A SMOOTHING MODEL AND ITS ASYMPTOTICS WITH APPLICATIONS TO HEALTH STUDIES AND SOCIAL RESEARCH

A Dissertation Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Shujiao Huang

December 2016

A SMOOTHING MODEL AND ITS ASYMPTOTICS WITH APPLICATIONS TO HEALTH STUDIES AND SOCIAL RESEARCH

Shujiao Huang

APPROVED:

Dr. Wenjiang Fu, Chairman
Department of Mathematics, University of Houston

Dr. Robert Azencott
Department of Mathematics, University of Houston

Dr. Manos Papadakis
Department of Mathematics, University of Houston

Dr. Yipeng Yang
Department of Mathematics and Statistics,
University of Houston-Clear Lake

Dean, College of Natural Sciences and Mathematics

Acknowledgements

I wish to acknowledge my advisor Dr. Wenjiang Fu for his excellent guidance. Also, I would like to thank Dr. Robert Azencott, Dr. Manos Papadakis, and Dr. Yipeng Yang for serving on my committee and their valuable comments. Lastly, I would like to thank my family for all their love and encouragement.

A SMOOTHING MODEL AND ITS ASYMPTOTICS WITH APPLICATIONS TO HEALTH STUDIES AND SOCIAL RESEARCH

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Mathematics
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Shujiao Huang
December 2016

Abstract

Smoothing is a data-driven technique in statistical modeling. It has many desirable properties, and can be applied to modeling complex data. In this dissertation, a smoothing cohort model is considered as an effective alternative to address the identifiability problem in age-period-cohort analysis, in which multiple estimators are induced by a linear dependence of covariates: Period - Age = Cohort in the regression model of APC analysis. The smoothing cohort model yields consistent estimation of age and period effects, but cohort effect estimation is biased. Hence, the second stage model aims to correct the bias by setting a constraint using the consistent estimation of age or period effect from the first stage. Selection of constraints in the second stage is studied through simulations. The large sample behavior of the model parameter estimation is examined. The method is applied to cancer-incidence rate, mortality rate, and homicide-arrest rate data and yields sensible trend estimation in age, period, and cohort.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Background and Motivation

Disease or event rates, such as the incidence or mortality rate of chronic diseases (cancer, cardiovascular diseases, obesity, or diabetes, etc), homicide, suicide, and death rate, are important measure of public health for disease monitoring and health-program evaluation. [Coleman et al., 1993] pointed out that "Cancer mortality has been widely accepted as the most important measure of progress against cancer, since it reflects the impact of cancer on people, and has been considered less subject to distortion than incidence or survival, although this is open to question. Cancer mortality also reflects trends in incidence and survival to a greater or lesser extent."

As an example, Table 1.1.1 displays cervical cancer-incidence rate (per 10^5 person-year) in Ontario women in Canada [Fu, 2000]. There are 14 rows and 7 columns, and each row represents an age group and each column represents a period group. In addition, on each diagonal are the people who have the same birth cohort year.

Table 1.1.1: Cervical Cancer-incidence Rate (per 10^5 person-year) in Ontario Women in Canada by Age and Period

	Period						
Age	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94
20-24	3.89	3.24	2.90	2.05	2.19	1.76	1.73
25-29	16.01	11.18	8.92	9.74	8.48	7.43	7.54
30-34	26.02	21.14	16.32	15.84	14.54	13.67	12.71
35-39	38.84	25.09	21.07	18.74	18.80	18.04	18.18
40-44	47.65	32.50	22.71	20.01	18.78	16.19	18.12
45-49	51.48	36.69	22.15	19.20	17.74	17.29	18.31
50-54	49.12	37.26	25.51	18.41	16.66	15.41	14.07
55-59	51.48	40.87	34.70	21.83	16.97	17.69	13.73
60-64	47.68	42.80	29.76	22.71	20.16	17.69	16.94
65-69	40.44	39.17	31.44	28.79	23.35	19.26	19.16
70-74	42.40	35.32	27.78	24.31	20.27	20.19	14.95
75-79	42.44	36.68	28.75	25.22	21.17	21.08	19.43
80-84	41.50	29.74	31.54	22.31	20.14	15.25	21.28
85-89	30.79	32.43	37.10	19.81	16.42	14.87	12.06

Indeed temporal trend of cancer incidence and mortality has been frequently studied in the literature. [Vercelli et al., 2000, Quinn and Babb, 2002] studied patterns and trends in prostate-cancer incidence and mortality. [Gonzalez-Diego et al., 2000] presented time trends in ovarian-cancer mortality. [Pérez-Farinós et al., 2006] discussed time trend and age, period, and cohort effects on kidney-cancer mortality in Europe from 1981-2000.

[Znaor et al., 2015] examined trends in renal cell carcinoma incidence and mortality. [Boyar Cetinkaya et al., 2015] showed trends in incidence of neuroendocrine neoplasms in Norway. [Antoni et al., 2016] provided incidence and mortality trends of bladder cancer.

1.1.1 Age-Period-Cohort Analysis

For given Age-period-cohort (APC) data like the above, how to figure out the pattern or trends of mortality or incidence rate over age, period, and cohort is an interesting question. [Kupper et al., 1985] gave a review of APC analysis, and emphasized descriptive approach to the APC analysis. The rates are plotted in various ways as a function of the age, period, and cohort groups, see Figure 1.1.1, illustrating the need of a logarithm transformation of the rate. It also was pointed out that descriptive APC analysis is helpful in obtaining general qualitative features about age, period, and cohort patterns, but it has major limitations. For example, in Figure 1.1.1, the shape of the period curve is affected both by varying age effects and by varying cohort effects. Furthermore, a quantitative assessment of the way in which these age and cohort effects interact to influence the shape of this period curve cannot be obtained by visual examination of graphs like Figure 1.1.1. It can only be achieved via the use of valid statistical modeling procedures.

A three-factor model, the APC model (1.1.1) below, was first studied by [Greenberg et al., 1950]. It has been so far the most commonly used model in practice.

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij}, \tag{1.1.1}$$

where y_{ij} is the log-transformed rate in the i^{th} age and j^{th} period, with $i = 1, 2, ..., a; j =$

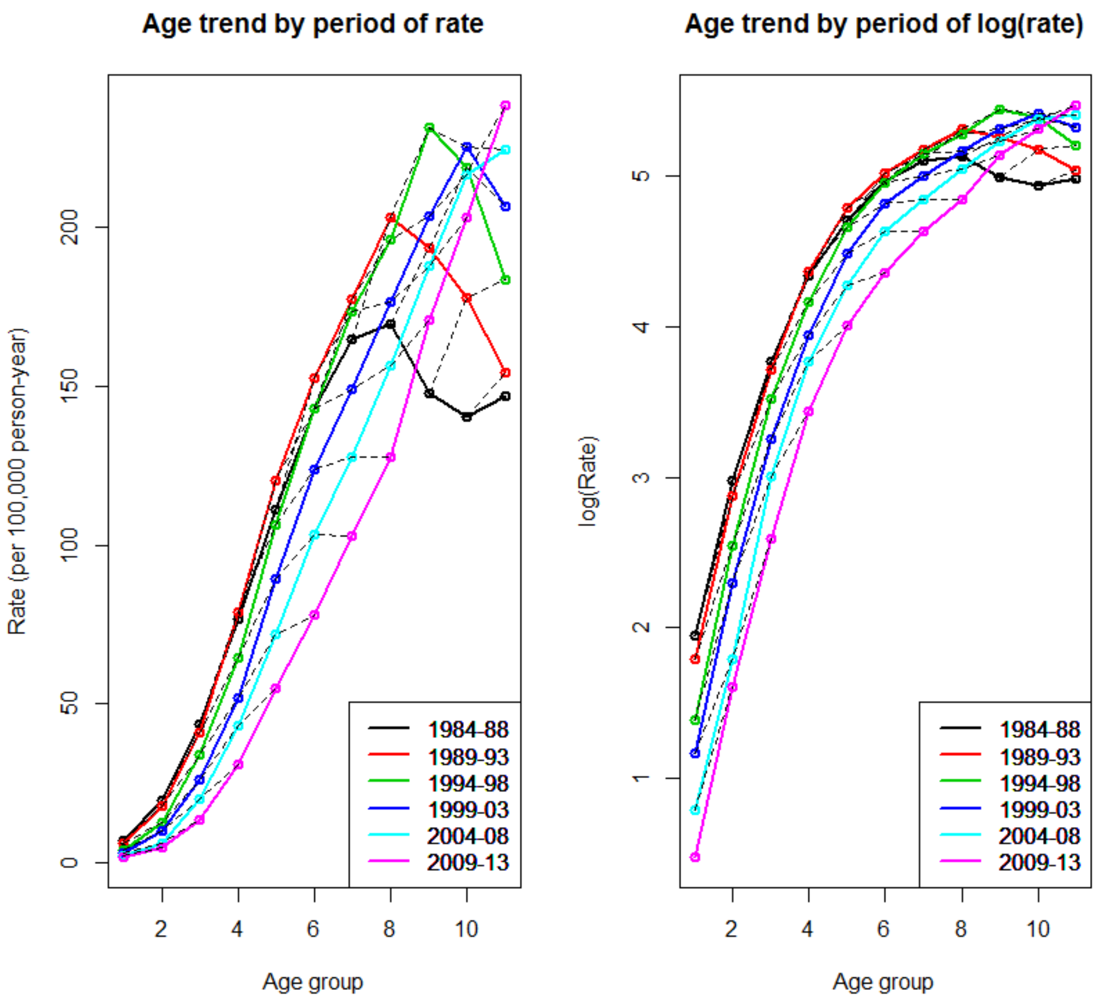

Figure 1.1.1: Descriptive APC Analysis for Male-mortality Rate of Liver Cancer in Korea (See Table 5.1.1)

$1,2,...,p$; and $k = 1,\cdots,a+p-1$. μ is the model intercept, α_i is the i^{th} age effect, β_j is the j^{th} period effect, and γ_k is the k^{th} cohort effect with $k = a-i+j$. And ε_{ij} are independently identically distributed Gaussian random errors with mean zero and common variance σ^2 (i.e. $\varepsilon_{ij} \sim N(0,\sigma^2)$).

[Kupper et al., 1985] also reviewed two-factor models (usually, age at occurrence and birth cohort). This approach assumes that one of the three factors in model (1.1.1) is "unimportant", in which case a two-factor model could reasonably provide a valid approximation of the data. A general strategy to choose a two-factor model is to examine "the goodness of fit". However, [Kupper et al., 1983] demonstrated that the above two-factor models can be seriously misleading.

1.1.2 Age-Period-Cohort Model and The Identifiability Problem

Writing the APC model in a matrix form, we have

$$EY = X\mathbf{b} = \begin{pmatrix} 1 & A & P & C \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{pmatrix}, \tag{1.1.2}$$

where $X = (1,A,P,C)$ is called design matrix, $\mathbf{b}^T = (\mu,\alpha^T,\beta^T,\gamma^T)$ is a vector of model parameters to estimate. Like in the analysis of variance (ANOVA) models, the parameters need a side-condition, either by setting a reference level, such as $\alpha_1 = \beta_1 = \gamma_1 = 0$ or by a parameter centralization below, which we prefer for the reason to provide variation assessment to each effect estimate

$$\sum_{i=1}^{a}\alpha_i = \sum_{j=1}^{p}\beta_j = \sum_{k=1}^{a+p-1}\gamma_k = 0. \tag{1.1.3}$$

For example, if the data is of 3×4, then the APC model can be written in a matrix form as

$$EY = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{pmatrix}$$

The estimates of the model parameters help to estimate the age, period, and cohort trend. However, the model has a fundamental problem, the identifiability problem as described below and has been discussed by [Kupper et al., 1983, 1985] and [Holford, 1985].

In this APC model, there exists a linear dependence, **Period** $-$ **Age** $=$ **Cohort**. In terms of linear algebra, this means the matrix $X^T X$ is singular. Consequently, the model yields multiple estimators, resulting in an identifiability problem.

Consider random response variable Y and predictive variables $X = (X_1, \cdots, X_p)$. The linear regression model is given by

$$EY = Xb = b_0 + b_1 X_1 + \cdots + b_p X_p. \tag{1.1.4}$$

The parameters $\mathbf{b}$ are estimated by minimizing the following sum of squares of the errors,

$$\begin{aligned} \min \quad & \sum_i \Big(y_i - (b_0 + b_1 X_{i1} + \cdots + b_p X_{ip}) \Big)^2 \\ = \quad & (\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - X\mathbf{b}), \end{aligned} \tag{1.1.5}$$

and if the matrix X is of full rank, the ordinary least-squares estimator is given by

$$\hat{\mathbf{b}} = (X^T X)^{-1} X^T \mathbf{y}.$$

For the APC model, the design matrix $X = (1, A, P, C)$ has linearly dependent columns, leading to a singular matrix $X^T X$ and resulting in multiple estimators of the parameters. Each estimator corresponds to one generalized inverse of the matrix $X^T X$.

1.1.3 Methods Used for APC Model

Many methods have been discussed to address the identifiability problem, including constraint method, estimable function method, smoothing method, see [Robertson and Boyle, 1986, Clayton and Schifflers, 1987a,b, Heuer, 1997, Robertson et al., 1999, O'Brien, 2000].

A common method to address this identifiability problem is using a constraint. The popular ones are linear constraints, given by $l^T\mathbf{b} = 0$, where $\mathbf{b}$ is the vector of parameters, and l is a vector. In practice, equality constraints are often specified, such as $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$ or $\gamma_1 = \gamma_2$. For example, $\alpha_1 = \alpha_2$ assuming disease-mortality rate does not vary in early ages. However, different constraint may lead to different trend estimation in age, period, and cohort effects. Such constraints highly rely on prior information. It is thus, difficult to verify that a constraint is satisfied by true parameter values before the parameters are accurately estimated, leading to an extremely difficult paradox. Furthermore different constraints lead to different estimation and thus, there is no uniqueness of the parameter estimation. For example, parameter estimation by different constraints for the cervical cancer-incidence rate data is illustrated in Figure 1.1.2. Clearly, it indicates that the constraint approach still leads to multiple trends and there is no unique trend estimation.

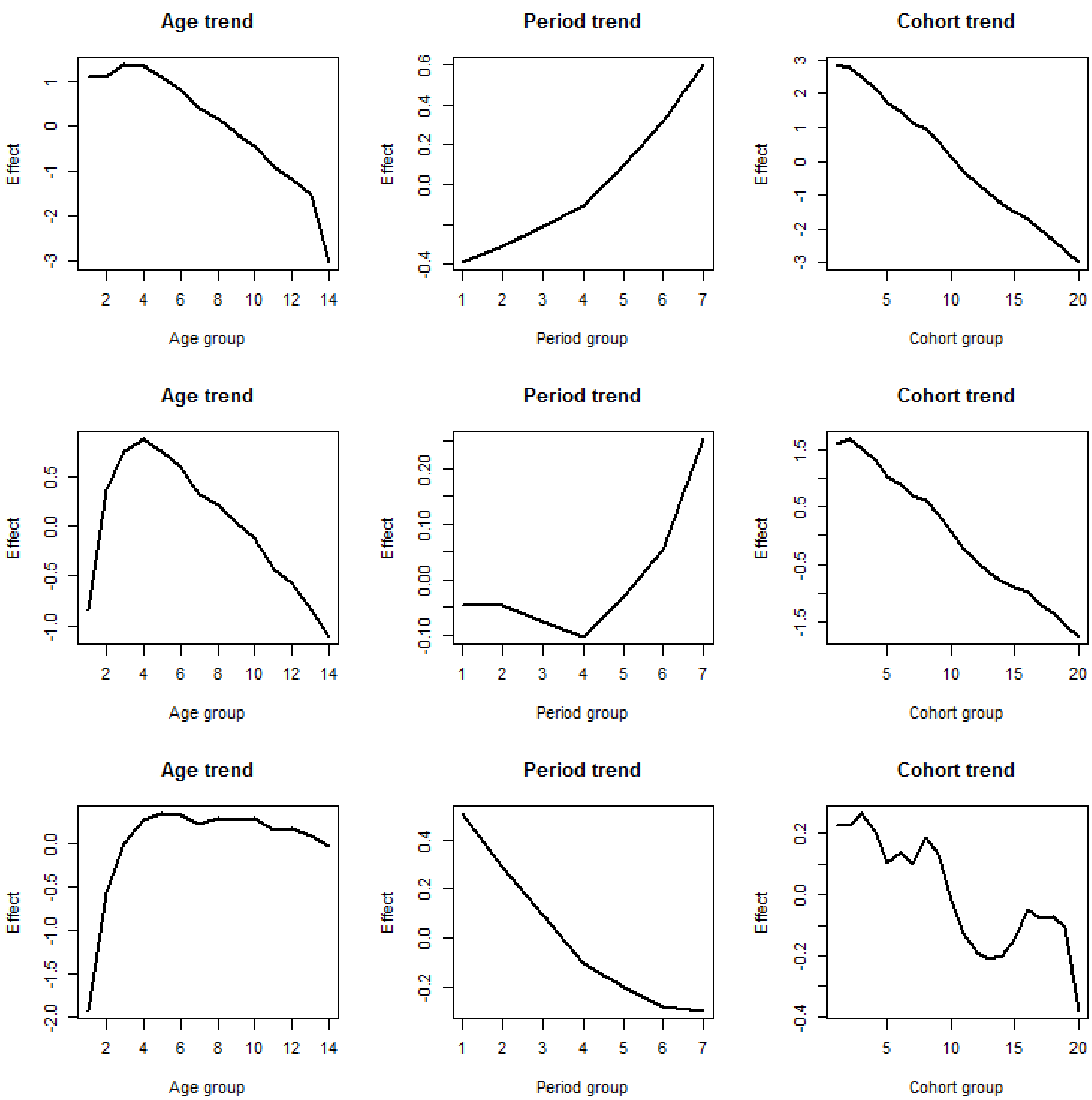

Figure 1.1.2: Comparison of Age, Period, and Cohort Trends Estimated with the linear Model by Different Constraints for the Cervical Cancer-incidence Rate. Upper panels: $\alpha_1 = \alpha_2$; middle panels: $\beta_1 = \beta_2$; lower panels: $\gamma_1 = \gamma_2$.

Another popular approach is to use estimable functions. A linear combination $l^T\mathbf{b}$ of the model parameters $\mathbf{b}$ is estimable if it has a linear unbiased estimate, i.e., a^TY for some a such that $E[a^TY] = l^T\mathbf{b}$ for all $\mathbf{b}$. Estimates of APC model cannot be uniquely determined, but non-linear characteristics of the age, period, and cohort trends, like the curvature [(age2-age1)-(age3-age2)], are estimable and can be uniquely estimated. But the linear trend varies with the estimator and has been claimed non-estimable. Many studies tried to search for the linear estimable functions, but failed, see [Clayton and Schifflers, 1987b, Rodgers, 1982, Holford, 1983, 1985, 1991, Kupper et al., 1985, OBrien and Stockard, 2009, Robertson et al., 1999, Tarone and Chu, 1996].

[Lee and Lin, 1996] studied an autoregressive APC model, where cohort effect is modeled as an AR(1) process. The estimates are found to be stable, but the assumption in the model needs to be carefully considered and the stationary requirement of the cohort effect may often be violated. [Heuer, 1997] used smoothing spline to model APC data, where all age, period, and cohort effects are smoothed by smoothing spline.

[Fu, 2000] proposed the intrinsic estimator, to address the identifiability problem in APC analysis. Because of the linear dependency, $X^T X$ is 1-less than full rank. Denote the unit null vector of $X^T X$ by B_0. The special one among the multiple estimators that is perpendicular to B_0 is called the intrinsic estimator B. It has been proved that the intrinsic estimator converges to true parameter values as the number of periods (columns) p tends to infinity [Fu, 2016].

1.2 Main Results of the Dissertation

This dissertation addresses two major issues in APC analysis, one computational and one theoretical.

In Chapter 1, the background of APC analysis is provided and previous methods to address the identifiability problem in APC analysis are reviewed.

In Chapter 2, a two-stage smoothing model is studied. Stage 1 takes fixed effects of age and period, with smoothing cohort effect. Stage 2 estimates age, period, and cohort effects using a constraint approach based on the consistent estimation of age or period effects from the first stage. Although the smoothing cohort model has been shown previously to be an important alternative approach with reasonably accurate estimation, it still remains unknown theoretically if it resolves the identifiability problem. More specifically, if the estimates of the intercept and the age effects converge to the true parameter values as p tends to infinity. This motivates our theoretical study of the asymptotic behavior of the estimator.

In Chapter 3, for efficient selection of the constraint, several methods have been studied through simulations, including constraint of large ratio, ratio of small variance, and constraint of small variance.

In Chapter 4, large sample behavior of the parameter estimator is studied. Consistency of smoothing cohort model, i.e., its estimation converges to the true parameter, is proved as smoothing bandwidth $h \to 0$ and $ph \to \infty$. The limit of the convergence equals to the limit of the intrinsic estimator.

Chapter 5 provides applications to real world examples using cancer-incidence and mortality rates to demonstrate the usefulness of the two-stage smoothing cohort model. The results are further compared with the intrinsic estimator method, to show that reasonable results can be expected from the two-stage smoothing cohort model.

Lastly, the conclusion from this dissertation is given in Chapter 6.

CHAPTER 2

Two-stage Smoothing Cohort Model

2.1 Smoothing Method

In order to model the dependence of the response variable Y on predictor X, in the nonlinear fashion, a more general model is given as follows

$$E(Y|X) = f(X), \tag{2.1.1}$$

where $X = (X_1, \cdots, X_p)^T$ is fixed p vector of explanatory variables, and $f(\cdot)$ is an unknown function from $\mathbb{R}^p$ to $\mathbb{R}^1$.

For the case $p > 1$, an additive model has been studied [Hastie and Tibshirani, 1990],

$$E(Y|X) = f_1(X_1) + \cdots + f_p(X_p), \tag{2.1.2}$$

where $f_i(\cdot)$ are arbitrary nonlinear functions.

Smoothing is a non-parametric, data-driven method to estimate $f(x)$ in (2.1.1). Let $s(\cdot)$ be a smoother, which is a function with same domain as the values in X, then $\hat{f}(\cdot) = s(\cdot)$. [Hastie and Tibshirani, 1990] and [Härdle et al., 2012] provide details about smoother, additive model, nonparametric, and semi-parametric models. A simple case for smoothing which helps to understand this concept is running-mean smoother.

Running-mean Smoother

A running-mean smoother to estimate $f(x_i)$ is to average Y values at x_i, as well as the k points to the left and k points to the right of x_i. Denote the indices of these points by $N^S(x_i)$, the running mean

$$s(x_i) = ave_{j \in N^S(x_i)}(y_i). \tag{2.1.3}$$

If it is not possible to take k points to the left or right of x_i, we take as many data points as possible. A formal definition of a symmetric nearest neighbor is

$$N^S(x_i) = \{\max(i-k,1), \cdots, i-1, i, i+1, \cdots, \min(i+k,n)\} \tag{2.1.4}$$

Next kernel and smoothing spline are presented, which are the most popular smoothers.

Kernel Smoother

First define a kernel $K(\cdot)$ as a non-negative continuous, bounded, and symmetric real-valued function which integrates to one, i.e., $\int K(u)du = 1$. And

$$K_h(x - x_i) = K(\frac{x - x_i}{h}), \tag{2.1.5}$$

where the parameter h is the window-width, also known as the bandwidth.

[Nadaraya, 1964, Watson, 1964] studied an estimator

$$s_{NW}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)}, \tag{2.1.6}$$

which is called Nadaraya-Waston (NW) estimator. It has been proved under mild assumption, when $h \to 0$ and $nh \to \infty$,

$$s_{NW}(x) \xrightarrow{p} f(x).$$

Details of convergence will be discussed in Chapter 4.

Smoothing Spline

Among all functions $f(x)$ with continuous second order derivative, smoothing spline minimizes the penalized residual sum of squares

$$\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_a^b [f''(t)]^2 dt, \tag{2.1.7}$$

where $a \leq x_1 \leq \cdots \leq x_n \leq b$. λ is a smoothing parameter that can be selected to tune the smoothness of the spline function.

Smoothing spline can be expressed as,

$$\hat{f} = (I + \lambda K)^{-1}\mathbf{y}, \tag{2.1.8}$$

where $K \in \mathbb{R}^{n\times n}$ and $K_{ij} = \int f_i''(x) f_j''(x)dx$, see [Hastie and Tibshirani, 1990, Härdle et al., 2012].

Denote $S_\lambda = (I + \lambda K)^{-1}$, then $\hat{f} = S_\lambda \mathbf{y}$. The degrees of freedom of a smoothing spline is defined by

$$df = Tr(S_\lambda) = \sum_{i=1}^{n} \frac{1}{1+\lambda d_i}, \tag{2.1.9}$$

where d_i are the eigenvalues of the matrix K. There is a strictly monotone relationship between λ and the degrees of freedom.

[McCullagh and Nelder, 1989] pointed out that "as for a cubic smoothing spline, it can be shown that as $n \to \infty$ and the smoothing parameter $\lambda \to 0$, under certain regularity conditions, $\hat{f}(x) \to f(x)$. This says as we get more and more data, the smoothing-spline estimate will converge to the true regression function $E(Y|X = x)$."

Linear smoother is important in smoothing, since some common smoothers, like running means, locally weighted running lines, kernel smoothers, smoothing splines, bin smoothers, and even the least-squares line, are linear smoothers.

Definition 2.1. A linear smoother is written as

$$\hat{\mathbf{y}} = S\mathbf{y}, \tag{2.1.10}$$

where $\mathbf{y} = (y_1, \cdots, y_n)^T$ and the $n \times n$ matrix S, called a smoother matrix, depends on some smoothing parameter, λ, and also on the data $\mathbf{x}$, but not $\mathbf{y}$.

It has been shown that both NW kernel smoother and smoothing spline are linear smoother, e.g. kernel smoother

$$\hat{\mathbf{y}} = \begin{pmatrix} \frac{K_h(x_1-x_1)}{\sum_{i=1}^n K_h(x_1-x_i)} & \cdots & \frac{K_h(x_1-x_n)}{\sum_{i=1}^n K_h(x_1-x_i)} \\ \vdots & \ddots & \vdots \\ \frac{K_h(x_n-x_1)}{\sum_{i=1}^n K_h(x_n-x_i)} & \cdots & \frac{K_h(x_n-x_n)}{\sum_{i=1}^n K_h(x_n-x_i)} \end{pmatrix} \mathbf{y}.$$

In smoothing, there is a fundamental trade-off between the bias and variance of the estimate, and this trade-off is governed by the smoothing parameter. Over smoothing may increase the bias, while under smoothing may increase the variance.

Smoothing has special properties that we can apply to the APC model. First, smoothing is data-driven and provides local approximation. It takes near neighbors into consideration and thus yields stable estimation when data becomes sparse, though it introduces bias in the estimation as well. The estimation is nonlinear and not additive with linear term, so it can be employed to break up the dependence in Period-Age=Cohort.

2.2 Two-stage Smoothing Cohort Model – A Semiparametric Approach

2.2.1 The First Stage of Smoothing Cohort Model

An alternative semiparametric approach to APC analysis, which is called two-stage smoothing cohort model, was proposed by [Fu, 2008]. Smoothing cohort stage 1 is given by,

$$Y_{ij} = \mu + \alpha_i + \beta_j + S\gamma_k + \varepsilon_{ij}, \tag{2.2.1}$$

where $i = 1, ..., a$; $j = 1, ..., p$ and $k = 1, ..., a + p - 1$. μ is the model intercept, α_i is i^{th} age effect, β_j is j^{th} period effect, $S\gamma_k = S(k; \gamma_1, \cdots, \gamma_{a+p-1})$ is k^{th} smoothed cohort effect with $k = a - i + j$ and $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

Parameter centralization is required in this model.

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{p} \beta_j = 0. \tag{2.2.2}$$

Consider a generalized linear model with a response variable Y following an exponential family distribution, with the link function $g(\cdot)$,

$$g(EY_{ij}) = \mu + \alpha_i + \beta_j + S\gamma_k. \tag{2.2.3}$$

Consider the cubic smoothing spline (2.1.8), the model may be written as

$$g(Ey) = \begin{pmatrix} 1 & A & P & (I+\lambda K)^{-1}C \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{pmatrix}, \tag{2.2.4}$$

where $\lambda \geq 0$ is a smoothing tuning parameter.

There are three main reasons for us to study this model.

1. The APC data with a age groups and p period groups has overlap between consecutive cohorts, which requires cohort effects to be estimated with contributions from nearest neighbors. For example, in Table 2.2.1, there are 4 years overlap between consecutive cohorts.

Table 2.2.1: Cohort Expression of Cervical Cancer-incidence Rate Data

	Period			
Age	1960-64	1965-69	$\cdots$	1990-94
20-24	1936-1944	1941-1949	$\cdots$	1966-1974
25-29	1931-1939	1936-1944	$\cdots$	1961-1969
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
50-54	.	.	$\cdots$	1936-1944
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
80-84	1876-1884	1881-1889	$\cdots$	1906-1914
85-89	1871-1879	1876-1884	$\cdots$	1901-1909

2. Smoothing breaks up linear dependence. Since the smoothing estimates take the nearest neighbor effects into consideration in a non-linear kernel fashion, it leads to a design matrix effectively non-singular, thus breaks up the linear dependence. It can be

summarized below.

Proposition 2.2. *[Fu, 2008] The design matrix* $\tilde{X} = (1, A, P, (I + \lambda K)^{-1}C)$ *is of full rank.*

3. It reduces instability in parameter estimation, especially for the extreme cohorts (oldest and youngest cohorts), where only one or two observations are on each cohort, leading to unstable cohort effect estimation. Smoothing takes neighborhood effects into consideration and thus makes the estimation more robust for the extreme cohorts.

2.2.2 Algorithm

Backfitting is an iterative method for estimating unknown components of an additive model, which was proposed by [Friedman and Stuetzle, 1981] for projection-pursuit regression, and by [Wecker and Ansley, 1982] and [Buja et al., 1989] for fitting additive spline models.

Algorithm 2.3. Backfitting Algorithm for Smoothing Cohort Model

1. Initialize smoothed cohort effects to 0, $\hat{S}\gamma_k = 0$ for $k = 1, \cdots, a + p - 1$.

2. Compute the residuals $U_{ij} = Y_{ij} - \hat{S}\gamma_k$ with $k = a - i + j$.

3. Fit an age-period model,

$$U_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

   and compute the residuals $Z_{ij} = Y_{ij} - \hat{U}_{ij}$.

4. Smoothing the residuals Z_{ij} against the cohort diagonals with a predetermined parameter λ for spline smoothing,

$$S\gamma_k = S^{\lambda}(k; Z_{11}, \cdots, Z_{ap}).$$

5. Iterate until convergence of the estimators of all α, β, and $S\gamma$ with a convergence criterion:

$$\sqrt{\sum_i (\hat{\alpha}_i^{old} - \hat{\alpha}_i^{new})^2 + \sum_j (\hat{\beta}_j^{old} - \hat{\beta}_j^{new})^2 + \sum_k (\hat{S}\gamma_k^{old} - \hat{S}\gamma_k^{new})^2} \leq \Delta,$$

where $\Delta > 0$ is a predetermined convergence threshold.

2.2.3 The Second Stage of Smoothing Cohort Model

Smoothing cohort model (2.2.3) can break up linear dependence, and yield unique estimation of the age, period, and cohort effects. In a special case of linear model (2.2.1), it also yields consistent estimation of the age or period effect (See Proposition 4.4). However it introduces bias in cohort estimation. Given that the smoothing cohort model yields consistent estimates of the age or period effects, such estimates can be further used to specify a constraint that leads to consistent estimation of the age, period, and cohort model, leading to bias correction. Furthermore, it has been proved that a non-contrast constraint may also yield consistent estimates, which can be further used for consistent estimation through a constraint. Note that a linear constraint $l^T \mathbf{b} = 0$ on the parameters $\mathbf{b}$ is contrast if the sum $\sum_i l_i = 0$.

CHAPTER 3

Second Stage in Smoothing Cohort Model

3.1 Introduction

In this chapter, a bias correction method will be discussed, which uses consistent estimation from the first stage of smoothing cohort model to set an age or period-effect constraint.

Usually a constraint with the following form is used in APC analysis:

$$\alpha_i - c\alpha_j = 0 \quad or \quad \beta_i - c\beta_j = 0, \tag{3.1.1}$$

where c is assumed by subjective assumption or prior knowledge. But this constraint often leads to biased estimation [Kupper et al., 1985, Fu, 2016]. In this dissertation, an estimate from another model is used to set the value of c. Let $c = \frac{\hat{\alpha}_i}{\hat{\alpha}_j}, i, j = 1, \cdots, a$; or $c = \frac{\hat{\beta}_i}{\hat{\beta}_j}, i, j = 1, \cdots, p$. $\hat{\alpha}_i$ and $\hat{\beta}_j$ are the age- and period-effect estimates from stage 1 smoothing cohort model.

For convenience, denote by τ the effect of age α or period β. Rewrite equation (3.1.1) as

$$\tau_i - c\tau_j = 0, \tag{3.1.2}$$

where $c = \frac{\hat{\tau}_i}{\hat{\tau}_j}$.

Although it has been proved that a consistent estimation can be obtained either through the smoothing cohort model or the non-contrast constraint method with p being large, different constraints may yield different convergence speed. This motivates the study of the selection of the second stage constraint.

3.2 Method 1: Constraint with Large Ratio

By the result from [Fu, 2016] that non-contrast constraint yields consistent estimates, it may be expected that the bigger the ratio, the better the convergence speed. Hence the following constraint is studied

$$\tau_i - c\tau_j = 0 \quad s.t. \quad (i, j) = \arg\max_{p,q} \frac{\hat{\tau}_p}{\hat{\tau}_q},$$

where $c = \frac{\hat{\tau}_i}{\hat{\tau}_j}$.

Taking the sign of the ratio into consideration, another constraint will also be considered:

$$\tau_i - c\tau_j = 0 \quad s.t. \quad (i, j) = \arg\max_{p,q} \left| \frac{\hat{\tau}_p}{\hat{\tau}_q} \right|,$$

where $c = \frac{\hat{\tau}_i}{\hat{\tau}_j}$.

In the simulation, a few other ratios, such as $c = \frac{\hat{\tau}_{(k)}}{\hat{\tau}_{(1)}}$, $c = \frac{\hat{\tau}_{(k)}}{\hat{\tau}_{(k-1)}}$, and $c = \frac{\hat{\tau}_{(2)}}{\hat{\tau}_{(1)}}$, are also examined to compare the convergence. Here $\hat{\tau}_{(i)}$ are sorted stage 1 estimates, $\hat{\tau}_{(1)} \leq \hat{\tau}_{(2)} \leq \cdots \leq \hat{\tau}_{(j)} < 0 < \hat{\tau}_{(j+1)} \leq \cdots \leq \hat{\tau}_{(k-1)} \leq \hat{\tau}_{(k)}$.

3.3 Method 2: Constraint with Ratio of Small Variance

Rewrite the linear constraint $\tau_i - c\tau_j = 0$ in ratio form:

$$c = \frac{\tau_i}{\tau_j}.$$

Here c is expected to be a constant which represents the ratio of the true parameters. In the stage 2 constraint, $\frac{\hat{\tau}_i}{\hat{\tau}_j}$ is used to estimate this constant c. So the smaller of the variance of $\frac{\hat{\tau}_i}{\hat{\tau}_j}$, the more likely the ratio will behave like a constant.

$$\tau_i - c\tau_j = 0 \quad s.t. \quad (i, j) = \arg\min_{p,q} Var(\frac{\hat{\tau}_p}{\hat{\tau}_q}).$$

In order to calculate and compare $Var(\frac{\hat{\tau}_p}{\hat{\tau}_q})$, the following results of ratio distribution are needed.

Proposition 3.1. *[Hayya et al., 1975] Suppose $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$ and that X and Y are not necessarily statistically independent. For an untransformed ratio of random variables, $W = Y/X$, the second-order Taylor expansion gives the approximations,*

$$V(W) \simeq \frac{\sigma_x^2 \mu_y^2}{\mu_x^4} + \frac{\sigma_y^2}{\mu_x^2} - 2\rho \frac{\sigma_x \sigma_y \mu_y}{\mu_x^3}.$$

Assume $\tau_i \sim N(\mu_{\tau_i}, \sigma_{\tau_i}^2)$ and $\tau_j \sim N(\mu_{\tau_j}, \sigma_{\tau_j}^2)$, the variance of $c_{pq} = \frac{\tau_p}{\tau_q}$ could be approximated by

$$\hat{\sigma}_{c_{pq}}^2 \simeq \frac{\hat{\sigma}_{\tau_q}^2 \hat{\tau}_p^2}{\hat{\tau}_q^4} + \frac{\hat{\sigma}_{\tau_p}^2}{\hat{\tau}_q^2} - 2\hat{\rho}_{pq} \frac{\hat{\sigma}_{\tau_p} \hat{\sigma}_{\tau_q} \hat{\tau}_p}{\hat{\tau}_q^3}, \tag{3.3.1}$$

where $\hat{\rho}_{pq} = cov(\hat{\tau}_p, \hat{\tau}_q)$.

Algorithm 3.2. Constraint selection with the ratio of the smallest variance

1. For $\forall i, j = 1, \cdots, a$ with age-effect constraint or $\forall i, j = 1, \cdots, p$ with period-effect constraint, compute the ratio $R_{ij} = \left(\frac{\hat{\tau}_i}{\hat{\tau}_j}\right)$ and its variance $Var(R)$.

2. Choose $(p, q) = \arg\min_{i,j} Var(R_{ij})$.

3. Set non-contrast constraint $\tau_p - c\tau_q = 0$, where $c = \frac{\hat{\tau}_p}{\hat{\tau}_q}$ not equal to 1.

3.4 Method 3: Constraint of Small Variance

Based on a similar argument to the previous method, the constraint of small variance is also considered.

$$\tau_p - c\tau_q = 0 \text{ s.t. } (p,q) = \arg\min_{i,j} Var(\hat{\tau}_i - c\hat{\tau}_j).$$

Suppose $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$ and that X and Y are not necessarily statistically independent. Then

$$Var(X - cY) = Var(X) - 2Cov(X,Y)c + Var(Y)c^2.$$

Assume $\hat{\tau}_i \sim N(\tau_i, \sigma_{\tau_i}^2)$ and $\hat{\tau}_j \sim N(\tau_j, \sigma_{\tau_j}^2)$, the variance of $\tau_p - c\tau_q$ could be approximated by

$$\hat{\sigma}_{\tau_p - c\tau_q}^2 = \sigma_{\hat{\tau}_p}^2 - 2\hat{\rho}_{pq}c + \sigma_{\hat{\tau}_q}^2 c^2, \tag{3.4.1}$$

where $\hat{\rho}_{pq} = cov(\hat{\tau}_p, \hat{\tau}_q)$.

Algorithm 3.3. Constraint selection with the smallest variance

1. For $\forall i, j = 1, \cdots, a$ with age-effect constraint or $\forall i, j = 1, \cdots, p$ with period-effect constraint, compute the ratio $R_{ij} = \left(\frac{\hat{\alpha}_i}{\hat{\alpha}_j}\right)$ and variance of constraint

$$Var(\hat{\alpha}_i - R_{ij}\hat{\alpha}_j) = var(\hat{\alpha}_i) - 2R_{ij}cov(\hat{\alpha}_i, \hat{\alpha}_j) + R_{ij}^2 Var(\hat{\alpha}_j).$$

2. Choose $(p,q) = \arg\min_{i,j} Var(\hat{\alpha}_i - R_{ij}\hat{\alpha}_j)$.

3. Set non-contrast constraint $\tau_p - c\tau_q = 0$, where $c = \frac{\hat{\tau}_p}{\hat{\tau}_q}$ not equal to 1.

3.5 Numerical Simulation 1

In this simulation, a table of 10 age groups and 5 period groups are assigned. The specified parameters are given in Table 3.5.6. The simulation parameter has N-shaped age trend, and in general an increasing-age trend is expected in cancer-mortality rates and a decreasing-age trend is expected in crime rates. A W-shaped period trend and a concave down shape for cohort trend are specified. The commend 'smooth.spline' is used in R with $df = 10$, and 1000 simulate runs are conducted. A signal-noise ratio $var(E)/\sigma^2 = 3$ is used, where $E = 1\mu + A\alpha + P\beta + C\gamma$ is the expected effects without noise.

Figure 3.5.1 shows plots of given parameters and stage 1 smoothing cohort model estimates, where the solid lines are the specified parameters (as in [Fu, 2008]), and dash line is the simulated mean estimates with the stage 1 model. Age and period trend estimated by smoothing cohort model are close to the specified parameters, but the cohort-effect estimates deviate from the specified trend, indicating certain bias.

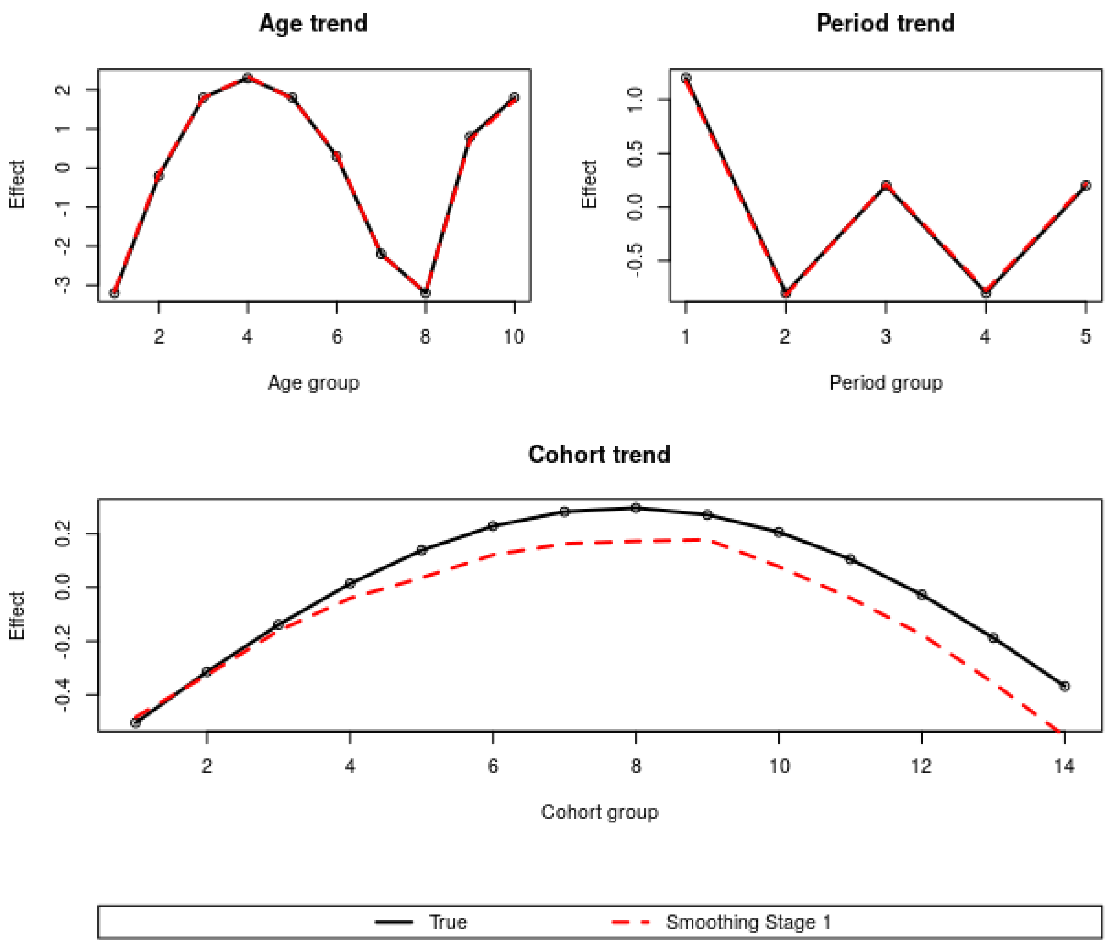

Figure 3.5.1: Age, Period, and Cohort Effects with Smoothing Cohort Stage 1 Model in Simulation I

3.5.1 Age-effect Constraints Compared with Period-effect Constraints

Based on asymptotics, constraints on smaller group-size effects will make the estimator close to large sample behavior, as the larger group size would make it more likely to behave as the asymptotics. In this simulation $p < a$, so period-effect constraint is expected to yield less biased estimation. First, the behavior of the estimation with period-effect constraints compare with age-effect constraints is examined. Parameter estimates are displayed in Table 3.5.1 and plotted in Figure 3.5.2. In Figure 3.5.2, the upper panels show the trends estimated with constraint of the largest ratio; middle panels show the trends estimated with constraints of the smallest variance ratio; and lower panels show the trends estimated with constraints of the smallest variance. The solid red line represents trends estimated with age-effect constraints and the dashed green line represents trends estimated with period-effect constraints. In all three methods, the period-effect constraints yield smaller bias than the age-effect constraints, except for the middle upper panel. The period-effect constraint yields slightly larger bias than the age-effect constraint in the middle plot of the upper panels, which indicates that although the age-effect constraint with the largest ratio (non-contrast constraint) yields consistent estimation of period effects, and also the period-effect constraint with the largest ratio (non-contrast constraint) yields consistent estimation, it does not warrant anything about the effect estimates of the other factors. In fact, the largest ratio would yield poor effect estimates of the other factor by the simulation results in the upper panel, which further illustrates the weakness of the non-contrast constraint methods in [Fu, 2016]. In the middle panels, the constraint with the smallest variance of ratio shows much less biased estimation for all age, period, and cohort effect than the other two constraints. The simulation estimates, MSE and bias are given in Table 3.5.1.

In conclusion, period-effect constraints yield smaller bias. This may be explained by the fact that period has smaller group size than age effect ($p < a$) in this simulation. In the following, simulations are conducted with different constraint selection rules on period effects only.

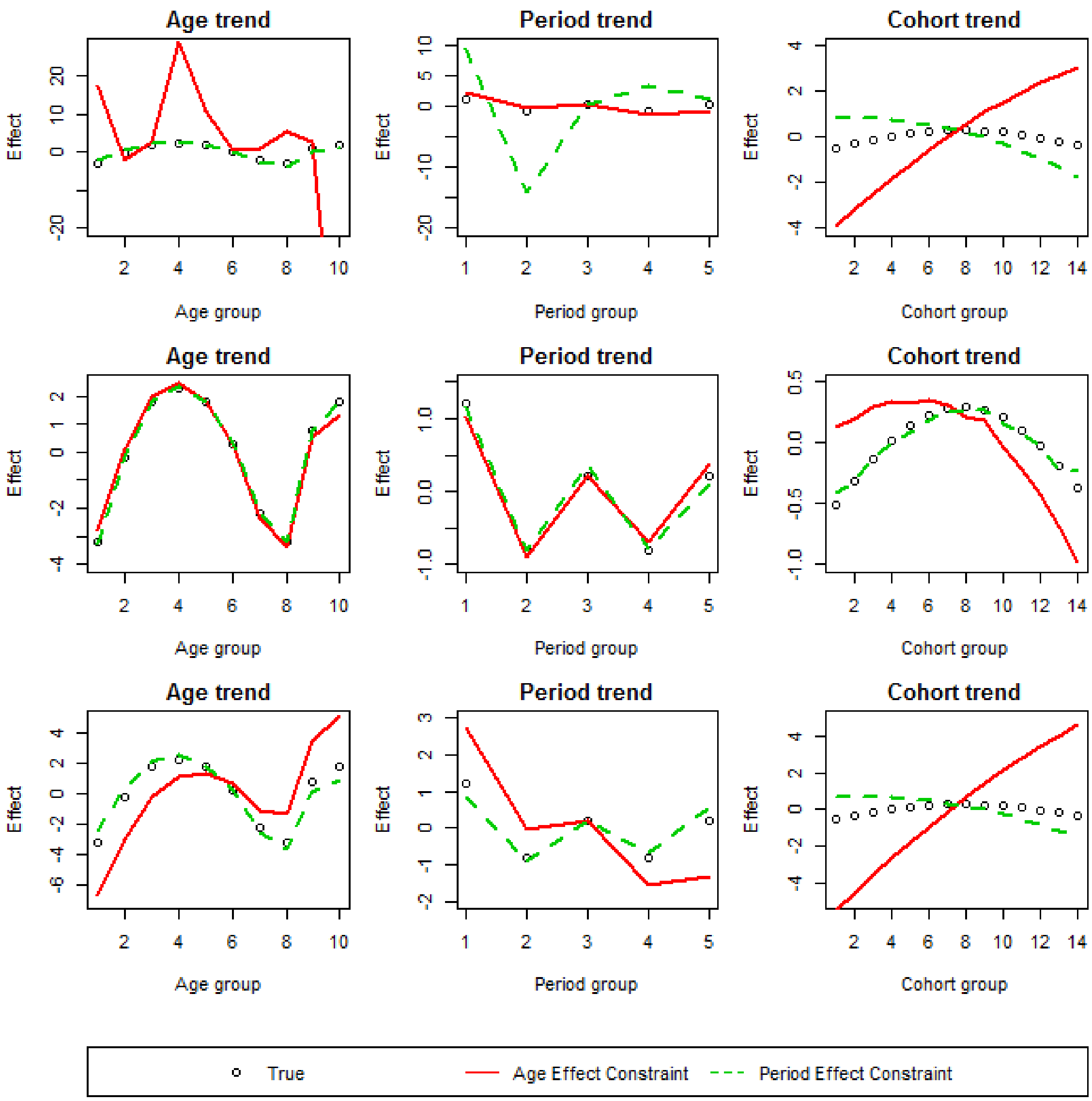

Figure 3.5.2: Comparison of Age- and Period-effect Constraints by Estimation with Three Constraint-selection Methods: Upper panels: the Largest Ratio; Middle panels: the Smallest Variance of Ratio; Lower panels: the Smallest Variance of Constraint

Table 3.5.1: Comparison of Smoothing Cohort Stage 2 Constraints on Age and Period Effects

	Method 1[a] age			Method 1 period			Method 2 [b] age			Method 2 period			Method 3[c] age			Method 3 period		
	Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
	1.0092	0.0001	0.0092	1.0086	0.0001	0.0086	1.0092	0.0001	0.0092	1.0172	0.0003	0.0172	1.0092	0.0001	0.0092	1.0096	0.0001	0.0096
Age	17.6006	507.3273	20.8006	-2.2786	0.8608	0.9214	-2.8018	0.1619	0.3982	-3.238	0.0022	-0.038	-6.6542	11.9728	-3.4542	-2.407	0.6442	0.7930
	-2.0026	3.2813	-1.8026	0.5253	0.5337	0.7253	0.1276	0.1092	0.3276	-0.1721	0.0013	0.0279	-2.8956	7.2917	-2.6956	0.4316	0.4082	0.6316
	2.9034	1.5145	1.1034	2.2783	0.2327	0.4783	1.9925	0.0383	0.1925	1.7829	0.0007	-0.0171	-0.1542	3.8318	-1.9542	2.2047	0.1688	0.4047
	29.1352	771.6743	26.8352	2.6207	0.1046	0.3207	2.4508	0.0234	0.1508	2.3389	0.0019	0.0389	1.2038	1.2063	-1.0962	2.5803	0.0805	0.2803
	10.725	87.128	8.925	1.8726	0.0058	0.0726	1.8156	0.0006	0.0156	1.7747	0.001	-0.0253	1.4234	0.1426	-0.3766	1.8662	0.0049	0.0662
	0.6113	0.098	0.3113	0.2446	0.0036	-0.0554	0.316	0.0006	0.016	0.3234	0.0009	0.0234	0.7302	0.1858	0.4302	0.2549	0.0026	-0.0451
	1.024	14.5169	3.224	-2.5019	0.0928	-0.3019	-2.3255	0.0164	-0.1255	-2.1798	0.0008	0.0202	-1.0313	1.3706	1.1687	-2.4588	0.0689	-0.2588
	5.5657	79.9026	8.7657	-3.6761	0.2306	-0.4761	-3.3919	0.0381	-0.1919	-3.1758	0.001	0.0242	-1.2864	3.6744	1.9136	-3.6053	0.1693	-0.4053
	2.6171	3.3317	1.8171	0.0575	0.5587	-0.7425	0.5068	0.0881	-0.2932	0.7597	0.0022	-0.0403	3.4719	7.1644	2.6719	0.1563	0.4236	-0.6437
	-68.1797	5033.022	-69.9797	0.8575	0.9004	-0.9425	1.3097	0.2444	-0.4903	1.7861	0.0009	-0.0139	5.1925	11.5494	3.3925	0.977	0.6924	-0.8230
Period	2.2444	1.1008	1.0444	9.3535	70.9611	8.1535	1.0099	0.0369	-0.1901	1.1539	0.0023	-0.0461	2.7337	2.3605	1.5337	0.8374	0.1339	-0.3626
	-0.2799	0.2731	0.5201	-14.2052	183.1652	-13.4052	-0.8972	0.0098	-0.0972	-0.8113	0.0003	-0.0113	-0.0353	0.587	0.7647	-0.913	0.0135	-0.1130
	0.211	0.0002	0.011	0.1981	0.0001	-0.0019	0.211	0.0002	0.011	0.3442	0.0209	0.1442	0.211	0.0002	0.011	0.1892	0.0002	-0.0108
	-1.3162	0.2691	-0.5162	3.3661	18.4471	4.1661	-0.6989	0.0105	0.1011	-0.7795	0.0006	0.0205	-1.5609	0.5811	-0.7609	-0.6899	0.0128	0.1101
	-0.8593	1.132	-1.0593	1.2874	1.3954	1.0874	0.3753	0.0315	0.1753	0.0927	0.0116	-0.1073	-1.3486	2.4062	-1.5486	0.5763	0.1444	0.3763
Cohort	-3.8795	11.4951	-3.3749	0.8763	1.9318	1.3809	0.1329	0.4146	0.6375	-0.4149	0.0104	0.0897	-5.4697	24.7384	-4.9651	0.6852	1.4477	1.1898
	-3.2026	8.4212	-2.8887	0.8179	1.2991	1.1318	0.1924	0.2618	0.5063	-0.3086	0.0013	0.0053	-4.5482	17.9918	-4.2343	0.6663	0.9831	0.9802
	-2.4776	5.5211	-2.3389	0.8182	0.9275	0.9569	0.3001	0.1961	0.4388	-0.119	0.0013	0.0197	-3.5786	11.8747	-3.4399	0.6862	0.6957	0.8249
	-1.8192	3.3916	-1.8333	0.7525	0.553	0.7384	0.3413	0.1094	0.3272	-0.0001	0.0008	-0.0142	-2.6755	7.2587	-2.6896	0.6492	0.4131	0.6351
	-1.2207	1.8623	-1.3589	0.6116	0.2283	0.4734	0.3225	0.0352	0.1843	0.088	0.003	-0.0502	-1.8324	3.896	-1.9706	0.5413	0.1676	0.4031
	-0.5801	0.6603	-0.8088	0.5197	0.0864	0.291	0.3458	0.0144	0.1171	0.1879	0.0021	-0.0408	-0.9471	1.3874	-1.1758	0.474	0.0622	0.2453
	-0.0036	0.0827	-0.2857	0.3623	0.007	0.0802	0.3051	0.001	0.023	0.2629	0.0008	-0.0192	-0.1259	0.1674	-0.408	0.3525	0.0056	0.0704
	0.5273	0.0543	0.231	0.159	0.0194	-0.1373	0.2186	0.0065	-0.0777	0.2501	0.0026	-0.0462	0.6496	0.1257	0.3533	0.1841	0.0132	-0.1122
	1.1054	0.703	0.8349	0.0079	0.0707	-0.2626	0.1795	0.0091	-0.091	0.2711	0.0004	0.0006	1.4724	1.4494	1.2019	0.0562	0.0481	-0.2143
	1.5124	1.7223	1.3064	-0.3096	0.2698	-0.5156	-0.0308	0.0573	-0.2368	0.1635	0.0022	-0.0425	2.124	3.6917	1.918	-0.2496	0.2127	-0.4556
	1.9451	3.4158	1.8399	-0.6201	0.5335	-0.7253	-0.2154	0.1051	-0.3206	0.0923	0.0008	-0.0129	2.8013	7.2942	2.6961	-0.5098	0.3876	-0.6150
	2.3536	5.7227	2.3814	-0.9352	0.8353	-0.9074	-0.4241	0.1605	-0.3963	-0.0338	0.0009	-0.006	3.4545	12.1685	3.4823	-0.8019	0.6145	-0.7741
	2.7014	8.4234	2.8892	-1.3272	1.3163	-1.1394	-0.6937	0.2612	-0.5059	-0.2126	0.0019	-0.0248	4.0469	17.9956	4.2347	-1.1756	0.998	-0.9878
	3.0382	11.711	3.4065	-1.7332	1.8889	-1.3649	-0.9741	0.3749	-0.6058	-0.2269	0.0221	0.1414	4.6284	25.054	4.9967	-1.5582	1.4483	-1.1899

[a] Constraints with the largest ratio

[b] Constraints with ratio of the smallest variance

[c] Constraints with the smallest variance

3.5.2 Simulation on Constraint Selection with Different Ratio

In this simulation, several constraints (the largest ratio, the largest absolute ratio, the largest difference, and two small difference constraints) on only period effects are considered and compared. The motivation to examine the various ratio is to make sure the constraint is non-contrast constraint so that the ratio is far from 1. Monte Carlo estimates of age, period, and cohort effect, MSE, and bias are given in Table 3.5.2, and the mean effect of 1000 runs are plotted in Figure 3.5.3.

In Figure 3.5.3, the constraint $c = \frac{\hat{\beta}_{(2)}}{\hat{\beta}_{(1)}}$ gives the best results in this simulation. It may be explained as this largest difference ratio $\frac{\hat{\beta}_{(p)}}{\hat{\beta}_{(1)}}$ may have large variance so the ratio may not behave like a constant, consequently the constraint defined by the ratio may not be valid.

Table 3.5.2: Comparison of Smoothing Cohort Stage 2 Constraints on Period Effect by Ratio

	Max vs Min [a]			Min2 vs Min1 [b]			Max1 vs Max2 [c]			Max [d]			Abs Max [e]		
	Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
	1.0094	0.0001	0.0094	1.0100	0.0001	0.0100	1.0100	0.0002	0.0100	1.0086	0.0001	0.0086	1.0103	0.0002	0.0103
Age	-0.6424	6.5432	2.5576	-3.0094	0.0452	0.1906	-0.6611	6.4479	2.5389	-2.2786	0.8608	0.9214	-1.7348	2.1596	1.4652
	1.7993	3.9984	1.9993	-0.0384	0.0317	0.1616	1.7901	3.9617	1.9901	0.5253	0.5337	0.7253	0.9531	1.3374	1.1531
	3.1910	1.9357	1.3910	1.8755	0.0087	0.0755	3.1804	1.9064	1.3804	2.2783	0.2327	0.4783	2.5820	0.6159	0.7820
	3.1677	0.7535	0.8677	2.3807	0.0077	0.0807	3.1628	0.7450	0.8628	2.6207	0.1046	0.3207	2.8044	0.2562	0.5044
	2.0520	0.0639	0.2520	1.7898	0.0006	-0.0102	2.0511	0.0634	0.2511	1.8726	0.0058	0.0726	1.9328	0.0181	0.1328
	0.0618	0.0571	-0.2382	0.3281	0.0013	0.0281	0.0651	0.0556	-0.2349	0.2446	0.0036	-0.0554	0.1869	0.0133	-0.1131
	-3.0482	0.7199	-0.8482	-2.2598	0.0048	-0.0598	-3.0458	0.7160	-0.8458	-2.5019	0.0928	-0.3019	-2.6791	0.2313	-0.4791
	-4.5889	1.9299	-1.3889	-3.2675	0.0075	-0.0675	-4.5769	1.8968	-1.3769	-3.6761	0.2306	-0.4761	-3.9833	0.6180	-0.7833
	-1.2158	4.0649	-2.0158	0.6167	0.0390	-0.1833	-1.2042	4.0181	-2.0042	0.0575	0.5587	-0.7425	-0.3735	1.3851	-1.1735
	-0.7764	6.6400	-2.5764	1.5843	0.0553	-0.2157	-0.7614	6.5625	-2.5614	0.8575	0.9004	-0.9425	0.3117	2.2279	-1.4883
Period	1.1869	0.0011	-0.0131	1.1217	0.0080	-0.0783	3.9152	7.7864	2.7152	9.3535	70.9611	8.1535	1.2419	7.0440	0.0419
	-1.3681	0.3230	-0.5681	-1.3519	0.3049	-0.5519	-4.7113	15.7063	-3.9113	-14.2052	183.1652	-13.4052	-2.5802	7.0603	-1.7802
	0.1834	0.0004	-0.0166	0.2412	0.0018	0.0412	0.1851	0.0003	-0.0149	0.1981	0.0001	-0.0019	0.2032	0.0001	0.0032
	-0.2249	0.3309	0.5751	-1.2614	0.2132	-0.4614	-0.2302	0.3249	0.5698	3.3661	18.4471	4.1661	-1.1596	0.6427	-0.3596
	0.2228	0.0014	0.0228	1.2504	1.1055	1.0504	0.8412	0.4228	0.6412	1.2874	1.3954	1.0874	2.2948	7.0326	2.0948
Cohort	3.2212	13.8856	3.7258	-0.1623	0.1372	0.3423	3.2025	13.7466	3.7071	0.8763	1.9318	1.3809	1.6598	4.7105	2.1644
	2.8197	9.8227	3.1336	-0.0524	0.0816	0.2615	2.8016	9.7091	3.1155	0.8179	1.2991	1.1318	1.4926	3.2824	1.8065
	2.4559	6.7344	2.5946	0.0985	0.0655	0.2372	2.4407	6.6558	2.5794	0.8182	0.9275	0.9569	1.3683	2.2837	1.5070
	2.0245	4.0431	2.0104	0.1780	0.0322	0.1639	2.0123	3.9943	1.9982	0.7525	0.5530	0.7384	1.1724	1.3500	1.1583
	1.5207	1.9122	1.3825	0.2015	0.0069	0.0633	1.5087	1.8791	1.3705	0.6116	0.2283	0.4734	0.9113	0.6022	0.7731
	1.0682	0.7053	0.8395	0.2760	0.0035	0.0473	1.0620	0.6949	0.8333	0.5197	0.0864	0.2910	0.7011	0.2250	0.4724
	0.5494	0.0719	0.2673	0.2809	0.0005	-0.0012	0.5480	0.0711	0.2659	0.3623	0.0070	0.0802	0.4256	0.0212	0.1435
	-0.0194	0.1001	-0.3157	0.2494	0.0027	-0.0469	-0.0199	0.1004	-0.3162	0.1590	0.0194	-0.1373	0.1004	0.0390	-0.1959
	-0.5401	0.6577	-0.8106	0.2443	0.0020	-0.0262	-0.5358	0.6507	-0.8063	0.0079	0.0707	-0.2626	-0.1752	0.2005	-0.4457
	-1.2190	2.0314	-1.4250	0.0864	0.0171	-0.1196	-1.2110	2.0088	-1.4170	-0.3096	0.2698	-0.5156	-0.6214	0.6890	-0.8274
	-1.8852	3.9632	-1.9904	-0.0557	0.0314	-0.1609	-1.8705	3.9050	-1.9757	-0.6201	0.5335	-0.7253	-1.0378	1.3145	-1.1430
	-2.5706	6.4680	-2.5428	-0.2285	0.0496	-0.2007	-2.5555	6.3912	-2.5277	-0.9352	0.8353	-0.9074	-1.4855	2.1375	-1.4577
	-3.3244	9.8413	-3.1366	-0.4491	0.0819	-0.2613	-3.3096	9.7486	-3.1218	-1.3272	1.3163	-1.1394	-1.9965	3.2900	-1.8087
	-4.1009	13.9373	-3.7326	-0.6669	0.1079	-0.2986	-4.0734	13.7329	-3.7051	-1.7332	1.8889	-1.3649	-2.5151	4.6358	-2.1468

[a] Smoothing cohort stage 2 model using constraint with ratio $c = \hat{\beta}_{(p)}/\hat{\beta}_{(1)}$.

[b] Smoothing cohort stage 2 model using constraint with ratio $c = \hat{\beta}_{(2)}/\hat{\beta}_{(1)}$.

[c] Smoothing cohort stage 2 model using constraint with ratio $c = \hat{\beta}_{(p)}/\hat{\beta}_{(p-1)}$.

[d] Smoothing cohort stage 2 model using constraint with ratio $c = \hat{\beta}_i/\hat{\beta}_j$, where $(i,j) = \arg\max_{p \neq q} \hat{\beta}_p/\hat{\beta}_q$.

[e] Smoothing cohort stage 2 model using constraint with ratio $c = \hat{\beta}_i/\hat{\beta}_j$, where $(i,j) = \arg\max_{p \neq q} \left|\hat{\beta}_p/\hat{\beta}_q\right|$.

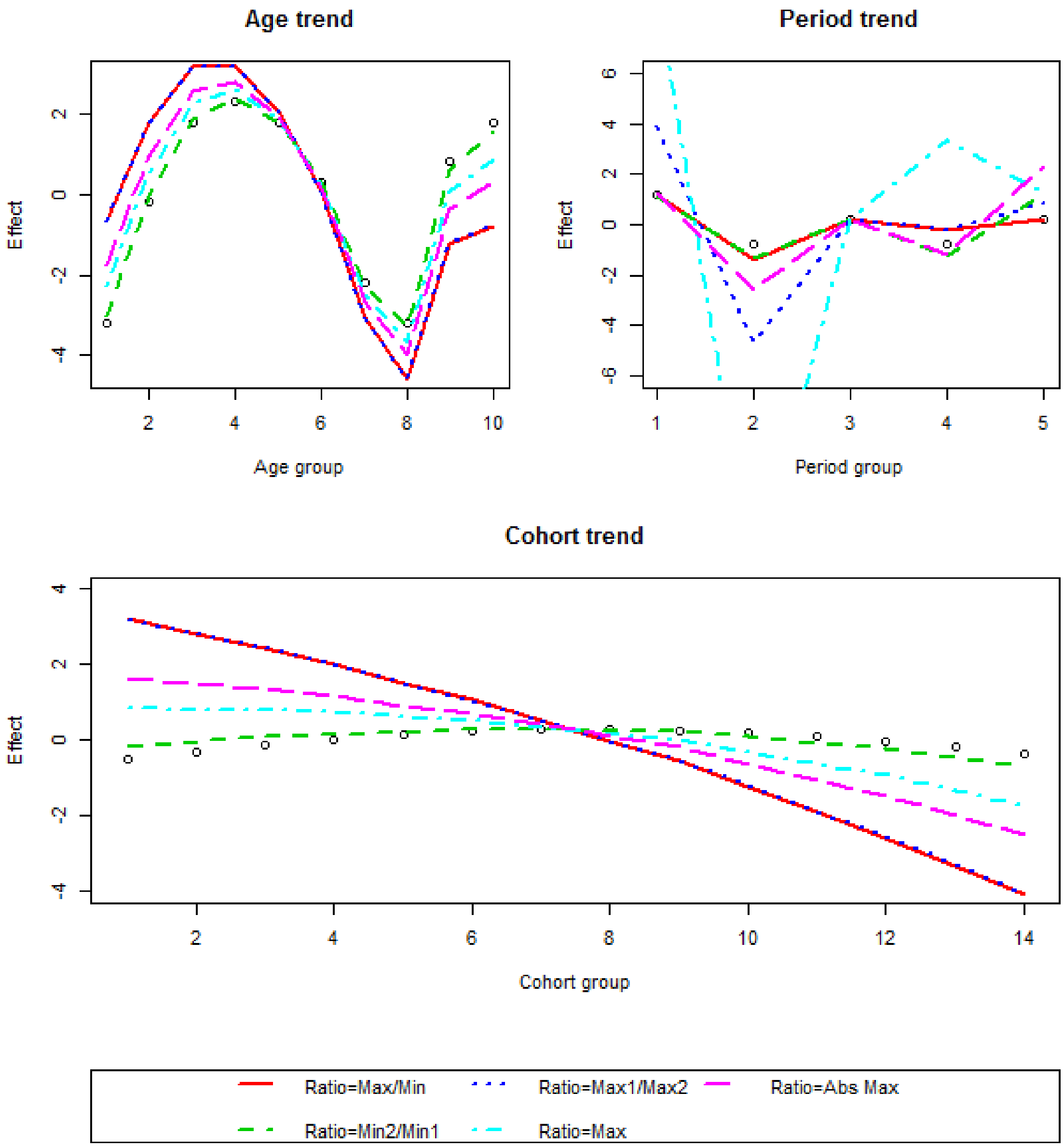

Figure 3.5.3: Smoothing Cohort Stage 2 Constraints on Period Effect by Ratio of Estimates

3.5.3 Simulation on Constraint Selection by the Variance of Ratio

Monte Carlo estimates of age, period, and cohort effects by the average of the estimates of 1000 runs, the MSE, and the bias are given in Table 3.5.3, and the estimated trends are plotted in Figure 3.5.4. From the plot, it is observed that the smaller the variance of the ratio, the smaller the estimates deviate from the specified parameter values for all age, period, and cohort effects. When the variance is small enough, there is not much difference, indicating consistent estimation.

Table 3.5.3: Comparison of Smoothing Cohort Stage 2 Constraints on Period Effect by Ratio of Variance

The Smallest Variance [a]			5^{th} Variance [b]			10^{th} Variance [c]			The Largest Variance [d]		
Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
1.0172	0.0003	0.0172	1.0130	0.0002	0.0130	1.0074	0.0001	0.0074	1.0103	0.0002	0.0103
-3.2380	0.0022	-0.0380	-3.1002	0.0145	0.0998	-2.4394	0.5865	0.7606	-1.4716	3.0010	1.7284
-0.1721	0.0013	0.0279	-0.1041	0.0121	0.0959	0.3875	0.3501	0.5875	1.1577	1.8518	1.3577
1.7829	0.0007	-0.0171	1.8053	0.0016	0.0053	2.1946	0.1583	0.3946	2.7282	0.8662	0.9282
2.3389	0.0019	0.0389	2.3278	0.0015	0.0278	2.5802	0.0797	0.2802	2.8921	0.3525	0.5921
1.7747	0.0010	-0.0253	1.7777	0.0009	-0.0223	1.8533	0.0033	0.0533	1.9620	0.0268	0.1620
0.3234	0.0009	0.0234	0.3101	0.0005	0.0101	0.2609	0.0020	-0.0391	0.1577	0.0208	-0.1423
-2.1798	0.0008	0.0202	-2.2263	0.0015	-0.0263	-2.4462	0.0618	-0.2462	-2.7668	0.3232	-0.5668
-3.1758	0.0010	0.0242	-3.1756	0.0022	0.0244	-3.5997	0.1624	-0.3997	-4.1295	0.8687	-0.9295
0.7597	0.0022	-0.0403	0.7003	0.0127	-0.0997	0.2030	0.3614	-0.5970	-0.5782	1.9080	-1.3782
1.7861	0.0009	-0.0139	1.6850	0.0177	-0.1150	1.0059	0.6388	-0.7941	0.0485	3.0814	-1.7515
1.1539	0.0023	-0.0461	1.1584	0.0026	-0.0416	1.1906	0.0015	-0.0094	-0.2397	14.7029	-1.4397
-0.8113	0.0003	-0.0113	-1.1286	0.1084	-0.3286	-1.2737	0.2251	-0.4737	-2.7604	7.7394	-1.9604
0.3442	0.0209	0.1442	0.2503	0.0027	0.0503	0.2102	0.0002	0.0102	0.2032	0.0001	0.0032
-0.7795	0.0006	0.0205	-0.7891	0.0006	0.0109	-0.7605	0.0024	0.0395	-0.1244	2.0379	0.6756
0.0927	0.0116	-0.1073	0.5090	0.0961	0.3090	0.6333	0.1894	0.4333	2.9212	16.7037	2.7212
-0.4149	0.0104	0.0897	-0.2982	0.0537	0.2064	0.6449	1.3384	1.1495	2.0399	6.5021	2.5445
-0.3086	0.0013	0.0053	-0.2091	0.0177	0.1048	0.6131	0.8713	0.9270	1.8142	4.5493	2.1281
-0.1190	0.0013	0.0197	-0.0379	0.0147	0.1008	0.6400	0.6144	0.7787	1.6314	3.1471	1.7701
-0.0001	0.0008	-0.0142	0.0762	0.0067	0.0621	0.6211	0.3737	0.6070	1.3771	1.8667	1.3630
0.0880	0.0030	-0.0502	0.1447	0.0015	0.0065	0.5060	0.1379	0.3678	1.0575	0.8499	0.9193
0.1879	0.0021	-0.0408	0.2335	0.0008	0.0048	0.4742	0.0615	0.2455	0.7889	0.3157	0.5602
0.2629	0.0008	-0.0192	0.2626	0.0008	-0.0195	0.3450	0.0045	0.0629	0.4549	0.0304	0.1728
0.2501	0.0026	-0.0462	0.2655	0.0014	-0.0308	0.1796	0.0141	-0.1167	0.0711	0.0513	-0.2252
0.2711	0.0004	0.0006	0.3059	0.0020	0.0354	0.0586	0.0462	-0.2119	-0.2629	0.2865	-0.5334
0.1635	0.0022	-0.0425	0.1594	0.0038	-0.0466	-0.2372	0.1990	-0.4432	-0.7676	0.9525	-0.9736
0.0923	0.0008	-0.0129	0.0426	0.0068	-0.0626	-0.5101	0.3836	-0.6153	-1.2425	1.8248	-1.3477
-0.0338	0.0009	-0.0060	-0.1004	0.0099	-0.0726	-0.7835	0.5791	-0.7557	-1.7487	2.9749	-1.7209
-0.2126	0.0019	-0.0248	-0.3673	0.0395	-0.1795	-1.0918	0.8293	-0.9040	-2.3182	4.5583	-2.1304
-0.2269	0.0221	0.1414	-0.4777	0.0223	-0.1094	-1.4599	1.2094	-1.0916	-2.8952	6.4146	-2.5269

Note: Variance of ratio could be defined as a set $S_{ij} = \{Var(\hat{\beta}_i/\hat{\beta}_i), \forall i, j = 1, \cdots, p, and\, i \neq j\}$ and $s_{(n)}$ be the n^{th} smallest element of S_{ij}.

[a] Smoothing cohort stage 2 model using constraint with ratio corresponding $s_{(1)}$.

[b] Smoothing cohort stage 2 model using constraint with ratio corresponding $s_{(5)}$.

[c] Smoothing cohort stage 2 model using constraint with ratio corresponding $s_{(10)}$.

[d] Smoothing cohort stage 2 model using constraint with ratio corresponding $\max S$.

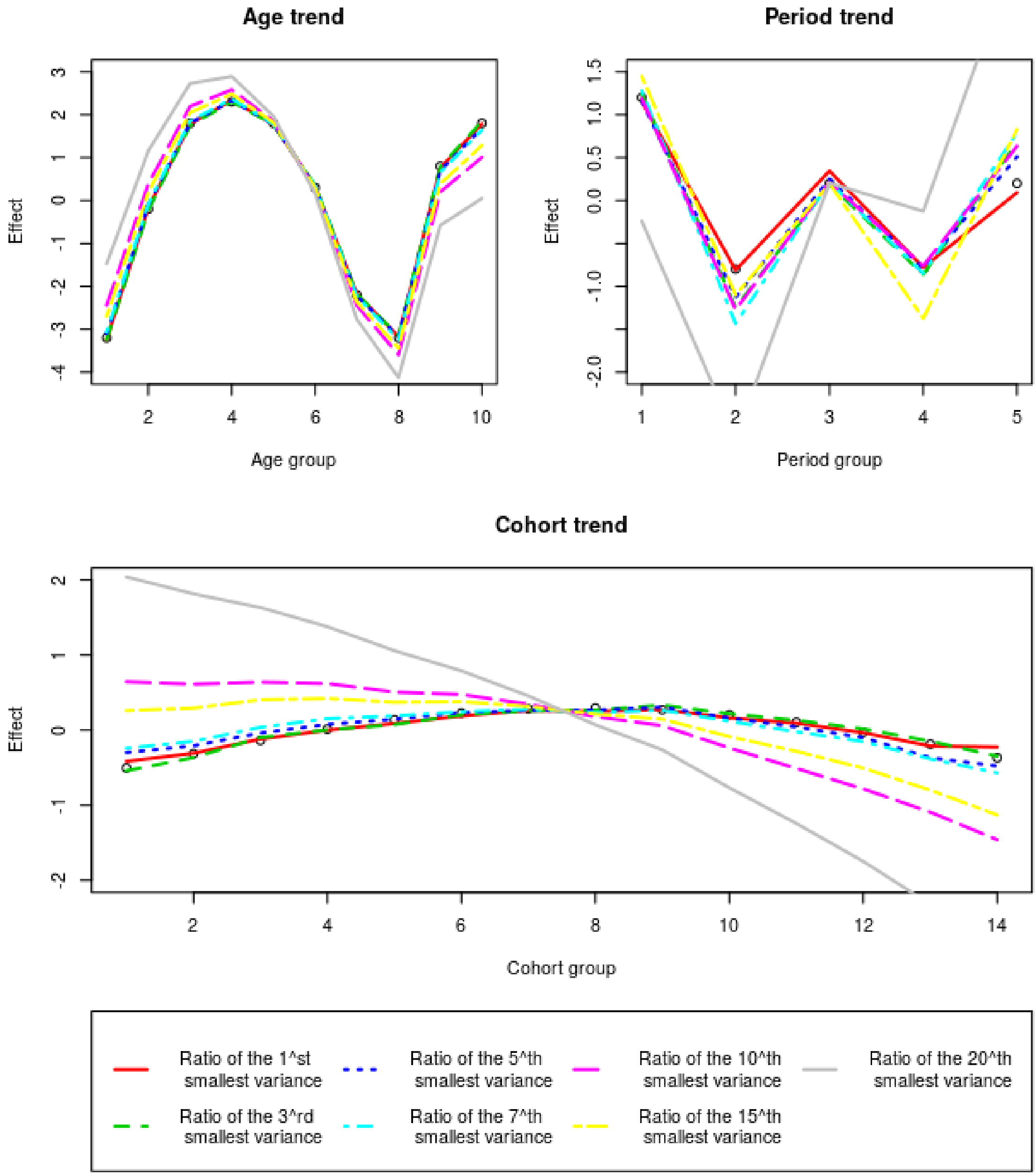

Figure 3.5.4: Smoothing Cohort Stage 2 Constraints on Period Effect by Ratio of Variance.

Note: The small variance of the ratio in 1^{st}, 3^{rd}, 5^{th} and 7^{th} leads to small deviation from specified value.

3.5.4 Simulation on Constraint Selection by the Variance of Constraint

Monte Carlo estimates of age, period, and cohort effects by the average of the estimates of 1000 runs, the MSE, and the bias are given in Table 3.5.4, and the estimated trends are plotted in Figure 3.5.5.

In Figure 3.5.5, it can be observed that the larger the variance of the constraint, the larger the estimates deviate from the specified parameter values for period effect. Estimated age effects are all close to the specified effects. But behavior of cohort-effect estimation is not predictable. The period effect also presents unpredictable pattern. So the small variance constraint approach cannot provide good constraint selection. This may be explained by the fact that even when the variance of the constraint is small, the linear combination of the parameters $\tau_i - c\tau_j$ may behave like a constant, but not necessarily equal to zero, indicating the ratio constraint may not be valid.

Table 3.5.4: Comparison of Smoothing Cohort Stage 2 Constraints on Period Effect by Variance of Constraints

The Smallest Variance [a]			5^{th} Variance [b]			10^{th} Variance [c]			20^{th} Variance [d]		
Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
1.0096	0.0001	0.0096	1.0098	0.0001	0.0098	1.0085	0.0001	0.0085	1.0156	0.0003	0.0156
-2.4070	0.6442	0.7930	-2.0963	1.2308	1.1037	-2.6817	0.2818	0.5183	-3.2631	0.0048	-0.0631
0.4316	0.4082	0.6316	0.6708	0.7661	0.8708	0.2051	0.1722	0.4051	-0.1827	0.0009	0.0173
2.2047	0.1688	0.4047	2.3831	0.3442	0.5831	2.0550	0.0693	0.2550	1.7667	0.0016	-0.0333
2.5803	0.0805	0.2803	2.6839	0.1491	0.3839	2.4924	0.0388	0.1924	2.3124	0.0005	0.0124
1.8662	0.0049	0.0662	1.8969	0.0099	0.0969	1.8115	0.0006	0.0115	1.7688	0.0013	-0.0312
0.2549	0.0026	-0.0451	0.2313	0.0052	-0.0687	0.2913	0.0006	-0.0087	0.3199	0.0008	0.0199
-2.4588	0.0689	-0.2588	-2.5668	0.1363	-0.3668	-2.3680	0.0300	-0.1680	-2.1799	0.0008	0.0201
-3.6053	0.1693	-0.4053	-3.7819	0.3426	-0.5819	-3.4529	0.0683	-0.2529	-3.1426	0.0037	0.0574
0.1563	0.4236	-0.6437	-0.0961	0.8109	-0.8961	0.3714	0.1918	-0.4286	0.7835	0.0009	-0.0165
0.9770	0.6924	-0.8230	0.6750	1.2784	-1.1250	1.2759	0.2881	-0.5241	1.8171	0.0011	0.0171
0.8374	0.1339	-0.3626	0.8713	0.1102	-0.3287	1.1865	0.0028	-0.0135	13.1393	205.2986	11.9393
-0.9130	0.0135	-0.1130	-1.1457	0.1201	-0.3457	-1.5635	0.5841	-0.7635	-4.2621	18.8781	-3.4621
0.1892	0.0002	-0.0108	0.1994	0.0001	-0.0006	0.1930	0.0002	-0.0070	0.3387	0.0194	0.1387
-0.6899	0.0128	0.1101	-0.6084	0.0373	0.1916	-0.7756	0.0015	0.0244	-12.4118	196.8118	-11.6118
0.5763	0.1444	0.3763	0.6834	0.2358	0.4834	0.9596	0.5790	0.7596	3.1958	16.8236	2.9958
0.6852	1.4477	1.1898	1.1463	2.7515	1.6509	0.2477	0.5944	0.7523	-0.4563	0.0049	0.0483
0.6663	0.9831	0.9802	1.0594	1.9049	1.3733	0.3086	0.4072	0.6225	-0.3401	0.0022	-0.0262
0.6862	0.6957	0.8249	1.0089	1.3294	1.1476	0.4080	0.3124	0.5467	-0.1527	0.0012	-0.0140
0.6492	0.4131	0.6351	0.8894	0.7743	0.8753	0.4359	0.1862	0.4218	-0.0182	0.0017	-0.0323
0.5413	0.1676	0.4031	0.7185	0.3411	0.5803	0.3804	0.0630	0.2422	0.0694	0.0052	-0.0688
0.4740	0.0622	0.2453	0.5805	0.1256	0.3518	0.3868	0.0268	0.1581	0.1759	0.0032	-0.0528
0.3525	0.0056	0.0704	0.3790	0.0100	0.0969	0.3165	0.0017	0.0344	0.2728	0.0005	-0.0093
0.1841	0.0132	-0.1122	0.1360	0.0262	-0.1603	0.2111	0.0078	-0.0852	0.2484	0.0028	-0.0479
0.0562	0.0481	-0.2143	-0.0560	0.1084	-0.3265	0.1521	0.0160	-0.1184	0.2956	0.0011	0.0251
-0.2496	0.2127	-0.4556	-0.4224	0.3991	-0.6284	-0.0743	0.0830	-0.2803	0.1798	0.0011	-0.0262
-0.5098	0.3876	-0.6150	-0.7633	0.7621	-0.8685	-0.2991	0.1717	-0.4043	0.1222	0.0010	0.0170
-0.8019	0.6145	-0.7741	-1.1240	1.2141	-1.0962	-0.5282	0.2634	-0.5004	0.0044	0.0020	0.0322
-1.1756	0.9980	-0.9878	-1.5587	1.8984	-1.3709	-0.8148	0.4128	-0.6270	-0.1906	0.0015	-0.0028
-1.5582	1.4483	-1.1899	-1.9935	2.6691	-1.6252	-1.1308	0.6103	-0.7625	-0.2108	0.0271	0.1575

Note: Variance of ratio could be defined as a set $T = \{Var(\hat{\beta}_i - c\hat{\beta}_i), \forall i, j = 1, \cdots, p, and\, i \neq j\}$ and $t_{(n)}$ be the n^{th} smallest element of T.

[a] Smoothing cohort stage 2 model using constraint with ratio corresponding $t_{(1)}$.

[b] Smoothing cohort stage 2 model using constraint with ratio corresponding $t_{(5)}$.

[c] Smoothing cohort stage 2 model using constraint with ratio corresponding $t_{(10)}$.

[d] Smoothing cohort stage 2 model using constraint with ratio corresponding $t_{(20)}$.

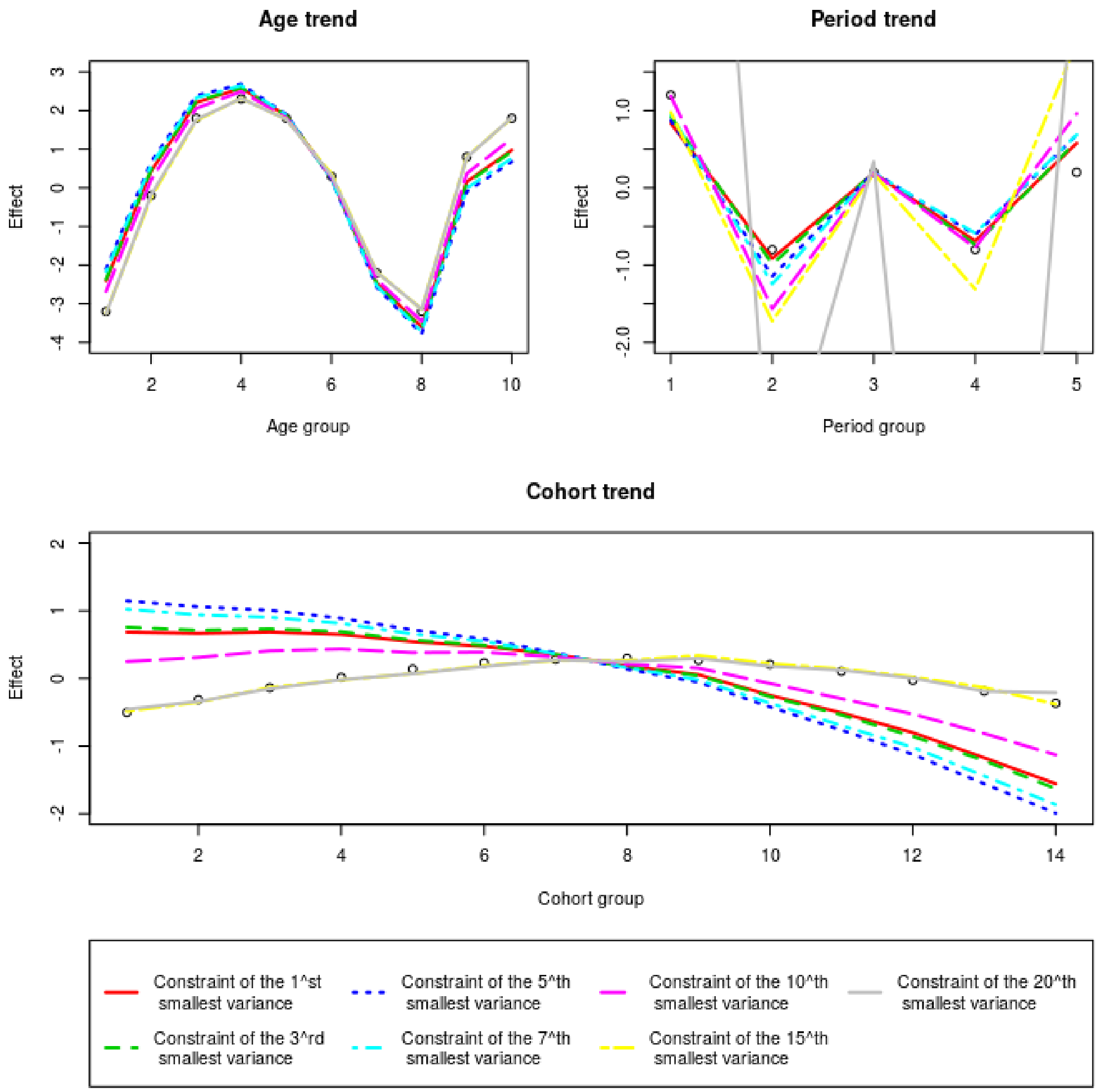

Figure 3.5.5: Smoothing Cohort Stage 2 Constraints on Period Effect by Variance

3.5.5 Optimal Constraint Selection Method

From the simulation on constraint selection with large ratio, it has been observed that $c = \frac{\hat{\beta}_{(2)}}{\hat{\beta}_{(1)}}$ yields the smallest bias. So we compare this with the constraint of the smallest variance of the ratio, as well as the constraint of the smallest variance to reveal which constraint selection method performs the best. Figure 3.5.6 shows estimated age, period, and cohort effects with constraints from three different selection rules. The one with the smallest variance of the ratio yields the least bias among all three. So the constraint chosen with the ratio of the smallest variance is recommended in the second stage of the smoothing cohort model.

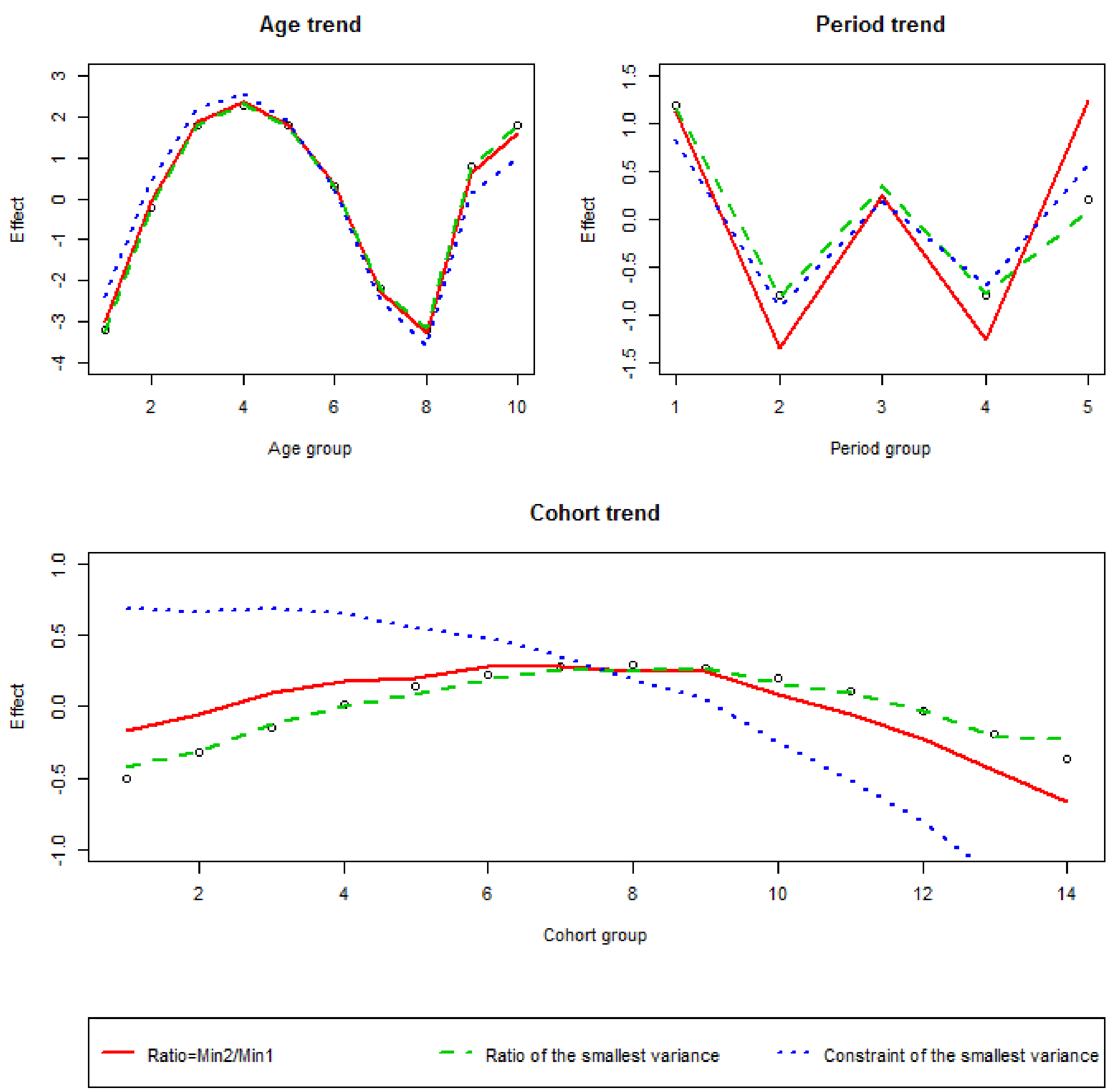

Figure 3.5.6: Smoothing Cohort Stage 2 Model Estimation by Period-effect Constraints with Three Constraint Selection Methods

Descriptive statistics of the standard error of ratios are given in Table 3.5.5. Figure 3.5.7 shows the standard error of the ratios by the period-effect constraints from three different selection rules. There are three outliers with huge values of 2119161, 265886

Table 3.5.5: Standard Error of Ratios by Period-effect Constraints with Three Constraint Selection Methods

	Min	25th Percentile	Median	Mean	75th Percentile	Max
Constraint with the Largest Ratio	0.50	3.60	12.60	2931.00	60.50	2119000
Constraint with the Largest Absolute Ratio	0.73	4.94	15.20	6661.01	75.88	5636209
Constraint with Ratio of the Smallest Variance	0.11	0.21	0.25	0.26	0.30	0.69
Constraint with the Smallest Variance	0.73	2.91	5.29	218.40	11.03	120100

and 120917 in constraint with the largest ratio, three outliers with large values of 5636209, 265886 and 156744 in constraint with the largest absolute ratio, and three outliers with values of 120068, 48356 and 20425 in constraint with the smallest variance. The majority of the standard error of constraint with the largest ratio, the largest absolute ratio, and the smallest variance are not small, and can be huge as well in extreme case. This could explain the fact that the constraint chosen with the largest ratio, the largest absolute ratio, and the smallest variance yield larger bias, since the ratios do not behave like constants.

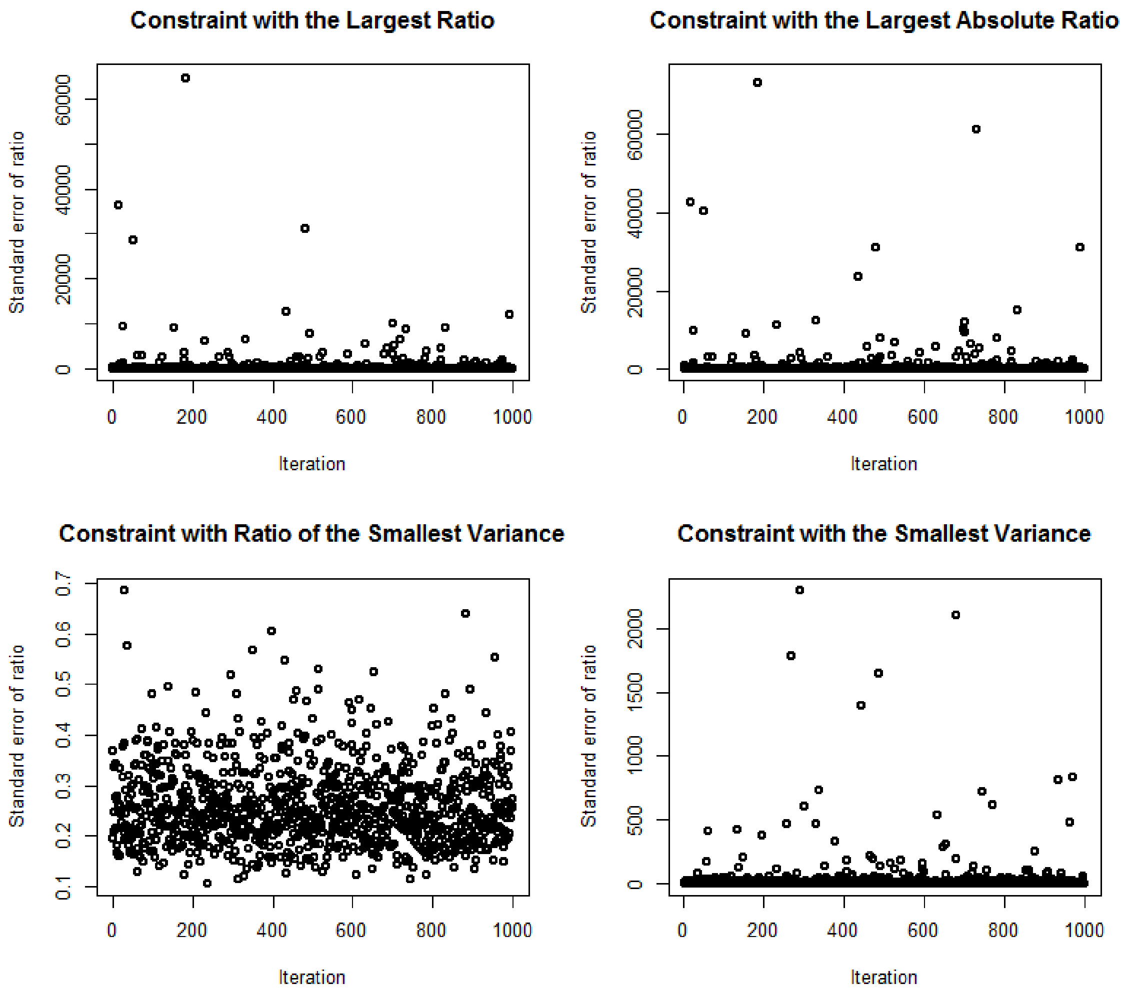

Figure 3.5.7: Standard Error of Ratios by Period-effect Constraints with Three Constraint-selection Methods

3.5.6 Optimal Constraint Estimation Compared with Intrinsic Estimator

Comparison of the smoothing cohort stage 1, the smoothing cohort stage 2 with the optimal constraint-selection method, as well as the intrinsic estimator is given in Table 3.5.6 and

Figure 3.5.8. For age and period effects, all smoothing stage 1, smoothing stage 2 and intrinsic estimator yield consistent estimation. But for cohort effect, the smoothing stage 2 and the intrinsic estimator yield smaller bias than the smoothing stage 1 model. So the two-stage smoothing cohort model is a competitive APC analysis tool, as well as the intrinsic estimator.

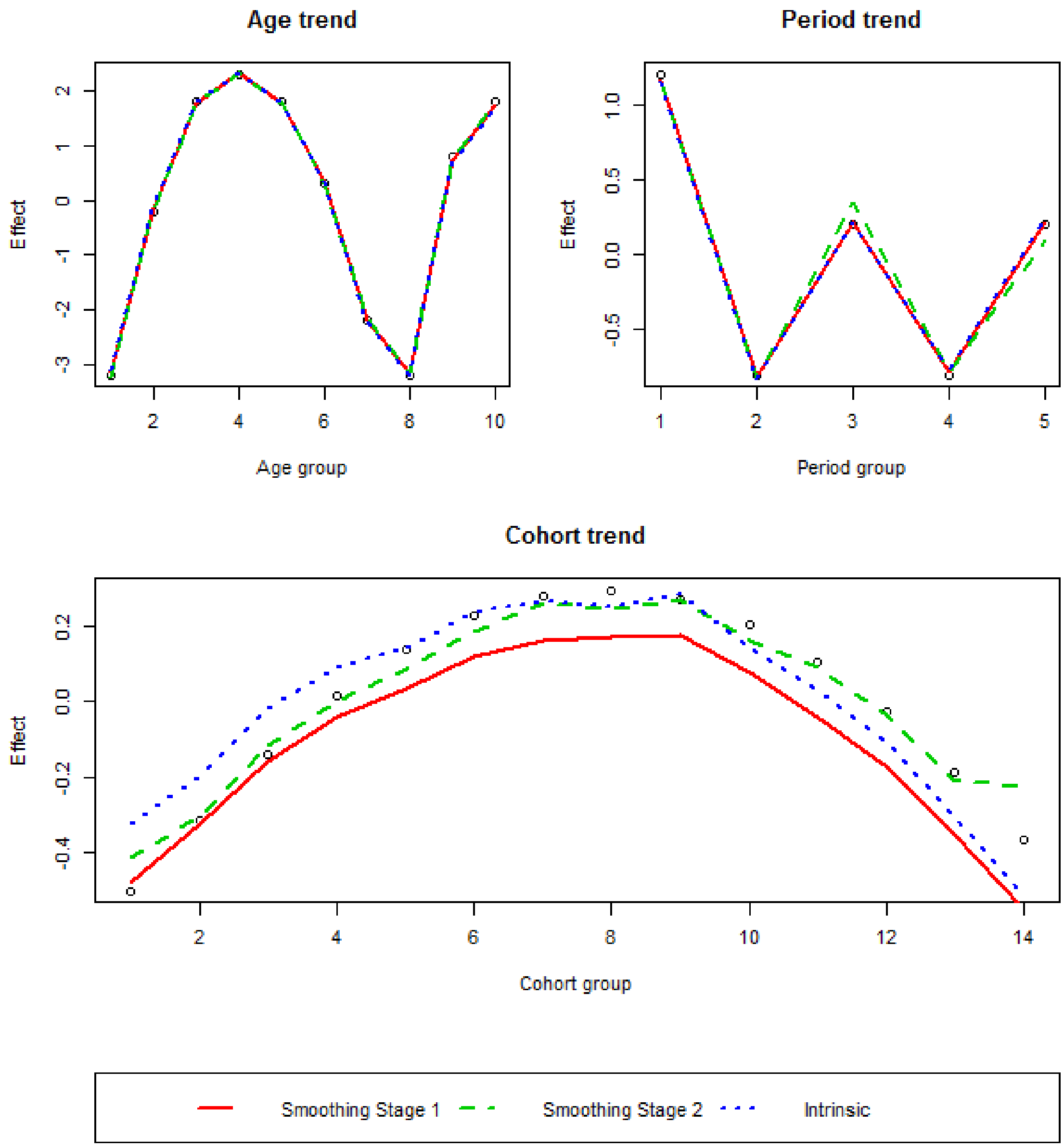

Figure 3.5.8: Comparison of the Two-stage Smoothing Cohort Model with the Period-effect Constraint of the Smallest Variance of the Ratio and the Intrinsic Estimator for Simulation I

Table 3.5.6: Comparison of the Two-stage Smoothing Cohort Model by the Smallest Variance of Period Ratio and the Intrinsic Estimator for Simulation I

		Smoothing Stage 1			Smoothing Stage 2			Intrinsic		
	Specified	Estimate	MSE	Bias	Estimtate	MSE	Bias	Estiamte	MSE	Bias
μ	1.0000	1.1090	0.0119	0.1090	1.0172	0.0003	0.0172	1.0092	0.0001	0.0092
α_1	-3.2000	-3.1562	0.0023	0.0438	-3.2380	0.0022	-0.0380	-3.1165	0.0074	0.0835
α_2	-0.2000	-0.1485	0.0029	0.0515	-0.1721	0.0013	0.0279	-0.1224	0.0063	0.0776
α_3	1.8000	1.7928	0.0004	-0.0072	1.7829	0.0007	-0.0171	1.8128	0.0005	0.0128
α_4	2.3000	2.3324	0.0014	0.0324	2.3389	0.0019	0.0389	2.3448	0.0023	0.0448
α_5	1.8000	1.7762	0.0009	-0.0238	1.7747	0.0010	-0.0253	1.7792	0.0008	-0.0208
α_6	0.3000	0.3434	0.0022	0.0434	0.3234	0.0009	0.0234	0.3380	0.0018	0.0380
α_7	-2.2000	-2.2088	0.0004	-0.0088	-2.1798	0.0008	0.0202	-2.2195	0.0007	-0.0195
α_8	-3.2000	-3.1923	0.0004	0.0077	-3.1758	0.0010	0.0242	-3.2152	0.0005	-0.0152
α_9	0.8000	0.7314	0.0050	-0.0686	0.7597	0.0022	-0.0403	0.7040	0.0095	-0.0960
α_{10}	1.8000	1.7297	0.0053	-0.0703	1.7861	0.0009	-0.0139	1.6947	0.0115	-0.1053
β_1	1.2000	1.1673	0.0012	-0.0327	1.1539	0.0023	-0.0461	1.1512	0.0025	-0.0488
β_2	-0.8000	-0.8190	0.0005	-0.0190	-0.8113	0.0003	-0.0113	-0.8265	0.0008	-0.0265
β_3	0.2000	0.2124	0.0003	0.0124	0.3442	0.0209	0.1442	0.2110	0.0002	0.0110
β_4	-0.8000	-0.7774	0.0006	0.0226	-0.7795	0.0006	0.0205	-0.7696	0.0010	0.0304
β_5	0.2000	0.2167	0.0004	0.0167	0.0927	0.0116	-0.1073	0.2339	0.0013	0.0339
γ_1	-0.5046	-0.4836	0.0018	0.0210	-0.4149	0.0104	0.0897	-0.3266	0.0330	0.1780
γ_2	-0.3139	-0.3264	0.0005	-0.0125	-0.3086	0.0013	0.0053	-0.1964	0.0145	0.1175
γ_3	-0.1387	-0.1607	0.0007	-0.0220	-0.1190	0.0013	0.0197	-0.0180	0.0151	0.1207
γ_4	0.0141	-0.0405	0.0032	-0.0546	-0.0001	0.0008	-0.0142	0.0939	0.0068	0.0798
γ_5	0.1382	0.0356	0.0107	-0.1026	0.0880	0.0030	-0.0502	0.1457	0.0004	0.0075
γ_6	0.2287	0.1211	0.0118	-0.1076	0.1879	0.0021	-0.0408	0.2398	0.0005	0.0111
γ_7	0.2821	0.1623	0.0145	-0.1198	0.2629	0.0008	-0.0192	0.2697	0.0006	-0.0124
γ_8	0.2963	0.1722	0.0156	-0.1241	0.2501	0.0026	-0.0462	0.2540	0.0022	-0.0423
γ_9	0.2705	0.1779	0.0088	-0.0926	0.2711	0.0004	0.0006	0.2855	0.0006	0.0150
γ_{10}	0.206	0.0774	0.0167	-0.1286	0.1635	0.0022	-0.0425	0.1459	0.0039	-0.0601
γ_{11}	0.1052	-0.0396	0.0211	-0.1448	0.0923	0.0008	-0.0129	0.0320	0.0058	-0.0732
γ_{12}	-0.0278	-0.1756	0.0221	-0.1478	-0.0338	0.0009	-0.0060	-0.1060	0.0066	-0.0782
γ_{13}	-0.1878	-0.3558	0.0286	-0.1680	-0.2126	0.0019	-0.0248	-0.3049	0.0144	-0.1171
γ_{14}	-0.3683	-0.5546	0.0361	-0.1863	-0.2269	0.0221	0.1414	-0.5146	0.0233	-0.1463

3.6 Numerical Simulation 2

It is also of interest to examine the behavior of different methods when the cohort effect is not smooth. In this simulation, 25 age groups and 10 period groups are specified, and 1000 simulation runs are conducted. By the given parameter values, the age effect has an increasing trend, the period effect has a U-shaped trend, and the cohort effect is periodic with overall concave up trend. Figure 3.6.1 shows the trend estimated by smoothing cohort stage 1, smoothing cohort stage 2, and the intrinsic estimator, compared with specified values. Table 3.6.1 and Table 3.6.2 display the estimates of the age, period, and cohort effects, their SE and Bias of smoothing cohort stage 1, smoothing cohort stage 2, and the intrinsic estimator for this simulation study.

All three methods yield accurate estimation for age and period effects. Both the smoothing cohort stage 1 and stage 2 models yield accurate estimation for the cohort effect. However, the intrinsic estimator shows the periodic shape but overall biased cohort effect. This may be explained by the fact that large sample behavior warrants convergence of the period effect in this study, because $a > p$, but cannot warrant the convergence of the age and cohort effects as profile estimates, see Chapter 4.

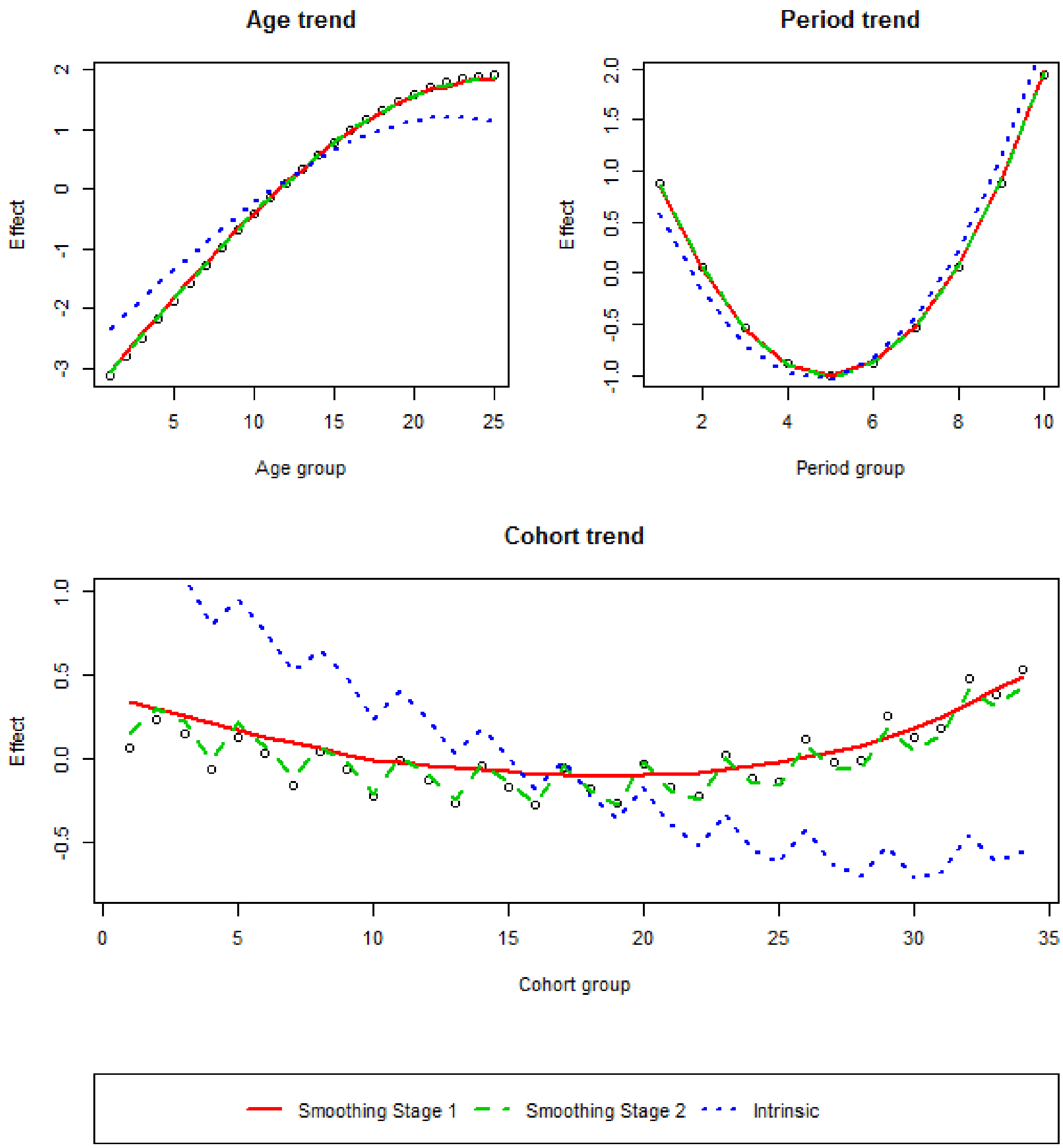

Figure 3.6.1: Comparison of the Two-stage Smoothing Cohort Model with the Smallest Variance of Ratio Period-effect Constraint and the Intrinsic Estimator for Simulation II

Table 3.6.1: Age and Period Effects Comparison of the Two-stage Smoothing Cohort Model by the Smallest Variance of Period Ratio and the Intrinsic Estimator for Simulation II

		Smoothing Cohort Stage 1			Smoothing Cohort Stage 2			Intrinsic		
	Specified	Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
μ	1.0000	0.9344	0.0043	-0.0656	1.0021	0.0000	0.0021	1.0021	0.0000	0.0021
α_1	-3.1131	-3.0534	0.0037	0.0597	-3.0635	0.0030	0.0496	-2.3381	0.6008	0.7750
α_2	-2.8000	-2.7399	0.0037	0.0601	-2.7446	0.0035	0.0554	-2.0797	0.5190	0.7203
α_3	-2.4869	-2.4173	0.0050	0.0696	-2.4384	0.0027	0.0485	-1.8339	0.4265	0.6530
α_4	-2.1752	-2.1277	0.0024	0.0474	-2.1243	0.0029	0.0509	-1.5803	0.3541	0.5949
α_5	-1.8662	-1.8159	0.0026	0.0503	-1.8146	0.0030	0.0516	-1.3310	0.2866	0.5353
α_6	-1.5614	-1.5237	0.0015	0.0377	-1.5434	0.0006	0.0180	-1.1202	0.1947	0.4411
α_7	-1.2620	-1.2470	0.0003	0.0150	-1.2442	0.0005	0.0178	-0.8815	0.1449	0.3805
α_8	-0.9694	-0.9573	0.0003	0.0120	-0.9597	0.0003	0.0096	-0.6575	0.0974	0.3119
α_9	-0.6848	-0.6575	0.0009	0.0272	-0.6752	0.0003	0.0096	-0.4334	0.0633	0.2514
α_{10}	-0.4095	-0.3975	0.0003	0.0121	-0.3899	0.0005	0.0196	-0.2086	0.0405	0.2010
α_{11}	-0.1448	-0.1351	0.0002	0.0098	-0.1352	0.0002	0.0097	-0.0143	0.0172	0.1306
α_{12}	0.1082	0.1220	0.0003	0.0138	0.1080	0.0001	-0.0002	0.1684	0.0038	0.0602
α_{13}	0.3484	0.3006	0.0024	-0.0478	0.3114	0.0015	-0.0370	0.3114	0.0015	-0.0370
α_{14}	0.5748	0.5667	0.0002	-0.0081	0.5668	0.0002	-0.0080	0.5063	0.0048	-0.0685
α_{15}	0.7863	0.7941	0.0002	0.0078	0.7811	0.0002	-0.0052	0.6602	0.0160	-0.1261
α_{16}	0.9820	0.9711	0.0002	-0.0109	0.9863	0.0002	0.0043	0.8050	0.0315	-0.1770
α_{17}	1.1611	1.1468	0.0003	-0.0143	1.1492	0.0003	-0.0118	0.9074	0.0644	-0.2536
α_{18}	1.3227	1.3012	0.0006	-0.0214	1.2929	0.0011	-0.0298	0.9906	0.1104	-0.3320
α_{19}	1.4661	1.4335	0.0012	-0.0327	1.4500	0.0005	-0.0162	1.0873	0.1436	-0.3789
α_{20}	1.5908	1.5656	0.0007	-0.0252	1.5703	0.0007	-0.0205	1.1471	0.1969	-0.4437
α_{21}	1.6961	1.6746	0.0006	-0.0215	1.6696	0.0010	-0.0266	1.1860	0.2604	-0.5102
α_{22}	1.7817	1.7181	0.0041	-0.0636	1.7393	0.0021	-0.0424	1.1952	0.3441	-0.5865
α_{23}	1.8470	1.8003	0.0023	-0.0467	1.8066	0.0020	-0.0404	1.2021	0.4161	-0.6449
α_{24}	1.8919	1.8443	0.0024	-0.0476	1.8450	0.0027	-0.0469	1.1801	0.5069	-0.7119
α_{25}	1.9162	1.8332	0.0070	-0.0830	1.8566	0.0041	-0.0596	1.1312	0.6163	-0.7850
β_1	0.8824	0.8488	0.0012	-0.0336	0.8573	0.0007	-0.0251	0.5849	0.0885	-0.2974
β_2	0.0588	0.0470	0.0002	-0.0119	0.0503	0.0001	-0.0086	-0.1671	0.0511	-0.2260
β_3	-0.5294	-0.5559	0.0007	-0.0265	-0.5564	0.0008	-0.0269	-0.7054	0.0310	-0.1760
β_4	-0.8824	-0.9022	0.0004	-0.0198	-0.8964	0.0002	-0.0140	-0.9871	0.0110	-0.1047
β_5	-1.0000	-1.0112	0.0002	-0.0112	-1.0167	0.0003	-0.0167	-1.0469	0.0022	-0.0469
β_6	-0.8824	-0.8662	0.0003	0.0161	-0.8660	0.0003	0.0164	-0.8346	0.0023	0.0477
β_7	-0.5294	-0.5181	0.0002	0.0113	-0.5180	0.0002	0.0114	-0.4237	0.0112	0.1057
β_8	0.0588	0.0753	0.0003	0.0165	0.0687	0.0001	0.0099	0.2189	0.0257	0.1600
β_9	0.8824	0.9154	0.0011	0.0330	0.9126	0.0010	0.0302	1.1242	0.0585	0.2418
β_{10}	1.9412	1.9673	0.0007	0.0261	1.9645	0.0006	0.0234	2.2369	0.0875	0.2957

Table 3.6.2: Cohort Effect Comparison of the Two-stage Smoothing Cohort Model by the Smallest Variance of Period Ratio and the Intrinsic Estimator for Simulation II

		Smoothing Cohort Stage 1			Smoothing Cohort Stage 2			Intrinsic		
	Specified	Estimate	MSE	Bias	Estimtate	MSE	Bias	Estimtate	MSE	Bias
γ_1	0.0584	0.3382	0.0786	0.2798	0.1463	0.0097	0.0879	1.1437	1.1791	1.0854
γ_2	0.2300	0.2979	0.0048	0.0679	0.2977	0.0058	0.0677	1.2347	1.0100	1.0047
γ_3	0.1482	0.2543	0.0114	0.1062	0.2200	0.0060	0.0718	1.0965	0.8996	0.9483
γ_4	-0.0627	0.2109	0.0749	0.2736	-0.0176	0.0028	0.0451	0.7985	0.7419	0.8611
γ_5	0.1237	0.1722	0.0024	0.0485	0.2059	0.0074	0.0822	0.9616	0.7022	0.8379
γ_6	0.0301	0.1292	0.0099	0.0991	0.0710	0.0022	0.0409	0.7662	0.5420	0.7361
γ_7	-0.1582	0.0880	0.0606	0.2462	-0.1108	0.0027	0.0474	0.5239	0.4655	0.6821
γ_8	0.0414	0.0561	0.0002	0.0147	0.0762	0.0016	0.0348	0.6505	0.3711	0.6091
γ_9	-0.0633	0.0205	0.0071	0.0838	-0.0233	0.0019	0.0400	0.4906	0.3069	0.5539
γ_{10}	-0.2277	-0.0108	0.0471	0.2170	-0.2201	0.0003	0.0076	0.2333	0.2126	0.4610
γ_{11}	-0.0164	-0.0250	0.0001	-0.0085	0.0101	0.0009	0.0265	0.4030	0.1761	0.4195
γ_{12}	-0.1314	-0.0433	0.0078	0.0881	-0.0996	0.0012	0.0318	0.2329	0.1328	0.3643
γ_{13}	-0.2705	-0.0610	0.0439	0.2095	-0.2439	0.0009	0.0265	0.0281	0.0893	0.2986
γ_{14}	-0.0490	-0.0679	0.0004	-0.0189	-0.0342	0.0004	0.0147	0.1773	0.0514	0.2263
γ_{15}	-0.1731	-0.0819	0.0084	0.0913	-0.1451	0.0009	0.0280	0.0060	0.0322	0.1791
γ_{16}	-0.2855	-0.0947	0.0364	0.1907	-0.2717	0.0003	0.0138	-0.1810	0.0110	0.1045
γ_{17}	-0.0552	-0.0969	0.0018	-0.0417	-0.0473	0.0002	0.0079	-0.0171	0.0016	0.0382
γ_{18}	-0.1871	-0.1048	0.0068	0.0823	-0.1924	0.0002	-0.0053	-0.2226	0.0014	-0.0355
γ_{19}	-0.2713	-0.1065	0.0272	0.1648	-0.2781	0.0002	-0.0069	-0.3688	0.0097	-0.0975
γ_{20}	-0.0333	-0.0964	0.0040	-0.0631	-0.0239	0.0002	0.0095	-0.1750	0.0202	-0.1417
γ_{21}	-0.1710	-0.0948	0.0058	0.0762	-0.1898	0.0005	-0.0188	-0.4014	0.0532	-0.2303
γ_{22}	-0.2253	-0.0873	0.0191	0.1381	-0.2508	0.0008	-0.0254	-0.5228	0.0886	-0.2975
γ_{23}	0.0198	-0.0665	0.0075	-0.0863	-0.0016	0.0007	-0.0214	-0.3341	0.1254	-0.3539
γ_{24}	-0.1211	-0.0507	0.0050	0.0704	-0.1563	0.0015	-0.0352	-0.5492	0.1835	-0.4282
γ_{25}	-0.1430	-0.0267	0.0135	0.1163	-0.1670	0.0009	-0.0240	-0.6204	0.2280	-0.4774
γ_{26}	0.1099	0.0078	0.0105	-0.1022	0.0836	0.0010	-0.0263	-0.4302	0.2919	-0.5401
γ_{27}	-0.0297	0.0353	0.0042	0.0649	-0.0683	0.0019	-0.0386	-0.6426	0.3758	-0.6129
γ_{28}	-0.0148	0.0699	0.0072	0.0847	-0.0696	0.0034	-0.0547	-0.7043	0.4755	-0.6895
γ_{29}	0.2496	0.1197	0.0169	-0.1299	0.1731	0.0064	-0.0766	-0.5221	0.5958	-0.7718
γ_{30}	0.1202	0.1767	0.0032	0.0566	0.0392	0.0072	-0.0810	-0.7165	0.7001	-0.8366
γ_{31}	0.1826	0.2473	0.0043	0.0648	0.1314	0.0033	-0.0512	-0.6847	0.7524	-0.8673
γ_{32}	0.4723	0.3284	0.0208	-0.1439	0.4105	0.0048	-0.0618	-0.4661	0.8809	-0.9384
γ_{33}	0.3779	0.4089	0.0012	0.0311	0.3102	0.0059	-0.0676	-0.6268	1.0099	-1.0046
γ_{34}	0.5257	0.4889	0.0017	-0.0368	0.4364	0.0100	-0.0893	-0.5610	1.1822	-1.0867

3.7 Summary

Based on the above simulation studies and comparison of different methods, we conclude that

1. Smoothing cohort stage 1 model yields consistent estimation for age and period effects, but not cohort effects.

2. Bias in cohort-effect estimation can be corrected in the second stage with a constraint specified using consistent age- or period-effect estimates based on the first stage estimation.

3. Since the consistency is for age (row) effect as $p \to \infty$, often, a larger $p > a$ is preferred. In general, the constraint on the effect of smaller group size yields smaller bias. If $a < p$, age-effect constraints are recommended; otherwise, period-effect constraints are recommended.

4. The constraint chosen with the ratio of the smallest variance is optimal in the second stage of the smoothing cohort model.

CHAPTER 4

Asymptotic Analysis

In this chapter, we study the asymptotics of the estimator of the smoothing cohort model. Let $\mathbf{b}^T = (\theta^T, \xi^T)$, where $\theta^T = (\mu, \alpha_1, ..., \alpha_{a-1})$ represents the intercept and row effects, $\xi^T = (\beta_1, \cdots, \beta_{p-1}, \gamma_1, \cdots, \gamma_{a+p-2})$ represents the column and diagonal effects. In the APC model, due to human life limitation, a is limited to be finite. To study large sample behavior, we set $p \to \infty$. So θ^T is the parameter of primary interest and ξ^T a nuisance parameter, as its number of parameters diverges to infinity.

4.1 Introduction

Assume the random variable Y_{ij} with $i = 1,...,a$ and $j = 1,...,p$ follows a distribution in the exponential family [McCullagh and Nelder, 1989], then the log-likelihood function can be written as

$$l(\mathbf{b};\mathbf{y}) = \frac{\mathbf{y}\zeta(\mathbf{b}) - \psi(\zeta(\mathbf{b}))}{\kappa(\phi)} + c(\mathbf{y},\phi), \tag{4.1.1}$$

where ζ is a parameter, $\kappa(\phi)$ is a dispersion parameter, and link function $g(EY) = \eta$ for the model $\eta = X'\mathbf{b}$.

The log-likelihood function for the smoothing model can be written as

$$l_S(\mu,\alpha,\beta,\gamma;y) = \sum_{i=1}^{a}\sum_{j=1}^{p} l(y_{ij},\mu,\alpha,\beta,S\gamma). \tag{4.1.2}$$

In order to study the consistency of the estimator for θ as $p \to \infty$, the profiled log-likelihood function for θ is considered since there will be infinitely many β_j and γ_k.

Define a profile log-likelihood function for smoothing model,

$$Pl_S(\theta) = l_S(\theta,\xi_S(\theta)), \tag{4.1.3}$$

where $\xi_S(\theta) = \arg\max_{\xi} l_S(\theta,\xi;\mathbf{y})$. Denote the maximum profile likelihood estimator (MaPLE) of the profile log-likelihood function $Pl_S(\theta)$ as $\tilde{\theta}_S^T = (\tilde{\mu}_S,\tilde{\alpha}_S^T)$.

The log-likelihood function for the intrinsic estimator can be written as

$$l_I(\mu,\alpha,\beta,\gamma;y) = \sum_{i=1}^{a}\sum_{j=1}^{p} l(y_{ij},\mu,\alpha,\beta,\gamma) - \delta(\mathbf{b}^T B_0)^2 \quad \text{for} \quad \delta > 0. \quad (4.1.4)$$

Similarly, define the profile log-likelihood functions for $l(\theta,\xi)$ and $l_I(\theta,\xi)$:

$$Pl(\theta) = l(\theta,\xi(\theta)), \quad (4.1.5)$$

and

$$Pl_I(\theta) = l_I(\theta,\xi_I(\theta)), \quad (4.1.6)$$

where $\xi(\theta) = \arg\max l(\theta,\xi;y)$, $\xi_I(\theta) = \arg\max_{\xi} l_I(\theta,\xi;y)$. Denote by $\tilde{\theta}_I = (\tilde{\mu}_I,\tilde{\alpha}_I^T)$ the MaPLE of $Pl_I(\theta)$.

Note that the basic idea for the intrinsic estimator is to find the estimator B s.t. $B \perp B_0$. This can be achieved by adding a penalty term to the profile log-likelihood function of $Pl_I(\theta)$ with $\delta > 0$

$$\begin{aligned} Pl_I(\theta) &= l_I(\theta,\xi_I(\theta)) \\ &= Pl(\theta) - \delta(\theta_0^T\theta + \xi_0^T\xi(\theta))^2 \\ &= l(\theta,\xi(\theta)) - \delta(\theta_0^T\theta + \xi_0^T\xi(\theta))^2, \end{aligned} \quad (4.1.7)$$

where θ_0 is the intercept and age-effect vector and ξ_0 is the period- and cohort-effect vector in B_0.

Properties of the Intrinsic Estimator

We need the following regularity conditions for the asymptotics studies.

The Regularity Conditions:

1. The parameter space $\mathbb{D} = D_1 \times D_2 \times \cdots \times D_{2a+2p-3} \in \mathbb{R}^{2a+2p-3}$ for model parameters $\mathbf{b}$ has its subspaces D_j uniformly bounded with respect to p for $j = 1, \cdots, 2a+2p-3$.

2. The log-likelihood $l(\mathbf{b};\mathbf{y})$ of the exponential family is continuously differentiable with continuous derivatives $\frac{\partial^2 l}{\partial b_i \partial b_j}$ and a uniform bound $\left|\frac{\partial l}{\partial b_j}\right| < M < \infty$ for $\mathbf{b} \in \mathbb{D}, i, j = 1, \cdots, 2a+2p-3$.

Lemma 4.1. *[Fu, 2016] Under the regularity conditions 1 & 2, there exists a unique set of true parameters of intercept, row and column effects. Denote these true parameters as μ^∞, α^∞ and β^∞, which are the limits of the intrinsic estimator.*

Proposition 4.2. *[Fu, 2016] Under the regularity conditions 1 & 2, consider a linear constraint on row effects $l^T\alpha = 0$:*

1. *If $l^T\theta^\infty = 0$, that is, the constraint is satisfied by true row effect parameters, the row parameter estimates $\tilde{\alpha}_p$ is $\sqrt{p}$-consistent, i.e., $\sqrt{p}(\tilde{\alpha}_p - \alpha^\infty) \to N(0, \Sigma_r)$ as $p \to \infty$ for some positive definite $(a-1)\times(a-1)$ variance-covariance matrix Σ_r.*

2. *If $l^T\theta^\infty \neq 0$, that is, the constraint is not satisfied by true parameters, the constraint yields asymptotic bias.*

Proposition 4.3. *[Fu, 2016] Under the regularity conditions 1 & 2, consider a linear constraint on column or diagonal effects $l^T\xi = 0$:*

1. *If the constraint is satisfied by true parameter values, row effect estimation $\tilde{\alpha}_p$ is $\sqrt{p}-$consistent.*

2. *If the constraint is not satisfied by true parameters, a contrast constraint yields asymptotic bias, which does not depend on p.*

3. *If the constraint is not satisfied by true parameters, a non-contrast constraint yields consistent estimates with the asymptotic bias of the order of $O(p^{-1}) \to 0$ as $p \to \infty$.*

Propositions 4.2 and 4.3 are the ideas we followed when doing stage 2 smoothing cohort constraint selection.

Properties of Smoothing Cohort Model

Proposition 4.4. *[Fu, 2008] Assume the following bounded cohort condition:*

$$|S\gamma_k| \leq C(p), \quad for \quad k = 1, 2, ... \tag{4.1.8}$$

1. *If $C(p) = o(p)$, the smoothing cohort model yields consistent estimation for model intercept μ and age effects $\alpha_1, \cdots, \alpha_a$ as the number of period groups p goes to infinity.*

2. *If $C(p) = o(\sqrt{p})$, the $\sqrt{p}$ consistency also holds for the above estimation.*

Proposition 4.4 shows that smoothing cohort stage 1 model yields consistent age- or period-effect estimation, which guarantees constraint in smoothing cohort stage 2 model will not be 'far' away from true parameters, so the stage 2 constraint estimator is consistent, hence corrects the bias.

Lemma 4.5. *[Nadaraya, 1964]Under assumptions*

1. $\int |K(u)|\, du < \infty$

2. $uK(u) \to 0$, *as* $|u| \to \infty$

3. $EY^2 < \infty$

$$s_{NW}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)} \xrightarrow{p} f(x) \tag{4.1.9}$$

as $h \to 0$ *and* $nh \to \infty$, *where* $f(\cdot)$ *is the function satisfying* $Y = f(X)$.

4.2 Asymptotics of Smoothing Cohort Model Estimation

Lemma 4.6. *The MaPLE* $\tilde{\theta}_{S,p}$ *of the profile log-likelihood* $Pl_S(\theta; y_p)$ *satisfies*

$$\frac{Pl_S(\tilde{\theta}_{S,p})}{p} \xrightarrow{p} -\frac{2}{a}\kappa(\phi) \quad as \quad p \to \infty \tag{4.2.1}$$

Proof. We abuse the notation a little without confusion and write the model deviance $Dev(b; y) = 2l_S(y; y) - 2l_S(b; y)$ as $Dev(b; y) = -2l_S(b; y)$. The deviance is asymptotically distributed as $\kappa(\phi)\chi_d^2$ with d degrees of freedom and dispersion parameter $\kappa(\phi)$ [Murphy and Van der Vaart, 2000]. With the submodel having a parameters under the profile log-likelihood, $d = (ap - a)$. Hence, the deviance of the submodel $Dev(\theta; y)$ satisfies

$$\frac{Dev(\theta; y)}{p} = \frac{Dev(\theta, \xi_S(\theta); y)}{p} \sim \frac{\kappa(\phi)\chi_{ap-a}^2}{p} \quad as \quad p \to \infty.$$

Let $Z_p \sim \chi^2_{ap-a}$. By Chebyshev's inequality on the random variable $Z_p\kappa(\phi)/p$ with any fixed $\varepsilon > 0$,

$$P\Big(\Big|\frac{\kappa(\phi)Z_p}{p} - E\big[\frac{\kappa(\phi)Z_p}{p}\big]\Big| > \varepsilon\Big) \leq var\big[\frac{\kappa(\phi)Z_p}{p}\big]/\varepsilon^2.$$

Since $E\big[\frac{\kappa(\phi)Z_p}{p}\big] = \kappa(\phi)(ap-a)/p \to a\kappa(\phi)$ and $var\big[\frac{\kappa(\phi)Z_p}{p}\big] = 2(ap-a)\kappa^2(\phi)/p^2$,

$$\begin{aligned} P\Big(\Big|\frac{\kappa(\phi)Z_p}{p} - a\kappa(\phi)\Big| > \varepsilon\Big) &\leq 2(ap-a)\kappa^2(\phi)/(p^2\varepsilon^2) \\ &< 2a\kappa^2(\phi)/(p\varepsilon^2) \to 0 \quad as \quad p \to \infty. \end{aligned} \tag{4.2.2}$$

It is followed by the convergence in probability $\kappa(\phi)Z_p/p \to_p a\kappa(\phi)$ as $p \to \infty$. Hence $\frac{Pl_S(\tilde{\theta}_{S,p})}{p} \to_p -\frac{a}{2}\kappa(\phi)$. □

Lemma 4.7. *For any bandwidth $h > 0$,*

$$\frac{\partial Pl_S(\theta)}{\partial \theta} = \frac{\partial l_S(\theta,\xi)}{\partial \theta}\Big|_{\xi=\xi_S(\theta)}, \quad \forall\theta \tag{4.2.3}$$

and the limit of the partial derivative of the profile log-likelihood exists

$$-\frac{1}{p}\frac{\partial^2 Pl_S(\theta;y)}{\partial \theta^2} \xrightarrow{p} C_1 \quad as \quad p \to \infty, \tag{4.2.4}$$

where C_1 is a positive definite $a \times a$ Fisher information matrix of the row effect model, and is also independent of $h > 0$.

Proof. By the chain rule $\frac{\partial l_S(\theta,\xi)}{\partial \xi}|_{\xi=\xi_S(\theta)} = 0$ for $\forall\theta$ and $\frac{\partial^2 l_S(\theta,\xi;y)}{\partial \xi \partial \theta}|_{\xi=\xi_S(\theta)} = 0$ for smooth likelihood.

$$
\begin{aligned}
\frac{\partial Pl_S(\theta)}{\partial \theta} &= \frac{\partial l_S(\theta, \xi_S(\theta))}{\partial \theta} \\
&= \frac{\partial l_S(\theta, \xi)}{\partial \theta}|_{\xi=\xi_S(\theta)} + \frac{\partial l_S(\theta, \xi)}{\partial \xi}\frac{\partial \xi_S(\theta)}{\partial \theta}|_{\xi=\xi_S(\theta)} \\
&= \frac{\partial l_S(\theta, \xi)}{\partial \theta}|_{\xi=\xi_S(\theta)}.
\end{aligned}
$$

For large p,

$$
-\frac{\partial^2 Pl_S(\theta)}{\partial \theta^2} = -\frac{\partial^2 l_S(\theta, \xi; y)}{\partial \theta^2} \approx pC_1, \forall \theta \tag{4.2.5}
$$

where C_1 is the Fisher information matrix for the row effect model $l_S(\theta; y)$ with parameters θ, and is positive-definite. Since C_1 is derived from only the second order partial derivative of the log-likelihood function with respect to θ, it does not involve the cohort effects and thus is also independent of the smoothing bandwidth $h > 0$. □

Theorem 4.8. *Under the regularity conditions 1 & 2, the limit of the MaPLE $\tilde{\theta}_{S,h,p}$ exists for any smoothing bandwidth $h > 0$ as $p \to \infty$, i.e.,*

$$
\tilde{\theta}_{S,h,p} \xrightarrow{p} \theta^{\infty}_{S,h} \quad as \quad p \to \infty \quad and \quad \forall h > 0,
$$

for some $\theta^{\infty}_{S,h}$.

Proof. For any bandwidth $h > 0$, consider a Cauchy sequence $\{\tilde{\theta}_{S,p+n}\}$ of the MaPLE $\tilde{\theta}_{S,p}$ for arbitrarily large number p and any finite number n. Take Taylor expansion of the profile log-likelihood $Pl_{S,p}(\theta; y)$ with y in a table of a rows and an increasing number

$(p+n)$ columns, where the log-likelihood $Pl_{S,p}(\theta;y)$ assumes the values for y in the first p columns.

$$\begin{aligned}
Pl_{S,p}(\tilde{\theta}_{S,p+n};y) &= Pl_{S,p}(\tilde{\theta}_{S,p};y) + \frac{\partial Pl_{S,p}(\theta;y)}{\partial\theta}|_{\theta=\tilde{\theta}_{S,p}}(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}) \\
&+ \frac{1}{2}(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p})^T \frac{\partial^2 Pl_{S,p}(\theta;y)}{\partial\theta^2}|_{\theta=\tilde{\theta}_{S,p}}(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}) \\
&+ o(||\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}||^2) \\
&= Pl_{S,p}(\tilde{\theta}_{S,p};y) + \frac{1}{2}(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p})^T \frac{\partial^2 Pl_{S,p}(\theta;y)}{\partial\theta^2}|_{\theta=\tilde{\theta}_{S,p}}(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}) \\
&+ o(||\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}||^2). \qquad (4.2.6)
\end{aligned}$$

The likelihood functions satisfy $l_{S,p+n}(\tilde{\theta}_{S,p+n},\xi_S(\tilde{\theta}_{p+n});y) \le l_{S,p}(\tilde{\theta}_{S,p+n},\xi_S(\tilde{\theta}_{S,p+n});y) \le l_{S,p}(\tilde{\theta}_{S,p},\xi_S(\tilde{\theta}_{S,p});y)$, following the discussion in [Fu, 2016]. For large p, $\forall h>0, n>0$,

$$\begin{aligned}
& \frac{1}{p}\left|Pl_{S,p+n}(\tilde{\theta}_{p+n};y) - Pl_{S,p}(\tilde{\theta}_{S,p+n};y)\right| \\
\le & \frac{1}{p}\left|l_{S,p+n}(\tilde{\theta}_{S,p+n},\xi_S(\tilde{\theta}_{p+n});y) - l_{S,p}(\tilde{\theta}_{S,p+n},\xi_S(\tilde{\theta}_{S,p+n});y)\right| \\
\le & \frac{1}{p}\left|l_{S,p+n}(\tilde{\theta}_{S,p+n},\xi_S(\tilde{\theta}_{p+n});y) - l_{S,p}(\tilde{\theta}_{S,p},\xi_S(\tilde{\theta}_{S,p});y)\right| \\
= & \left|\frac{p+n}{p}\frac{1}{p+n}Pl_{S,p+n}(\tilde{\theta}_{S,p+n};y) - \frac{1}{p}Pl_{S,p}(\tilde{\theta}_{S,p};y)\right| \\
\xrightarrow{p} & \; 0 \quad as \quad p\to\infty \qquad (4.2.7)
\end{aligned}$$

by Lemma 4.6. Then by Equation (4.2.6) and Lemma 4.7,

$$(\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p})^T C_1 (\tilde{\theta}_{S,p+n}-\tilde{\theta}_{S,p}) \le \frac{2}{p}\left|Pl_{S,p+n}(\tilde{\theta}_{S,p+n};y) - Pl_{S,p}(\tilde{\theta}_{S,p};y)\right|$$

$$
\begin{aligned}
&\leq \frac{2}{p}\left|Pl_{S,p}(\tilde{\theta}_{S,p+n}) - Pl_{S,p+n}(\tilde{\theta}_{S,p+n})\right| + 2\left|\frac{p+n}{p}\frac{1}{p+n}Pl_{S,p+n}(\tilde{\theta}_{S,p+n};y) - \frac{1}{p}Pl_{S,p}(\tilde{\theta}_{S,p};y)\right| \\
&\xrightarrow{p} 0 \quad as \quad p \to \infty,
\end{aligned}
$$

which implies the convergence in probability of the Cauchy sequence $\tilde{\theta}_{S,p+n}$ as $p \to \infty$. Hence there exists an a-dimensional vector $\theta_{S,h}^{\infty}$ such that $\tilde{\theta}_{S,h,p} \xrightarrow{p} \theta_{S,h}^{\infty}$ as $p \to \infty$. □

Theorem 4.9. *Under the regularity conditions:*

1. *The parameter space $\mathbb{D} = D_1 \times D_2 \times \cdots \times D_{2a+2p-3} \in \mathbb{R}^{2a+2p-3}$ for model parameters $\mathbf{b}$ has its subspaces D_j uniformly bounded with respect to p for $j = 1, \cdots, 2a+2p-3$,*
2. *The log-likelihood $l(\mathbf{b};\mathbf{y})$ of the exponential family is continuously differentiable with continuous derivatives $\frac{\partial^2 l}{\partial b_i \partial b_j}$ and a uniform bound $\left|\frac{\partial l}{\partial b_j}\right| < M < \infty$ for $\mathbf{b} \in \mathbb{D}, i, j = 1, \cdots, 2a+2p-3$,*
3. $\int |K(u)|\, du < \infty$,
4. $uK(u) \to 0$, *as* $|u| \to \infty$,

as $h \to 0$ and $ph \to \infty$, there exists an a-dimensional nonrandom vector θ_S^{∞} s.t.

$$\tilde{\theta}_{S,h,p} \xrightarrow{p} \theta_S^{\infty},$$

and

$$Pl_S(\tilde{\theta}_{S,h,p}) \xrightarrow{p} Pl(\theta_S^{\infty}).$$

Meanwhile, θ_S^{∞} satisfies $Pl(\theta_S^{\infty}) = \max Pl(\theta)$.

Proof. As $h \to 0$ and $ph \to \infty$, Lemma 4.5 provides $s_{NW}(\gamma) \to S(\gamma)$. So there exists a function of θ that is $\xi_S^\infty(\theta)$ s.t. $\xi_{S,h,p}(\theta) \to \xi_S^\infty(\theta)$, as $h \to 0$ and $ph \to \infty$. Define $\theta_S^\infty = \arg\max l_S(\theta, \xi_S^\infty(\theta))$.

By the property of smoothing method,

$$l_S(\theta,\xi) = l(\theta, S\xi) \to l(\theta,\xi), \quad as \quad h \to 0, \tag{4.2.8}$$

where $S\xi^T = (\beta^T, S\gamma^T)$.

Thus

$$Pl_S(\theta) \to Pl(\theta), \quad as \quad h \to 0,$$

and

$$Pl_S(\tilde{\theta}_S) = \max Pl_S(\theta) \to \max Pl(\theta), \quad as \quad h \to 0.$$

Therefore

$$Pl(\theta_S^\infty) = \max Pl(\theta).$$

□

Remark 4.10. *This proof is based on the convergence of NW kernel smoother, but it can be applied to the smoothing spline cohort model, since [Silverman, 1984] illustrated spline smoothing and kernel method are essentially asymptotically equivalent.*

Theorem 4.11. *Under the regularity conditions:*

1. *The parameter space* $\mathbb{D} = D_1 \times D_2 \times \cdots \times D_{2a+2p-3} \in \mathbb{R}^{2a+2p-3}$ *for model parameters* $\mathbf{b}$ *has its subspaces* D_j *uniformly bounded with respect to* p *for* $j = 1, \cdots, 2a+$

$2p-3$,

2. *The log-likelihood* $l(\mathbf{b};\mathbf{y})$ *of the exponential family is continuously differentiable with continuous derivatives* $\frac{\partial^2 l}{\partial b_i \partial b_j}$ *and a uniform bound* $\left|\frac{\partial l}{\partial b_j}\right| < M < \infty$ *for* $\mathbf{b} \in \mathbb{D}, i, j = 1, \cdots, 2a+2p-3$,

3. $\int |K(u)|\, du < \infty$,

4. $uK(u) \to 0$*, as* $|u| \to \infty$,

as $ph \to \infty$ *and* $h \to 0$,

$$\tilde{\theta}_S - \tilde{\theta}_I \to_p 0,$$

i.e., $\theta_S^\infty = \theta^\infty$*, where* θ_S^∞ *is the limit of MaPLE* $\tilde{\theta}_S$*, and* θ^∞ *is the unique true parameters, the limit of the intrinsic estimator.*

Proof. The intrinsic estimator B maximizes the following penalized likelihood [Fu, 2016],

$$l_I(\mu,\alpha,\beta,\gamma;y) = \sum_{i=1}^{a}\sum_{j=1}^{p} l(y_{ij},\mu,\alpha,\beta,\gamma) - \delta(\mathbf{b}^T B_0)^2 \quad \text{for} \quad \delta > 0. \tag{4.2.9}$$

Since B is perpendicular to the B_0, the penalty term achieves 0 at B and thus δ can be arbitrarily small. We requires $\delta = o(\frac{1}{p^\rho})$.

For the intrinsic estimator, we have

$$-\frac{\partial^2 Pl_I(\theta)}{\partial \theta^2} \quad = \quad -\frac{\partial^2 l_I(\theta,\xi)}{\partial \theta^2} \approx pC_1 + 2\delta\theta_0 \otimes \theta_0, \forall\theta, \tag{4.2.10}$$

where C_1 is the Fisher information matrix for the row effect model $l(\theta,\xi;y)$ with parameters θ, and is positive-definite and independent of δ.

Note $\tilde{\theta}_I$ satisfies

$$(\tilde{\theta}_I, \xi(\tilde{\theta}_I)) \perp B_0.$$

So

$$Pl_I(\tilde{\theta}_I) = Pl(\tilde{\theta}_I) = \max Pl(\theta),$$

and

$$Pl_I(\tilde{\theta}_S) = Pl(\tilde{\theta}_S) + \delta(\theta_0^T \tilde{\theta}_S + \xi_0^T \xi(\tilde{\theta}_S))^2.$$

Since $(B_0^T \mathbf{b})^2 = O(p)$. Let $\delta = o(\frac{1}{p^\rho}), \rho > 1$, then $\delta(B_0^T \mathbf{b})^2 = o(1)$.

$$Pl_I(\tilde{\theta}_S) = Pl(\tilde{\theta}_S) + o(1),$$

$$Pl_I(\tilde{\theta}_I) - Pl_I(\tilde{\theta}_S) = Pl(\tilde{\theta}_I) - Pl(\tilde{\theta}_S) + o(1). \tag{4.2.11}$$

Let $h \to 0$ and $ph \to \infty$, by Theorem 4.9,

$$Pl_I(\theta^\infty) - Pl_I(\theta_S^\infty) = Pl(\theta^\infty) - Pl(\theta_S^\infty) = \max Pl(\theta) - \max Pl(\theta) = 0.$$

Take Taylor expansion,

$$\begin{aligned} Pl_I(\theta_S^\infty) &= Pl_I(\theta^\infty) + \frac{\partial Pl_I}{\partial \theta}|_\theta (\theta_S^\infty - \theta^\infty) \\ &+ (\theta_S^\infty - \theta^\infty)^T \frac{\partial^2 Pl_I}{\partial \theta^2}|_{\theta^\infty} (\theta_S^\infty - \theta^\infty) \\ &= Pl_I(\theta^\infty) + (\theta_S^\infty - \theta^\infty)^T \frac{\partial^2 Pl_I}{\partial \theta^2}|_{\theta^\infty} (\theta_S^\infty - \theta^\infty) \\ &= Pl_I(\theta^\infty) - (\theta_S^\infty - \theta^\infty)^T [pC_1 + o(1)](\theta_S^\infty - \theta^\infty). \end{aligned}$$

Since C_1 is positive definite matrix, let h go to 0, by Lemma 4.7, we have

$$\theta_S^\infty = \theta^\infty,$$

that is,

$$\tilde{\theta}_S - \theta^\infty \xrightarrow{p} 0$$

as $h \to 0$ and $ph \to \infty$. □

4.3 Simulation Analysis for Consistency of Smoothing Cohort Model

Simulations are conducted in this section to show the consistency of the smoothing cohort model with intrinsic estimation. In the simulation, difference between $\hat{\theta}_S$ and $\hat{\theta}_I$ are defined as L_2-norm. The norm tends to be increasing over the value of a. So we focus on the mean difference, $\frac{1}{a}\left|\left|\hat{\theta}_S - \hat{\theta}_I\right|\right|_2$.

In this example, a data set in a table of 20×150 dimension is generated, with specified parameter values in Figure 4.3.1. Age effect follows an increasing trend, period effect follows a decreasing trend, and cohort effect has an increasing-then-decreasing shape. One hundred runs are conducted in this simulation study.

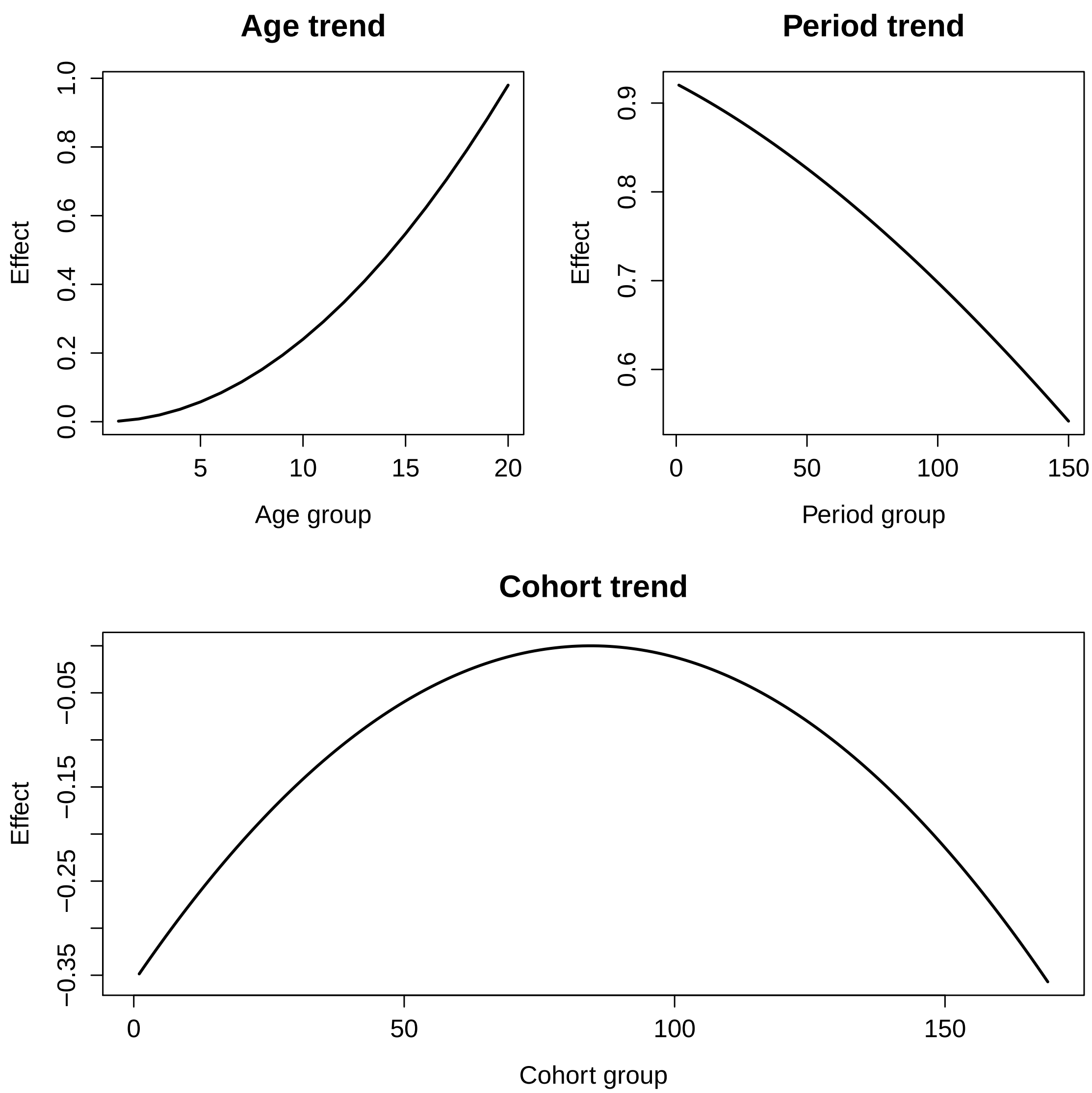

Figure 4.3.1: The Row, Column, and Diagonal Effects Specified

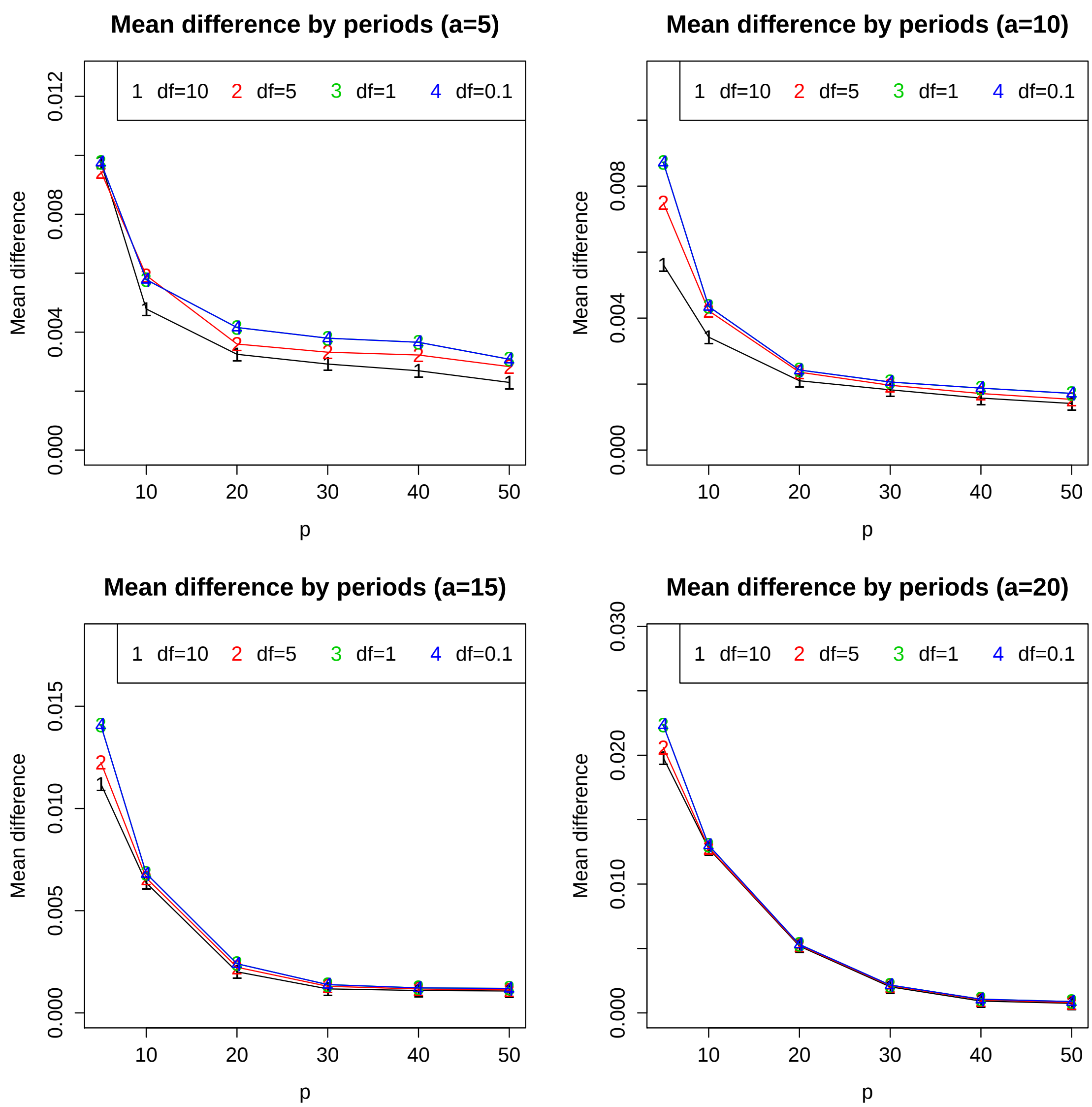

Figure 4.3.2: Mean Difference by Periods for Varying Degrees of Freedom with Fixed Age Group Size

In Figure 4.3.2, smoothing splines with different degrees of freedom (df) are simulated to reveal the influence of df onto the consistency of $\hat{\theta}_S$ and $\hat{\theta}_I$ with different values of fixed a. For $a = 5, a = 10, a = 15$, and $a = 20$, different degrees of freedom do not yield large difference (about 10^{-2}).

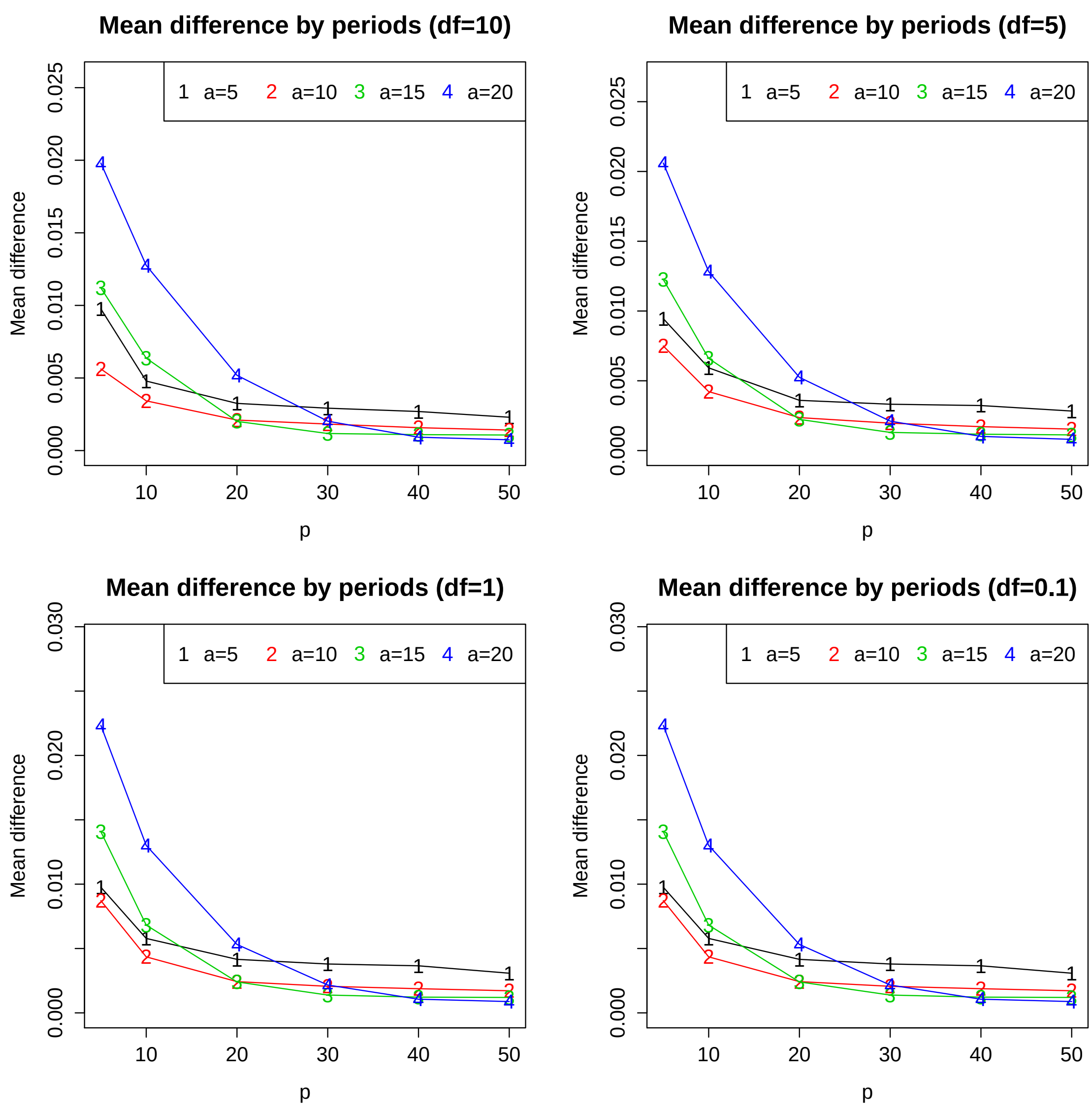

Figure 4.3.3: Mean Difference by Periods for Varying Age Group Size with Fixed Degrees of Freedom

In Figure 4.3.3, the df is fixed to reveal the influence of a onto the consistency of $\hat{\theta}_S$ and $\hat{\theta}_I$. When both a and p are small, the difference is large due to small sample size.

When p is small, the larger difference with larger a is due to the constraint specified on the age effects which has a relatively larger group size than the period size, as discussed before.

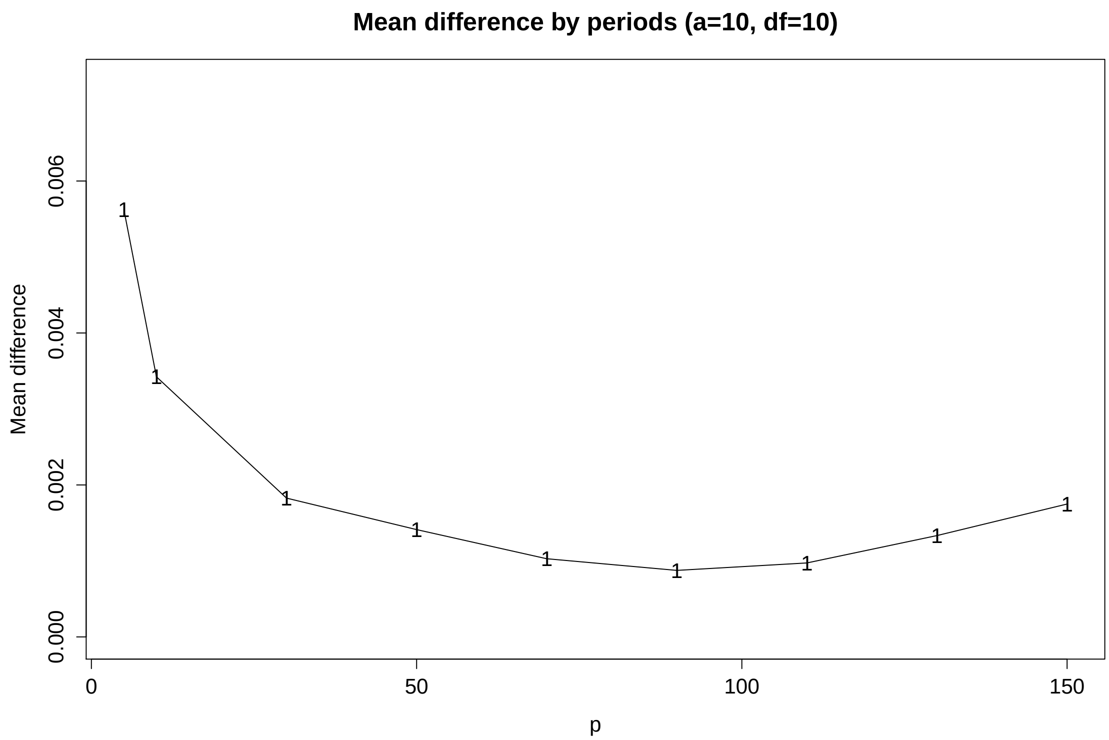

Figure 4.3.4: Mean Difference by Periods with $a = 10$ and $df = 10$

In Figure 4.3.4, to reveal the asymptotic behavior, an example with large $p > 50$ is considered. As p increases, the difference between $\hat{\theta}_S$ and $\hat{\theta}_I$ decreases to a negligible value (about 10^{-3}).

4.4 Summary

Asymptotic properties of the smoothing cohort model are studied in this chapter. Under mild regularity conditions, when bandwidth $h \to 0$ and $ph \to \infty$, smoothing cohort model estimation converges to the true parameter. Two simulation studies were conducted, which illustrate the convergence of the smoothing cohort model estimation and the intrinsic estimator. The consistency establishes the result that the two models (two-stage smoothing and intrinsic estimator) yield age estimates converging to the same true parameter values, and further addresses the identifiability problem.

CHAPTER 5

Application

In this chapter, the smoothing cohort stage 2 model with the smallest variance of ratio constraint is applied to several examples. Comparison with the intrinsic estimator is also conducted.

5.1 Male-mortality Rate of Liver Cancer in Korea

Liver cancer is one of the most common causes of death in the world. In this example, data for the mortality rate of liver cancer in Korean adults (30 years and older) were obtained from the Korean Statistical Information Service for the period from 1984-2013. The data is given in the Table 5.1.1, [Park and Jee, 2015]. Note the last row (age > 80) is included in application, even the cohort of the last row is different from the previous one, but the number of observations with age 85+ was small. We use the two-stage smoothing cohort model with ratio of the smallest variance constraint, and compare with the intrinsic estimator.

Table 5.1.1: Male-mortality Rate (per 10^5 person-year) of Liver Cancer in Korea by Age and Period

	Period					
Age	1984-1988	1989-1993	1994-1998	1999-2003	2004-2008	2009-2013
30-34	7	6	4	3.2	2.2	1.6
35-39	19.6	17.8	12.7	9.9	6	5
40-44	43.6	40.9	34	26	20.2	13.4
45-49	76.7	78.7	64.5	51.8	43.3	31.1
50-54	111	120.2	106.3	89.2	72	54.9
55-59	143.1	152.5	142.9	123.8	103.5	78.1
60-64	164.9	177.5	173.3	149.2	127.9	102.8
65-69	169.6	203.1	196.2	176.7	156.5	127.6
70-74	147.9	193.5	231.3	203.5	187.9	170.9
75-79	140.2	177.7	219	225.4	216.7	203
80+	147	154.1	183.4	206.4	224.3	238.4

This study has $a = 11$ age groups and $p = 6$ period groups. Based on the simulation study, the period-effect constraint should yield more accurate estimation than the age-effect constraint. Figure 5.1.1 shows estimates and bootstrapped 95% confidence intervals from the intrinsic estimator and the smoothing cohort stage 2 model with both the age- and the period-effect constraints. Estimated values are displayed in Table 5.1.2. The two-stage smoothing cohort model with the period-effect constraint gives close trend estimation with the intrinsic estimator, and close estimates of variance component 0.001161 and 0.001227, which are smaller than variance component of the age-effect constraint. The standard error of the intrinsic estimator is smaller than bootstrapped standard errors of the period- and age-effect constraint, but the period-effect constraint gives comparatively small standard error close to that of the intrinsic estimator. Both intrinsic estimator and the period-effect constraint give reasonable age, period, and cohort trend estimation, with increasing age-effect trend, flat period trend, as well as increasing-then-decreasing trend for cohort effect. At the ends of cohort effect, there are some small difference between intrinsic estimates and smoothing period constraint estimates due to few observations in the extreme young and old cohorts. [Park and Jee, 2015] discussed this age-period-cohort analysis of liver-cancer mortality among Korean men from 1984 to 2013. It has been demonstrated that the national vaccination program against HBV may have contributed to the reduction of liver-cancer mortality in Korean children and adolescents. The period effect following the implementation of the national vaccination program in 1995 was significant after accounting for age and cohort effects. The decrease of HBV infection is limited to the younger population and viral persistence remains in the middle-aged and older population. HBsAg (also known as the Australia antigen, which is the surface antigen of the HVB) seroprevalence has been reduced by half during the past 30 years, probably due to the combined effect of vaccination and antiviral therapy.

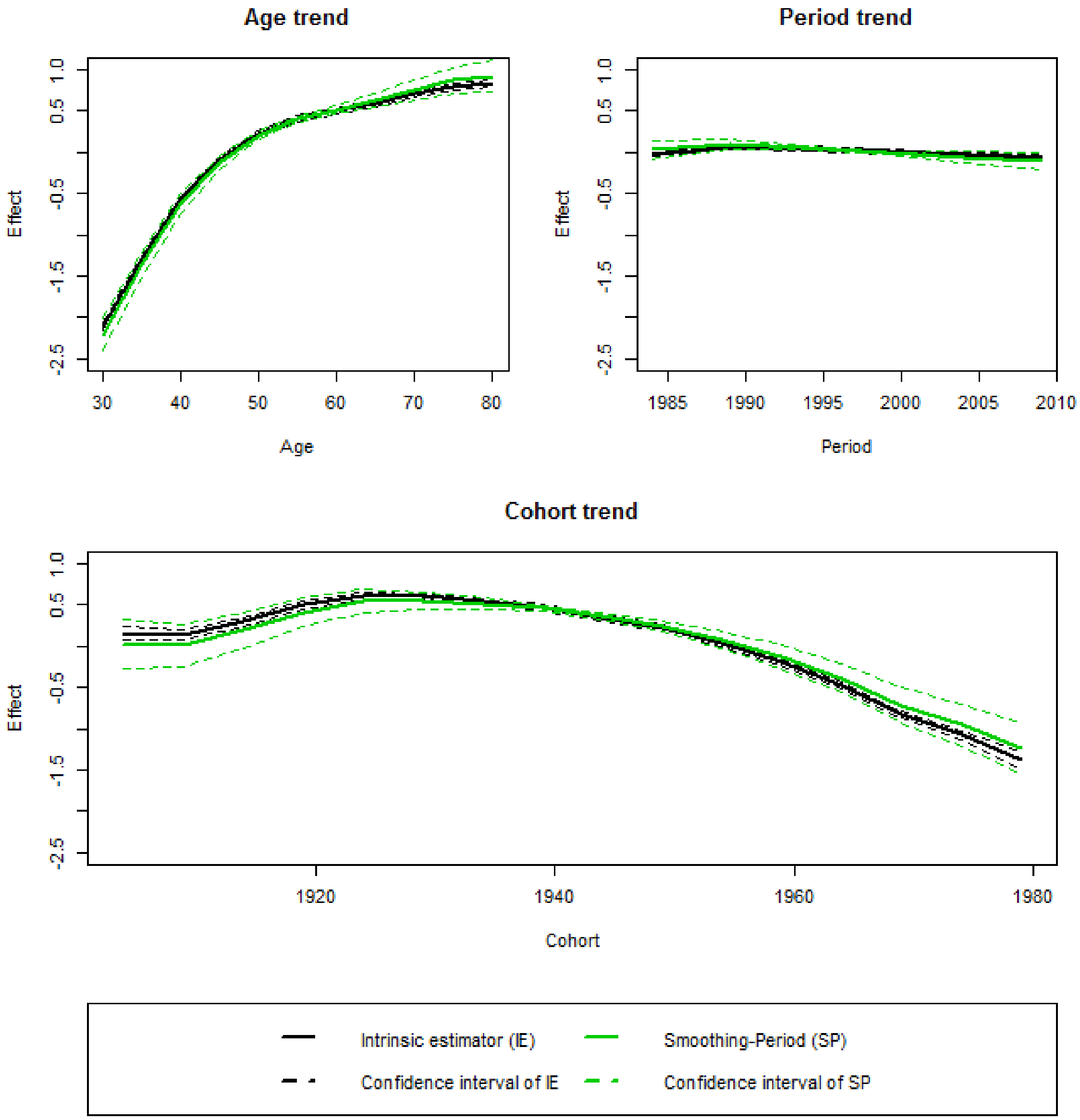

Figure 5.1.1: Smoothing Cohort Stage 2 Model (Constraint with the Smallest Variance of Ratio) and the Intrinsic Estimator with Application to Male-mortality Rate of Liver Cancer in Korea

Table 5.1.2: Age, Period, and Cohort Trends Estimated with the Smoothing Cohort Stage 2 Model and the Intrinsic Estimator for Male-mortality Rate of Liver Cancer in Korea

	Intrinsic		Period-effect Constraint		Age-effect Constraint	
	Estimate	SE	Estimate	Bootstrapped SE	Estimate	Bootstrapped SE
$\hat{\mu}$	4.0233	0.0074	4.0233	0.0057	4.0233	0.0121
$\hat{\alpha}_1$	-2.1155	0.0200	-2.2094	0.1016	-3.2609	0.9704
$\hat{\alpha}_2$	-1.2904	0.0183	-1.3655	0.0812	-2.2067	0.7792
$\hat{\alpha}_3$	-0.5649	0.0186	-0.6213	0.0610	-1.2522	0.5830
$\hat{\alpha}_4$	-0.0879	0.0188	-0.1255	0.0424	-0.5461	0.3106
$\hat{\alpha}_5$	0.2291	0.0190	0.2103	0.0243	0.1534	0.0485
$\hat{\alpha}_6$	0.4020	0.0190	0.4020	0.0142	0.4020	0.0310
$\hat{\alpha}_7$	0.4984	0.0190	0.5172	0.0248	0.7275	0.1991
$\hat{\alpha}_8$	0.5963	0.0189	0.6339	0.0429	1.0545	0.3930
$\hat{\alpha}_9$	0.7040	0.0187	0.7603	0.0624	1.3912	0.5895
$\hat{\alpha}_{10}$	0.7967	0.0184	0.8719	0.0805	1.7131	0.7862
$\hat{\alpha}_{11}$	0.8321	0.0194	0.9261	0.1002	1.8242	0.6897
$\hat{\beta}_1$	-0.0287	0.0128	0.0324	0.0562	0.5440	0.4885
$\hat{\beta}_2$	0.0636	0.0130	0.0918	0.0301	0.4072	0.2930
$\hat{\beta}_3$	0.0448	0.0130	0.0542	0.0130	0.1593	0.0998
$\hat{\beta}_4$	0.0094	0.0130	-0.0141	0.0089	-0.1051	0.1010
$\hat{\beta}_5$	-0.0300	0.0128	-0.0582	0.0334	-0.3736	0.2930
$\hat{\beta}_6$	-0.0591	0.0134	-0.1061	0.0533	-0.6318	0.4875
$\hat{\gamma}_1$	0.1637	0.0426	0.0228	0.1504	-1.5544	1.4683
$\hat{\gamma}_2$	0.1352	0.0309	0.0130	0.1301	-1.3539	1.2713
$\hat{\gamma}_3$	0.3020	0.0261	0.1986	0.1104	-0.9580	1.0735
$\hat{\gamma}_4$	0.5016	0.0234	0.4170	0.0927	-0.5293	0.8790
$\hat{\gamma}_5$	0.6181	0.0215	0.5523	0.0730	-0.1837	0.6853
$\hat{\gamma}_6$	0.6034	0.0199	0.5564	0.0523	0.0307	0.4886
$\hat{\gamma}_7$	0.5415	0.0207	0.5133	0.0354	0.1979	0.2912
$\hat{\gamma}_8$	0.4682	0.0209	0.4588	0.0193	0.3537	0.0990
$\hat{\gamma}_9$	0.3522	0.0208	0.3616	0.0184	0.4668	0.1083
$\hat{\gamma}_{10}$	0.2151	0.0202	0.2433	0.0328	0.5588	0.2980
$\hat{\gamma}_{11}$	0.0369	0.0193	0.0839	0.0513	0.6097	0.4923
$\hat{\gamma}_{12}$	-0.1901	0.0206	-0.1243	0.0710	0.6118	0.6878
$\hat{\gamma}_{13}$	-0.4694	0.0223	-0.3848	0.0902	0.5615	0.8794
$\hat{\gamma}_{14}$	-0.8231	0.0249	-0.7197	0.1098	0.4369	1.0709
$\hat{\gamma}_{15}$	-1.0769	0.0298	-0.9547	0.1341	0.4122	1.2639
$\hat{\gamma}_{16}$	-1.3787	0.0489	-1.2377	0.1574	0.3395	1.4556
$\hat{\sigma}^2$	0.001160969		0.001227093		0.005437853	

Note: Standard errors were computed by simulation based on 1000 runs of bootstrapped residuals.

5.2 Homicide-Arrest Rate

The homicide-arrest rate [O'Brien, 2000] data is given in the Table 5.2.1. Each age group covers a range of 5 years and the span between the periods is 5 years. The age, period, and cohort trend of this data has been discussed by [O'Brien, 2000, Fu, 2008].

Table 5.2.1: Homicide-Arrest Rate by Age and Period

	Period							
Age	1960-1964	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999
15-19	8.89	9.07	17.22	17.54	18.02	16.32	36.52	35.24
20-24	14.00	15.18	23.76	25.62	23.95	21.11	29.10	32.34
25-29	13.45	14.69	20.09	21.05	18.91	16.79	17.99	16.75
30-34	10.73	11.70	16.00	15.81	15.22	12.59	12.44	10.05
35-39	9.37	9.76	13.13	12.83	12.31	9.60	9.38	7.27
40-44	6.48	7.41	10.10	10.52	8.79	7.50	6.81	5.48
45-49	5.71	5.56	7.51	7.32	6.76	5.31	5.17	3.67

This sample has $a = 7$ age groups and $p = 8$ period groups. Age-effect group size is similar with period-effect group size. So we consider both age- and period-effect constraints. In Figure 5.2.1, the period-effect constraint yields closer estimation to intrinsic estimation than the age-effect constraint. Estimated values are displayed in Table 5.2.2. Difference between the intrinsic estimation and the period-effect constraint estimation is around 10^{-2}. Also the model variance components of the intrinsic estimator and smoothing cohort estimation with the optimal period-effect constraint are close, which are 0.001709 and 0.001970, and both standard errors are small and close to each other. The age-effect constraint yields different trend estimation from the period-effect constraint and intrinsic estimation, with larger estimate of variance component (0.00303) and standard error. The intrinsic estimator and the period-effect constraint yield an increasing age trend from age 15 to age 20, then decreasing age trend after age 20, a rapidly increasing period trend from 1960 to 1970 and decreasing trend after 1970, as well as a slowly decreasing then sharp increasing trend for cohort effect. This result met the investigator's expectation [O'Brien, 2000].

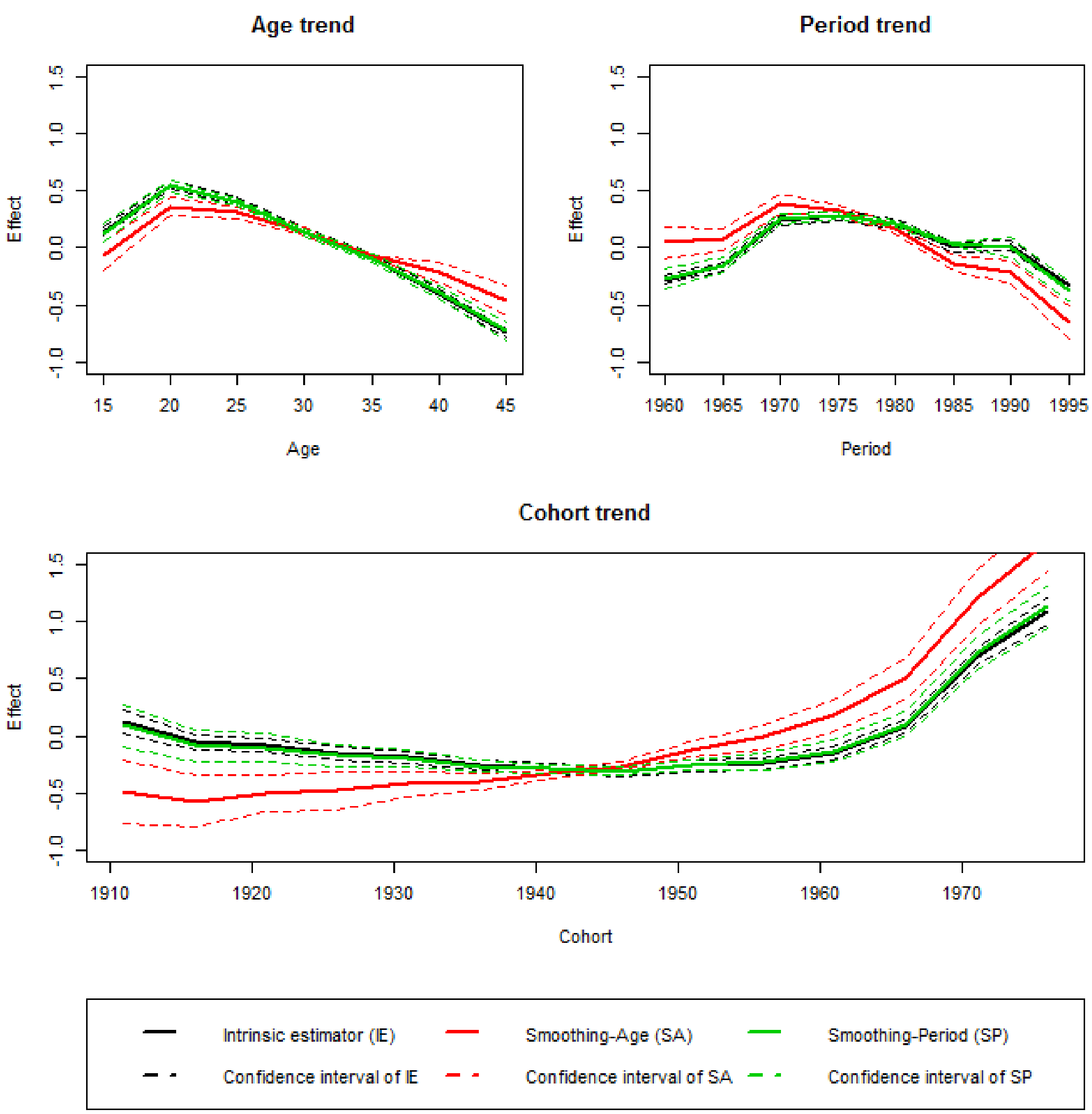

Figure 5.2.1: Smoothing Cohort Stage 2 Model (Constraints with the Smallest Variance of Ratio) and the Intrinsic Estimator with Application to Homicide-Arrest Rate

Table 5.2.2: Age, Period, and Cohort Trends Estimated with the Smoothing Cohort Stage 2 Model and the Intrinsic Estimator for Homicide-Arrest Rate

	Intrinsic		Period-effect Constraint		Age-effect Constraint	
	Estimate	SE	Estimate	Bootstrapped SE	Estimate	Bootstrapped SE
μ	2.6468	0.0098	2.6468	0.0077	2.6468	0.0082
α_1	0.1446	0.0198	0.1299	0.0405	-0.0692	0.0644
α_2	0.5513	0.0186	0.5415	0.0285	0.3634	0.0407
α_3	0.4066	0.0190	0.4017	0.0185	0.3126	0.0258
α_4	0.1417	0.0191	0.1417	0.0147	0.1417	0.0177
α_5	-0.0940	0.0191	-0.0890	0.0186	-0.0680	0.0035
α_6	-0.4036	0.0188	-0.3938	0.0291	-0.2157	0.0436
α_7	-0.7466	0.0188	-0.7318	0.0398	-0.4647	0.0668
β_1	-0.2832	0.0196	-0.2660	0.0442	0.0457	0.0715
β_2	-0.1601	0.0205	-0.1478	0.0330	0.0748	0.0494
β_3	0.2436	0.0208	0.2510	0.0233	0.3845	0.0437
β_4	0.2842	0.0208	0.2866	0.0168	0.3312	0.0224
β_5	0.2145	0.0207	0.2120	0.0179	0.1675	0.0189
β_6	0.0074	0.0204	0.0323	0.0085	-0.1336	0.0368
β_7	0.0193	0.0200	0.0070	0.0439	-0.2156	0.0530
β_8	-0.3256	0.0224	-0.3752	0.0450	-0.6545	0.0740
γ_1	0.1252	0.0493	0.0932	0.0929	-0.4856	0.1431
γ_2	-0.0579	0.0370	-0.0850	0.0729	-0.5747	0.1151
γ_3	-0.0800	0.0321	-0.1021	0.0603	-0.5028	0.0836
γ_4	-0.1537	0.0292	-0.1709	0.0472	-0.4826	0.0842
γ_5	-0.1879	0.0270	-0.2002	0.0368	-0.4228	0.0549
γ_6	-0.2581	0.0250	-0.2655	0.0268	-0.3991	0.0394
γ_7	-0.2885	0.0227	-0.2910	0.0189	-0.3355	0.0208
γ_8	-0.3126	0.0223	-0.3101	0.0185	-0.2656	0.0168
γ_9	-0.2615	0.0239	-0.2542	0.0253	-0.1206	0.0352
γ_{10}	-0.2436	0.0254	-0.2314	0.0372	-0.0087	0.0552
γ_{11}	-0.1521	0.0273	-0.1349	0.0476	0.1767	0.0715
γ_{12}	0.0791	0.0302	0.1013	0.0572	0.5020	0.0930
γ_{13}	0.6954	0.0355	0.7225	0.0764	1.2122	0.1281
γ_{14}	1.0963	0.0597	1.1283	0.0927	1.7070	0.1359
$\hat{\sigma}^2$	0.00171		0.00197		0.00303	

Note: Standard errors were computed by simulation based on 1000 runs of bootstrapped residuals.

5.3 Incidence Rate of Testis Cancer

This study fits the APC model with incidence rate of testis cancer (per 1000 male person-years) in Denmark 1943-1996 in 5-year classes, [Carstensen, 2007]. The data set is given in Table 5.3.1.

Table 5.3.1: Incidence Rate of Testis Cancer (per 10^3 person-year) in Denmark

	Period										
Age	43-47	48-52	53-57	57-62	63-67	68-72	73-77	78-82	83-87	88-92	93-96
15-19	4.3	3.1	5.5	4.5	4.8	11.4	12.2	12.2	16.3	18.3	20.4
20-24	12.3	13.9	21.2	21.2	19.1	26.9	37.9	49.0	49.4	49.1	49.1
25-29	23.2	26.4	29.0	39.1	37.9	35.7	48.3	69.7	74.6	86.6	77.4
30-34	23.4	28.4	35.5	41.2	49.0	53.7	56.9	66.0	72.9	89.9	101.9
35-39	23.0	23.8	24.6	29.4	46.4	58.9	61.7	53.2	60.5	73.4	86.8
40-44	16.8	20.8	28.2	28.4	29.5	39.9	49.3	52.4	42.9	50.5	55.7
45-49	15.5	14.8	16.7	23.9	20.3	28.6	30.1	32.2	41.3	31.0	39.1
50-54	9.9	15.5	11.2	12.6	20.9	16.7	23.0	24.2	30.8	29.6	24.2
55-59	4.2	9.1	9.3	13.4	12.7	13.8	13.7	20.9	21.8	18.1	27.5
60-64	7.4	9.2	7.7	8.2	11.6	9.6	14.5	12.2	13.8	8.5	7.2

This sample has $a = 10$ age groups and $p = 11$ period groups. Comparison of the intrinsic estimator and smoothing cohort stage 2 model are given in Figure 5.3.1. Estimated values are displayed in Table 5.3.2. For this case, the intrinsic estimator gives the smaller standard error and variance component than the period-effect constraint. The age-effect constraint has the largest bootstrapped standard error, as well as $\hat{\sigma}^2$. From the figure, the intrinsic estimator and the period-effect constraint yield similar pattern of trend estimation. This age-period-cohort trend estimation is similar with the result in [Carstensen, 2007].

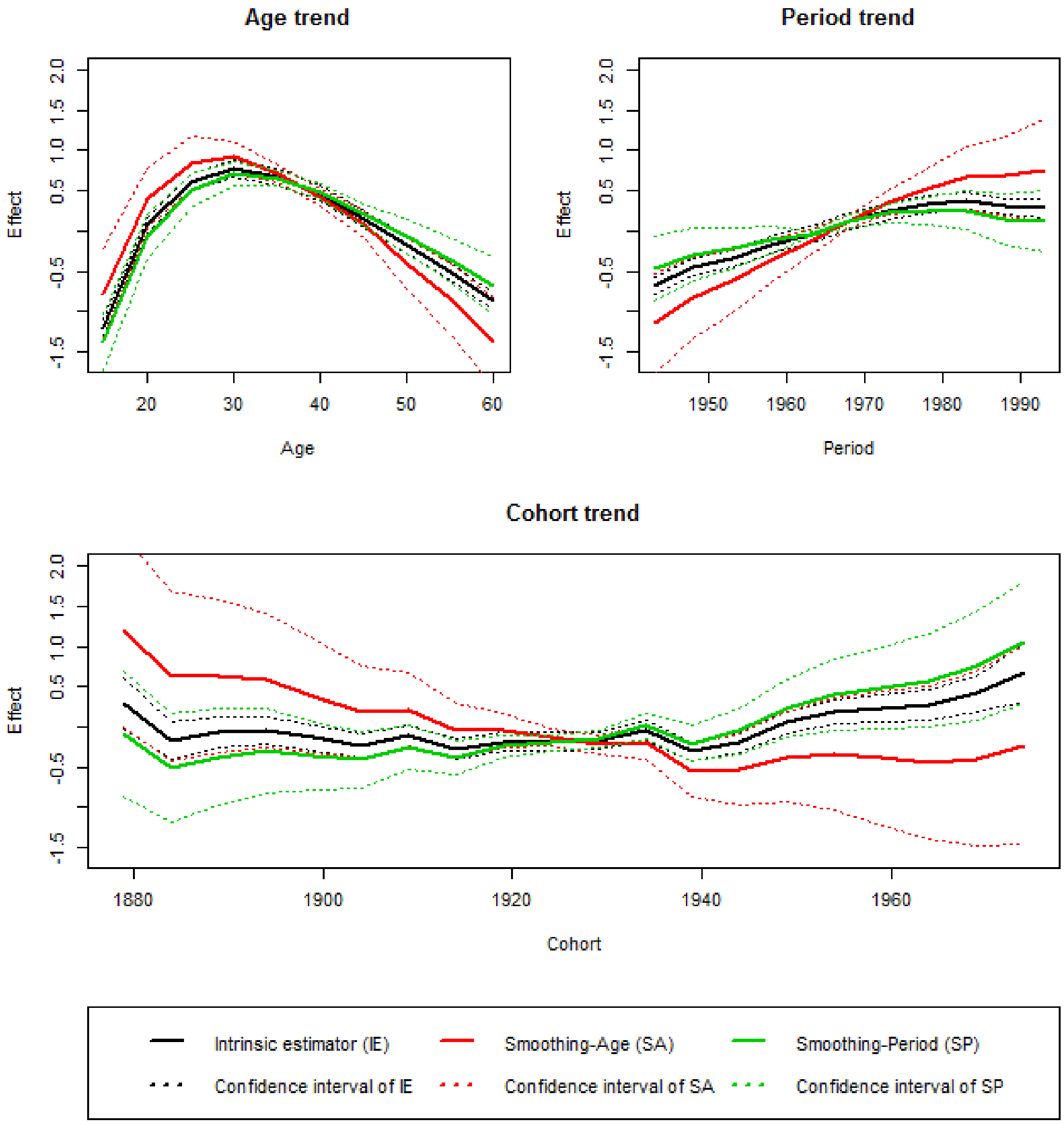

Figure 5.3.1: Smoothing Cohort Stage 2 Model (Constraints with the Smallest Variance of Ratio) and the Intrinsic Estimator with Application to Testis Cancer-incidence Rate

Table 5.3.2: Age, Period, and Cohort Trends Estimated with the Smoothing Cohort Stage 2 Model and the Intrinsic Estimator of Testis Cancer-incidence Rate

	Intrinsic		Period-effect Constraint		Age-effect Constraint	
	Estimate	SE	Estimate	Bootstrapped SE	Estimate	Bootstrapped SE
μ	3.2543	0.0228	3.2543	0.0186	3.2543	0.0194
α_1	-1.2026	0.0529	-1.3825	0.1784	-0.7755	0.2859
α_2	0.0753	0.0504	-0.0646	0.1400	0.4074	0.1908
α_3	0.6113	0.0508	0.5113	0.1078	0.8485	0.1655
α_4	0.7737	0.0510	0.7138	0.0723	0.9161	0.1043
α_5	0.6749	0.0511	0.6549	0.0472	0.7223	0.0544
α_6	0.4685	0.0512	0.4885	0.0475	0.4211	0.0545
α_7	0.1424	0.0512	0.2023	0.0713	0.0762	0.0838
α_8	-0.1710	0.0510	-0.0711	0.1050	-0.4083	0.1626
α_9	-0.5070	0.0506	-0.3671	0.1419	-0.8391	0.2266
α_{10}	-0.8654	0.0511	-0.6855	0.1817	-1.3686	0.2789
β_1	-0.6722	0.0526	-0.4723	0.2038	-1.1467	0.3190
β_2	-0.4505	0.0537	-0.2906	0.1645	-0.8301	0.2584
β_3	-0.3353	0.0541	-0.2154	0.1253	-0.6200	0.1965
β_4	-0.1798	0.0542	-0.0999	0.0868	-0.3696	0.1355
β_5	-0.0400	0.0542	-0.0419	0.0165	-0.1349	0.0802
β_6	0.1206	0.0540	0.1206	0.0401	0.1206	0.0451
β_7	0.2595	0.0539	0.2196	0.0560	0.3544	0.0767
β_8	0.3302	0.0537	0.2503	0.0876	0.5200	0.1346
β_9	0.3797	0.0534	0.2598	0.1195	0.6644	0.1976
β_{10}	0.3028	0.0530	0.1429	0.1587	0.6824	0.2577
β_{11}	0.2850	0.0577	0.1270	0.1968	0.7595	0.3229
γ_1	0.2899	0.1602	-0.0898	0.3988	1.1915	0.6105
γ_2	-0.1744	0.1161	-0.5141	0.3433	0.6323	0.5419
γ_3	-0.0723	0.0976	-0.3721	0.3029	0.6394	0.4831
γ_4	-0.0379	0.0869	-0.2977	0.2708	0.5790	0.4232
γ_5	-0.1363	0.0797	-0.3561	0.2168	0.3857	0.3590
γ_6	-0.2320	0.0743	-0.4118	0.1810	0.1951	0.2925
γ_7	-0.1171	0.0698	-0.2570	0.1431	0.2150	0.2320
γ_8	-0.2765	0.0658	-0.3764	0.1104	-0.0392	0.1696
γ_9	-0.1816	0.0620	-0.2415	0.0738	-0.0392	0.1080
γ_{10}	-0.1747	0.0580	-0.1947	0.0468	-0.1273	0.0576
γ_{11}	-0.1719	0.0576	-0.1519	0.0533	-0.2193	0.0575
γ_{12}	-0.0449	0.0607	0.0150	0.0783	-0.1873	0.1061
γ_{13}	-0.3062	0.0638	-0.2063	0.1138	-0.5435	0.1638
γ_{14}	-0.1932	0.0672	-0.0533	0.1502	-0.5254	0.2245
γ_{15}	0.0530	0.0713	0.2329	0.1841	-0.3741	0.2881
γ_{16}	0.1819	0.0765	0.4017	0.2245	-0.3401	0.3566
γ_{17}	0.2282	0.0837	0.4880	0.2613	-0.3886	0.4223
γ_{18}	0.2725	0.0948	0.5723	0.2967	-0.4392	0.4821
γ_{19}	0.4151	0.1141	0.7548	0.3407	-0.3916	0.5504
γ_{20}	0.6783	0.1823	1.0580	0.3852	-0.2233	0.6291
$\hat{\sigma}^2$	0.01994		0.02026		0.02110	

Note: Standard errors were computed by simulation based on 1000 runs of bootstrapped residuals.

5.4 Summary

The two-stage smoothing cohort model yields accurate estimation of the age, period, and cohort effects in the study of incidence rate, mortality rate, and homicide-arrest rate data, with meaningful trend estimation. Most often the constraint specified on the effects of smaller group size yields estimation close to intrinsic estimation, indicating the consistency of the method.

CHAPTER 6

Conclusion

Estimating the trend of incidence or mortality rate in age, period, and cohort is of crucial importance to public health. The identifiability problem in the APC model has been studied for 40 years. Although many methods have been proposed, none of them have addressed the problem appropriately. [Fu, 2016] showed the expected value of the intrinsic estimator is an estimable function and is the only estimable function that determines the parameters. Further the true parameter values can be identified by taking the limit of the intrinsic estimator under mild regularity condition.

This dissertation provides the asymptotic results for an alternative semiparametric approach – a two-stage smoothing cohort model. It has been proved that the intercept and age-effect estimates converge to the true parameter values as $h \to 0$ and $ph \to \infty$, indicating that this alternative method also provides consistent estimation and an alternative to address the identifiability problem.

Another significant contribution of this dissertation is to identify the optimal constraint to correct the bias for finite samples, which yields accurate estimation for the age, period, and cohort effects. Furthermore, it also has been shown that the two-stage smoothing cohort model with a constraint set up on the effects of the factor that has the smaller group size with the smallest variance of the ratio yields less-biased estimation on the cohort effects than the intrinsic estimator.

The smoothing cohort model with optimal constraint possesses desirable properties by the large sample theory, and works as valid alternative approach to addressing the identifiability problem. Both simulation results and application demonstrate the usefulness of this method.

Bibliography

U. Amato and A. Antoniadis. Adaptive wavelet series estimation in separable nonparametric regression models. *Statistics and Computing*, 11(4):373–394, 2001.

U. Amato, A. Antoniadis, and I. De Feis. Fourier series approximation of separable models. *Journal of Computational and Applied Mathematics*, 146(2):459–479, 2002.

C. Ansley and R. Kohn. Convergence of the backfitting algorithm for additive models. *Journal of the Australian Mathematical Society (Series A)*, 57(03):316–329, 1994.

S. Antoni, J. Ferlay, I. Soerjomataram, A. Znaor, A. Jemal, and F. Bray. Bladder cancer incidence and mortality: a global overview and recent trends. *European Urology*, 71(1): 96–108, 2016.

G. Boente, X. He, and J. Zhou. Robust estimates in generalized partially linear models. *The Annals of Statistics*, 34(6):2856–2878, 2006.

R. Boyar Cetinkaya, B. Aagnes, E. Thiis-Evensen, S. Tretli, D. Bergestuen, and S. Hansen. Trends in incidence of neuroendocrine neoplasms in norway: a report of 16,075 cases from 1993 through 2010. *Neuroendocrinology*, 104(1):1–10, 2015.

A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, 17(2):453–555, 1989.

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

B. Carstensen. Age–period–cohort models for the lexis diagram. *Statistics in medicine*, 26(15):3018–3045, 2007.

X. Chang and L. Qu. Wavelet estimation of partially linear models. *Computational statistics & data analysis*, 47(1):31–48, 2004.

H. Chen and J. Shiau. A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*, 27(2):187–201, 1991.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. i: age–period and age–cohort models. *Statistics in medicine*, 6(4):449–467, 1987a.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. ii: age–period–cohort models. *Statistics in medicine*, 6(4):469–481, 1987b.

M. Coleman, J. Esteve, P. Damiecki, A. Arslan, and H. Renard. Trends in cancer incidence and mortality. *IARC scientific publications*, (121):1–806, 1993.

J. Cuzick. Semiparametric additive regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3):831–843, 1992.

R. Engle, C. Granger, J. Rice, and A. Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394):310–320, 1986.

J. Fan and J. Zhang. Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics*, 32(5):1858–1907, 2004.

J. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American statistical Association*, 76(376):817–823, 1981.

W. Fu. Ridge estimator in singulah oesiun with application to age-period-cohort analysis of disease rates. *Communications in statistics-Theory and Methods*, 29(2):263–278, 2000.

W. Fu. A smoothing cohort model in age–period–cohort analysis with applications to homicide arrest rates and lung cancer mortality rates. *Sociological methods & research*, 36(3):327–361, 2008.

W. Fu. Constrained estimators and consistency of a regression model on a lexis diagram. *Journal of the American Statistical Association*, 111(513):180–199, 2016.

I. Gannaz. Robust estimation and wavelet thresholding in partially linear models. *Statistics and Computing*, 17(4):293–310, 2007.

P. Gonzalez-Diego, G. Lopez-Abente, M. Pollan, and M. Ruiz. Time trends in ovarian cancer mortality in europe (1955–1993): effect of age, birth cohort and period of death. *European Journal of Cancer*, 36(14):1816–1824, 2000.

P. Green and B. Yandell. Semi-parametric generalized linear models. In *Generalized linear models*, pages 44–55. Springer, 1985.

B. Greenberg, J. Wright, and C. Sheps. A technique for analyzing some factors affecting the incidence of syphilis. *Journal of the American Statistical Association*, 45(251): 373–399, 1950.

P. Hall, S. Lahiri, and Y. Truong. On bandwidth choice for density estimation with dependent data. *The Annals of Statistics*, 23(6):2241–2263, 1995.

S. Hamilton and Y. Truong. Local linear estimation in partly linear models. *Journal of Multivariate Analysis*, 60(1):1–19, 1997.

W. Härdle and P. Hall. On the backfitting algorithm for additive regression models. *Statistica neerlandica*, 47(1):43–57, 1993.

W. Härdle, E. Mammen, and M. Müller. Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, 93 (444):1461–1474, 1998.

W. Härdle, H. Liang, and J. Gao. *Partially linear models*. Springer, 2000.

W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and semiparametric models*. Springer Science & Business Media, 2012.

T. Hastie and R. Tibshirani. *Generalized additive models*. CRC Press, 1990.

T. Hastie and R. Tibshirani. Bayesian backfitting (with comments and a rejoinder by the authors. *Statistical Science*, 15(3):196–223, 2000.

J. Hayya, D. Armstrong, and N. Gressis. A note on the ratio of two normally distributed variables. *Management Science*, 21(11):1338–1341, 1975.

C. Heuer. Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics*, 53(1):161–177, 1997.

T. Holford. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39 (2):311–324, 1983.

T. Holford. An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases*, 38(10):831–836, 1985.

T. Holford. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual review of public health*, 12(1):425–457, 1991.

T. Holford. Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in medicine*, 25(6):977–993, 2006.

B. Jiang and C. Keumhee. Age-period-cohort models using smoothing splines: a generalized additive model approach. *Statistics in Medicine*, 33(4):595–606, 2014.

D. Kuang, B. Nielsen, and J. Nielsen. Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, 95(4):987–991, 2008a.

D. Kuang, B. Nielsen, and J. Nielsen. Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, 95(4):979–986, 2008b.

L. Kupper, J. Janis, I. Salama, C. Yoshizawa, and B. Greenberg. Age-period-cohort analysis: an illustration of the problems in assessing interaction in one observation per cell data. *Communications in Statistics - Theory and Methods*, 12(23):2779–2807, 1983.

L. Kupper, J. Janis, A. Karmous, and B. Greenberg. Statistical age-period-cohort analysis: a review and critique. *Journal of chronic diseases*, 38(10):811–830, 1985.

W. Lee and R. Lin. Autoregressive age-period-cohort models. *Statistics in Medicine*, 15 (3):273–281, 1996.

P. Li, A. Znaor, I. Holcatova, E. Fabianova, D. Mates, M. Wozniak, J. Ferlay, and G. Scelo. Regional geographic variations in kidney cancer incidence rates in european countries. *European urology*, 67(6):1134–1141, 2015.

O. Linton and J. Nielsen. A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1):93–100, 1995.

E. Mammen and S. van de Geer. Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics*, 25(3):1014–1035, 1997.

E. Mammen, O. Linton, and J. Nielsen. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27(5): 1443–1490, 1999.

E. Mammen, J. Marron, B. Turlach, and M. Wand. A general projection framework for constrained smoothing. *Statistical Science*, 16(3):232–248, 2001.

P. McCullagh and J. Nelder. *Generalized linear models*. CRC press, 1989.

M. Müller. Estimation and testing in generalized partial linear modelsa comparative study. *Statistics and Computing*, 11(4):299–309, 2001.

S. Murphy and A. Van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465, 2000.

E. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1): 141–142, 1964.

J. Nielsen and S. Sperlich. Smooth backfitting in practice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):43–61, 2005.

R. O'Brien. Age period cohort characteristic models. *Social science research*, 29(1): 123–139, 2000.

R. OBrien. Estimable functions in age-period-cohort models: a unified approach. *Quality & Quantity*, 48(1):457–474, 2014.

R. OBrien and J. Stockard. Can cohort replacement explain changes in the relationship between age and homicide offending? *Journal of Quantitative Criminology*, 25(1): 79–101, 2009.

J. Opsomer. Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73(2):166–179, 2000.

J. Opsomer and D. Ruppert. Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25(1):186–211, 1997.

J. Opsomer and D. Ruppert. A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, 93(442):605–619, 1998.

J. Opsomer and D. Ruppert. A root-n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computational and Graphical Statistics*, 8(4):715–732, 1999.

J. Park and Y. Jee. Age-period-cohort analysis of liver cancer mortality in korea. *Asian Pacific journal of cancer prevention: APJCP*, 16(18):8589–8594, 2015.

N. Pérez-Farinós, G. López-Abente, and R. Pastor-Barriuso. Time trend and age-period-cohort effect on kidney cancer mortality in europe, 1981–2000. *BMC Public Health*, 6 (1):6–119, 2006.

M. Quinn and P. Babb. Patterns and trends in prostate cancer incidence, survival, prevalence and mortality. part i: international comparisons. *BJU international*, 90(2):162–173, 2002.

J. Rice. Convergence rates for partially splined models. *Statistics & Probability Letters*, 4 (4):203–208, 1986.

C. Robertson and P. Boyle. Age, period and cohort models: the use of individual records. *Statistics in Medicine*, 5(5):527–538, 1986.

C. Robertson, S. Gandini, and P. Boyle. Age-period-cohort models: a comparative study of available methodologies. *Journal of clinical epidemiology*, 52(6):569–583, 1999.

W. Rodgers. Estimable functions of age, period, and cohort effects. *American sociological review*, 47(6):774–787, 1982.

S. Sardy and P. Tseng. Amlet, ramlet, and gamlet: automatic nonlinear fitting of additive models, robust and generalized, with wavelets. *Journal of Computational and Graphical Statistics*, 13(2):283–309, 2004.

S. Sardy, A. Bruce, and P. Tseng. Block coordinate relaxation methods for nonparametric wavelet denoising. *Journal of computational and graphical statistics*, 9(2):361–379, 2000.

M. Schimek. Estimation and inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference*, 91(2):525–540, 2000.

T. Severini and J. Staniswalis. Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association*, 89(426):501–511, 1994.

T. Severini and W. Wong. Profile likelihood and conditionally parametric models. *The Annals of statistics*, 20(4):1768–1802, 1992.

B. Silverman. Spline smoothing: the equivalent variable kernel method. *The Annals of Statistics*, 12(3):898–916, 1984.

P. Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):413–436, 1988.

G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM review*, 31(4): 614–627, 1989.

R. Tarone and K. Chu. Evaluation of birth cohort patterns in population disease rates. *American Journal of Epidemiology*, 143(1):85–91, 1996.

D. Tjøstheim and B. Auestad. Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, 89(428):1398–1409, 1994.

M. Vercelli, A. Quaglia, E. Marani, and S. Parodi. Prostate cancer incidence and mortality trends among elderly and adult europeans. *Critical reviews in oncology/hematology*, 35 (2):133–144, 2000.

M. Wand. A central limit theorem for local polynomial backfitting estimators. *Journal of Multivariate Analysis*, 70(1):57–65, 1999.

G. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

W. Wecker and C. Ansley. Nonparametric multiple regression by the alternating projection method. *Proceedings ASA Business and Economic Statistics Section*, pages 311–316, 1982.

Y. Yang, S. Schulhofer-Wohl, W. Fu, and K. Land. The intrinsic estimator for age-period-cohort analysis: What it is and how to use it1. *American Journal of Sociology*, 113(6): 1697–1736, 2008.

A. Yatchew. *Semiparametric regression for the applied econometrician*. Cambridge University Press, 2003.

H. Yousof and A. Gad. Bayesian estimation and inference for the generalized partial linear model. *International Journal of Probability and Statistics*, 4(2):51–64, 2015.

A. Znaor, J. Lortet-Tieulent, M. Laversanne, A. Jemal, and F. Bray. International variations and trends in renal cell carcinoma incidence and mortality. *European urology*, 67(3): 519–530, 2015.