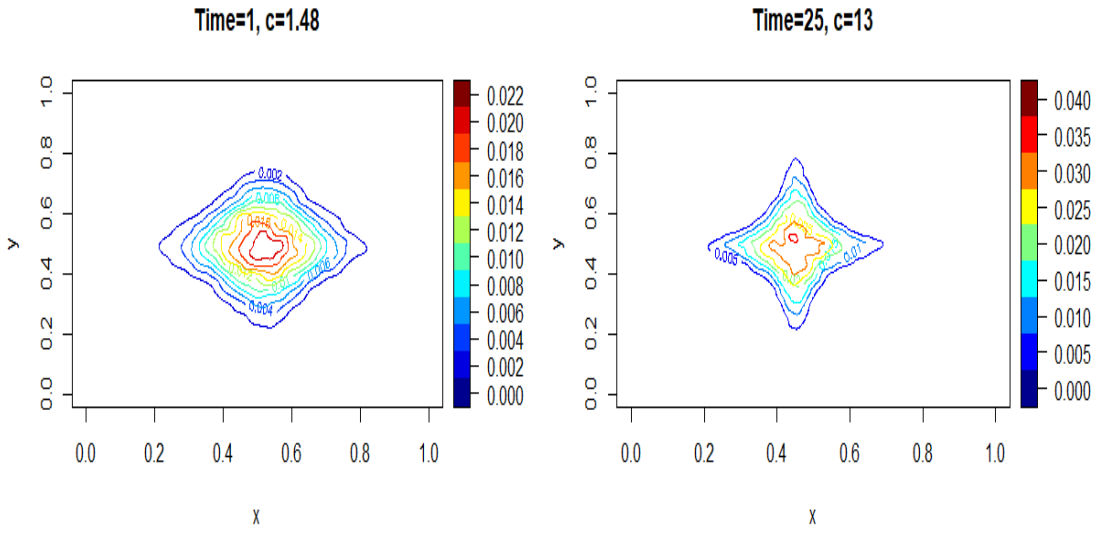
$$c(t) = 13 - (t-25)^2/50 \tag{5.10}$$

Step 5 for Scenario 3: The correlations between X(t) and Z(t) over time is linear :

$$\rho_{xz}(t) = 0.008t + 0.2 \tag{5.11}$$

Step 6 for Scenario 3: At each time point t, we generate 20000 samples. $X(t)$ and $Y(t)$ are simulated by the Gibbs sampling method, similar in scenario 2. The third variable $Z(t)$ is generated separately by assigning the correlation values with variable X(t) according to step

5. The density contour plots at time points t=1 and t=25 are shown in the following figures.

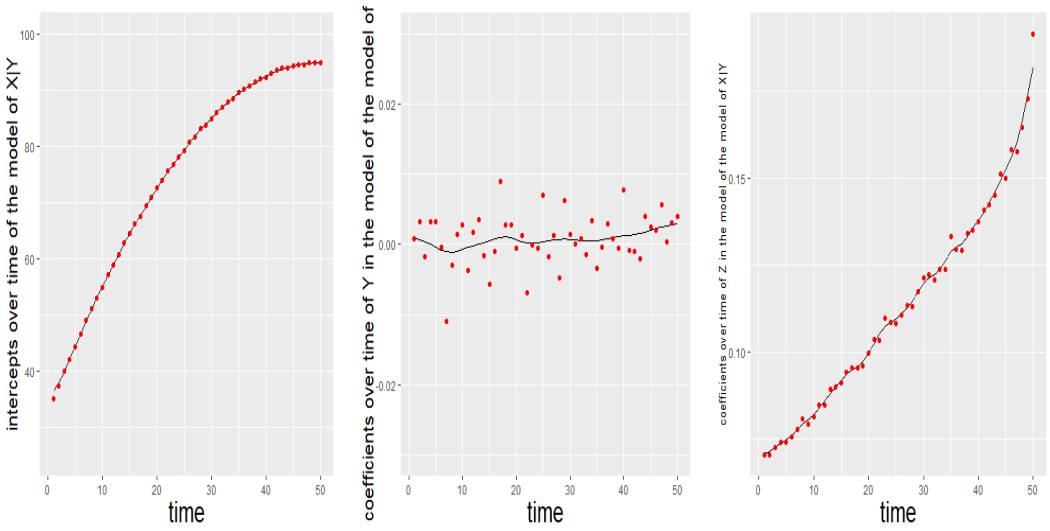


(a) (b)

Figure 5.16: Scenario 3: Density contours for simulated x and y at time point t= 1 and t= 25

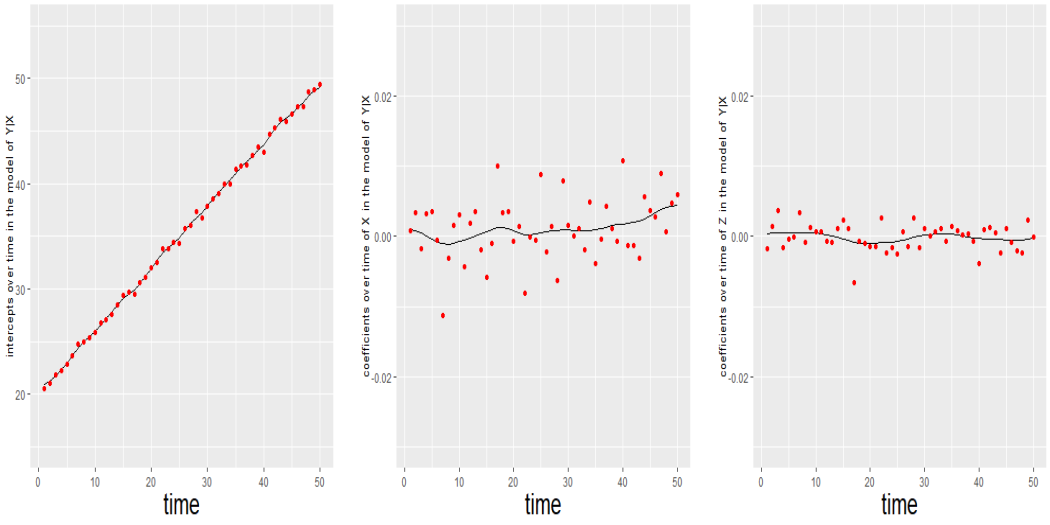
5.4.3.2 Estimation Results for Scenario 3

As in previous scenarios, the samples are generated without any missing data. Again, the estimation steps are the same as we discussed in previous chapters.



(a) (b) (c)

Figure 5.17: Scenario 3: Two-step time-varying coefficients over time for univariate model $X(t) = \alpha_0(t) + \beta_1(t)Y(t) + \beta_2(t)Z(t)$. Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points.

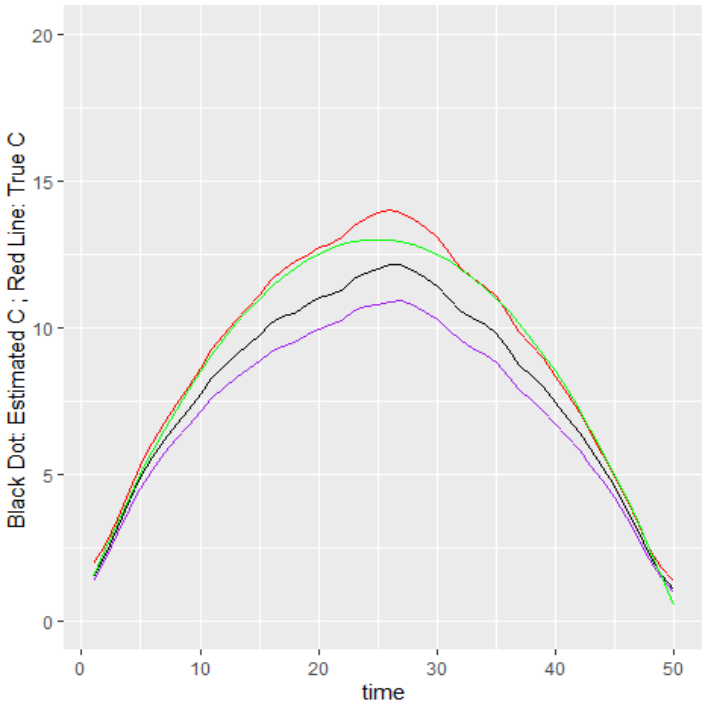


(a) (b) (c)

Figure 5.18: Scenario 3: Two-step time-varying coefficients over time for univariate model $Y(t) = \alpha_1(t) + \gamma_1(t)X(t) + \gamma_2(t)Z(t)$.Black curves: estimated estimated time-varying coefficient curves. Red dots: raw coefficients at different time points.

t	$c(t)$	$\hat{c}(t)$	$\hat{k}(c(t))$
1	1.48	1.493929	1.330134
10	8.5	7.726321	1.856361
20	12.5	11.04196	2.03187
30	12.5	11.39735	2.048764
40	8.5	7.442971	1.839543
50	0.5	1.037703	1.260081

Table 5.7: Scenario 3: Estimated $\hat{c}(t)$ versus True c(t) at different time points

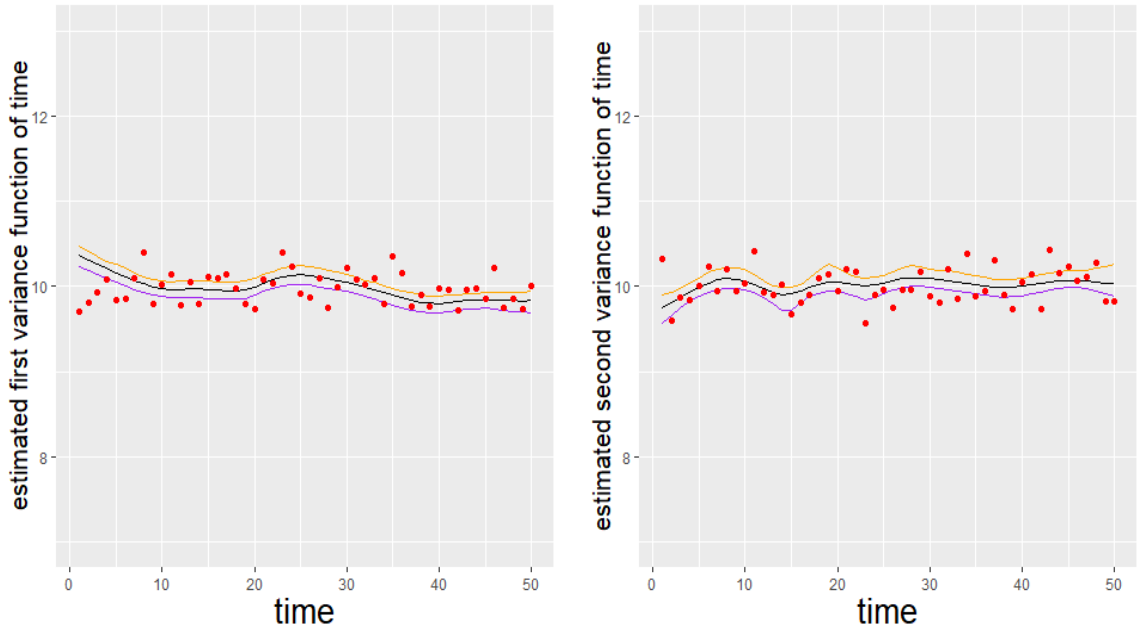


(a)

Figure 5.19: Scenario 3: Estimated dependent parameter c(t). Black curve: estimated c(t), Green curve: true c(t), Red and Purple curves: 95% Bootstrap pointwise confidence intervals

At time points: t=1, 10, 20, 30, 40 and 50, the estimated MME: $\hat{c}(t)$ for scenario 3 are listed in Table 5.7, also a visualized comparison between the $\hat{c}(t)$ and $c(t)$ is shown in

Figure 5.19. The estimated curve has a similar shape of the true setting. Some bias exists especially when t is close to 25, this is caused by the small sample size of our sample data at each time point. Due to the limitation of computational power of CPU, we will consider using HPC to generate large sample data in future studies. The corresponding $\hat{k}(c(t))$ are also calculated.



(a) (b)

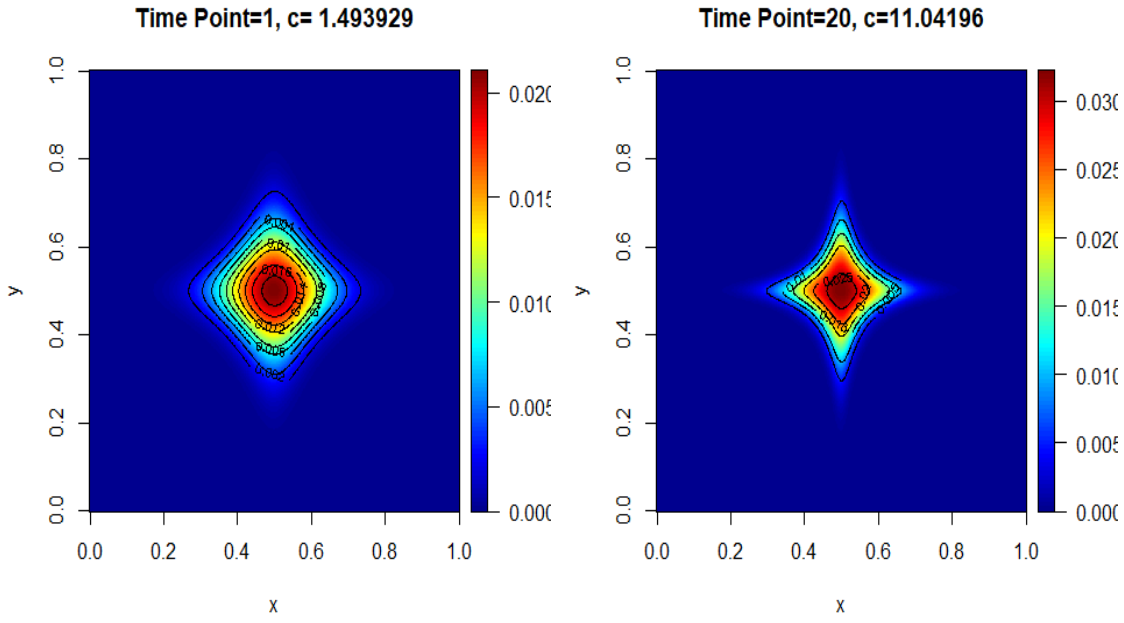
Figure 5.20: Scenario 3: Dynamic MME of variances over time and bootstrap point wise confidence intervals for the X and Y conditioning on the mean value of Z. Black: estimated MME; Purple and Orange: 95% Bootstrap pointwise confidence intervals

t	$\sigma_x^2(t)$	MME of $\hat{\sigma_x^2}(t)$	Bootstrap 95% point wise CI of $\hat{\sigma_x^2}(t)$	$\lvert\hat{\sigma_x^2}(t) - \sigma_x^2(t)\rvert$
1	9.703886	10.364870	(10.233478,10.471943)	0.660984
10	10.026225	9.975733	(9.881174,10.063927)	0.050492
20	9.737145	9.997695	(9.893961,10.10113)	0.26055
30	10.222742	10.046615	(9.946881,10.143845)	0.176127
40	9.982104	9.801399	(9.695031,9.888891)	0.180705
50	10.005500	9.833072	(9.694587,9.945578)	0.172428

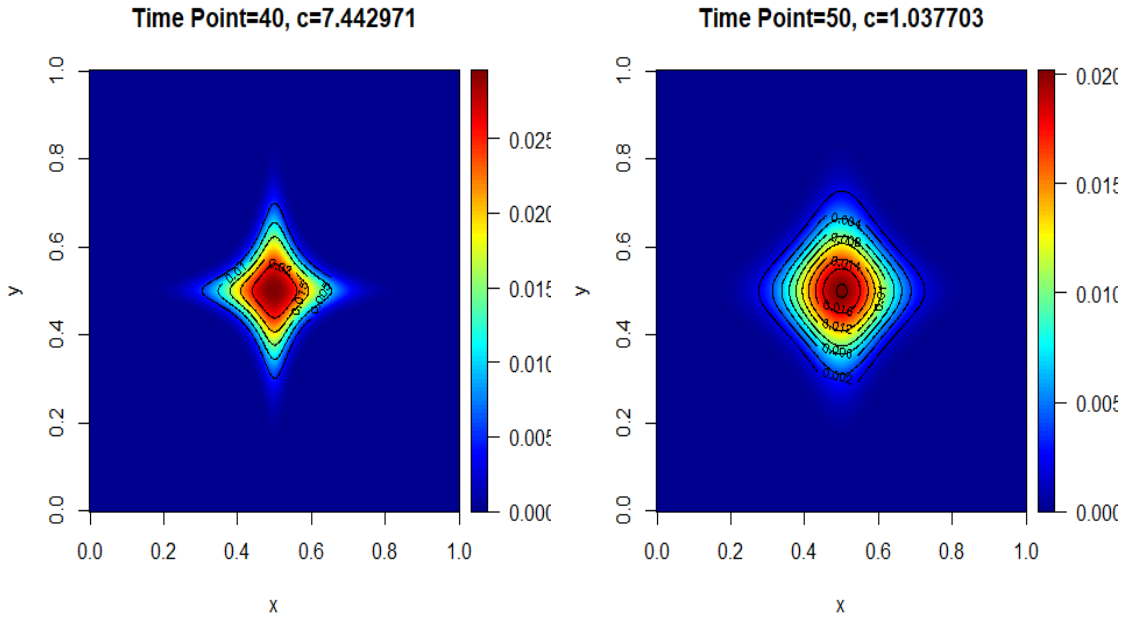
Table 5.8: Scenario 3: Estimated $\hat{\sigma_x^2}(t)$, 95% Bootstrap pointwise confidence intervals based on 500 Bootstrap samples and absolute errors of $\hat{\sigma_x^2}(t)$

t	$\sigma_y^2(t)$	MME of $\hat{\sigma_y^2}(t)$	Bootstrap 95% point wise CI of $\hat{\sigma_y^2}(t)$	$\lvert\hat{\sigma_y^2}(t) - \sigma_y^2(t)\rvert$
1	10.315434	9.742862	(9.567784,9.899037)	0.572572
10	10.029160	10.061346	(9.963132,10.202525)	0.032186
20	9.949992	10.054426	(9.939710,10.224198)	0.104434
30	9.891622	10.093957	(9.987424,10.206085)	0.202335
40	10.051407	10.000473	(9.891027,10.096751)	0.050934
50	9.818517	10.032072	(9.892064,10.265188)	0.213555

Table 5.9: Scenario 3: Estimated $\hat{\sigma_y^2}(t)$, 95% Bootstrap pointwise confidence intervals based on 500 Bootstrap samples and absolute errors of $\hat{\sigma_y^2}(t)$



(a) (b)



(c) (d)

Figure 5.21: Scenario 3: Simulated density contours for x and y at time point t= 1, 20, 40 and 50

In Figure 5.20, Table 5.8 and Table 5.9, the estimated MME variances $\hat{\sigma}_x^2(t)$ and $\hat{\sigma}_y^2(t)$

are compared with the original variance $\sigma_x^2(t)$ and $\sigma_y^2(t)$ for scenario 3. Again, similar boundary bias of the kernel smoothing exists at time t=1 as in scenario 2. However, the absolute errors $|\hat{\sigma}_x^2(t) - \sigma_x^2(t)|$ and $|\hat{\sigma}_y^2(t) - \sigma_y^2(t)|$ listed in the right columns are all < 0.7, demonstrating the precision of the estimation results.

In the contour plots in Figure 5.21, at time point t=1 and t=50, c(t) are close to 1, the density contours are rhombus. When t=20 and 40, c= 11.04 and 7.45, the contours are all cross-shaped. All contour plots are not elliptical, telling that at all the 4 time points, given mean(z(t)), X(t) and Y(t) are conditional normal, but not bivariate normally distributed.

5.5 Results and Discussion

In this section, we explore the feasibility of our approach using simulated data. In scenario 1, we simulate one subject of three covariates from multivariate normal distributions over time, given specified mean and variance functions. At each time point, the multivariate data has a dimension of 800×3. All the three covariates are independent with each other. Hence the c(t) are all equal to 0.

In the scenario 2, we simulate one subject of three covariates from conditional normal distributions by time, also given specified mean and variance functions as in the scenario 1. The first two covariates X(t) and Y(t) are generated by a widely applicable MCMC method: the Gibbs sampling with a linearly distributed dependent variable over time, and the third covariate Z(t) is generated separately of a normal distribution. Note that, the Gibbs sampling is based on large sample theory, we find that until generating up to 300000 samples at each time point.

In the third scenario, different with scenario 2, the dependent variable c(t) over time is a polynomial curve, and the correlations between variable Z(t) and X(t) are linearly increasing over time. Due to the limitation of computational power of the CPU, we only generate 20000 samples at each time point.

Note that in order to minimize the bias, we perform the two-step estimation by un-

dersmoothing in all scenarios. The final estimated dynamic variances of the bivariate distributions over time are all very close to the original parameters, justifying by comparing the absolute difference between estimated and the true parameters. The contour plots also make a visualized demonstration of the joint distributions at different time points. The simulation study demonstrates the effectiveness with good accuracy of our approaches.

Chapter 6: Asymptotic Property

In our thesis study, we are making estimations basically based on two processes. In the first process, we applied the two-step time-varying estimation method to estimate the time-varying coefficients of univariate models. In the second process, we also use two steps to estimate the dynamic parameters over time for joint density distributions. To investigate the asymptotic properties of final estimators at each step, first we look at the consistency property in the first process.

6.1 Consistency of the two-step time-varying estimators of univariate models

In the first step of the two-step estimation methods, we use the MLE estimators as raw estimators which are unbiased estimator of $\beta_l(t_j)$. The asymptotic property is presented in the Appendix A.1 to A.3.

In the second step, we use the kernel smoothing methods. The kernel estimator is also known as the Nadaraya-Watson estimator or the local constant estimator. It is the special case of local polynomial smoother with p=0. Note that in the second step, the problem is actually an univariate smoothing problem and in fact they are the linear combination form of their corresponding raw estimators. And the bias and variance-covariance structure of the estimator does not depend on the correlation structures within subjects over time under some conditions [18, 48]. We apply the bandwidth selection procedures by leave-one-out cross-validation in the second smoothing step.

The consistency properties have been derived and presented by [18] and [48]. The following is the theorem: First, to estimate the coefficients of time based on the raw coefficients $\tilde{\beta}_l(t_j)$, the notation is as following:

$$\tilde{\beta}_l(t) = (\tilde{\beta}_l(t_1), \tilde{\beta}_l(t_2), ...\tilde{\beta}_l(t_j))^T$$

$$b_0(t) = (b_{00}(t), ...b_{k0}(t))^T$$

$$W = diag(K_h(t_1 - t), ...K_h(t_j - t))$$

the weighted least square criteria:

$$J(\beta) = \sum_{j=1}^{J} \left\{ \tilde{\beta}_l(t_j) - b_{l0}(t) \right\}^2 K(\frac{t_j - t}{h})$$

And the second step estimator is:

$$\hat{\beta}_l(t) = \sum_{j=1}^{J} \frac{\tilde{\beta}_l(t_j) K(\frac{t_j - t}{h})}{\sum_{j=1}^{J} K(\frac{t_j - t}{h})}$$

Under the assumptions in Appendix A.1, [18] and [48]. have proved that (Appendix A.4) :

$$\begin{aligned} E(\hat{\beta}_l^{(q)}(t)) &= \sum_{j=1}^{J} w_{q,p+1}(t_j, t)\beta_l(t_j) \\ &= \sum_{j=1}^{J} w_{q,p+1}[\sum_{k=0}^{p+1} \beta_l^k(t) \frac{(t_j - t)^k}{k!} + o(t_j - t)^{p+1}] \\ &= \beta_l^{(q)}(t) + \frac{q!\beta_l^{p+1}(t) h^{p-q+1}}{(p+1)!} B_{p+1}(K_{q,p+1}) 1 + o_p(1)] \end{aligned}$$

Since we are using kernel estimation which is the local constant fit, then set p=0, then $B_{p+1}(K_{q,p+1}) = \int K(u)udu = 0$ so that in the above equations, we will have $E(\hat{\beta}_l^{(q)}(t)) = \beta_l^{(q)}(t)$, let q=0, we have $E(\hat{\beta}_l(t)) = \beta_l(t)$, so that the kernel smoothing estimators are also unbiased. For the asymptotic variance property, under all the assumptions in Appendix A.1, the asymptotic variance are not affected by the covariance structures, following the proved theorem in [48], showing in Appendix A.4, by setting q=0 and p=0 we have:

$$Var[\hat{\beta}_l^{(q)}(t)] = \frac{2(q!)^2}{3cnJh^{2q+1}\pi(t)}\sigma_l^2(t)V(K_{q,p+1})[1+\circ(1)]$$
$$= \frac{2}{3cnJh\pi(t)}\sigma_l^2(t)V(K_{0,1})[1+\circ(1)]$$

where $V(K) = \int K^2(u)du.$ and $\sigma_l^2(t)$ is the l^{th} diagonal element of $\Lambda(t,t)$.

6.2 Consistency of MME for joint models

In the second process, we uses methods of moments estimators as the "pseudo observations" and then making kernel smoothing on them. Remember that the method of moments almost always produce asymptotically unbiased estimators. Then by looking at the asymptotic properties on those kernel smoothing estimators, we can get some insight into the reliability of our estimation procedures.

Let's start with simple model structures, assume we only have three variables, X, Y and Z over time t. Suppose at any time point t, they are absolutely continuous with respect to some product measure $\mu_1 \times \mu_2 \times \mu_3$ on the support set: $A \times B \times C$. Our goal is to estimate the dynamic parameters of joint density distributions of Y_1, Y_2 at different time point t and $t \in \tau$ within a bounded subset $[0, +\infty)$.

Our estimation steps starts from making two-step estimations of the time-varying coefficients in our conditional models of $X(t)|Y(t)$ and $Y(t)|X(t)$:

$$X(t)|Y(t) = \hat{\beta}_{0(t)} + \hat{\beta}_1(t)Y(t) + \hat{\beta}_2(t)Z(t) + e_1(t) \tag{6.1}$$

$$Y(t)|X(t) = \hat{\gamma}_0(t) + \hat{\gamma}_1(t)X(t) + \hat{\gamma}_2(t)Z(t) + e_2(t) \tag{6.2}$$

Since the two-step estimators are unbiased, $e_1(t)$ and $e_2(t)$ follow standard bivariate normal distribution, take the expectations to both sides of the above two models, let's define

$\mu_1(X(t)|Y(t)=y(t))$ and $\mu_2(Y(t)|X(t)=x(t))$ by

$$\begin{aligned}\mu_1(X(t)|Y(t)=y(t)) &= E(X(t)|Y(t)=y(t),Z(t)=z(t)) \\ &= E(\hat{\beta}_0(t)+\hat{\beta}_1(t)y(t)+\hat{\beta}_2(t)z(t)+e_1(t)) \\ &= \beta_0(t)+\beta_1(t)y(t)+\beta_2(t)z(t)\end{aligned} \tag{6.3}$$

the same, we can also have $\mu_2(Y(t)|X(t)=x(t))$. In the centering steps, we will center $X(t)|Y(t)$ and $Y(t)|X(t)$ by subtracting $\mu_1(X(t)|Y(t)$ and $\mu_2(Y(t)|X(t)=x(t))$ from them. Since the centering steps are using two-step smoothing which are all unbiased, we will only take care of the other estimators like dynamic variances of time for the joint density distributions.

6.2.1 Consistency of raw MMEs over time

A set of derivation steps need to be done before getting the final raw estimators. After centering, we have centered X(t) and Y(t) noted as $X'(t)$ and $Y'(t)$. We assume the normal conditionals, according to (Appendix A.3.1), we have :

$$X'(t) \sim N(0, \sigma^2_{x'}(y'(t))) \tag{6.4}$$

$$Y'(t) \sim N(0, \sigma^2_{y'}(x'(t))) \tag{6.5}$$

We have at any time point t, the joint distribution of $f_{X'(t),Y'(t)}(x'(t),y'(t))$ is of this form:

$$k(c(t))\frac{1}{2\pi\sigma_{x'}(t)\sigma_{y'}(t)}exp\{-\frac{1}{2}[(\frac{x'(t)}{\sigma_{x'}(t)})^2+(\frac{y'(t)}{\sigma_{y'}(t)})^2+c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2(\frac{y'(t)}{\sigma_{y'}(t)})^2]\} \tag{6.6}$$

From above equations, we can calculate the moments by the equation $M_{X'^2(t),Y'^2(t)}(s,w) = E\left(e^{sX'^2(t)+wY'^2(t)}\right)$:

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{k(c(t))}{2\pi\sigma_{x'}(t)\sigma_{y'}(t)} exp\{-\frac{1}{2}[(\frac{x'(t)}{\sigma_{x'}(t)})^2 + (\frac{y'(t)}{\sigma_{y'}(t)})^2 + c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2(\frac{y'(t)}{\sigma_{y'}(t)})^2 - 2sX'^2(t) - 2wY'^2(t)]\}dx'(t)dy'(t)$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{k(c(t))}{2\pi\sigma_{x'}(t)\sigma_{y'}(t)} exp\{-\frac{1}{2}[(\frac{1}{\sigma_{x'}^2(t)} - 2s)x'^2(t) + (\frac{1}{\sigma_{y'}^2(t)} - 2w)y'^2(t) + c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2(\frac{y'(t)}{\sigma_{y'}(t)})^2]\}dx'(t)dy'(t)$

in order to guarantee the convergence of the integrand, s and w must satisfy:

$$\frac{1}{\sigma_{x'}^2(t)} - 2s > 0 \tag{6.7}$$

$$\frac{1}{\sigma_{y'}^2(t)} - 2w > 0 \tag{6.8}$$

Substitute some terms in above, let

$$(\frac{1}{\sigma_{x'}^2(t)} - 2s)x'^2(t) = m^2 \tag{6.9}$$

$$(\frac{1}{\sigma_{y'}^2(t)} - 2w)y'^2(t) = k^2 \tag{6.10}$$

so that

$$(\frac{x'^2(t)}{\sigma_{x'}^2}) = \frac{m^2}{1-2s\sigma_{x'}^2}$$
$$(\frac{y'^2(t)}{\sigma_{y'}^2}) = \frac{k^2}{1-2s\sigma_{y'}^2}$$

and

$$dx'(t) = \frac{\sigma_{x'}^2}{(1-2w\sigma_{y'}^2)^{\frac{1}{2}}}dk$$
$$dy'(t) = \frac{\sigma_{y'}^2}{(1-2s\sigma_{x'}^2)^{\frac{1}{2}}}dm$$

then, the above moment functions will be

$$= k(c(t)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_{x'}(t)\sigma_{y'}(t)} exp\{-\frac{1}{2}[m^2+k^2+c(t)\frac{k^2}{1-2w\sigma_{y'}^2(t)}\frac{m^2}{1-2s\sigma_{x'}^2(t)}]\}dx'(t)dy'(t)$$

$$= k(c(t)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi(1-2w\sigma_{y'}^2(t))^{\frac{1}{2}}(1-2s\sigma_{x'}^2(t))^{\frac{1}{2}}} exp\{-\frac{1}{2}[m^2+k^2+c(t)\frac{k^2}{1-2w\sigma_{y'}^2(t)}\frac{m^2}{1-2s\sigma_{x'}^2(t)}]\}dkdm$$

$$= \frac{k(c(t))(1-2w\sigma_{y'}^2(t))^{-\frac{1}{2}}(1-2s\sigma_{x'}^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2w\sigma_{y'}^2(t))^{-1}(1-2s\sigma_{x'}^2(t))^{-1}]}$$

Make the function $\delta(c(t)) = \frac{dlogk(c(t))}{dc(t)} = \frac{k'(c(t))}{k(c(t))}$ and we have derived:

$$M_{X'^2(t),Y'^2(t)}(s,w) = E(e^{sX'^2(t)+wY'^2(t)}) = \frac{k(c(t))(1-2w\sigma_{y'}^2(t))^{-\frac{1}{2}}(1-2s\sigma_{x'}^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2w\sigma_{y'}^2(t))^{-1}(1-2s\sigma_{x'}^2(t))^{-1}]} \quad (6.11)$$

Let $w = 0$, we can get :

$$E(e^{sX'^2(t)}) = \frac{k(c(t))(1-2s\sigma_{x'}^2(t))^{-\frac{1}{2}}}{k[c(t)(1-2s\sigma_{x'}^2(t))^{-1}]} \quad (6.12)$$

Take the derivative with respect to s and set $s = 0$, we can get

$$E(x'^2(t)) = \frac{dE(e^{sX'^2(t)})}{ds}|_{s=0}$$

$$= \frac{\sigma_{x'}^2(t)(1-2s\sigma_{x'}^2(t))^{-\frac{3}{2}}k(c(t))k[c(t)(1-2s\sigma_{x'}^2(t))^{-1} - 2c(t)\sigma_{x'}^2(t)(1-2s\sigma_{x'}^2(t))^{-\frac{5}{2}}k(c(t))k'[c(t)(1-2s\sigma_{x'}^2(t))^{-1}]}{(k[c(t)(1-2s\sigma_{x'}^2(t))^{-1}])^2}|_{s=0}$$

$$= \frac{\sigma_{x'}^2(t)(k(c(t)))^2 - 2c_t\sigma_{x'}^2(t)(k(c(t)))^2\delta(c(t))}{(k(c(t)))^2}$$

$$= \sigma_{x'}^2(t)[1-2c(t)\delta(c(t))]$$

Using the same steps, we can easily get $E(y'^2(t)) = \sigma_{y'}^2(t)[1-2c(t)\delta(c(t))]$. Since we assume all orders of moments exist , $E(X'^n) < \infty$ and $E(Y'^n) < \infty$. By the Strong Law of

large numbers, when n goes to infinity:

$$\overline{X'^2(t)} = \frac{1}{n}\sum_{i=1}^{n} X_i'^2(t) \xrightarrow{\text{a.s.}} \sigma_{x'}^2(t)(1-2c(t)\delta(c(t))) \tag{6.13}$$

$$\overline{Y'^2(t)} = \frac{1}{n}\sum_{i=1}^{n} Y_i'^2(t) \xrightarrow{\text{a.s.}} \sigma_{y'}^2(1-2c(t)\delta(c(t))) \tag{6.14}$$

In 2.2.2, we have already derived the marginal densities of :

$$f_{X'(t)}(x'(t)) = k(c(t))\frac{1}{\sigma_{x'}(t)\sqrt{2\pi}}[1+c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{x'(t)}{\sigma_{x'}(t)})^2] \tag{6.15}$$

$$f_{Y'(t)}(y'(t)) = k(c(t))\frac{1}{\sigma_{y'}(t)\sqrt{2\pi}}[1+c(t)(\frac{y'(t)}{\sigma_{y'}(t)})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{y'(t)}{\sigma_{y'}(t)})^2] \tag{6.16}$$

The n^{th} order moments can be derived by

$$E((X'(t))^n) = \int_{-\infty}^{\infty} k(c(t))\frac{(x'(t))^n}{\sigma_{x'}(t)\sqrt{2\pi}}[1+c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{x'(t)}{\sigma_{x'}(t)})^2] \tag{6.17}$$

$$E((Y'(t))^n) = \int_{-\infty}^{\infty} k(c(t))\frac{(y'(t))^n}{\sigma_{y'}(t)\sqrt{2\pi}}[1+c(t)(\frac{y'(t)}{\sigma_{y'}(t)})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{y'(t)}{\sigma_{y'}(t)})^2] \tag{6.18}$$

When n is an odds number, then the expectation equals to 0. When n is an even number, we can write it as:

$$E((X'(t))^n) = \int_{0}^{\infty} 2k(c(t))\frac{(x'(t))^n}{\sigma_{x'}(t)\sqrt{2\pi}}[1+c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2]^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\frac{x'(t)}{\sigma_{x'}(t)})^2]\ \}, \tag{6.19}$$

and let $w = c(t)(\frac{x'(t)}{\sigma_{x'}(t)})^2$ and $z = \frac{1}{2c(t)}$, then $dx'(t) = \frac{\sigma_{x'}^2(t)}{2c(t)x'(t)}dw$, by substituting, we can have :

$$E((X'(t))^n) = \int_{0}^{\infty} \frac{k(c(t))}{2\pi}\sigma_{x'}^n(t)c^{-\frac{(n+1)}{2}}w^{\frac{n}{2}}(1+w)^{-\frac{1}{2}} exp\{-wz\}dw \tag{6.20}$$

$$= \frac{k(c(t))}{2\pi} c^{-\frac{(n+1)}{2}}(t)\sigma_{x'}^{n}(t)\int_0^{\infty} w^{\frac{n}{2}}(1+w)^{-\frac{1}{2}} exp\{-wz\}dw \tag{6.21}$$

because $\frac{1}{\Gamma(\frac{n+1}{2})}\int_0^{\infty} w^{\frac{n}{2}}(1+w)^{-\frac{1}{2}} exp\{-wz\}dw$ is a density function of the confluent hypergeometric function : $U(\frac{n+1}{2}, \frac{n}{2}+1, \frac{1}{2c(t)})$, so it can be writen:

$$E((X'(t))^n) = \frac{k(c(t))}{2\pi} c^{-\frac{(n+1)}{2}}(t)\sigma_{x'}^{n}(t)\Gamma(\frac{n+1}{2})U(\frac{n+1}{2}, \frac{n}{2}+1, \frac{1}{2c(t)}) \tag{6.22}$$

By using the recurrence properties of the confluent bypergeometric function [1, 2], we can have

$$\frac{\sigma_{x'}^2(t)(n-1)}{c(t)} E((X'(t))^{n-2}) = (\frac{1}{c(t)} - n)E((X'(t))^n) + \frac{1}{\sigma_{x'}^2(t)} E((X'(t))^{n+2}) \tag{6.23}$$

where n is an even number. According to the theorem, let $n = 4$:

$$\frac{\sigma_{x'}^2(t)}{c(t)} - (\frac{1}{c(t)} - 2)E((X'(t))^2) = \frac{1}{\sigma_{x'}^2(t)} E((X'(t))^4)$$

$$E((X'(t))^4) = \frac{\sigma_{x'}^4(t)}{c(t)} - (\frac{1}{c(t)} - 2)\sigma_{x'}^4(t)[1 - 2c(t)\delta(c(t))]$$

since:

$$\begin{aligned} &V((X'(t))^2) = E((X'(t))^4) - (E(X'(t))^2))^2 \\ &= \frac{\sigma_{x'}^4(t)}{c(t)} - (\frac{1}{c(t)} - 2)\sigma_{x'}^4(t)[1 - 2c(t)\delta(c(t))] - \sigma_{x'}^4(t)[1 - 2c(t)\delta(c(t))]^2 \\ &= \sigma_{x'}^4(t) - 4\sigma_{x'}^4(t)c^2(t)\delta^2(c(t)) + 2\sigma_{x'}^4(t)\delta(c(t)) \end{aligned}$$

Also we can get: $V((Y'(t))^2)$. So, in summary:

$$V((X'(t))^2) = \sigma_{x'}^4(t) - 4\sigma_{x'}^4(t)c^2(t)\delta^2(c(t)) + 2\sigma_{x'}^4(t)\delta(c(t))$$

$$V((Y'(t))^2) = \sigma_{y'}^4(t) - 4\sigma_{y'}^4(t)c^2(t)\delta^2(c(t)) + 2\sigma_{y'}^4(t)\delta(c(t))$$

The $E((X'(t))^2(Y'(t))^2)$ can be derived according to [2], detail steps is in the Appendix. We can get

$$E((X'(t))^2(Y'(t))^2) = 2\sigma_{x'}^2(t)\sigma_{y'}^2(t)\delta(c(t))$$

According to the Rules for Probability Limits [29]. If x_n and y_n are random variables with plim $x_n = c$ and plim $y_n = d$, then

$$plim(x_n + y_n) = c + d, (sum\ rule)$$
$$plim(x_n y_n) = cd, (product\ rule)$$
$$plim(\frac{x_n}{y_n}) = \frac{c}{d}, (ratio\ rule)$$

Hence according to all above, and by the strong law of large sample, when n goes to infinity:

$$\overline{X'^2(t)} = \frac{1}{n}\sum_{i=1}^{n} X_i'^2(t) \xrightarrow{\text{a.s.}} \sigma_{x'}^2(t)(1 - 2c(t)\delta(c(t))) \tag{6.24}$$

$$\overline{Y'^2(t)} = \frac{1}{n}\sum_{i=1}^{n} Y_i'^2(t) \xrightarrow{\text{a.s.}} \sigma_{y'}^2(1 - 2c(t)\delta(c(t))) \tag{6.25}$$

$$S_{X'^2(t)}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i'^2(t) - \overline{X'^2(t)})^2 \xrightarrow{\text{a.s.}} \sigma_{x'}^4(t) - 4\sigma_{x'}^4(t)c^2(t)\delta^2(c(t)) + 2\sigma_{x'}^4(t)\delta(c(t))$$
$$S_{Y'^2(t)}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i'^2(t) - \overline{Y'^2(t)})^2 \xrightarrow{\text{a.s.}} \sigma_{y'}^4(t) - 4\sigma_{y'}^4(t)c^2(t)\delta^2(c(t)) + 2\sigma_{y'}^4(t)\delta(c(t))$$
$$S_{X'^2(t)Y'^2(t)}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i'^2(t) - \overline{X'^2(t)})(Y_i'^2(t) - \overline{Y'^2(t)}) \xrightarrow{\text{a.s.}} \sigma_{y'}^2(1 - 2c(t)\delta(c(t)))$$

and by the product and ratio rules for probability limits, we can get :

$$T = \frac{S_{X'^2(t)Y'^2(t)}^2}{\overline{X'^2(t)Y'^2(t)}} \xrightarrow{\text{a.s.}} \frac{1 + 2\delta(c(t)) - 4c^2(t)\delta^2(c(t))}{(1 - 2c(t)\delta(c(t))^2}$$

$$R = \frac{S^2_{X'^2(t)Y'^2(t)}}{S^2_{X'^2(t)} S^2_{Y'^2(t)}} \xrightarrow{\text{a.s.}} \frac{1-2\delta-4c(t)\delta(c(t))+4c(t)^2\delta^2(c(t))}{-1-2\delta(c(t))+4c(t)^2\delta^2(c(t))} \tag{6.26}$$

to get the strongly almost surely consistent estimators at time point t, we can solve above equations by equating the above statistics to the almost limits, then the following are obtained:

$$\hat{c}(t) = \frac{1}{4}[1+RT-2T-\frac{1-T^2}{1+RT}]$$

$$\hat{\sigma}^2_{x'}(t) = \frac{\overline{X'^2(t)}}{1-2\hat{c}(t)\hat{\delta}(c(t))}$$

$$\hat{\sigma}^2_{y'}(t) = \frac{\overline{Y'^2(t)}}{1-2\hat{c}(t)\hat{\delta}(c(t))}$$

6.2.2 Consistency of dynamic MMEs over time

In this step, we treat the raw MMEs as pseudo observations and we also use kernel smoothing to make the dynamic MMEs. The consistency property should be the same as the same as in 6.1.

Chapter 7: Summary and Future Work

In our thesis project, we proposed a new method to study the dynamic joint density distributions over time for multivariate longitudinal data. To investigate both population mean based and sub-population based characteristics, for example, if we want to study individuals at higher risk of CVD with abnormal high blood pressure levels SBP and DBP simultaneously, we devise a model scheme which allows us incorporate both the conditional mean and conditional distribution information. To avoid the limitation of previous direct estimation on two-dimensional response variable, such as Copula and joint kernel estimations, we build the models by a conditional specified approach and our estimation steps integrate both mathematical derivations of joint density distributions from a general family and the nonparametric smoothing technique: the two-step time-varying estimation. The two-step time-varying estimation fits a smoothed curve through the scatter points to best represent the dynamic trend of varying coefficients over time. It is consist of two steps. The first step is attempt to generate "raw estimators" which keeps track of the shapes and degrees of smoothness of coefficients over time. And the second step is accomplished by kernel smoothing on those raw estimators to get the dynamic coefficients functions over time.

As longitudinal study is often designed to make records for subjects at fixed time points and may result in significant loss of observations due to various reasons. We used the time point binning approach to group observations within a small neighborhood of some time points to reduce the effects of minor observation errors. To fit the assumption better, we also making data transformation on the outcome variables when necessary.

Both the applications in FHS and NGHS demonstrate the feasibility of our approach and provide us useful insights in the two studies. To access the generalizability and evaluate the estimation performance of our methods, we examined different simulation scenarios in the simulation study. All estimation results have great accuracy and demonstrate the

practicability and effectiveness or our approach.

However, there are many limitations in our study and further work needed to be continued in future. First, the analysis data in FHS and NGHS used in our study are of revised versions. We exclude all the records with missing data which might lead to biased estimators. In real world longitudinal studies, the analysis is often impeded by the presence of missing data and hindered by the unknown missing data mechanism such as missing completely at random (MCAR) and the missing at random (MAR). Multiple Imputation (MI) and the Expectation-Maximization Algorithm (EM) may be considered for analyzing the multivariate longitudinal data.

Second, our model assumption is joint distribution of conditional normal. Although conditional normal can be easily checked or be achievable after data preprocessing such us making data transformations, for example, the log-transformation, the square root and the Box-Cox transformations, we may need to consider more general family of distributions such as the exponential family. Third, we used kernel smoothing to estimate the smoothed curves of unknown parameters. One problem of kernel smoothing is that they are subject to boundary bias, which can be clearly seen in our simulation study results (Figure 5.6(b)). Hence, the kernel smoothing with adaptive bandwidth selection or local polynomial smoothing can be tried to avoid the problem in future work.

Forth, in the simulation, we devised three different scenarios, we included dynamic correlation structure between the bivariate response variables (X(t) and Y(t)) and also between the response variable and the third covariate (X(t) and Z(t)). However, in practice, more covariates and more complex correlation structures should be considered. Also due to the limitation of the computation power of CPU, the empirical coverage probabilities (CP) of the 95% bootstrap confidence intervals have not been calculated and the Gibbs sampling algorithm was used but only 20000 samples were generated at each time point in the third simulation scenario. Since the Gibbs sampling and the CP calculation are based on large sample theory, the small sample size may lower accuracy of our estimation result. We would

like to try the computing on GPUs for HPC later. Last but not least, there are also limitations of our current consistency results. Our joint density estimators are estimated by the method of moment and the consistency property of those MME estimators are directly ensured by the strong law of large sample and the rules for probability limits. However, the estimators have very complicate structures, we haven't assessed the other asymptotic properties such as the converging rate, limiting distributions etc. These asymptotic behavior should also be studied in our future work.

Bibliography

[1] M Stegun Abramowitz and IA Stegun. Ia.(1964), handbook of mathematical functions. *Washington: National Bureau of Standards*, page 923, 1999.

[2] Barry C Arnold, Enrique Castillo, and Jose Maria Sarabia. Conditionally specified distributions: An introduction (with comments and a rejoinder by the authors). *Statistical Science*, 16(3):249–274, 2001.

[3] Barry C Arnold, Enrique Castillo, José-Mariá Sarabia, and Jose M Sarabia. *Conditional Specification of Statistical Models*. Springer Science & Business Media, 1999.

[4] Barry C Arnold and S James Press. Compatible conditional distributions. *Journal of the American Statistical Association*, 84(405):152–156, 1989.

[5] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

[6] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[7] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[8] Leo Breiman. Probability. *Addison Wesley Series in Statistics, Reading, Mass.: Addison Wesley, 1968*, 1968.

[9] Laurent Briollais, Anjela Tzontcheva, and Shelley Bull. Multilevel modeling for the analysis of longitudinal blood pressure data in the framingham heart study pedigrees. In *BMC Genetics*, volume 4, page S19. Springer, 2003.

[10] Zongwu Cai and Xiaoping Xu. Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608, 2008.

[11] Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

[12] E Castillo and J Galambos. Conditional distributions and the bivariate normal distribution. *Metrika*, 36(1):209–214, 1989.

[13] NGHS Coordinating Center. Nhlbi growth and health study (nghs) data monitoring report. *Baltimore: Maryland Medical Research*, 1998.

[14] Mohammed Chowdhury, Colin Wu, and Reza Modarres. Local Box-Cox transformation on time-varying parametric models for smoothing estimation of conditional CDF with longitudinal data. *Journal of Statistical Computation and Simulation*, 87(15):2900–2914, 2017.

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

[16] Jianqing Fan. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Routledge, 2018.

[17] Jianqing Fan and J-T Zhang. Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):303–322, 2000.

[18] Jianqing Fan and J-T Zhang. Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):303–322, 2000.

[19] Stanley S Franklin, Shehzad A Khan, Nathan D Wong, Martin G Larson, and Daniel Levy. Is pulse pressure useful in predicting risk for coronary heart disease? the framingham heart study. *Circulation*, 100(4):354–360, 1999.

[20] Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[21] Andrew Gelman and TP Speed. Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):185–188, 1993.

[22] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *Journal of Applied Statistics*, 20(5-6):25–62, 1993.

[23] Christian Gourieroux and Alain Monfort. *Statistics and Econometric Models*, volume 1. Cambridge University Press, 1995.

[24] Arsalane Chouaib Guidoum. Kernel estimator and bandwidth selection for density and its derivatives. *The Kedd Package, Version*, 1, 2015.

[25] James Hobert and George Casella. The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473, 1996.

[26] James P Hobert and George Casella. The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473, 1996.

[27] Jianhua Z. Huang, Colin O. Wu, and Lan Zhou. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128, 2002.

[28] Jianhua Z. Huang, Colin O. Wu, and Lan Zhou. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788, 2004.

[29] David R Hunter. Notes for a graduate-level course in asymptotics for statisticians. *Penn State University, Pennsylvania*, 2014.

[30] Galin L Jones, James P Hobert, et al. Sufficient burn-in for gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2):784–817, 2004.

[31] Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004.

[32] John E Kolassa et al. Convergence and accuracy of gibbs sampling for conditional distributions in generalized linear models. *The Annals of Statistics*, 27(1):129–142, 1999.

[33] Peter Kraft, Lara Bauman, Jin Ying Yuan, and Steve Horvath. Multivariate variance-components analysis of longitudinal blood pressure measurements from the framingham heart study. In *BMC Genetics*, volume 4, page S55. Springer, 2003.

[34] David P Mason, Jeevanantham Rajeswaran, Sudish C Murthy, Ann M McNeill, Marie M Budev, Atul C Mehta, Gösta B Pettersson, and Eugene H Blackstone. Spirometry after transplantation: How much better are two lungs than one? *The Annals of Thoracic Surgery*, 85(4):1193–1201, 2008.

[35] Sean P Meyn and Richard L Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[36] Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2007.

[37] Esa Nummelin. *General Irreducible Markov Chains and Non-negative Operators*, volume 83. Cambridge University Press, 2004.

[38] Eva Obarzanek, Colin O Wu, Jeffrey A Cutler, Rae-Ellen W Kavey, Gail D Pearson, and Stephen R Daniels. Prevalence and incidence of hypertension in adolescent girls. *The Journal of Pediatrics*, 157(3):461–467, 2010.

[39] Prakash Panangaden. The category of markov kernels. *Electronic Notes in Theoretical Computer Science*, 22:171–187, 1999.

[40] Michael K Pitt and Neil Shephard. Analytic convergence rates and parameterization issues for the gibbs sampler applied to state space models. *Journal of Time Series Analysis*, 20(1):63–85, 1999.

[41] Jeffrey S Simonoff. *Smoothing Methods in Statistics*. Springer Science & Business Media, 2012.

[42] Kesar Singh, Minge Xie, William E Strawderman, et al. Confidence distribution (cd)–distribution estimator of a parameter. In *Complex Datasets and Inverse Problems*, pages 132–150. Institute of Mathematical Statistics, 2007.

[43] Geert Verbeke, Steffen Fieuws, Geert Molenberghs, and Marie Davidian. The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, 23(1):42–59, 2014.

[44] Lie Wang, Lawrence D Brown, T Tony Cai, et al. A difference based approach to the semiparametric partial linear model. *Electronic Journal of Statistics*, 5:619–641, 2011.

[45] Claudia Werthenbach and Eva Herrmann. A fast and stable updating algorithm for bivariate nonparametric curve estimation. *Journal of Computational and Graphical Statistics*, 7(1):61–76, 1998.

[46] Colin O. Wu. A cross-validation bandwidth choice for kernel density estimates with selection biased data. *Journal of Multivariate Analysis*, 61(1):38–60, 1997.

[47] Colin O Wu and Xin Tian. Nonparametric estimation of conditional distributions and rank-tracking probabilities with time-varying transformation models in longitudinal studies. *Journal of the American Statistical Association*, 108(503):971–982, 2013.

[48] Colin O Wu and Xin Tian. *Nonparametric Models for Longitudinal Data: With Implementation in R*. Chapman and Hall/CRC, 2018.

[49] Colin O Wu, Xin Tian, and Jarvis Yu. Nonparametric estimation for time-varying transformation models with longitudinal data. *Journal of Nonparametric Statistics*, 22(2):133–147, 2010.

[50] Xiaoqin Xiong and Joel A Dubin. A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statistics in Medicine*, 29(18):1919–1931, 2010.

Appendix A: Appendix

A.1 Asymptotic Assumptions for Two-step Time-varying Estimators

To evaluate the asymptotic expression for bias and variance of Two-step Time-varying Estimators, we need several theorems proved by [48] list below. In this thesis, we impose the following assumptions.

A_1. The number of total subjects $n \to \infty$ and the number of subjects at time point t_j, $n_j \to \infty$, where the time points $t_j \in \{t_1, t_2, ...t_J\}$ and J is the total number of time points.

A_2. The kernel bandwidth $h \to 0$, $n^{\frac{1}{2}}h \to \infty$, $nJh \to \infty$ and $\frac{nh}{Jn_{j_1 j_2}} \to \infty$.

A_3. The coefficient functions $\beta_l(t)$ are (q+1) times continuously differentiable with respect to t.

A_4. The kernel function K(.) is a bounded symmetric probability density function with a bounded support.

A_5. The time points $t_1, t_2, ...t_J$ are a random sample from the probability density f and t is a continuous point of f in the interior of the support of f.

A_6. The covariance matrix $\Lambda(t_j)$ is nonsingular and the $(p_1, p_2)^{th}$ element of the $\Sigma_{p_1,p_2}(t_{j_1}, t_{j_2})$ are $(q+1)$ times continuously differentiable for both t_{j_1} and t_{j_2} for some q.

A.1.1 Notations

$$A(t_j) = (A_1(t_j), ...A_p(t_j))^T = [n_j(n_j-1)]^{-1}\Lambda^{-1}(t_j)\sigma_{i_1 \neq i_2 \in \tau_j} U_{i_1,i_2}[\beta(t_j)]$$

$$\Lambda(t_j) = -E(\frac{\partial U_{i_1,i_2}[\beta(t_j)]}{\partial[\beta(t_j)]})$$

$$U_{i_1,i_2}[\tilde{\beta}(t_j)] = [X^T_{i_1,i_2}(t_j)\tilde{\beta}(t_j)]X_{i_1,i_2}(t_j)\Delta_{i_1,i_2}[X_{i_1,i_2}(t_j)]$$

$$Z_{i_1,i_2}(t_j) = 1_{[Y_{i_1}(t_j) \geqslant Y_{i_2}(t_j)]}$$

$$X_{i_1,i_2}(t_j) = X_{i_1}(t_j) - X_{i_2}(t_j)$$

$$\Delta_{i_1,i_2}[X_{i_1,i_2}(t_j)] = Z_{i_1,i_2}(t_j) - \xi[X^T_{i_1,i_2}(t_j)\beta(t_j)]$$

$$\xi(s) = \int_{-\infty}^{\infty}[1 - G(t+s)[dG(t), where\, G(.) = 1 - g^{-1}(.)$$

$$\Sigma(t_{j_1}, t_{j_2}) = \Lambda^{-1}(t_{j_1})\rho(t_{j_1}, t_{j_2})[\Lambda^{-1}(t_{j_2})]^T$$

$$r_{j_1,j_2} = [\frac{4n_{j_1 j_2}}{n_{j_1}n_{j_2}(n_{j_1}-1)(n_{j_2}-1)}]\{(n_{j_1} - n_{j_1 j_2})(n_{j_2} - n_{j_1 j_2}) + \frac{1}{2}(n_{j_1 j_2} - 1)[(n_{j_1} + n_{j_2} - 2n_{j_1 j_2}) + \frac{1}{3}(n_{j_1 j_2} - 2)]\}$$

$$V(K) = \int K^2(u)du$$

$$B_{p+1}(K) = \int K(u)u^{p+1}du$$

A.2 Asymptotic properties of the raw estimators in the two-step time-varying estimation

Under the assumption A_1, for $p = 1, 2, ...P$ and fixed t_j, the asymptotic bias of $\tilde{\beta}_p(t_j)$ is 0 and the asymptotic variance $Var(\tilde{\beta}_p(t_j)) = \frac{2}{3n_j}\sigma^2_p(t_j)[1 + \circ_p(1)] \to 0$ when $n_j \to \infty$, and when $n_{j_1} \to \infty, n_{j_2} \to \infty$ and $n_{j_1,j_2} \to \infty$, the covariance matrix $r_{j_1,j_2}\Sigma(t_{j_1}, t_{j_2})$ goes to zero in probability but the converging rate depends on how fast the $n_{j_1} \to \infty$, $n_{j_2} \to \infty$ and $n_{j_1 j_2} \to \infty$, all the above are proved by [48] :

A.3 Asymptotic properties of the smoothing estimators in the two-step time-varying estimation

Under the assumption all the assumptions above, proved by [18], [48] , the following are the asymptotic bias and variance:

$$B[\hat{\beta}_l^{(q)}(t)] = \frac{h^{p-q+1}(q!)}{(p+1)!}\beta_l^{p+1}(t)B_{p+1}(K_{q,p+1})[1+\circ(1)]$$
$$Var[\hat{\beta}_l^{(q)}(t)|X] = \frac{2(q!)^2}{3cnJh^{2q+1}f(t)}\sigma_l^2(t)V(K_{q,p+1})[1+\circ(1)]$$

A.4 Distribution Theory for Joint Models

A.4.1 Assumptions of Error Terms in Condition Models

Definition 1 [2] A nonlinear normal standard error model ia a model involving functions $\mu_1(.), \mu_2(.), \sigma_1(.), \sigma_2(.)$ such that:

$$\frac{X_1 - \mu_1(X_2)}{\sigma_1(X_2)} \stackrel{d}{=} \varepsilon_1$$
$$\frac{X_2 - \mu_2(X_1)}{\sigma_2(X_1)} \stackrel{d}{=} \varepsilon_2$$

where $(\varepsilon_1, \varepsilon_2)$ has a bivariate normal distribution (perhaps assuming that $(\varepsilon_1, \varepsilon_2)$ are independent standard normal variables). This is the general normal standard error model. And although the classical bivariate normals setting of $(\varepsilon_1, \varepsilon_2)$, the density of X_1, X_2 does not generally have normal marginals or conditionals.

A.4.2 Moment Generation Functions

$$\frac{Y_{1t}}{\sigma_{1t}}[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]^{-\frac{1}{2}} \sim N(0,1)$$
$$\frac{Y_{2t}}{\sigma_{2t}}[1+c_t(\frac{y_{1t}}{\sigma_{1t}})^2]^{-\frac{1}{2}} \sim N(0,1)$$

Notice that the the two variables left hand side are independent of Y_{1t}, Y_{2t}. Then the following relations among the moments of Y_{1t} *and* Y_{2t} can be obtained:

$$E\{(\frac{y_{1t}}{\sigma_{1t}})^n[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]^{\frac{n}{2}}\} = E\{(\frac{y_{2t}}{\sigma_{2t}})^n[1+c_t(\frac{y_{1t}}{\sigma_{1t}})^2]^{\frac{n}{2}}\} = E(Z^n)$$

where for n even $E(Z^n) = (n-1)(n-3)...1$, since Z denotes an $N(0,1)$ random variable. Because the random variable $Z = (\frac{Y_{1t}}{\sigma_{1t}})(1+c_t(\frac{Y_{2t}}{\sigma_{2t}})^2)^{\frac{1}{2}}$ is $N(0,1)$ and independent of Y_{2t}, if $V = \sigma_{1t}(1+c_t(\frac{Y_{2t}}{\sigma_{2t}})^2)^{\frac{1}{2}}$, it follows that:

$$E(Z^nV^n) = E(Z^n)E(V^n) = E((\frac{y_{1t}}{\sigma_{1t}})^n)[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]^{\frac{n}{2}}\sigma_{1t}^n(1+c_t(\frac{Y_{2t}}{\sigma_{2t}})^2)^{\frac{-n}{2}}) = E[(\frac{y_{1t}}{\sigma_{1t}})^n]$$

Hence:

$$\frac{E[(\frac{y_{1t}}{\sigma_{1t}})^n]}{E\{[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]^{\frac{-n}{2}}\}} = \frac{E[(\frac{y_{2t}}{\sigma_{2t}})^n]}{E\{[1+c_t(\frac{y_{1t}}{\sigma_{1t}})^2]^{\frac{-n}{2}}\}} = E(Z^n)$$

By the following theory [2]: According to the equation:

$$E\{(\frac{y_{1t}}{\sigma_{1t}})^n[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]^{\frac{n}{2}}\} = E\{(\frac{y_{2t}}{\sigma_{2t}})^n[1+c_t(\frac{y_{1t}}{\sigma_{1t}})^2]^{\frac{n}{2}}\} = E(Z^n)$$

let $n = 2$, we can get:

$$E\{(\frac{y_{1t}}{\sigma_{1t}})^2[1+c_t(\frac{y_{2t}}{\sigma_{2t}})^2]\} = E(Z^2) = 1$$
$$E(y_{1t}^2) + \frac{c_t}{\sigma_{2t}^2}E(y_{1t}^2 y_{2t}^2) = \sigma_{1t}^2$$
$$E(y_{1t}^2 y_{2t}^2) = 2\sigma_{1t}^2\sigma_{2t}^2\delta(c)$$

A.4.3 Parameter k and δ

As an alternative, the values of k and δ can be expressed as [2] :

$$k^{-1}(c) = E\{(1+cZ^2)^{-\frac{1}{2}}\}$$
$$\delta(c) = \frac{1}{2}\frac{E\{Z^2(1+cZ^2)^{-\frac{3}{2}}\}}{E\{(1+cZ^2)^{-\frac{1}{2}}\}}$$

where Z follows a standard normal distribution. The expressions allow straightforward numerical approximation using draws from an standard normal distribution. $k(c)$ is a monotone increasing function of c, and $\delta(c)$ is a decreasing function of c, and both of them are all positive. And from above, let the n equals to 2, we have

$$\begin{aligned} E(Y_1'^2(t)) &= \sigma_1^2(t)[1-2c(t)\delta_c(t)] \\ E(Y_2'^2(t)) &= \sigma_2^2(t)[1-2c(t)\delta_c(t)] \end{aligned} \tag{A.1}$$

and from $\frac{E[(\frac{y_1(t)}{\sigma_1(t)})^2]}{E\{[1+c(t)(\frac{y_2(t)}{\sigma_2(t)})^2]^{\frac{-2}{2}}\}} = \frac{E[(\frac{y_2(t)}{\sigma_2(t)})^2]}{E\{[1+c(t)(\frac{y_1(t)}{\sigma_1(t)})^2]^{\frac{-2}{2}}\}} = E(Z^2) = 1$, we have

$$E\{(\frac{Y_1'(t)}{\sigma_1(t)})^2[1+c(t)(\frac{Y_2'(t)}{\sigma_2(t)})^2]\} = E(Z^2) = 1$$
$$E(Y_1'^2(t)) + \frac{c(t)}{\sigma_2^2(t)}E(Y_1'^2(t)Y_2^2(t)) = \sigma_1^2(t) \tag{A.2}$$
$$E(Y_1'^2(t)Y_2'^2(t)) = 2\sigma_1^2(t)\sigma_2^2(t)\delta(c)$$