STATISTICAL INFERENCE IN AGE, PERIOD, AND COHORT MODELS

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

Kuikui Gao

August 2019

STATISTICAL INFERENCE IN AGE, PERIOD, AND COHORT MODELS

Kuikui Gao

APPROVED:

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

Dr. Shanyu Ji,
Department of Mathematics, University of Houston

Dr. Tsorng-whay Pan,
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 would like to thank my advisor Dr. Wenjiang Fu for his continuous support, encouragement, and guidance throughout my Ph.D study. He gave me lots of helpful advice on my writing and presentation skills and taught me how to be professional in academia. I would like to thank Dr. Shanyu Ji, Dr. Tsorng-whay Pan, and Dr. Yipeng Yang, for serving my dissertation committee and their insightful comments and suggestions on my dissertation. Especially, I really appreciate Dr. Shanyu Ji for his advice on my family which have changed our life.

I would like to thank my peers, Shujiao Huang, Junyu Ding, Xiao Zhang, and Manyang Sun for their advice and discussions during my Ph.D study.

At the last, I want to thank my husband Dr. Wanli Cheng for his inspiration, understanding, and support both in my academia and life, and thank our little girl Arya Cheng for bringing us joy and happiness. Also, I really appreciate our parents for their support and love. I could not have done it without them.

STATISTICAL INFERENCE IN AGE, PERIOD, AND COHORT MODELS

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
Kuikui Gao
August 2019

Abstract

Age, Period, and Cohort (APC) models have been applied to analyzing disease incidence or mortality rates, and estimating trend in age, period, and birth cohort effects. However, the identification problem —the linear dependency among these three variables: $period - age = cohort$, induces a singular design matrix and thus yields multiple estimators. To address the problem, an intrinsic estimator to the identification problem has been proposed by Fu (2000), and later on been proven unbiased, estimable, and consistent with theoretical justification.

This dissertation addresses the issue of parameter estimation and its variance in APC models, and also derives a new F test statistic testing on the equality of trends of age effects among multiple populations with heteroscedasticity of variance. In Chapter 2, two important issues of parameter estimation are studied. One is to address the sensitivity of how the intrinsic estimator vary with side conditions by selecting the side condition to yield the efficient estimation through theoretical justification and simulation. Centralization is recommended for all these three effects. The other is to derive the variance formula of period and cohort effects by the Delta method to improve the default generalized linear model approach, which is unjustifiable due to the diverging number of parameters. The Delta method yields smaller variance of period and cohort effects than the PCA method through the simulation results. An F test was derived for testing on the equality of age trend across populations in Chapter 3. Simulation and applications of the F test are given in Chapter 4.

Contents

CHAPTER 1

Introduction

1.1 Motivation

The Age, Period, and Cohort (APC) models have been widely studied in modelling data arisen in many fields such as demography and sociology [1], epidemiology, psychology, and economics [48], to quantify the fixed effect of aging, period of time (e.g., measured by year of death), and cohort (e.g., measured by year of birth). Such as, the outcomes of interest includes changes in the party affiliations of Americans [24], and men's earning in the US over time [32] in demography and sociology; the incidence of syphilis [13], the incidence of breast cancer [16], and chronic disease [38] in

epidemiology; the onset of major depression [27] and psychological distress in the US in psychology [23]. There has been a shift from descriptive analysis, e.g., Frost (1939) [8], to a more analytically statistical modelling approach applied to the treatment of APC data, such as Mason et al. (1973) [32] and Fienberg (1979) [6]. Identification of the temporal trend in age, period, and cohort effects may provide clues to understanding the social momentum of events or biological dynamics of diseases and may help to identify risk factors for further studies.

The APC data, like the incidence or mortality of chronic diseases, is gathered on an individual or a population followed over time, which the time scale is usually discrete, like a year. Cohort effects are differences across a set of individuals who were born or shared an initial event in the same year, like marriage. The birth cohort often refers to the calendar year when an individual is born. The disease incidence may vary with a particular risk factor associated with the year of birth, i.e. a birth cohort effect. The period is the calendar year of the disease diagnosis. During that period, social events and factors are involved with period effect, such as world wars, infectious disease, and technology breakthroughs. Then, the age effect can be obtained by period and cohort, i.e. $age = period - cohort$, which means that we can get any one from the other two effects.

In a descriptive APC analysis, it is necessary to group periods of time into intervals with equal length (i.e. five- or ten-year ranges) for the calculation of the age- and period-specific rates. Then, cancer incidence or mortality rates are assembled within each cell of a two-way table with a rows representing categories of age groups (Age), p columns representing categories of year (Period), and $a + p - 1$ diagonals

representing cohort factors (for example, people on the same diagonal have the same birth cohort). An example of cervical cancer incidence rate in Ontario women in Canada from 1960 to 1994 is given in Table 1.1. These rates are based on 5-year age and period intervals, so that the people on the diagonals have the same birth cohort. Actually, it also allows different length of interval of age and period, especially when we don't have enough data along the period. The data set displays in a 14×7 table with fourteen rows of age groups from 20-24 to 85-89 with 5-year intervals in age, seven columns of periods from 1960-1964 to 1990-1994, and twenty diagonals of cohorts from 1871-1879 to 1966-1974. The data set is a typical APC data form, where the diagonals of the table represent birth cohorts. The study aims to identify temporal trend of cervical cancer incidence rate in age, period, and birth cohort in the study population.

The APC studies can be traced back to 1939, when Frost primarily applied these three effects to analyze the mortality rate of tuberculosis in a descriptive way with graphs and showed how the pattern of mortality along age groups changed across different periods or cohorts [8]. Even though it is useful for visualizing data with two-way graphs of age by period or age by cohort, it is not possible to quantify the assessment of age, period, and cohort effects and how these three variables operate in the source of change [13, 25]. Later, a so-called APC multiple classification model was proposed to measure these three factors simultaneously, which serves as a methodology in cohort analysis [32].

However, there is an identification problem in APC analysis caused by the linear dependence among age, period, and cohort: $period - age = cohort$. It means that

these three variables are logically confounded with one another, inducing a singular design matrix, and lead to multiple estimates when we fit the model by ordinary least square or maximum likelihood [32, 25]. Later on, many methods were proposed to address the identifiability issue, such as ignoring one of the three independent variable in the full APC model to a two-factor model (e.g., an age-period) [42, 6]. Further, Kupper (1983) explained the limitations of this approach by theory and example [26].

Another common traditional approach is provided by Mason et al. (1973), who suggest to impose constraints of two adjacent age, period, or cohort effects, like setting two adjacent age effects to be equal [32]. However, it does not resolve the problem and raises more issues about which constraint should be used because different constraint yields different estimate. Also, the choice of such a constraint relies on prior knowledge of the event, which is rarely exists or difficult to be justified [12].

Hence, identifying a unique solution among these multiple estimates opens discussion. Later on, many approaches have been proposed to resolve the identification problem, such as the shrinkage method (e.g., Osmond and Gardner (1982) [35]), the individual record approach (e.g., Robertson and Boyle (1986) [37]) forming a three-way table of age, period, and cohort, and other approaches based on estimable functions (e.g., Clayton and Shifflers (1987) [3], Holdford (1992) [15], Tarone and Chu (1992) [43]). A complete review and comparison of these methods can be found in the paper by Roberton et al. (1999) [39].

Most recent contributions include the principal component analysis (PCA) method (Fu (2000) [9]), smoothing on cohort effects (Fu (2008) [10]) and the mixed effects

model called the Hierarchical APC models (HAPC) (Yang and Land (2006) [49]). In particular, the intrinsic estimator (IE) was proposed by Fu (2000), and has been proved to be unbiased, estimable [9], and consistent with theoretical justification in 2016 [11]. In the paper, the asymptotic behavior of the multiple estimators, specially the constrained estimator, and the consistency of the intrinsic estimator were also studied by using framework of profile likelihood with large sample theory.

Even though we can get the intrinsic estimator of APC model by the PCA method, side conditions on the age, period, and cohort effects are still needed as in ANOVA models. However, Luo et al. (2013, 2014) raised concerns about the sensitivity issues that the trend of IE can be highly sensitive to side conditions, such as centralization or different reference levels [30, 31]. To address this issue, the selection of side conditions based on efficient estimation are discussed in this dissertation.

Another topic in the thesis is to find a more accurate variance estimation of the estimated period and cohort effects when fitting the log-linear model with data following a distribution from an exponential family except for Gaussian case. As we known, it usually yields biased estimates for all age, period, and cohort effects fitting with generalized linear model with small samples. That's why we need to consider the asymptotic property of biased estimates. As mentioned before, the intrinsic estimator corresponding to the intercept and age effect has been proved by Fu (2016) to be consistent [11]. However, both the estimated period and cohort effects are not, because the number of parameters for the period and cohort effects diverges to infinity when $p \to \infty$, which is the reason why a more accurate variance estimation of the period and cohort effects needs to be derived.

Among the factors influencing the incidence or mortality rate data, cancer screening plays an important role in detecting the disease at early diagnosis. For example, more people being at risk of cancer would be diagnosed at an earlier age than it was supposed to be. Therefore, the effect of screening can be reflected from the difference of age trends fitted with data before and after the cancer screening test, respectively. Meanwhile, researchers may have an interest of testing if there are differences in age trends across different populations, like gender or race. Identifying the difference of age trends may help to understand the social behavior. Therefore, the selection of an appropriate test is really important to detect the difference. In this dissertation, I discuss which test could be applied for different situations, such as the equality on trend of age effects between two or more populations with equal and unequal variance component. In particular, when testing across more than two populations with unequal variances, it can be a challenge, since there has been virtually no work done before. The derivation of such test is discussed in this dissertation.

Table 1.1: Cervical cancer incidence rate (per 10^5 person-year) and frequency 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
	51	57	58	44	47	35	13
25-29	16.01	11.18	8.92	9.74	8.48	7.43	7.53
	183	173	165	195	193	175	65
30-34	26.02	21.14	16.23	15.84	14.54	13.67	12.71
	289	266	258	295	308	334	129
35-39	38.84	25.09	21.07	18.74	18.80	18.04	18.18
	448	292	270	296	361	401	177
40-44	47.65	32.50	22.71	20.01	18.78	16.19	18.12
	564	382	268	258	304	325	158
45-49	51.48	36.69	22.15	19.20	17.74	17.29	18.31
	526	435	260	225	230	286	145
50-54	49.12	37.26	25.51	18.41	16.66	15.41	14.07
	436	386	302	216	196	202	85
55-59	51.48	40.87	34.70	21.83	16.97	17.69	13.73
	391	357	355	256	200	210	69
60-64	47.68	42.80	29.76	22.71	20.16	17.69	16.94
	306	322	255	229	234	207	80
65-69	40.44	39.17	31.44	28.79	23.35	19.26	19.16
	220	247	230	240	230	218	87
70-74	42.40	35.32	27.78	24.31	20.27	20.19	14.95
	190	177	161	165	157	186	63
75-79	42.44	36.68	28.75	25.22	21.17	21.08	19.43
	137	137	123	127	125	143	58
80-84	41.50	29.74	31.54	22.31	20.04	15.25	21.28
	81	71	88	74	79	71	45
85-89	30.79	32.43	37.10	19.81	16.42	14.87	12.06
	39	56	81	54	55	60	22

1.2 The Age, Period, and Cohort Model

The conventional linear-regression model, also known as the APC multiple classification model, was introduced to sociologists by Mason (1973) and serves as a general methodology for cohort analysis when all three of age, period, and cohort are potentially of interest [32].

To quantify the fixed effect of age, period, and cohort in analyzing cancer incidence or mortality rate data, which is usually fitted by a linear model as

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ij}, \tag{1.1}$$

or a log-linear model from generalized linear models [34] that can take various alternative forms,

$$\log(E_{ij}) = \log(N_{ij}) + \mu + \alpha_i + \beta_j + \gamma_k, \tag{1.2}$$

where $Y_{ij} = f(R_{ij})$ is a transformation by monotone function f (for example, Identity or Log function) of event rate $R_{ij} = r_{ij}/N_{ij}$ with event frequency, r_{ij}, and the total population-years of exposure, N_{ij}, in cell (i, j) for $i = 1, \ldots, a$, $j = 1, \ldots, p$, $k = 1, \ldots, (a + p - 1)$. μ is the intercept, α_i, the ith row effect (Age), β_j, the jth column effect (Period), γ_k, the kth diagonal effect (Cohort) with $k = a - i + j$. The random errors, ϵ_{ij}, are independently identically distributed (iid) with mean, 0, and constant variance, σ^2. E_{ij} is the expected number of events in cell (i, j) that is assumed to follow a Poisson distribution, and $\log(N_{ij})$ is the offset for the log-linear model in Equation (1.2).

Usually, model in Equation (1.1) can be fitted as a general linear model of fixed effects after reparameterization, by setting any reference levels for any given i_0, j_0, and k_0, where $1 \leq i_0 \leq a$, $1 \leq j_0 \leq p$ and $1 \leq k_0 \leq a + p - 1$

$$\alpha_{i_0} = 0, \beta_{j_0} = 0, \gamma_{k_0} = 0, \tag{1.3}$$

or parameter centralization,

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{p} \beta_j = 0, \sum_{k=1}^{a+p-1} \gamma_k = 0. \tag{1.4}$$

For the convenience of notation, we rewrite APC models in Equations (1.1) and (1.2) in a matrix form after reparameterization

$$\mathbf{Y} = X\mathbf{b} + \boldsymbol{\epsilon}, \tag{1.5}$$

and

$$\log(\mathbf{E}) = \log(\mathbf{N}) + X\mathbf{b}, \tag{1.6}$$

where either $\mathbf{b} = (\mu, \alpha_1, \ldots, \alpha_{a-1}, \beta_1, \ldots, \beta_{p-1}, \gamma_1, \ldots, \gamma_{a+p-2})^T$ is a vector of model parameters excluding the parameters for last row, last column, and last diagonal due to parameter centralization in Equation (1.4), or $\mathbf{b}$ is a vector of left model parameters excluding the reference levels in Equation (1.3). X is the design matrix, $\mathbf{Y}$ is a vector of the transformed rates, $\boldsymbol{\epsilon}$ is a vector of random variables with mean 0 and constant diagonal variance-covariance matrix, $\sigma^2 I$, with I denoting an identity

matrix. $\mathbf{E}$ and $\mathbf{N}$ are vectors of the expected number of events and the population-year exposure, respectively.

The ordinary least squares (OLS) estimator for the model Equation (1.5), if it exists, can be given by

$$\hat{\boldsymbol{b}} = (X^T X)^{-1} X^T \mathbf{Y}. \tag{1.7}$$

However, the inverse of the matrix $X^T X$ does not exist, because the design matrix X has one less than full column rank [25], which leads to multiple solutions when solving the linear Equation (1.7). This is the identification problem of APC analysis, which is caused by the linear dependency among these three variables: $period - age = cohort$ [40, 14].

Kupper et al. (1983) studied the APC model in Equation (1.5) and provided a closed form for the null vector $\mathbf{v_0}$ of the singular design matrix X [26], which is ${\mathbf{v_0}}^T = (0\, \mathbf{A}^T\, \mathbf{P}^T\, \mathbf{C}^T)$ with

$$\mathbf{A}^T = (1 - \frac{a+1}{2}, \ldots, (a-1) - \frac{a+1}{2}),$$

$$\mathbf{P}^T = (\frac{p+1}{2} - 1, \ldots, \frac{p+1}{2} - (p-1)),$$

and

$$\mathbf{C}^T = (1 - \frac{a+p}{2}, \ldots, (a+p-2) - \frac{a+p}{2}).$$

The null vector spans a null space, $\{t\mathbf{v_0}\}$, with an arbitrary real number, t, which vanishes by the design matrix X, i.e. $X^T X(t\mathbf{v_0}) = 0$. Further, if $\mathbf{v}$ is an estimator of the APC model in Equation (1.5) or (1.6), $\mathbf{v} + t\mathbf{v_0}$ also is an estimator of the model

for any real number t.

1.2.1 The Intrinsic Estimator

To address the identification problem, the conventional and most widely used approach is to add a constraint on coefficients, for example, setting adjacent fixed effect of age, period, or cohort to have equal value, such as $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$ and so on. In fact, such selection of a constraint relies on the investigators subjective judgment, which could introduce bias to model parameter estimates [25]. It yields different secular trends in row, column, and diagonal effects with different constraints. It is difficult to determine which estimator presents the right trends and can be used in trend estimation for a given data set.

The conventional constrained method has been widely used for over three decades until 2000, an innovative approach of estimation termed the intrinsic estimator (IE) was proposed and compared to the traditional constrained approach, which has been proven to be estimable, unbiased [9, 50], and consistent [11].

Since the design matrix X is one less than full column rank, the parameter space of model in Equation (1.5 or 1.6) can be decomposed as the direct sum of two linear sub-spaces that are orthogonal to each other. One subspace denotes the null space generated by the normalized eigenvector $\mathbf{B}_0$ corresponding to the only one zero eigenvalue of the matrix, $X^T X$, where $\mathbf{B}_0 = \mathbf{v_0}/||\mathbf{v_0}||$.

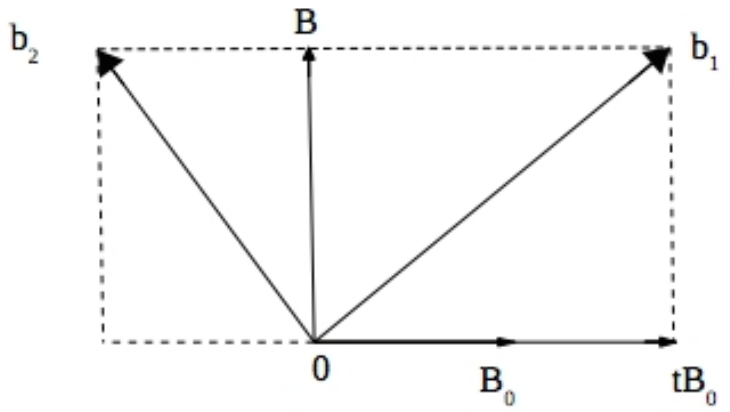

Figure 1.1: The geometric plot of projection of parameter estimates.

Any estimator $\hat{\mathbf{b}}$ of the unconstrained APC model can be decomposed as

$$\hat{\mathbf{b}} = \mathbf{B} + t\mathbf{B}_0, \tag{1.8}$$

where $\mathbf{B}_0 = \mathbf{v_0}/||\mathbf{v_0}||$ is the normalized eigenvector $\mathbf{v_0}$ of matrix, $X^T X$ and t an arbitrary real number. $\mathbf{B}$ is a special estimator orthogonal to $\mathbf{B}_0$ in the parameter space and is called the intrinsic estimator. It can be obtained by projecting any constrained estimator to the non-null parameter space, which is $\mathbf{B} = (I - \mathbf{B}_0\mathbf{B}_0^T)\hat{\mathbf{b}}$. The projection is illustrated in Figure 1.1. Another way to compute the IE is by principal component analysis (PCA) method [41].

CHAPTER 2

Selection of Side Conditions and Variance Estimation of Parameters in Age, Period, and Cohort Model

2.1 Selection of Side Conditions in Age, Period, and Cohort Model

2.1.1 Introduction

The analysis of variance (ANOVA), which was developed by Fisher (1992), is a collection of statistical models like one-way ANOVA or two-way ANOVA model,

used to compare the differences among group means [7]. Usually, side conditions are needed in order to get a unique parameter estimator, either by centralization or setting reference levels. So for APC models, it's a special case of two-way ANOVA model just treating the cohort effect as an interaction term between the age and period effects.

The intrinsic estimator (IE) has recently been shown to possess good properties: estimability, unbiasedness, robustness, consistency (large sample theory) [9, 11]. However, Luo et al. (2013, 2014) raised concerns about the sensitivity issues and thus questioned the usefulness of the IE. They found that trend of IE can be highly sensitive to side conditions, centralization or different reference levels [30, 31]. So a researcher would often reach different conclusions about the effects of age, period and cohort depending on the selection of side conditions.

Same issue can be found in ANOVA models, but, it does not make a difference because we are usually interested in differences among the treatment effects rather than their actual values, which is uniquely estimated regardless of which side condition would be used. Researchers select side conditions based on their practical needs, if there is one reference level to which they would like to compare with all the other levels. For example, the first reference level is assumed to be zero in R, while the last reference is used in SAS. However, it matters in APC model, since we are interested in interpreting the trend of age, period, and cohort effects. Therefore, the selection of side conditions plays an important role in estimating the model parameters, specially in APC models. Also, there has been virtually no criterion about how to select side conditions when to estimate parameters in APC models.

Given a variety of possible estimates, how would we choose which side condition to use? Qualitatively, it would be sensible to choose that estimate whose sampling distribution was most highly concentrated about the true parameter value [36]. Mean squared error (MSE) is the most commonly used measure of concentration. In searching for an optimal estimate, we might ask whether there is a lower bound for the MSE of any estimate. If such a lower bound exists, it would function as a benchmark against which estimates could be compared. The sampling variance of an estimator has been widely applied as a criterion of choosing method when to estimate parameters since the days of Laplace and Gauss [22]. But only in relatively recent times has it been established that, under fairly general conditions, there exists a lower bound on the variance of an estimator of a deterministic parameter, which is known as the Cramér-Rao inequality.

Based on statistics theory, efficient estimates are always preferred. A term applied in the context of comparing different methods of estimating the same parameter; the estimate with the lowest variance being regarded as the most efficient [5]. We select the side conditions on parameters yielding relatively efficient estimate. To illustrate the method, the one-way and two-way ANOVA models are studied first, since it is unlikely to derive a closed form for the parameter estimates in APC models.

2.1.2 Side Conditions for ANOVA Models

2.1.2.1 One-Way ANOVA Model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \tag{2.1}$$

where Y_{ij} is the response variable, μ is the intercept, and τ_i is the ith group effect. $\epsilon_{ij} \sim N(0, \sigma^2)$ are iid, for $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$, where n_i is the number of observations in ith group.

Parameter Estimate

By the least square method, we have

$$\min \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \mu - \tau_i)^2.$$

By optimal condition, which is taking the partial derivative with respect to μ and τ_i, respectively, then setting them equal to zero, we have

$$(\sum_{i=1}^{a} n_i)\mu + \sum_{i=1}^{a} n_i \tau_i = \sum_{i,j} Y_{ij},$$

$$n_i \mu + n_i \tau_i = \sum_{j=1}^{n_i} Y_{ij},\ i = 1, \ldots, a.$$

Here we have $a + 1$ unknown parameters $\mu, \tau_1, \ldots, \tau_a$, but we only have a independent equations, so we still need one more side condition on parameters to get the unique estimate, either by centralization

$$\sum_{i=1}^{a} \tau_i = 0, \tag{2.2}$$

or setting one reference level equal to zero, i.e.

$$\tau_{i_0} = 0, \tag{2.3}$$

where $i_0 \in \{1, 2, \ldots, a\}$.

Comparison of variances between different side conditions

The formula of parameter estimates and its variance for $\hat{\mu}, \hat{\tau}_1, \ldots, \hat{\tau}_a$ and a general contrast corresponding to each side condition, i.e. $\tau_1 = 0$, $\tau_2 = 0$, ..., $\tau_a = 0$ and $\sum_{i=1}^{a} \tau_i = 0$, are given in the Tables 2.1 and 2.2, respectively.

Table 2.1: Estimates and contrast by side condition

Side Condition	$\hat{\mu}$	$\hat{\tau}_j$	$L = \sum_{j=1}^{a} c_j \hat{\tau}_j$
$\tau_1 = 0$	$\overline{Y}_{1.}$	$\overline{Y}_{j.} - \overline{Y}_{1.}, j \neq 1$	$\sum_{j=1}^{a} c_j \overline{Y}_{j.}$
$\tau_i = 0$	$\overline{Y}_{i.}$	$\overline{Y}_{j.} - \overline{Y}_{i.}, j \neq i$	$\sum_{j=1}^{a} c_j \overline{Y}_{j.}$
$\tau_a = 0$	$\overline{Y}_{a.}$	$\overline{Y}_{j.} - \overline{Y}_{a.}, j \neq a$	$\sum_{j=1}^{a} c_j \overline{Y}_{j.}$
$\sum_{i=1}^{a} \tau_i = 0$	$\frac{1}{a} \sum_{i=1}^{a} \overline{Y}_{i.}$	$\overline{Y}_{j.} - \frac{1}{a} \sum_{i=1}^{a} \overline{Y}_{i.}$	$\sum_{j=1}^{a} c_j \overline{Y}_{j.}$

Table 2.2: Variance of estimates and the contrast by side condition

Side Condition	$Var(\hat{\mu})$	$Var(\hat{\tau}_j)$	$Var(L)$
$\tau_1 = 0$	$\frac{\sigma^2}{n_1}$	$\frac{\sigma^2}{n_j} + \frac{\sigma^2}{n_1}, j \neq 1$	$\sum_{j=1}^{a} c_j^2 \frac{\sigma^2}{n_j}$
$\tau_i = 0$	$\frac{\sigma^2}{n_i}$	$\frac{\sigma^2}{n_j} + \frac{\sigma^2}{n_i}, j \neq i$	$\sum_{j=1}^{a} c_j^2 \frac{\sigma^2}{n_j}$
$\tau_a = 0$	$\frac{\sigma^2}{n_a}$	$\frac{\sigma^2}{n_j} + \frac{\sigma^2}{n_a}, j \neq a$	$\sum_{j=1}^{a} c_j^2 \frac{\sigma^2}{n_j}$
$\sum_{i=1}^{a} \tau_i = 0$	$\frac{1}{a^2} \sum_{i=1}^{a} \frac{\sigma^2}{n_i}$	$(1 - \frac{2}{a}) \frac{\sigma^2}{n_j} + \frac{1}{a^2} \sum_{i=1}^{a} \frac{\sigma^2}{n_i}$	$\sum_{j=1}^{a} c_j^2 \frac{\sigma^2}{n_j}$

Remark: $L = \sum_{j=1}^{a} c_j \hat{\tau}_j$ is a contrast with $\sum_{j=1}^{a} c_j = 0$ in Tables 2.1 and 2.2.

Based on Tables 2.1 and 2.2, we conclude that

- Contrasts have the same value and the same variance regardless of side condition.
- With the same sample size for each group (i.e. $n_i = n$ for all $i = 1, \ldots, a$), centralization yields the smallest variance of estimates, because of

$$\underbrace{(1 - \frac{1}{a})\frac{\sigma^2}{n}}_{Var(\hat{\tau}_j)_{cen}} < \underbrace{\frac{2\sigma^2}{n}}_{Var(\hat{\tau}_j)_{ref}} . \tag{2.4}$$

- The reference level using the largest sample size yields smallest variance among all reference levels $\tau_1 = 0$, $\tau_2 = 0$, ..., $\tau_a = 0$.

$$\min_{\tau_1=0,\ldots,\tau_a=0} \{Var(\hat{\tau}_j)\} = \frac{\sigma^2}{n_j} + \frac{\sigma^2}{n_{i'}}, \tag{2.5}$$

where $i' = arg\max_k \{n_k\}$.

With different sample size for each treatment group, we compare the variance $Var(\hat{\tau}_j)$ using reference level having the largest sample size (i.e. $\tau_{i'}$=0) with the centralization in Equation (2.2) as follows,

$$\begin{aligned} Var(\hat{\tau}_j)_{ref} \geq Var(\hat{\tau}_j)_{cen} \quad &\Longleftrightarrow \quad (1 - \frac{1}{a^2})\frac{\sigma^2}{n_{i'}} \geq \frac{1}{a^2}\sum_{k\neq i}^{a} \frac{\sigma^2}{n_k} - \frac{2}{a}\frac{\sigma^2}{n_j} \\ &\Longleftrightarrow \quad \begin{cases} n_{i'} \leq \frac{a^2-1}{\sum_{k\neq i'}^{a} \frac{1}{n_k} - \frac{2a}{n_j}}, & \text{if } n_j > \frac{2a-1}{\sum_{k\neq i',j}^{a} \frac{1}{n_k}}, \\ \text{Always True}, & \text{Otherwise}, \end{cases} \end{aligned} \tag{2.6}$$

where $Var(\hat{\tau}_j)_{ref}$ and $Var(\hat{\tau}_j)_{cen}$ are the variances using the largest sample and the centralization, respectively. Based on Equation (2.6), we have

- If the sample size is about the same ($\approx$), centralization yields the smallest variance of estimates.
- If sample size varies largely, variance for some estimates using reference level may be smaller than that by the centralization.

Under this case, we couldn't figure out which side condition yields the efficient estimate. To address it, we introduce a minimax approach, which minimizes the largest variance of group effect among all side conditions $\tau_1 = 0$, $\tau_2 = 0$, ..., $\tau_a = 0$, and $\sum_{i=1}^{a} \tau_i = 0$, which is

$$\min_{\text{Side Cond.}} \max_{j=1,\ldots,a} \{Var(\hat{\tau}_j)\}. \tag{2.7}$$

- Minimizing the largest variance among all reference levels $\tau_1 = 0$, $\tau_2 = 0$, ..., $\tau_a = 0$

$$\begin{aligned} \min_{\tau_1=0,\ldots,\tau_a=0} \max_{j=1,\ldots,a} \{Var(\hat{\tau}_j)\} &= \min_{i=1,\ldots,a} \{\frac{\sigma^2}{\min_{j\neq i}\{n_j\}} + \frac{\sigma^2}{n_i}\} \\ &= \frac{\sigma^2}{\min_j\{n_j\}} + \frac{\sigma^2}{\max_i\{n_i\}}. \end{aligned} \tag{2.8}$$

- Under centralization

$$\max_{j=1,\ldots,a} \{Var(\hat{\tau}_j)\} = (1 - \frac{2}{a})\frac{\sigma^2}{\min_j\{n_j\}} + \frac{1}{a^2}\sum_{i=1}^{a}\frac{\sigma^2}{n_i}. \tag{2.9}$$

We conclude that the centralization achieves the minimum of the largest variance of estimates, because of Equation (2.9) $<$ (2.8),

$$\begin{aligned}(1-\frac{2}{a})\frac{\sigma^2}{\min\limits_j\{n_j\}}+\frac{1}{a^2}\sum_{i=1}^{a}\frac{\sigma^2}{n_i} &\leq (1-\frac{2}{a})\frac{\sigma^2}{\min\limits_j\{n_j\}}+\frac{1}{a^2}\sum_{i=1}^{a}\frac{\sigma^2}{\min\limits_i\{n_i\}}\\ &= (1-\frac{2}{a})\frac{\sigma^2}{\min\limits_j\{n_j\}}+\frac{1}{a}\frac{\sigma^2}{\min\limits_i\{n_i\}}\\ &= \frac{\sigma^2}{\min\limits_j\{n_j\}}-\frac{1}{a}\frac{\sigma^2}{\min\limits_j\{n_j\}}\\ &< \frac{\sigma^2}{\min\limits_j\{n_j\}}+\frac{\sigma^2}{\max\limits_i\{n_i\}}.\end{aligned}$$

2.1.2.2 Two-Way ANOVA Model

$$Y_{ijk}=\mu+\tau_i+\beta_j+\epsilon_{ijk}, \tag{2.10}$$

where $i=1,...,a;\ j=1,...,b;\ k=1,...,n_{ij}$, $\epsilon_{ijk}\sim N(0,\sigma^2)$ are iid, n_{ij} is the number of observations for each cell (i,j). Y_{ijk} is the response variable, μ is the intercept, and τ_i is the ith row effect, β_j is the jth column effect.

Parameter estimate

By the least square method, we have

$$\min\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(Y_{ijk}-\mu-\tau_i-\beta_j)^2. \tag{2.11}$$

After taking the partial derivative of the loss function above with respect to each

parameter and setting them to zero, we have $a+b+1$ normal equations as follows,

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}\mu+\sum_{i=1}^{a}(\sum_{j=1}^{b}n_{ij})\tau_i+\sum_{j=1}^{b}(\sum_{i=1}^{a}n_{ij})\beta_j=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}Y_{ijk},$$

$$\sum_{j=1}^{b}n_{ij}\mu+(\sum_{j=1}^{b}n_{ij})\tau_i+\sum_{j=1}^{b}n_{ij}\beta_j=\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}Y_{ijk}, i=1,\ldots,a,$$

$$\sum_{i=1}^{a}n_{ij}\mu+\sum_{i=1}^{a}n_{ij}\tau_i+(\sum_{i=1}^{a}n_{ij})\beta_j=\sum_{i=1}^{a}\sum_{k=1}^{n_{ij}}Y_{ijk}, j=1,\ldots,b,$$

where there are $a+b+1$ unknown parameters, but only $a+b-1$ independent equations, which means that two more side conditions are needed to get the unique estimate, either by centralization

$$\sum_{i=1}^{a}\tau_i=0,\ \sum_{j=1}^{b}\beta_j=0, \tag{2.12}$$

or by setting one reference level of row and column effect equal to zero, i.e.

$$\tau_{i_0}=0,\ \beta_{j_0}=0, \tag{2.13}$$

where $i_0\in\{1,2,\ldots,a\}$ and $j_0\in\{1,2,\ldots,b\}$.

Comparison of variances between different side conditions

Using similar approach as in one-way ANOVA models, the centralization or reference levels on both row and column effects simultaneously are considered for simplification instead of the combination of centralization on one factor and reference level on the other.

With the equal sample size across all cells (i.e. $n_{ij}=n$ for $i=1,\ldots,a$ and $j=1,\ldots b$), we have

Table 2.3: Estimates and the contrast by side condition

Estimates	Side condition	
	$\tau_i = 0, \beta_j = 0$	$\sum_{i=1}^{a} \tau_i = 0, \sum_{j=1}^{b} \beta_j = 0$
$\hat{\mu}$	$\overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$	$\overline{Y}_{...}$
$\hat{\tau}_l$	$\overline{Y}_{l..} - \overline{Y}_{i..}, l \neq i$	$\overline{Y}_{l..} - \overline{Y}_{...}, l = 1, \ldots, a$
$\hat{\beta}_m$	$\overline{Y}_{.m.} - \overline{Y}_{.j.}, m \neq j,$	$\overline{Y}_{.m.} - \overline{Y}_{...}, m = 1, \ldots, b$
$L_1 = \sum_{l=1}^{a} c_l \hat{\tau}_l$	$\sum_{l=1}^{a} c_l \overline{Y}_{l..}$	$\sum_{l=1}^{a} c_l \overline{Y}_{l..}$
$L_2 = \sum_{m=1}^{b} d_m \hat{\beta}_m$	$\sum_{m=1}^{b} d_m \overline{Y}_{.m.}$	$\sum_{m=1}^{b} d_m \overline{Y}_{.m.}$

Table 2.4: Variance of estimates and the contrast by side condition

Variance	Side condition	
	$\tau_i = 0, \beta_j = 0$	$\sum_{i=1}^{a} \tau_i = 0, \sum_{j=1}^{b} \beta_j = 0$
$Var(\hat{\mu})$	$\frac{(a+b-1)\sigma^2}{abn}$	$\frac{\sigma^2}{abn}$
$Var(\hat{\tau}_l)$	$\frac{2\sigma^2}{bn}, l \neq i$	$\frac{\sigma^2}{bn} - \frac{\sigma^2}{abn}, l = 1, \ldots, a$
$Var(\hat{\beta}_m)$	$\frac{2\sigma^2}{an}, m \neq j$	$\frac{\sigma^2}{an} - \frac{\sigma^2}{abn}, m = 1, \ldots, b$
$Var(L_1)$	$\frac{\sigma^2}{bn}(\sum_{l=1}^{a} c_l^2)$	$\frac{\sigma^2}{bn}(\sum_{i=1}^{a} c_i^2)$
$Var(L_2)$	$\frac{\sigma^2}{an}(\sum_{m=1}^{b} d_m^2)$	$\frac{\sigma^2}{an}(\sum_{j=1}^{b} d_j^2)$

Remark: $\sum_{l=1}^{a} c_l = 0$ and $\sum_{m=1}^{b} d_m = 0$. From Tables 2.3 and 2.4, we have:

- Contrasts have the same value and the same variance whichever side condition on row and column effects would be used.
- The centralization on row and column effects yields the smallest variance of estimates, thus achieving an efficient estimation, because

$$\underbrace{(1 - \frac{1}{a})\frac{\sigma^2}{bn}}_{Var(\hat{\tau}_l)_{cen}} < \underbrace{\frac{2\sigma^2}{bn}}_{Var(\hat{\tau}_l)_{ref}}, \tag{2.14}$$

$$\underbrace{(1-\frac{1}{b})\frac{\sigma^2}{bn}}_{Var(\hat{\beta}_m)_{cen}} < \underbrace{\frac{2\sigma^2}{an}}_{Var(\hat{\beta}_m)_{ref}} . \tag{2.15}$$

- We have the same variance of row effects and column effects, respectively, among any reference levels $\tau_i = 0, \beta_j = 0$ for $i \in \{1, \ldots, a\}$ and $j \in \{1, \ldots, b\}$.

With unequal sample size among each cell, because there is no explicit expression of estimates and its variance as we did in Tables 2.3 and 2.4, a simulation was conducted later in the simulation part, from which the centralization still yields the smallest variance of estimates. Also, the variance of a contrast remains the same regardless of side condition, which further emphasizes its estimablity.

2.1.3 Side Conditions for Age, Period, and Cohort Models

The APC models are special cases of two-way ANOVA models with cohort effects interpreted as an interaction effect between age and period effects. Following the data structure of APC models, there are p and a observations for each age and period effect, respectively, which is a balanced data design for these two variables. However, the number of observations for each cohort group varies from one observation on the oldest and youngest cohorts to the largest number of observations $(\min(a, p))$ on the middle cohorts.

The selection of side conditions on age, period, and cohort effects is important, because efficient estimation is preferred when interpreting the trend of these three effects. To get efficient estimation, it is recommended to use the centralization on age and period effects following the balanced data design.

For cohort effects, the selection of side conditions on cohort effects needs further study,

because the imbalance data design of the cohort effects, i.e. the number of observations varies with cohort effect. Following the similar procedure as what we did in one-way ANOVA models, the variance of estimates was compared with different side condition on cohort effects by simulation instead of theoretical justification, because it is very unlikely to derive an explicit formula of the parameter estimates.

From the simulation result, the centralization is still preferred, because it almost yields efficient estimation except for very few end points of parameters, such as the first age effect, the last period, and the last cohort effect. By comparing all reference levels, the reference level having smaller observations on cohort effect presents larger variance of estimates, which coincides with the conclusion in one-way ANOVA models.

2.1.4 Summary

Based on the theoretical justification and simulation results of ANOVA models, I conclude that the centralization is preferred as an efficient side condition for one-way ANOVA, two-way ANOVA, and APC models.

2.1.5 Simulation and Application

The sensitivity issue of APC model

In the following, I illustrate the sensitivity issue in Figure 2.1 on cervical cancer incidence rate data in the Table 1.1.

From Figure 2.1, different side conditions yield different IE trend estimations in age, period, and cohort effects, where estimations were obtained through side conditions of the

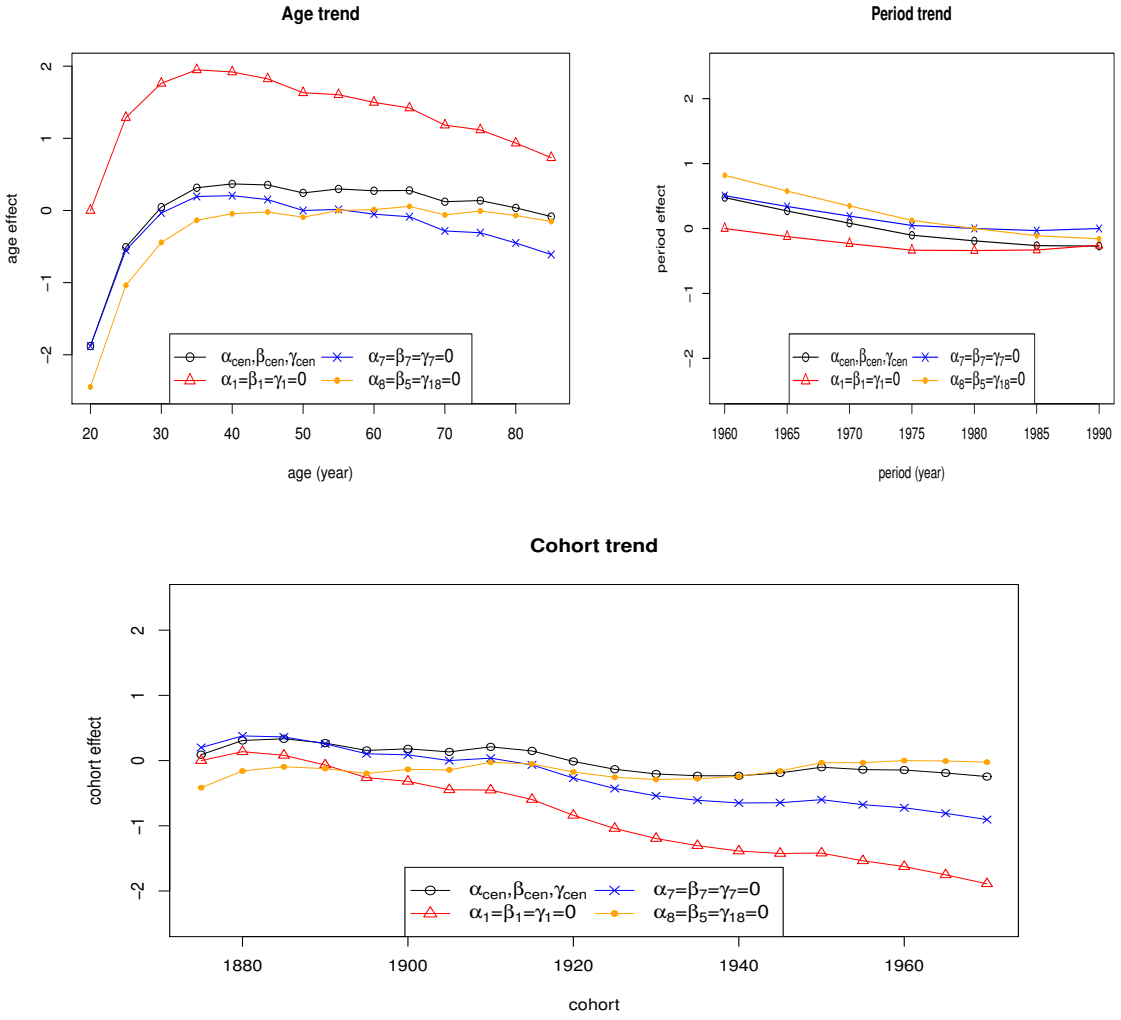

Figure 2.1: **The trend of age, period, and cohort effects of the cervical cancer incidence rate by side condition**. The figure displays the estimated age, period, and cohort effects by four different side conditions, which are $\sum_{i=1}^{14}\alpha_i = \sum_{j=1}^{7}\beta_j = \sum_{k=1}^{20}\gamma_k = 0$, $\alpha_1 = \beta_1 = \gamma_1 = 0$, $\alpha_7 = \beta_7 = \gamma_7 = 0$, and $\alpha_8 = \beta_5 = \gamma_{18} = 0$. Age represent age groups of 5-years interval from 20-24, 25-29, to 85-89, period represent period groups of 5-years interval from 1960-1964, 1965-1969, to 1990-1994, and cohort represent cohort groups of 9-years interval from 1871-1879, 1976-1984, to 1966-1974.

centralization, $\alpha_1 = \beta_1 = \gamma_1 = 0$, $\alpha_7 = \beta_7 = \gamma_7 = 0$, or $\alpha_8 = \beta_5 = \gamma_{18} = 0$, respectively. If all trends of IE are parallel, then it does not matter because we only care about trends in the prediction not the specific value. However, they are not parallel, especially from the third sub-figure, which means that we may have different interpretation of the trends with different side conditions.

Simulation for one-way ANOVA model

Data generation:
Let $\mu = 1.0$ be the intercept, all group effects were given by $\tau = (-3.2, -0.2, 2.1, 2.8, 1.8, 0.3, -1.2, -2.2)^T$, the sample size for each group were given in a vector $n = (10, 20, 30, 40, 50, 60, 70, 80)^T$. Each observation was generated by

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$, the variance component, σ^2, was determined by a given signal-noise ratio of 3. I computed variance of estimates for each generated data set, then repeated the procedure 1000 times and took the average of them, which is shown in Figure 2.2.

Figure 2.2 displays the variance of estimates by different side conditions, which were $\tau_1 = 0$, $\tau_4 = 0$, $\tau_8 = 0$ with the number of observations of 10, 40 and 80, respectively, and the centralization $\sum_{i=1}^{8} \tau_i = 0$. From Figure 2.2, centralization yields the smallest variance of estimates. The reference level with larger sample size yields smaller variance of estimates among all side condition of reference levels.

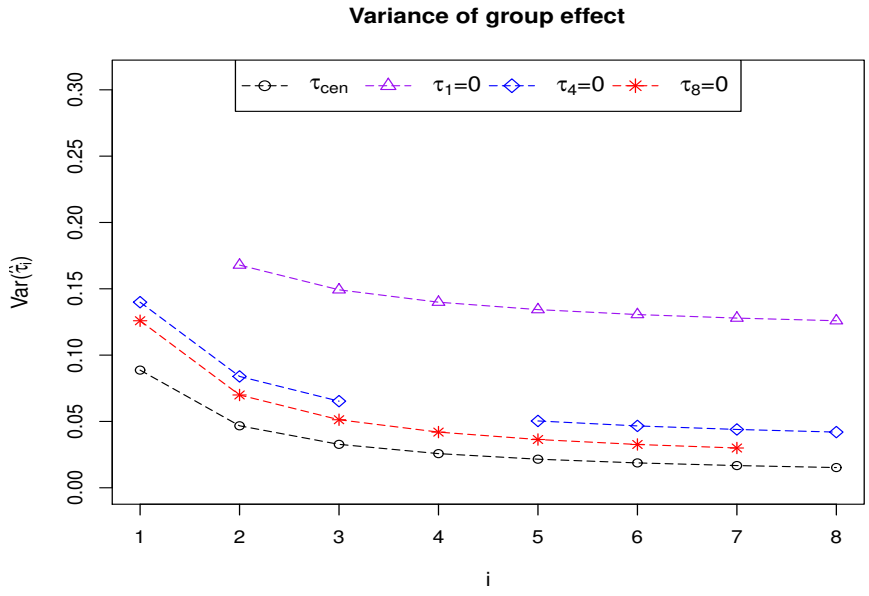

Figure 2.2: **Variance simulation in one-way ANOVA by side condition**. $\text{Var}(\hat{\tau}_i)$ represents the variance of each group effect for $i = 1, \ldots, 8$, by four different side conditions, which are $\sum_{i=1}^{8} \tau_i = 0$, $\tau_1 = 0$, $\tau_4 = 0$, and $\tau_8 = 0$.

Simulation for two-way ANOVA model

Data generation:

Let $\mu = 1.0$ be the intercept, $\tau = (-3.2, -0.2, 1.8, 2.3, 1.8, 0.3, -2.2, -3.2, 0.8, 1.8)^T$ be the row effects, and $\beta = (1.2, -0.8, 0.2, -0.8, 0.2, 0.5)^T$ be column effects. The number of observations for each cell is given in Table 2.5.

Each observation was generated by

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$, the variance component, σ^2, was determined by a given signal-noise

Table 2.5: The sample size for each cell in a two-way table

	n_{ij}						Total
1	80	80	20	60	30	40	310
2	80	60	10	90	80	80	400
3	40	100	20	30	90	60	340
4	40	90	20	20	30	90	290
5	40	100	50	90	60	80	420
6	20	90	80	10	50	100	350
7	60	90	90	100	60	90	490
8	10	20	90	90	80	90	380
9	100	40	90	80	70	30	410
10	20	80	20	70	40	20	250
Total	490	750	490	640	590	680	

ratio of 3. Variances of estimates by side condition were computed over 1000 times, then were taken for the average, which is shown in Figure 2.3.

Figure 2.3 displays the variance of estimates by different side conditions, which includes reference levels of $\tau_7 = \beta_2 = 0$, $\tau_3 = \beta_5 = 0$, $\tau_{10} = \beta_1 = 0$ varying from the largest number of observations to smallest number of observations for row and column effects and the centralization of $\sum_{i=1}^{10} \tau_i = 0$ and $\sum_{j=1}^{6} \beta_j = 0$. From Figure 2.3, the centralization yields the smallest variance of estimates among all side conditions. Among these reference levels, the reference level $\tau_7 = \beta_2 = 0$ yields the smallest variance for both row and column effects, because the row effect τ_7 and the column effect β_2 have the largest number of observations (490 and 750) among all row effects and column effects, receptively, which coincides with the theoretical justification in one-way ANOVA model.

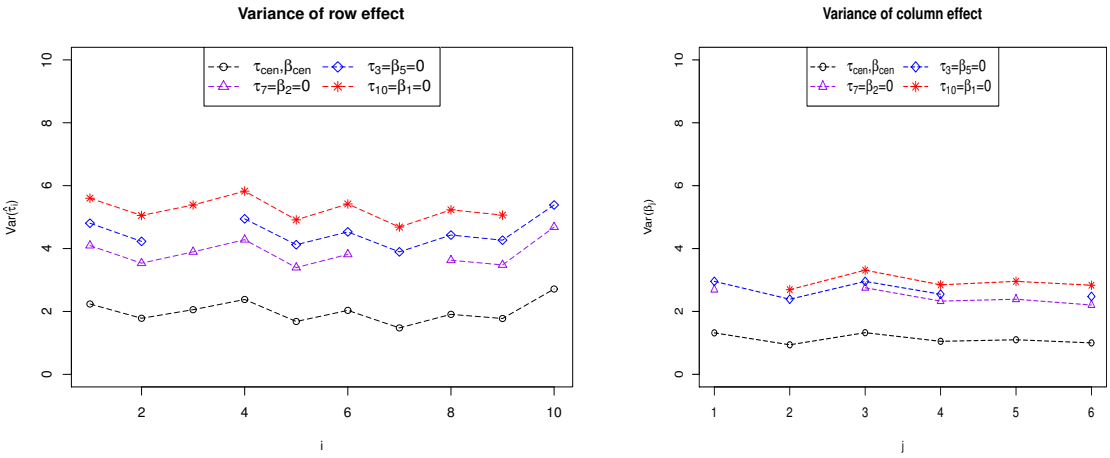

Figure 2.3: **Simulation on variance of estimates in two-way ANOVA model by side condition**. $\mathrm{Var}(\hat{\tau}_i\ )$ and $\mathrm{Var}(\hat{\beta}_j)$ represent the variance of row and column effect, respectively, for $i = 1, \ldots, 8$ and $j = 1, \ldots, 6$, by four different side conditions, which are $\sum_i \tau_i = \sum_j \beta_j = 0$, $\tau_7 = \beta_2 = 0$, $\tau_3 = \beta_5 = 0$, and $\tau_{10} = \beta_1 = 0$.

Simulation for APC Model

Here are the parameters specified for APC model, $\mu = 1$, $\alpha = (-3.2, -0.2, 1.8, 2.3, 1.8, 0.3, -2.2, -3.2, 0.8, 1.8)^T$, $\beta = (1.2, -0.8, 0.2, -0.8, 0.2)^T$, $\gamma = (-0.5046, -0.3139, -0.1387, 0.0141, 0.1382, 0.2287, 0.2821, 0.2963, 0.2705, 0.2060, 0.1052, -0.0278, -0.1878, -0.3687)^T$.

Figure 2.4 displays the variance of age, period, and cohort effects by different side conditions in APC model, which include three reference levels of $\gamma_1 = 0$, $\gamma_6 = 0$, and $\gamma_{11} = 0$ having 1, 5, and 4 observations for each cohort, respectively, and the centralization $\sum_{k=1}^{14} \gamma_k = 0$, in which the zero variance of reference level ($\gamma_1 = 0$, $\gamma_6 = 0$, and $\gamma_{11} = 0$) was dropped from the plot of the variance of cohort effects.

It is shown that the centralization almost yields the smallest variance of estimates except few end points of age, period, and cohort effects, like the first age effect, the last

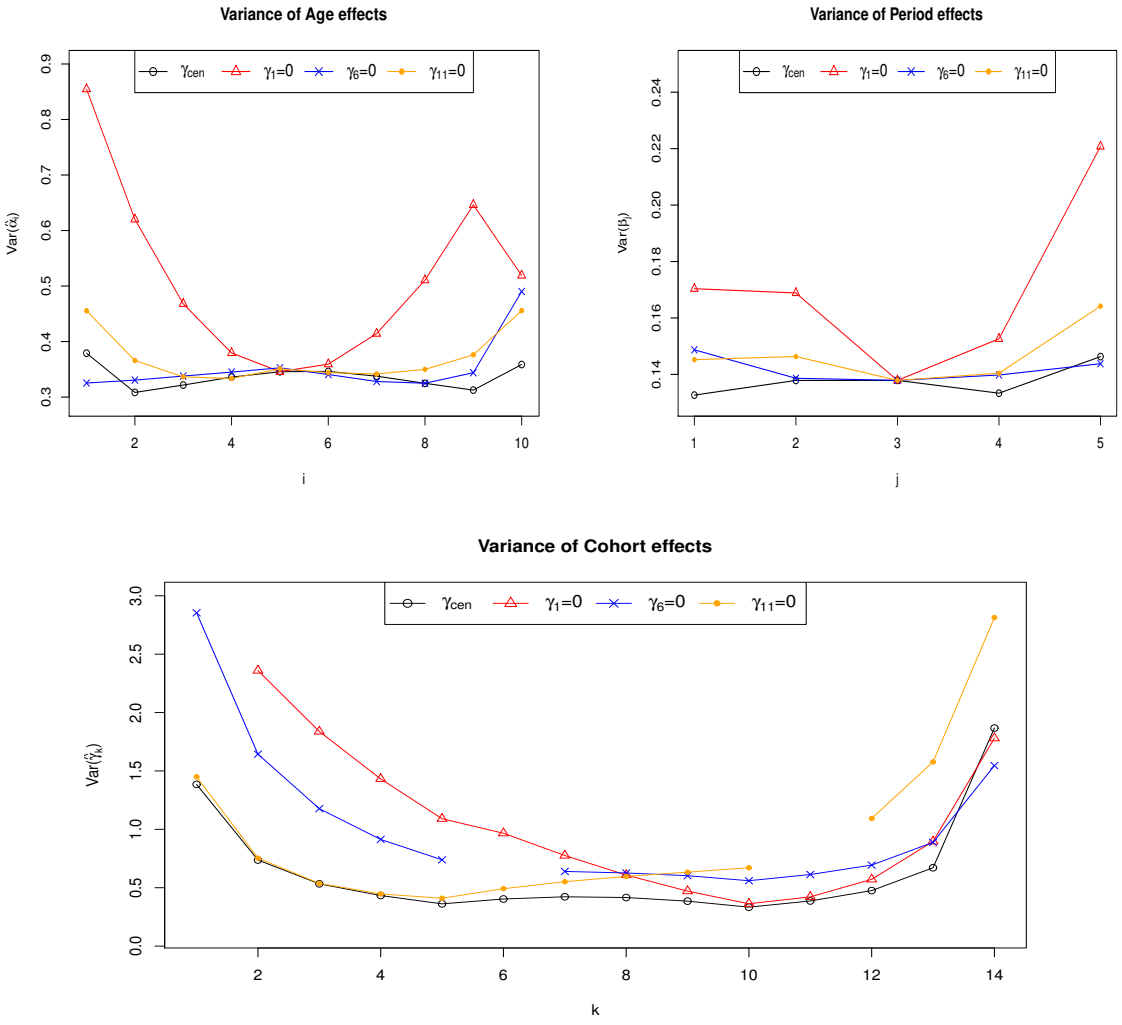

Figure 2.4: **Simulation on variance of age, period, and cohort effects by side condition on cohort effects**. $\mathrm{Var}(\hat{\alpha}_i)$, $\mathrm{Var}(\hat{\beta}_j)$, and $\mathrm{Var}(\hat{\gamma}_k)$ represent the variance of estimated age, period, and cohort effects, respectively, for $i = 1, \ldots, 10$, $j = 1, \ldots, 6$, and $k = 1, \ldots, 14$, by four different side conditions on cohort effects, which are $\sum_{k=1}^{14} \gamma_k = 0$, $\gamma_1 = 0$, $\gamma_6 = 0$, and $\gamma_{11} = 0$.

period, and the last cohort effect. Furthermore, the reference level having smaller observations on cohort effect almost yields the larger variance of estimates, specially for age and period effects. Similar result can also be found in Figure 2.6 for the real APC data of the cervical cancer incidence rate for women in Canada.

Parameters estimation and variance for the cervical Cancer incidence data

For APC models, the centralization is preferred on age, period, and cohort effects, the following example shows the variance of estimates by different side condition on cohort effects. Figure 2.5 displays the trends of age, period and cohort effects with 95% confidence interval with standard error computed in Figure 2.6, by different side conditions on cohort effects for the cervical cancer incidence rate data, respectively. It is shown that the trends of age, period, and cohort effects almost did not vary with side condition of cohort effects, but the centralization yielded the narrowest confidence interval, which means that it yielded an efficient estimation.

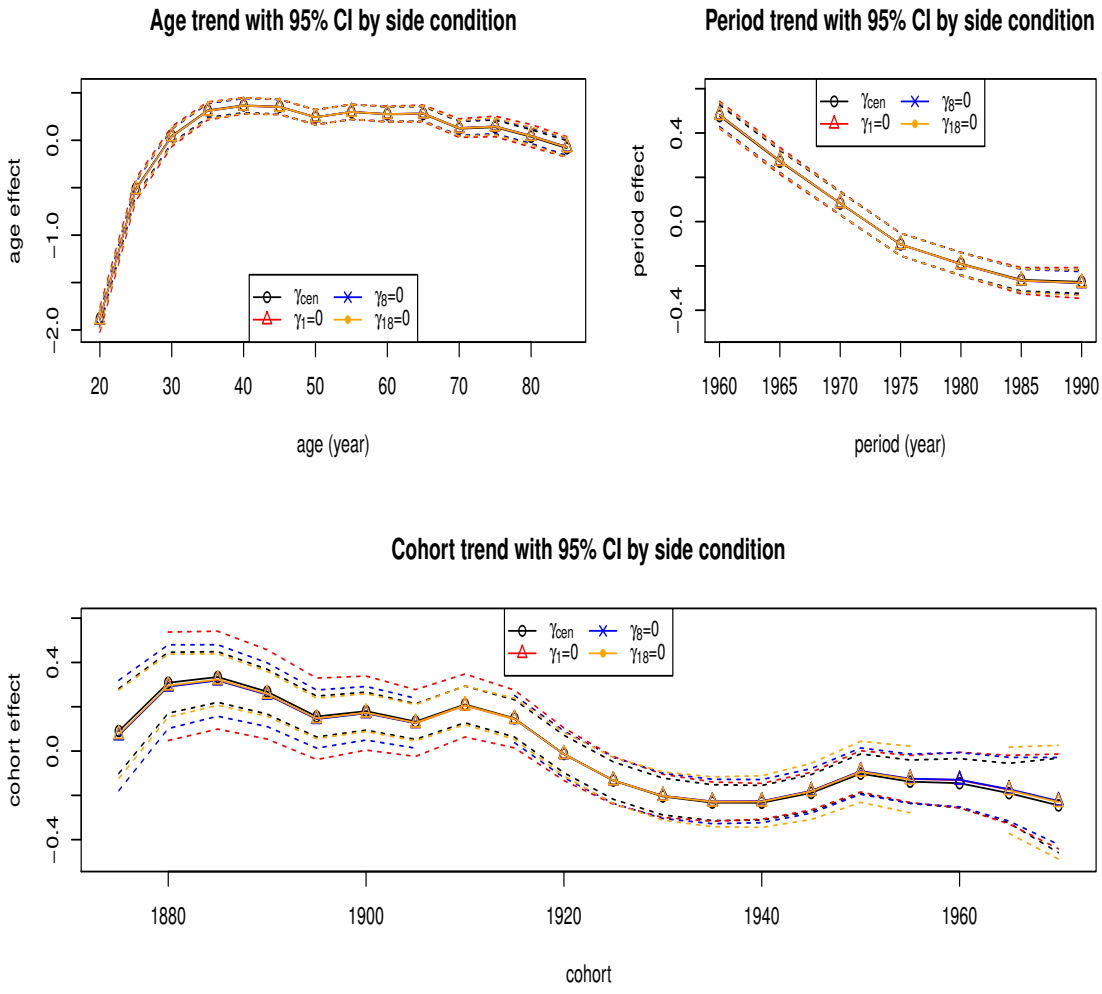

Figure 2.5: **The trend of age, period and cohort effects with 95% confidence interval by side condition for the cervical cancer data in Table 1.1**. The four different 95% confidence intervals were computed from the results in the Figure 2.6 by four different side conditions on cohort effects, which are which are $\sum_{k=1}^{20} \gamma_k = 0$, $\gamma_1 = 0$, $\gamma_8 = 0$, and $\gamma_{18} = 0$. Age represent age groups of 5-years interval from 20-24, 25-29, to 85-89, period represent period groups of 5-years interval from 1960-1964, 1965-1969, to 1990-1994, and cohort represent cohort groups of 9-years interval from 1871-1879, 1976-1984, to 1966-1974.

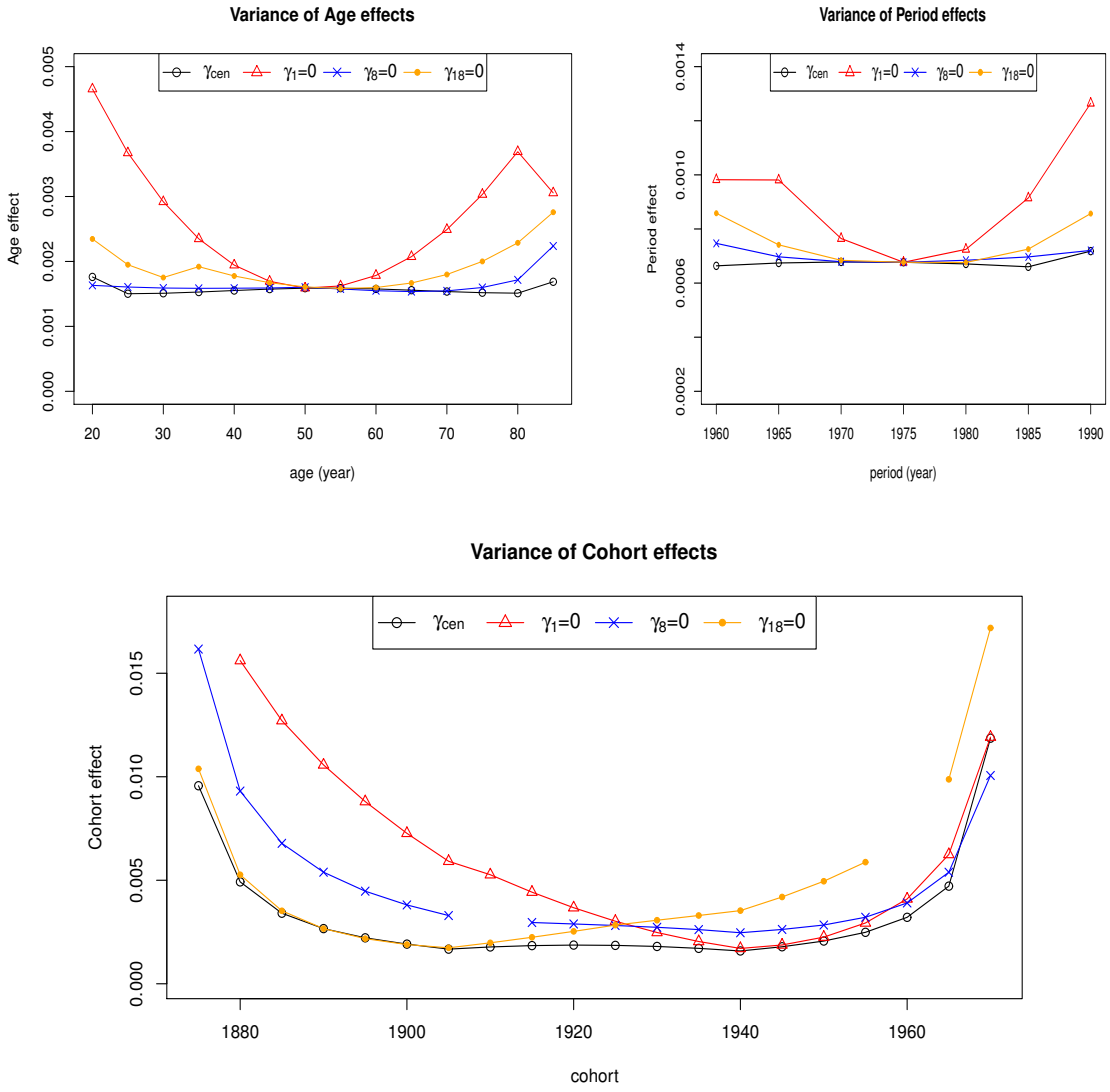

Figure 2.6: **Variance of estimates by side condition for the cervical cancer incidence data in Table 1.1**. The figure display the variance of estimated age, period, and cohort effects by four different side conditions on cohort effects, which are $\sum_{k=1}^{20} \gamma_k = 0$, $\gamma_1 = 0$, $\gamma_8 = 0$, and $\gamma_{18} = 0$. Age represent age groups of 5-years interval from 20-24, 25-29, to 85-89, period represent period groups of 5-years interval from 1960-1964, 1965-1969, to 1990-1994, and cohort represent cohort groups of 9-years interval from 1871-1879, 1976-1984, to 1966-1974.

2.2 Variance Estimation of Parameters in APC Models

2.2.1 Introduction

In the data setting of APC model, we have only one data point in each cell. To study the asymptotic property of the estimate, more samples are needed. Therefore, either we increase the number of rows and keep the number of columns fixed or increase the number of columns and keep the number of rows fixed. However, it is impossible for the first case to happen in reality, because people usually can live up to about 100 years old. Then, the second case with larger number of columns ($p \to \infty$) is considered, which ends up with infinite numbers of period and cohort effects. That's why the profile log-likelihood of the intercept and age effects is considered, treating the period and cohort effects as nuisance parameters of the model. It has been proven that the maximum profile log-likelihood estimator (MaPLE) corresponding to intercept and age effects is consistent when the number of columns goes to infinity [11].

In numerical computing, one requires the observed information matrix, which is the negative Hessian matrix of the second derivative of profile log-likelihood with respect to age effects, to obtain the parameter estimates by the Newton method [20], but instead uses Fisher-scoring algorithm, which is to use the expected value of the observed information matrix-the Fisher information matrix, due to its simplicity [28]. The algorithm is already implemented in the *glm* function in *R*, so the MaPLE can be computed with principle component analysis (PCA) method directly. The Fisher information matrix is needed not only in solving for the maximum profile likelihood estimates, but also in providing with

the asymptotic variance-covariance matrix of age effects.

Even though the PCA method can still provide the variance estimation for period and cohort effects, it may be invalid because both period and cohort effects don't have asymptotic property as the number of parameters for the period and cohort effects diverges to infinity when $p \to \infty$, which motivates me to find a more accurate variance estimation of the estimated period and cohort effects instead.

The Delta method is commonly used by statisticians and other scientists as a technique to approximate the variance of a function of a random variable or the limiting distribution of a function of an estimator, which has asymptotic property when sample size is large enough. JM Ver Hoef (2012) gave a quick review of all various definitions of the Delta method and showed that it is more likely that Dorfman (1938) [4] proposed the method [44]. In this section, I applied the Delta method to derive the explicit form of variance-covariance matrix of estimated period and cohort effects, since these effects are treated as a function of intercept and age effects under the setting of profile log-likelihood. Finally, I compared it with the PCA method through the simulation study.

2.2.2 Consistency of Estimator

The definition of the profile log-likelihood, which was borrowed to prove the asymptotic property of the intrinsic estimator of the intercept and age effects, is introduced as follows

Assume Y_{ij} follows a distribution in the exponential family with log-likelihood [33]

$$l(\zeta; y_{ij}) = \frac{y_{ij}\zeta - \psi(\zeta)}{k(\phi)} + c(y_{ij}, \phi), \tag{2.16}$$

where $i = 1, \ldots, a$; $j = 1, \ldots, p$, ζ is a parameter, ϕ is a dispersion parameter, and functions

$\psi(\cdot)$, $k(\cdot)$, and $c(\cdot)$ are known.

It yields multiple estimators by maximizing the log-likelihood in Equation (2.16), because the design matrix X is one less than full rank. Then the estimator is considered by a penalized log-likelihood as follows

$$l_p^{\lambda}(\mathbf{b}, \boldsymbol{y}) = \sum_{i=1}^{a} \sum_{j=1}^{p} l_{ij}(\zeta(\mathbf{b}); y_{ij}) - \lambda(\mathbf{b}^T \mathbf{B}_0)^2, \tag{2.17}$$

where

$$l_{ij}(\zeta(\mathbf{b}); y_{ij}) = \frac{y_{ij}\zeta(\mathbf{b}) - \psi(\zeta(\mathbf{b}))}{k(\phi)} + c(y_{ij}, \phi),$$

and $\lambda > 0$ is the tuning parameter and needs not to be selected, because $l_{ij}(\zeta(\mathbf{b}); y_{ij})$ can be maximized at $\boldsymbol{b} = \boldsymbol{B}$ with penalty $(\boldsymbol{B}^T \boldsymbol{B}_0)^2 = 0$, the link function $g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{\boldsymbol{ij}}^T \boldsymbol{b}$, where $\boldsymbol{b}^T = (\boldsymbol{\theta}^T, \boldsymbol{\xi}^T)$ with $\boldsymbol{\theta}^T = (\mu, \alpha_1, \ldots, \alpha_{a-1})$, $\boldsymbol{\xi}^T = (\beta_1, \ldots, \beta_{p-1}, \gamma_1, \ldots, \gamma_{a+p-2})$, and $\boldsymbol{x}_{\boldsymbol{ij}}^T = (x_{ij1}, \ldots, x_{ija}, \ldots, x_{ij(2a+2p-3)})$.

To study the consistency of $\boldsymbol{\theta}$, Fu (2016) applied the method using Penalized profile log-likelihood as follows

$$Pl_p^{\lambda}(\boldsymbol{\theta}; \boldsymbol{y}) := l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi}_{\boldsymbol{\theta}}^{\boldsymbol{\lambda}}; \boldsymbol{y}), \tag{2.18}$$

where $\boldsymbol{\xi}_{\boldsymbol{\theta}}^{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\xi}} l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{y})$ for each fixed $\boldsymbol{\theta}$ and $\lambda > 0$. He then considered the maximum profile likelihood estimator (MaPLE), $\tilde{\boldsymbol{\theta}}_p$, estimated from the profile log-likelihood Equation (2.18). It has been proven that $\tilde{\boldsymbol{\theta}}_p$ is consistent as $p \to \infty$. Some important lemmas and theorems in the paper are introduced as follows.

Lemma 2.1. *For any $\lambda > 0$, the limit of the partial derivative of the profile log-likelihood exists,*

$$-\frac{1}{p} \frac{\partial^2 l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi}_{\boldsymbol{\theta}}; \boldsymbol{y})}{\partial \boldsymbol{\theta}^2} \to_p C_1 \; as \; p \to \infty, \tag{2.19}$$

where C_1 is an $a \times a$ Fisher information matrix of the row effect model.

Theorem 2.2. *Under the regularity conditions, the MaPLE, $\tilde{\boldsymbol{\theta}}_p$, of the intrinsic estimator,* **B***, of the generalized linear model in Equation (2.16) with singular design matrix X on the rows, columns, and diagonals of the Lexis diagram of an $a \times p$ table follows*

$$\sqrt{p}\,(\tilde{\boldsymbol{\theta}}_p - \boldsymbol{\theta}^{\infty}) \to_d N(\mathbf{0},\, C_1^{-1}),\ as\ p \to \infty. \tag{2.20}$$

2.2.3 The Variance-Covariance Matrix of Period and Cohort Effects Under Profile Log-likelihood

Based on Equation (2.20), we could approximate the variance-covariance matrix of MaPLE, $\tilde{\boldsymbol{\theta}}_p$, by C_1^{-1}/p as $p \to \infty$. However, we still need to know the variance-covariance matrix of $\boldsymbol{\xi}_{\boldsymbol{\theta}}$.

By the Delta method, for fixed $\boldsymbol{\theta}$, we have

$$\boldsymbol{\xi}_{\boldsymbol{\theta}} \approx \boldsymbol{\xi}_{\tilde{\boldsymbol{\theta}}_{\boldsymbol{p}}} + \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_p). \tag{2.21}$$

By taking the covariance of the Equation (2.21), we have

$$Cov(\boldsymbol{\xi}_{\tilde{\boldsymbol{\theta}}_{\boldsymbol{p}}}) = \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} Cov(\tilde{\boldsymbol{\theta}}_{\boldsymbol{p}}) \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}^{T}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{2.22}$$

The formula of $\frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}$ and variance-covariance matrix, Cov($\tilde{\boldsymbol{\theta}}_{\boldsymbol{p}}$), of MaPLE $\tilde{\boldsymbol{\theta}}_p$, will be derived in the following sections.

2.2.3.1 The Formula for Partial Derivative of $\boldsymbol{\xi_\theta}$ w.r.t $\boldsymbol{\theta}$

For fixed $\boldsymbol{\theta}$, we have

$$\frac{\partial l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{y})}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\xi}=\boldsymbol{\xi_\theta}} = 0, \tag{2.23}$$

because $\boldsymbol{\xi_\theta}$ maximizes the log-likelihood $l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi})$.

Then differentiating it w.r.t $\boldsymbol{\theta}$, we have

$$\frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\theta}} + \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi}^2} \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} = 0. \tag{2.24}$$

Therefore,

$$\frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} = \Big[- \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi}^2} \Big]^{-1} \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\theta}}. \tag{2.25}$$

From the definition of $l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})$ in Equation (2.17), we have

$$- \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi}^2} = \underbrace{- \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi}^2}}_{:=H} + 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0, \tag{2.26}$$

$$\frac{\partial^2 l_p^\lambda(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\theta}} = \underbrace{\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\theta}}}_{:=G} - 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0, \tag{2.27}$$

Therefore, by plugging Equations (2.26) and (2.27) into the Equation (2.24), we have

$$G - 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0 + (H - 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0) \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} = 0. \tag{2.28}$$

Next, we need to figure out the matrix H and G. Let $H := \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \boldsymbol{\xi}^2}$, then

$$H = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_1^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_1 \partial \xi_{a+2p-3}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_{a+2p-3} \partial \xi_1} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_{a+2p-3}^2} \end{bmatrix}_{(a+2p-3)\times(a+2p-3)}.$$

Then we define

$$H_{ks} := \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_k \partial \xi_s} = \sum_{i=1}^{a}\sum_{j=1}^{p} \frac{\partial^2 l_{ij}(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_k \partial \xi_s} \text{ for } k, s = 1, \ldots, (a+2p-3), \tag{2.29}$$

where

$$\begin{aligned} \frac{\partial l_{ij}(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_k} &= \frac{\partial l_{ij}}{\partial \mu_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot \frac{\partial \eta_{ij}}{\partial \xi_k} \\ &= \frac{y_{ij}-\mu_{ij}}{V_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ij(a+k)}. \end{aligned} \tag{2.30}$$

Since

$$\begin{aligned} \frac{\partial l_{ij}}{\partial \mu_{ij}} &= \frac{\partial l_{ij}}{\partial \zeta} \Big/ \frac{\partial \mu_{ij}}{\partial \zeta} = \frac{y_{ij}-\psi'(\zeta)}{k(\phi)} \Big/ \frac{\partial \mu_{ij}}{\partial \zeta} \\ &= \frac{y_{ij}-\mu_{ij}}{k(\phi)} / \psi''(\zeta), \\ \frac{\partial \eta_{ij}}{\partial \xi_k} &= x_{ij(a+k)}, \end{aligned}$$

where $\mu_{ij} = \psi'(\zeta), V_{ij} := k(\phi)\psi''(\zeta)$. Therefore, by plugging Equation (2.30) into Equation (2.29), we have

$$H_{ks} = \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi}_{\boldsymbol{\theta}},\boldsymbol{y})}{\partial \xi_k \partial \xi_s} = \sum_{i=1}^{a}\sum_{j=1}^{p} \frac{\partial}{\partial \xi_s}\Big[\frac{y_{ij}-\mu_{ij}}{V_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ij(a+k)}\Big]$$

$$
\begin{aligned}
&= \sum_{i=1}^{a}\sum_{j=1}^{p}\Big[\frac{\partial}{\partial\xi_s}(y_{ij}-\mu_{ij})\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}}x_{ij(a+k)} + (y_{ij}-\mu_{ij})\frac{\partial}{\partial\xi_s}(\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}}x_{ij(a+k)})\Big] \\
&= \sum_{i=1}^{a}\sum_{j=1}^{p}\Big[\underbrace{-\frac{1}{V_{ij}}(\frac{d\mu_{ij}}{d\eta_{ij}})^2}_{:=W_{ij}}x_{ij(a+k)}x_{ij(a+s)} + x_{ij(a+k)}(y_{ij}-\mu_{ij})\frac{\partial}{\partial\xi_s}(\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}})\Big].
\end{aligned}
\tag{2.31}
$$

Since

$$
\frac{\partial}{\partial\xi_s}(y_{ij}-\mu_{ij}) = -\frac{\partial\mu_{ij}}{\partial\xi_s} = -\frac{d\mu_{ij}}{d\eta_{ij}}\cdot\frac{\partial\eta_{ij}}{\partial\xi_s} = -\frac{d\mu_{ij}}{d\eta_{ij}}\cdot x_{ij(a+s)}.
$$

In general, we use the expected information matrix $A_1 := -E(H)$ instead, the second term in Equation (2.31) would be cancelled, then

$$
\begin{aligned}
A_1 &:= \begin{bmatrix} \sum_{i,j} W_{ij}x_{ij(a+1)}^2 & \cdots & \sum_{i,j} W_{ij}x_{ij(a+1)}x_{ij(2a+2p-3)} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} W_{ij}x_{ij(2a+2p-3)}x_{ij(a+1)} & \cdots & \sum_{i,j} W_{ij}x_{ij(2a+2p-3)}^2 \end{bmatrix} \\
&= X_2^T W X_2,
\end{aligned}
\tag{2.32}
$$

where

$$
X_2 := \begin{bmatrix} x_{11(a+1)} & x_{11(a+2)} & \cdots & x_{11(2a+2p-3)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p(a+1)} & x_{1p(a+2)} & \cdots & x_{1p(2a+2p-3)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{a1(a+1)} & x_{a1(a+2)} & \cdots & x_{a1(2a+2p-3)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ap(a+1)} & x_{ap(a+2)} & \cdots & x_{ap(2a+2p-3)} \end{bmatrix}.
$$

X_2 is actually a part of design matrix X only corresponding to the period effects $\boldsymbol{\beta}$ and

cohort effects $\boldsymbol{\gamma}$, and $W := diag(W_{11}, \ldots, W_{1p}, W_{21}, \ldots, W_{2p}, \ldots, W_{a1}, \ldots, W_{ap})$ is called the weight matrix.

Let $G := \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\theta}}$, then it has a matrix form as follows

$$G := \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_1 \partial \theta_1} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_1 \partial \theta_a} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_{a+2p-3} \partial \theta_1} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_{a+2p-3} \partial \theta_a} \end{bmatrix}_{(a+2p-3)\times a}.$$

For each element in the matrix G, by the Equation (2.30)

$$\begin{aligned} G_{ks} &:= \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_k \partial \theta_s} = \sum_{i=1}^{a} \sum_{j=1}^{p} \frac{\partial}{\partial \theta_s} \Big[\frac{\partial l_{ij}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \xi_k} \Big] \qquad (2.33) \\ &= \sum_{i=1}^{a} \sum_{j=1}^{p} \frac{\partial}{\partial \theta_s} \Big[\frac{y_{ij} - \mu_{ij}}{V_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ij(a+k)} \Big] \\ &= \sum_{i=1}^{a} \sum_{j=1}^{p} \Big[\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}} x_{ij(a+k)} \frac{\partial}{\partial \theta_s} (y_{ij} - \mu_{ij}) + (y_{ij} - \mu_{ij}) \frac{\partial}{\partial \theta_s} (\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}} x_{ij(a+k)}) \Big] \\ &= \sum_{i=1}^{a} \sum_{j=1}^{p} \Big[- \underbrace{\frac{1}{V_{ij}} (\frac{d\mu_{ij}}{d\eta_{ij}})^2}_{:=W_{ij}} x_{ij(a+k)} x_{ijs} + x_{ij(a+k)} (y_{ij} - \mu_{ij}) \frac{\partial}{\partial \theta_s} (\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}}) \Big]. \end{aligned}$$

(2.34)

Since

$$\frac{\partial}{\partial \theta_s}(y_{ij} - \mu_{ij}) = -\frac{\partial \mu_{ij}}{\partial \theta_s} = -\frac{d\mu_{ij}}{d\eta_{ij}} \cdot \frac{\partial \eta_{ij}(\boldsymbol{b})}{\partial \theta_s} = -\frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ijs}.$$

In general, we use expected information matrix $A_2 := -E(G)$ instead, the second term

in Equation (2.33) would be cancelled as zero, we have

$$\begin{aligned} A_2 &:= \begin{bmatrix} \sum_{i,j} W_{ij} x_{ij(a+1)} x_{ij1} & \cdots & \sum_{i,j} W_{ij} x_{ij(a+1)} x_{ija} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} W_{ij} x_{ij(2a+2p-3)} x_{ij1} & \cdots & \sum_{i,j} W_{ij} x_{ij(2a+2p-3)} x_{ija} \end{bmatrix} \\ &= X_2^T W X_1, \end{aligned} \tag{2.35}$$

where

$$X_1 := \begin{bmatrix} x_{111} & x_{112} & \cdots & x_{11a} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p1} & x_{1p2} & \cdots & x_{1pa} \\ \vdots & \vdots & \ddots & \vdots \\ x_{a11} & x_{a12} & \cdots & x_{a1a} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ap1} & x_{ap2} & \cdots & x_{apa} \end{bmatrix},$$

which is one part of the design matrix corresponding to μ and age effects $\boldsymbol{\alpha}$, and the design matrix has the form as $X = (X_1\ X_2)$.

After taking the expectation of Equation (2.28) and plugging A_1 and A_2, we get

$$-A_2 - 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0 - (A_1 + 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0) \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = 0,$$

i.e.

$$-X_2^T W X_1 - 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0 - (X_2^T W X_2 + 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0) \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = 0.$$

So we have

$$\frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = -(X_2^T W X_2 + 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0)^{-1} (X_2^T W X_1 + 2\lambda \boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0). \tag{2.36}$$

If $\lambda \to 0$ in Equation (2.36), then

$$\frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} = -(X_2^T W X_2)^{-1}(X_2^T W X_1). \tag{2.37}$$

2.2.4 The Fisher Information Matrix of the Row Effects

Next, we try to figure out C_1. By Lemma 2.1, we have

$$-\frac{1}{p}\frac{\partial^2 Pl_p^\lambda(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}^2} \to_p C_1 \text{ as } p \to \infty,$$

so

$$-\frac{\partial^2 Pl_p^\lambda(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}^2} \approx pC_1 \text{ as } p \to \infty. \tag{2.38}$$

We differentiate the penalized profile log-likelihood function $Pl_p^\lambda(\boldsymbol{\theta};\boldsymbol{y})$ in Equation (2.18) w.r.t $\boldsymbol{\theta}$ by the chain rule, we have

$$\frac{\partial Pl_p^\lambda(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}} = \frac{\partial l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}} + \frac{\partial l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial \boldsymbol{\xi_\theta}} \cdot \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} \tag{2.39}$$

$$= \frac{\partial l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}}. \tag{2.40}$$

The first term in the above Equation (2.39) is the partial derivative of function $l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})$ only w.r.t $\boldsymbol{\theta}$ assuming that $\boldsymbol{\xi_\theta}$ is a constant, the second term would be zero, because $\boldsymbol{\xi_\theta}$ maximizes the log-likelihood $l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi})$, i.e. $\frac{\partial l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{y})}{\partial \boldsymbol{\xi}}|_{\boldsymbol{\xi}=\boldsymbol{\xi_\theta}} = 0$.

Then we differentiate the Equation (2.40) again w.r.t $\boldsymbol{\theta}$, we have

$$\frac{\partial^2 Pl_p^\lambda(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}^2} = \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}^2} + \frac{\partial^2 l_p^\lambda(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\xi_\theta}} \cdot \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}}. \tag{2.41}$$

The first term in the above Equation (2.41) is the second partial derivative of function $l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})$ only w.r.t $\boldsymbol{\theta}$ assuming that $\boldsymbol{\xi_\theta}$ is a constant.

Then

$$\frac{\partial^2 l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\theta}^2} = \underbrace{\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\theta}^2}}_{:=C} - 2\lambda \boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0, \tag{2.42}$$

$$\frac{\partial^2 l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\xi_\theta}} = \underbrace{\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\xi_\theta}}}_{:=D} - 2\lambda \boldsymbol{\theta}_0 \otimes \boldsymbol{\xi}_0, \tag{2.43}$$

since $l_p^{\lambda}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y}) = l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y}) - \lambda(\mathbf{b}^T \mathbf{B}_0)^2$.

Let $C := \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \boldsymbol{\theta}^2}$, then its element $C_{ks} := \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \theta_k \partial \theta_s} = \sum_{i=1}^{a} \sum_{j=1}^{p} \frac{\partial^2 l_{ij}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \theta_k \partial \theta_s}$ for $k, s = 1, \ldots, a$. Then

$$\begin{aligned} \frac{\partial l_{ij}(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \theta_k} &= \frac{\partial l_{ij}}{\partial \mu_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot \frac{\partial \eta_{ij}(\boldsymbol{b})}{\partial \theta_k} \\ &= \frac{y_{ij} - \mu_{ij}}{V_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ijk}, \end{aligned} \tag{2.44}$$

where $\mu_{ij} = \psi'(\zeta)$, $V_{ij} := k(\phi)\psi''(\zeta)$.

Similarly, by differentiating the Equation (2.44) w.r.t θ_s, assuming that $\boldsymbol{\xi_\theta}$ is a constant, we have

$$\begin{aligned} C_{ks} &:= \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\xi_\theta}, \boldsymbol{y})}{\partial \theta_k \partial \theta_s} = \sum_{i=1}^{a} \sum_{j=1}^{p} \frac{\partial}{\partial \theta_s} \Big[\frac{y_{ij} - \mu_{ij}}{V_{ij}} \cdot \frac{d\mu_{ij}}{d\eta_{ij}} \cdot x_{ijk} \Big] \\ &= \sum_{i=1}^{a} \sum_{j=1}^{p} \Big[\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}} x_{ijk} \frac{\partial}{\partial \theta_s} (y_{ij} - \mu_{ij}) + (y_{ij} - \mu_{ij}) \frac{\partial}{\partial \theta_s} (\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}} x_{ijk}) \Big] \\ &= \sum_{i=1}^{a} \sum_{j=1}^{p} \Big[- \underbrace{\frac{1}{V_{ij}} (\frac{d\mu_{ij}}{d\eta_{ij}})^2}_{:=W_{ij}} x_{ijk} \cdot x_{ijs} + (y_{ij} - \mu_{ij}) x_{ijk} \frac{\partial}{\partial \theta_s} (\frac{1}{V_{ij}} \frac{d\mu_{ij}}{d\eta_{ij}}) \Big]. \end{aligned} \tag{2.45}$$

Let $D := \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\xi_\theta}}$, then its element $D_{ks} := \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial\theta_k\partial\xi_s} = \sum_{i=1}^{a}\sum_{j=1}^{p}\frac{\partial^2 l_{ij}(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial\theta_k\partial\xi_s}$ for $k = 1,\ldots,a$ and $s = 1,\ldots,(a+2p-3)$.

Similarly, by differentiating the Equation (2.44) only w.r.t ξ_s, we have

$$\begin{aligned} D_{ks} &:= \frac{\partial^2 l(\boldsymbol{\theta},\boldsymbol{\xi_\theta},\boldsymbol{y})}{\partial\theta_k\partial\xi_s} = \sum_{i=1}^{a}\sum_{j=1}^{p}\frac{\partial}{\partial\xi_s}\Big[\frac{y_{ij}-\mu_{ij}}{V_{ij}}\cdot\frac{d\mu_{ij}}{d\eta_{ij}}\cdot x_{ijk}\Big] \\ &= \sum_{i=1}^{a}\sum_{j=1}^{p}\Big[\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}}x_{ijk}\frac{\partial}{\partial\xi_s}(y_{ij}-\mu_{ij}) + (y_{ij}-\mu_{ij})\frac{\partial}{\partial\xi_s}(\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}}x_{ijk})\Big] \\ &= \sum_{i=1}^{a}\sum_{j=1}^{p}\Big[\underbrace{-\frac{1}{V_{ij}}(\frac{d\mu_{ij}}{d\eta_{ij}})^2}_{:=W_{ij}}x_{ijk}\cdot x_{ij(a+s)} + (y_{ij}-\mu_{ij})x_{ijk}\frac{\partial}{\partial\xi_s}(\frac{1}{V_{ij}}\frac{d\mu_{ij}}{d\eta_{ij}})\Big]. \quad (2.46) \end{aligned}$$

Next, we take the expected value of the matrix C and D, both second terms in Equations (2.45) and (2.46) will be zero, we get

$$E(C) = -\begin{bmatrix} \sum_{i,j}W_{ij}x_{ij1}^2 & \cdots & \sum_{i,j}W_{ij}x_{ij1}x_{ija} \\ \vdots & \ddots & \vdots \\ \sum_{i,j}W_{ij}x_{ija}x_{ij1} & \cdots & \sum_{i,j}W_{ij}x_{ija}^2 \end{bmatrix} = -X_1^T W X_1, \quad (2.47)$$

$$\begin{aligned} E(D) &= -\begin{bmatrix} \sum_{i,j}W_{ij}x_{ij1}x_{ij(a+1)} & \cdots & \sum_{i,j}W_{ij}x_{ij(1)}x_{ij(2a+2p-3)} \\ \vdots & \ddots & \vdots \\ \sum_{i,j}W_{ij}x_{ija}x_{ij(a+1)} & \cdots & \sum_{i,j}W_{ij}x_{ija}x_{ij(2a+2p-3)} \end{bmatrix} \\ &= -X_1^T W X_2. \quad (2.48) \end{aligned}$$

Therefore, after taking the expected value of Equation (2.41) and plugging Equations

(2.42) and (2.43), we have

$$\begin{aligned} pC_1 &\approx -\frac{\partial^2 Pl_p^{\lambda}(\boldsymbol{\theta};\boldsymbol{y})}{\partial \boldsymbol{\theta}^2} \\ &= -(E(C) - 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0) - (E(D) - 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\xi}_0) \cdot \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \\ &= X_1^T W X_1 + X_1^T W X_2 \cdot \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\xi}_0 \cdot \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0. \end{aligned} \tag{2.49}$$

If $\lambda \to 0$ in Equation (2.49), and also plugging Equation (2.37) into Equation (2.49), we have

$$pC_1 \approx X_1^T W X_1 - X_1^T W X_2 \cdot (X_2^T W X_2)^{-1} \cdot X_2^T W X_1. \tag{2.50}$$

2.2.5 Summary

$$\begin{aligned} Cov(\boldsymbol{\xi}_{\tilde{\boldsymbol{\theta}}_{\boldsymbol{p}}}) &= \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} (C_1^{-1}/p) \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}^T, \\ pC_1 &\approx X_1^T W X_1 + X_1^T W X_2 \cdot \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\xi}_0 \cdot \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + 2\lambda\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0, \\ \frac{\partial \boldsymbol{\xi}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} &= -(X_2^T W X_2 + 2\lambda\boldsymbol{\xi}_0 \otimes \boldsymbol{\xi}_0)^{-1} (X_2^T W X_1 + 2\lambda\boldsymbol{\xi}_0 \otimes \boldsymbol{\theta}_0). \end{aligned} \tag{2.51}$$

2.2.6 Simulation

2.2.6.1 Specification of Age, Period, and Cohort Effects

Two sets of row effects were specified with $a = 9$ groups in each. The column effects were defined as $\beta_j = sin(j)$ for $j = 1, \ldots, p$, and the diagonal effects were defined by combining sine and cosine functions $\gamma_k = cos(k) + sin(k \cdot 10^9)$. All of them are shown in Figure 2.7. The intercept was specified to be $\mu = 1$.

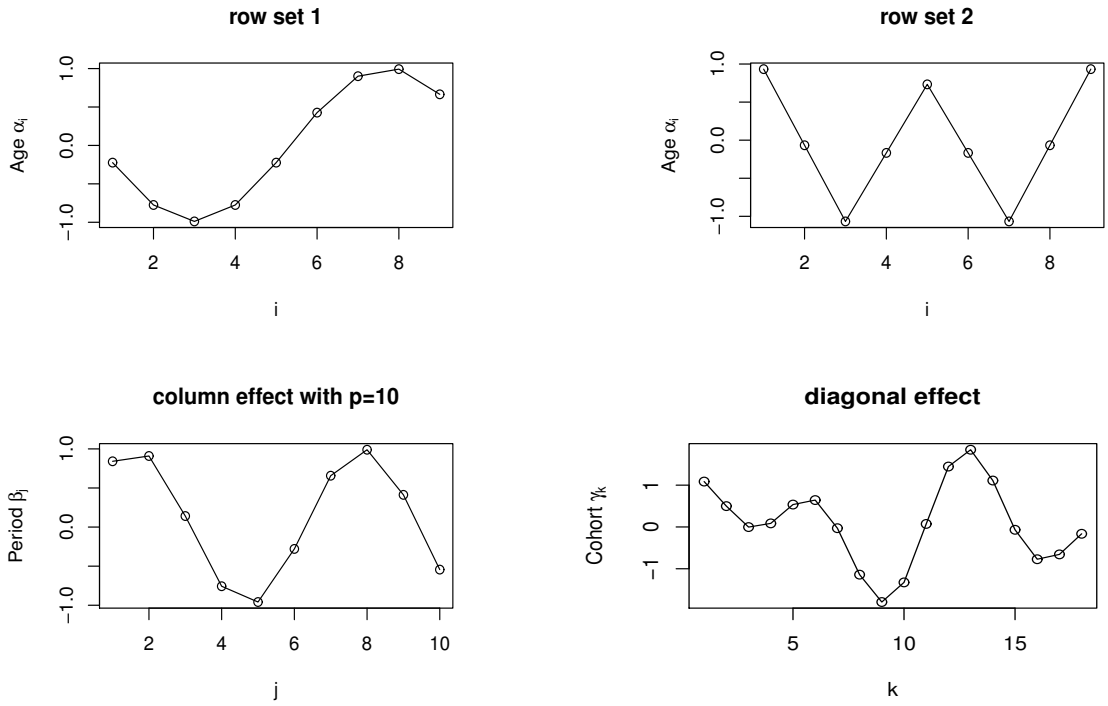

Figure 2.7: **Parameter specification for age, period, and cohort effects**. Top panel: Two different sets of row effects α_i are given for $i = 1, \ldots, 9$; Bottom panel: Column effects generated by $\beta_j = sin(j)$ for $j = 1, \ldots, 10$, and diagonal effects generated by $\gamma_k = cos(k) + sin(k \cdot 10^9)$ for $k = 1, \ldots, 18$.

2.2.6.2 Data Generation

For each set of specified parameters $(\mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_{a+p-1})$ with given $a = 9$ and $p = 10$ or 50, responses were generated by a linear model for the normal variables

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{a-i+j} + \epsilon_{ij},$$

or with a log-linear model for Poisson variables

$$y_{ij} = Pois[(\mu + \alpha_i + \beta_j + \gamma_{a-i+j}) * N_{ij}].$$

2.2.6.3 Simulation with data following Poisson distribution

For the Poisson distribution, we have $\theta = log\lambda$, $a(\phi) = 1$. The likelihood is

$$l(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^{ap} \{y_i log\lambda_i - \lambda_i + log(y_i!)\},$$

where $u_i = \lambda_i = e^{\theta_i}$, $b(\theta_i) = \mu_i = e^{\theta_i}$, and $V_i = e^{\theta_i}$, the canonical link function is $g(\mu_i) = logu_i = \theta_i = \eta_i$, so $W_i = V_i^{-1}(\frac{d\mu_i}{d\eta_i})^2 = N_i * e^{\eta_i}$, where N_i is the population for ith observation.

If the tuning parameter $\lambda \to 0$ in Equation (2.51), we have

$$\begin{aligned} \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} &= -(X_2^T W X_2)^{-1} \cdot X_2^T W X_1, \\ Cov(\boldsymbol{\xi}_{\tilde{\boldsymbol{\theta}}_p}) &= \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}} (C_1^{-1}/p) \frac{\partial \boldsymbol{\xi_\theta}}{\partial \boldsymbol{\theta}}^T. \end{aligned} \tag{2.52}$$

Here, assuming each cell to have the same population $N = 4000$ in the $a \times p$ table. I compared the standard error of estimates between the PCA and Delta methods by simulation on two data sets in Figure 2.7, which were shown in Figures 2.8 and 2.9. It is shown that the standard errors of estimates by the Delta method was mostly smaller than that by the PCA method, except for a few large values for the young and old cohort effects.

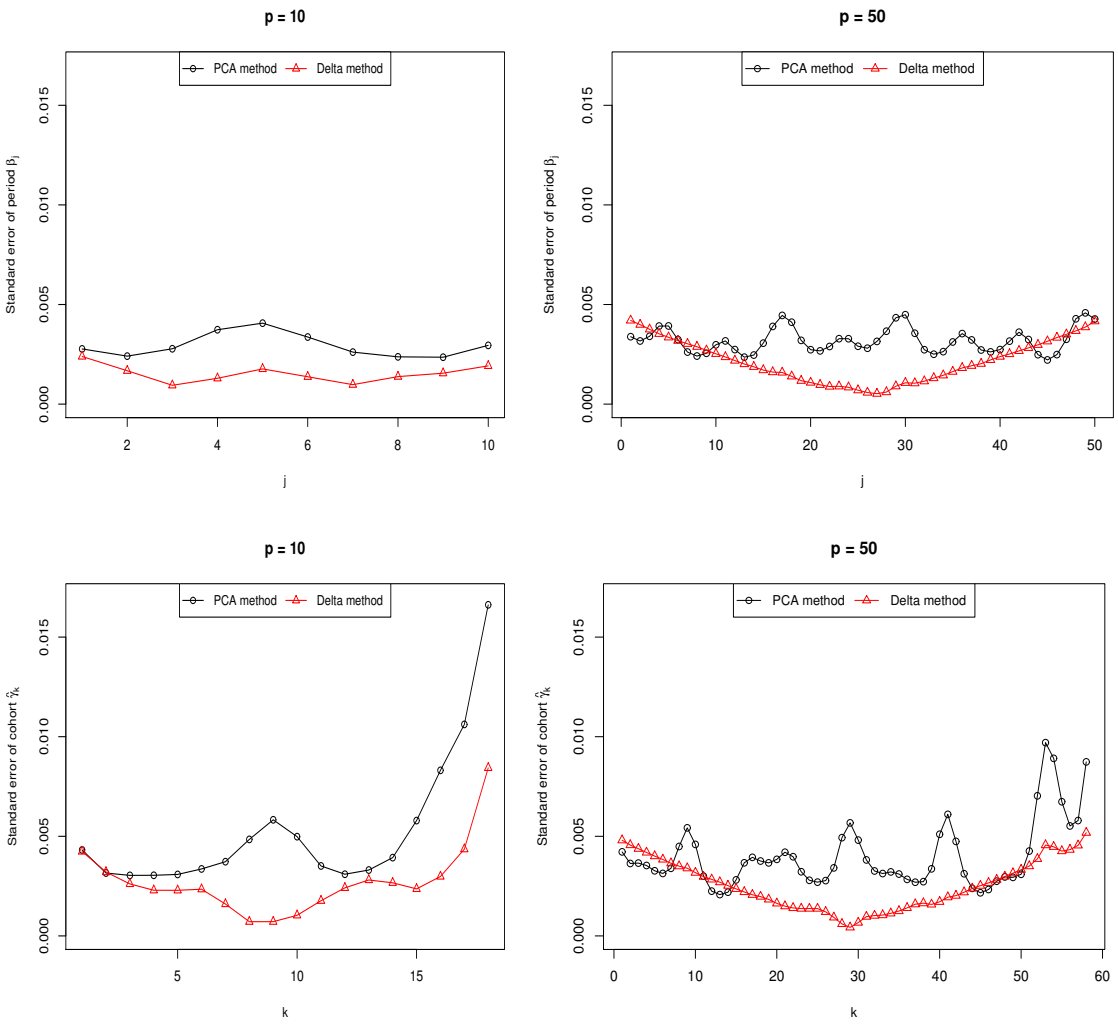

Figure 2.8: **Standard deviation of period and cohort effects by PCA and Delta methods with Poisson variables generated from data set 1**. The two subplots on the left panel are the standard deviation of the estimated period and cohort effects, $\text{sd}(\hat{\beta}_j)$ and $\text{sd}(\hat{\gamma}_k)$, for $j = 1, \ldots, p$ and $k = 1, \ldots, (a + p - 1)$ with $a = 9$ and $p = 10$, respectively; The two subplots on the right panel are the standard deviation of the estimated period and cohort effects, $\text{sd}(\hat{\beta}_j)$ and $\text{sd}(\hat{\gamma}_k)$, for $j = 1, \ldots, p$ and $k = 1, \ldots, (a + p - 1)$ with $a = 9$ and $p = 50$, respectively.

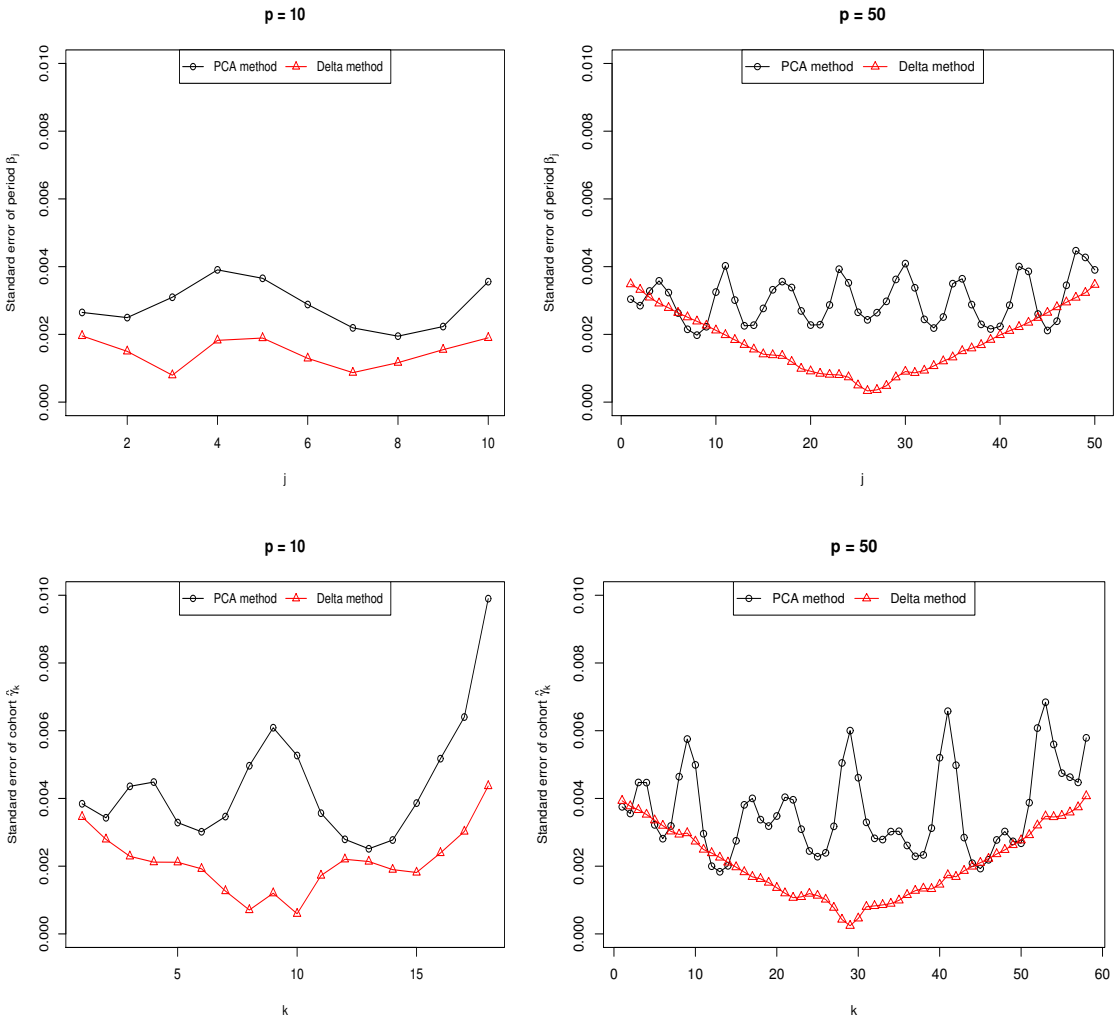

Figure 2.9: **Standard deviation of period and cohort effects by PCA and Delta methods with Poisson variables generated from data set 2**. The two subplots on the left panel are the standard deviation of the estimated period and cohort effects, $\text{sd}(\hat{\beta}_j)$ and $\text{sd}(\hat{\gamma}_k)$, for $j = 1, \ldots, p$ and $k = 1, \ldots, (a + p - 1)$ with $a = 9$ and $p = 10$, respectively; The two subplots on the right panel are the standard deviation of the estimated period and cohort effects, $\text{sd}(\hat{\beta}_j)$ and $\text{sd}(\hat{\gamma}_k)$, for $j = 1, \ldots, p$ and $k = 1, \ldots, (a + p - 1)$ with $a = 9$ and $p = 50$, respectively.

CHAPTER 3

Hypothesis Testing on Trend of Age Effects among Different Population

3.1 Introduction

Screening has been an important prevention against cancer, such as, breast, prostate, lung and ovarian, aiming to detect disease in early diagnosis. As a consequence, screening will make an impact on the incidence or mortality rate data, for example, which might lead to higher incidence rate at an earlier age. Within the population of the same cancer, the screening effect is possible to be detected by comparing the age trend between data of periods before and after the cancer screening. Also the screening participation can be

different across groups, such as gender, race, and socioeconomic status. In the US, there is evidence for racial disparities in screening uptake [45]. Further, in terms of interpreting the trend of age effects, researchers may have interest in whether there's difference of age or period trends between male and female, or across multiple racial groups like white, black, and other. Under both cases, a statistical test is needed to detect the difference for different situations, which will be covered in this section later.

For APC linear models, the error terms for each population are assumed to follow a normal distribution with a mean 0, and a variance component σ^2. Usually the derivation of a test under the assumption of homoscedasticity of variance is much easier (e.g., the F-test in ANOVA models). However, for APC models, it is very unlikely to have the same variance component across different populations.

To test the difference of vectors of mean, the simple case is only comparing between two populations assuming equal variance-covariance matrix. The traditional best-known two-sample test is the Hotelling T^2-test, which assumes common variance-covariance matrix for two groups [17]. Under the case of unequal variance-covariance matrix, which is also known as Behrens-Fisher problem, several approximate solutions based on T^2 have been proposed, such as James (1954) [19], Yao (1965) [51], and Johansen (1980) [21], which are similar to the Welch's (1947) approximate degrees of freedom solution to the univariate case [46]. Van der Merve (1986) proposed an approach to the multivariate Behrens-Fisher problem by approximating the distribution of the parameter matrix, which had the highest power among the methods whose Type I error were not inflated based on the comparison result [2].

In APC models, the variance-covariance matrix of age effects by the intrinsic estimator method is a special case of the general variance-covariance matrix form, where it has the

form of $\sigma^2 M$ with unknown variance component, σ^2. Under the case of equal variance component, the statistic of test can be easily achieved. However, under the case of testing the equality of age trends across more than two groups, no test procedure can be applied directly, which means a new test needs to be derived.

For the univariate case, Welch (1951) proposed Welch-F test to test the equality of group means if more than two groups are compared without assuming the homoscedasticity of variance, especially for the case when sample size is not large [47]. In this section, we borrowed the Taylor expansion strategy to extend Welch's scalar case to our vector case, when testing on the equality of age trends across multiple populations, like different race or gender, assuming unequal variance for different populations.

3.2 Testing on Equality of Age Trend Between Two Populations

Let $\boldsymbol{\alpha}_p := (\alpha_1, \ldots, \alpha_m)^T$ be an $m \times 1$ vector of age effects which excludes the last age effect α_a because of parameter centralization, where $m := a - 1$.

Given two data sets, which are shown in a $a_1 \times p_1$ and $a_2 \times p_2$ table, respectively, assuming the same dimension, i.e. $a_1 = a_2 = a$, $p_1 = p_2 = p$, then we have

$$\tilde{\boldsymbol{\alpha}}_p^1 \sim N(\boldsymbol{\alpha}^1, \Sigma_1) \text{ and } \tilde{\boldsymbol{\alpha}}_p^2 \sim N(\boldsymbol{\alpha}^2, \Sigma_2),$$

where $\tilde{\boldsymbol{\alpha}}_p^1$ and $\tilde{\boldsymbol{\alpha}}_p^2$ are m-dim vectors of age effects for data sets 1 and 2, respectively.

The null hypothesis is $H_0 : \boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^2$, and the alternative hypothesis is $H_1 : \boldsymbol{\alpha}^1 \neq \boldsymbol{\alpha}^2$.

Under the null hypothesis,

$$X^2 := (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)^T (\Sigma_1 + \Sigma_2)^{-1} (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2) \quad \sim \chi^2(a-1). \tag{3.1}$$

The variance-covariance matrix of estimated age effects by the intrinsic estimator method has the form as follows

$$\Sigma := \sigma^2 M, \tag{3.2}$$

where the matrix M is fixed, which can be calculated from the design matrix X.

The difference of variance-covariance matrix of estimated age effects between two populations comes only from the variance component, σ^2, assuming the same dimension of data set. Therefore,

$$X^2 := \frac{(\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)}{\sigma_1^2 + \sigma_2^2} \quad \sim \chi^2(a-1). \tag{3.3}$$

σ_1^2 and σ_2^2 are unknown parameters to be estimated. And we know that

$$\frac{nS_1^2}{\sigma_1^2} \quad \sim \chi^2(n) \quad \text{and} \quad \frac{nS_2^2}{\sigma_2^2} \quad \sim \chi^2(n),$$

where $S_1^2 := \frac{1}{n} r_1^2$, and $n = ap - a - (p-1) - (a+p-2) + 1 = (a-2)(p-2)$ is the degree of freedom. The sum of residuals of APC model is given by

$$r^2 = \sum_{i=1}^{a} \sum_{j=1}^{p} (y_{ij} - \mu - \alpha_i - \beta_j - \gamma_{a-i+j})^2.$$

3.2.1 Homoscedasticity

Under this case, we have

$$X^2 := \frac{(\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)}{2\sigma^2} \sim \chi^2(a-1), \tag{3.4}$$

$$Y^2 := \frac{n(S_1^2 + S_2^2)}{\sigma^2} \sim \chi^2(2n). \tag{3.5}$$

Therefore,

$$\begin{aligned} F &:= \frac{X^2/(a-1)}{Y^2/2n} \\ &= \frac{(\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)/(a-1)}{(S_1^2 + S_2^2)} \sim F(a-1, 2n). \end{aligned} \tag{3.6}$$

3.2.2 Heteroscedasticity

Under this case, by the Welch's student T method, an F statistic is given as follows:

$$F := \frac{(\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^1 - \tilde{\boldsymbol{\alpha}}_p^2)/(a-1)}{(S_1^2 + S_2^2)} \sim F(a-1, v). \tag{3.7}$$

The degrees of freedom v associated with the test is approximated using the Welch-Satterthwaite equation:

$$v := \frac{(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2})^2}{\frac{S_1^4}{N_1^2 n_1} + \frac{S_1^4}{N_2^2 n_2}} = \frac{(S_1^2 + S_2^2)^2}{(S_1^4 + S_2^4)/n}, \tag{3.8}$$

where $N_1 = N_2 = ap$ are sample size for two data sets, and $n_1 = n_2 = n$ are degrees of freedom associated with the variance estimates S_1^2 and S_2^2, respectively.

3.3 Testing on Equality of Age Trend Among K Populations

Here we have K data sets, which are shown in an $a_1 \times p_1$, $a_2 \times p_2, \ldots, a_K \times p_K$ table, respectively. We assume that these K ($K > 2$) data sets have the same number of age effects and period effects, i.e. $a_1 = a_2 = \cdots = a_K = a$, $p_1 = p_2 = \cdots = p_K = p$. As $p \to \infty$, we know that

$$\tilde{\boldsymbol{\alpha}}_p^1 \sim N(\boldsymbol{\alpha}^1, \Sigma_1) \ , \ \tilde{\boldsymbol{\alpha}}_p^2 \sim N(\boldsymbol{\alpha}^2, \Sigma_2), \ \cdots, \ \tilde{\boldsymbol{\alpha}}_p^K \sim N(\boldsymbol{\alpha}^K, \Sigma_K),$$

where $\tilde{\boldsymbol{\alpha}}_p^1$, $\tilde{\boldsymbol{\alpha}}_p^2, \ldots, \tilde{\boldsymbol{\alpha}}_p^K$ are m-dim vectors of age effects for data set $1, 2, \ldots, K$, respectively, where $m := a - 1$.

The null hypothesis is $H_0 : \boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^2 = \cdots = \boldsymbol{\alpha}^K$, and the alternative hypothesis is $H_1 : \boldsymbol{\alpha}^i \neq \boldsymbol{\alpha}^j$ for $i \neq j$.

From Equation (3.2), we also have

$$\Sigma_k := \sigma_k^2 M, \quad \text{for } \ k = 1, \ldots, K. \tag{3.9}$$

They have different variance component, σ_k^2, but the same matrix M across K populations.

For each $\tilde{\boldsymbol{\alpha}}_p^k$, it has variance-covariance matrix of Σ_k, all components in the vector $\tilde{\boldsymbol{\alpha}}_p^k$ are not independent, so we need to do the matrix transformation to make them independent with each other. The orthogonal decomposition of a real symmetric matrix M, is given by

$$M = Q \Lambda Q^T,$$

where Q is an orthogonal matrix whose ith column is the eigenvector of M and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, i.e. $\Lambda_{ii} = \lambda_i$ for $i = 1, \ldots, m$.

Let $\tilde{\boldsymbol{\nu}}_p^k := Q^T \tilde{\boldsymbol{\alpha}}_p^k$, then $\tilde{\boldsymbol{\nu}}_p^k \sim N(\boldsymbol{\nu}^k, \sigma_k^2 \Lambda)$ for $k = 1, \ldots, K$, where $\boldsymbol{\nu}^k := Q^T \boldsymbol{\alpha}^k$.

Multiplying the null hypothesis H_0 by Q^T on the left, it follows that

$$H_0 : \boldsymbol{\nu}^1 = \boldsymbol{\nu}^2 = \cdots = \boldsymbol{\nu}^K,$$

so the alternative hypothesis is changed to $H_1 : \boldsymbol{\nu}^i \neq \boldsymbol{\nu}^j$ for at least two indexes $i \neq j$.

Here we are testing the equivalence among vectors $\boldsymbol{\nu}^1, \boldsymbol{\nu}^2, \ldots$, and $\boldsymbol{\nu}^K$, it means that

$$\underbrace{\begin{pmatrix} \nu^1(1) \\ \nu^1(2) \\ \vdots \\ \nu^1(m) \end{pmatrix}}_{:\boldsymbol{\nu}^1}, \underbrace{\begin{pmatrix} \nu^2(1) \\ \nu^2(2) \\ \vdots \\ \nu^2(m) \end{pmatrix}}_{:\boldsymbol{\nu}^2}, \cdots, \underbrace{\begin{pmatrix} \nu^K(1) \\ \nu^K(2) \\ \vdots \\ \nu^K(m) \end{pmatrix}}_{:\boldsymbol{\nu}^K},$$

so we need to test the equivalence among corresponding components with the same index in vectors $\boldsymbol{\nu}^1, \boldsymbol{\nu}^2, \ldots$, and $\boldsymbol{\nu}^K$, i.e. $H_0 : \boldsymbol{\nu}^1(i) = \boldsymbol{\nu}^2(i) = \cdots = \boldsymbol{\nu}^K(i)$, for all $i = 1, \ldots, m$.

Similar problem has been discussed by G.S. James (1951), which is to test the equality of group means under the assumption of unequal variance [18]. We state the fact as follows

$$\sum_{k=1}^{K} \tau_{ik} \big(\tilde{\boldsymbol{\nu}}_p^k(i) - \overline{\boldsymbol{\nu}}_p(i)\big)^2 \ \sim \chi^2(K-1) \ \text{ for } \ k = i, \ldots, m, \tag{3.10}$$

where $\tau_{ik} := 1/(\sigma_k^2 \lambda_i)$, and $\overline{\nu}_p(i) = (\sum_{k=1}^{K} \tau_{ik} \tilde{\nu}_p^k(i))/\tau_{i\cdot}$ is the ith component of the weighted

averaged vector $\overline{\boldsymbol{\nu}}_p$, where $\tau_{i.} := \sum_{k=1}^{K} \tau_{ik} = \frac{1}{\lambda_i} \sum_{k=1}^{K} \frac{1}{\sigma_k^2}$.

Due to the independence between components, by summing Equation (3.10) up together, we have

$$X^2 := \sum_{i=1}^{m} \sum_{k=1}^{K} \tau_{ik} \big(\tilde{\boldsymbol{\nu}}_p^k(i) - \overline{\boldsymbol{\nu}}_p(i)\big)^2 \quad \sim \chi^2\big(m * (K-1)\big),$$

which can be written in a matrix form as

$$X^2 = \sum_{k=1}^{K} (\tilde{\boldsymbol{\nu}}_p^k - \overline{\boldsymbol{\nu}}_p)^T \Delta_k (\boldsymbol{\nu}_p^k - \overline{\boldsymbol{\nu}}_p) \quad \sim \chi^2\big(m * (K-1)\big), \tag{3.11}$$

where

$$\Delta_k := \text{diag}(\tau_{1k}, \ldots, \tau_{mk}) = \text{diag}(\frac{1}{\sigma_k^2 \lambda_1}, \frac{1}{\sigma_k^2 \lambda_2}, \ldots, \frac{1}{\sigma_k^2 \lambda_m}) = \frac{1}{\sigma_k^2} \Lambda^{-1},$$

$$\overline{\boldsymbol{\nu}}_p = (\overline{\nu}_p(1), \overline{\nu}_p(2), \ldots, \overline{\nu}_p(m))^T = (\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \boldsymbol{\nu}_p^k) / (\sum_{k=1}^{K} \frac{1}{\sigma_k^2}).$$

After transforming $\boldsymbol{\nu}_p^k$ back to $\boldsymbol{\alpha}_p^k$ in Equation (3.11), it follows

$$\begin{aligned} X^2 &= \sum_{k=1}^{K} (\tilde{\boldsymbol{\nu}}_p^k - \overline{\boldsymbol{\nu}}_p)^T \Delta_k (\tilde{\boldsymbol{\nu}}_p^k - \overline{\boldsymbol{\nu}}_p) = \sum_{k=1}^{K} \frac{1}{\sigma_k^2} (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)^T Q \Lambda^{-1} Q^T (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p) \\ &= \sum_{k=1}^{K} \frac{1}{\sigma_k^2} (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p) \\ &:= \sum_{k=1}^{K} \omega_k (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p) \quad \sim \chi^2\big(m * (K-1)\big), \end{aligned} \tag{3.12}$$

where $\omega_k := 1/\sigma_k^2$, $\overline{\boldsymbol{\alpha}}_p = (\sum_{k=1}^{K} \omega_k \tilde{\boldsymbol{\alpha}}_p^k)/(\sum \omega_i)$ is the weighted average of $\boldsymbol{\alpha}_p^k$ vector and $\tilde{\boldsymbol{\alpha}}_p^k \sim N(\boldsymbol{\alpha}^k, \frac{1}{\omega_k} M)$ for $k = 1, \ldots, K$.

3.3.1 Homoscedasticity

Under this case, the Chi-square statistic X^2 in Equation (3.12) is given by

$$X^2 = \frac{\sum_{k=1}^{K}(\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)^T M^{-1}(\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)}{\sigma^2} \sim \chi^2\big(m*(K-1)\big), \tag{3.13}$$

where

$$\overline{\boldsymbol{\alpha}}_p = (\sum_{k=1}^{K}\frac{1}{\sigma^2}\tilde{\boldsymbol{\alpha}}_p^k)/(\sum_{k=1}^{K}\frac{1}{\sigma^2}) = \frac{1}{K}\sum_{k=1}^{K}\tilde{\boldsymbol{\alpha}}_p^k.$$

And also,

$$Y^2 := \frac{(n-1)(S_1^2+S_2^2+\cdots+S_K^2)}{\sigma^2} \sim \chi^2\big(K*(n-1)\big), \tag{3.14}$$

then it follows Equations (3.13) and (3.14) by

$$\begin{aligned} F &:= \frac{X^2/\big(m*(K-1)\big)}{Y^2/\big(K*(n-1)\big)} \\ &= \frac{\sum_{k=1}^{K}(\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)^T M^{-1}(\tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p)}{(S_1^2+S_2^2+\cdots+S_K^2)/K} \sim F(m(K-1),K(n-1)). \end{aligned} \tag{3.15}$$

where $m = a-1$ and $n = (a-2)(p-2)$.

3.3.2 Heteroscedasticity

Since we have unequal variance component, σ_k^2, for K data sets, we use $\mathrm{w}_k = 1/S_k^2$ to replace $\omega_k = 1/\sigma_k^2$ in Equation (3.12), then we have the statistics T^2 as follows

$$T^2 := \sum_{k=1}^{K} \mathrm{w}_k (\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p), \tag{3.16}$$

where $\hat{\boldsymbol{\alpha}}_p := \big(\sum_{k=1}^{K} \mathrm{w}_k \tilde{\boldsymbol{\alpha}}_p^k\big)/(\sum_{i=1}^{K} \mathrm{w}_i)$.

When all the $\boldsymbol{\alpha}^k$ are equal, the statistic T^2 is χ^2 distributed with $m * (K-1)$ degrees of freedom for sufficiently large sample size (i.e. provided n is large). For sample size not large enough, Welch (1951) derived an approximate distribution to the similar statistic under the scalar case [47]. In this section, we extended it to the multidimensional case by applying the same procedure in Welch's paper.

Next, we approximate the distribution of T^2 by finding its moment generating function, which is not simple and after a certain order becomes infinite. If it exists, it is written as

$$M(u) = E[\exp(uT^2)] = E_1 E_2[\exp(uT^2)], \tag{3.17}$$

where E_2 denotes averaging over the joint distribution of S_k^2 and E_1 denotes averaging over the joint distribution of $\tilde{\boldsymbol{\alpha}}_p^k$.

Recall that $\exp(uT^2)$ is a function of $\mathbf{w} = (\mathrm{w}_1, \ldots, \mathrm{w}_K)^T$, so we take Taylor expansion to the first order at the point $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)^T$, which is given by

$$\exp(uT^2) \approx \exp(uT^2)|_{\mathbf{w}=\boldsymbol{\omega}} + \sum_{k=1}^{K} (\mathrm{w}_k - \omega_k) D_k \exp(uT^2)|_{\mathbf{w}=\boldsymbol{\omega}}$$

$$= [I + \sum_{k=1}^{K}(\mathrm{w}_k - \omega_k)D_k]\big(\exp(uT^2)\big) \approx \exp[\sum_{k}(\mathrm{w}_k - \omega_k)D_k]\big(\exp(uT^2)\big),$$

where $I = I|_{\mathbf{w}=\boldsymbol{\omega}}$ is an identity operator and $D_k = \frac{\partial}{\partial \mathrm{w}_k}|_{\mathbf{w}=\boldsymbol{\omega}}$. Then, we have

$$E_2[\exp(uT^2)] = \underbrace{E_2\big[\exp[\sum_{k}(\mathrm{w}_k - \omega_k)D_k]\big]}_{:\Phi}\big(\exp(uT^2)\big). \tag{3.18}$$

In Welch's paper, Φ is proceeded only to a certain order of $1/n_k$, which is given by

$$\Phi = I + 2\sum_{k}\frac{\omega_k D_k}{n_k} + \sum_{k}\frac{\omega_k^2 D_k^2}{n_k}, \tag{3.19}$$

where n_k is the number of degrees of freedom of $\frac{n_k S_k^2}{\sigma_k^2}$ for $k = 1, \ldots, K$.

From Equation (3.18), we have

$$\begin{aligned}
I\big(\exp(uT^2)\big) &= \exp(uX^2), \\
D_k\big(\exp(uT^2)\big) &= \Big[-\frac{2u}{\sum_{i=1}^{K}\omega_i}X^2 + u\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k\Big]\cdot\exp(uX^2), \\
D_k^2\big(\exp(uT^2)\big) &= \Big[-\frac{2u}{\sum_{i=1}^{K}\omega_i}X^2 + u\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k\Big]^2\cdot\exp(uX^2) \\
&\quad + \Big[\frac{4u}{(\sum_{i=1}^{K}\omega_i)^2}X^2 - \frac{2u}{\sum_{i=1}^{K}\omega_i}\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k + \cdots\Big]\cdot\exp(uX^2),
\end{aligned}$$

where $\boldsymbol{z}_k := \tilde{\boldsymbol{\alpha}}_p^k - \overline{\boldsymbol{\alpha}}_p$ for $k = 1, \ldots, K$. Therefore

$$\begin{aligned}
E_2[\exp(uT^2)] &= \exp(uX^2) \\
&\times \underbrace{\Big\{1 + 2u\sum_{k}\frac{\omega_k}{n_k}(1 - \frac{\omega_k}{\sum_{i=1}^{K}\omega_i})\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k + u^2\sum_{k}\frac{\omega_k^2}{n_k}\big[\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k\big]^2 + \ldots\Big\}}_{:\quad g(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K)}.
\end{aligned} \tag{3.20}$$

To obtain $M(u)$, we need to take expectation with respect to the right-hand side of Equation (3.20) over the joint distribution of the $\tilde{\boldsymbol{\alpha}}_p^k$, where $\tilde{\boldsymbol{\alpha}}_p^k \sim N(\boldsymbol{\alpha}^k, \frac{1}{\omega_k}M)$.

Since we only consider what happens under the null hypothesis which states the $\boldsymbol{\alpha}^k$ are equal, and Equation (3.20) does not depend on the origin of coordinates, without loss of generality, we take $\boldsymbol{\alpha}^1 = \ldots = \boldsymbol{\alpha}^K = \mathbf{0}$, and its density function is given by

$$p_{\tilde{\boldsymbol{\alpha}}_p^k}(\boldsymbol{x}_k) = (2\pi)^{-\frac{1}{2}m}\omega_k^{\frac{1}{2}m}(\det M)^{\frac{1}{2}} \cdot \exp\big(-\frac{1}{2}\omega_k \boldsymbol{x}_k^T M^{-1}\boldsymbol{x}_k\big). \tag{3.21}$$

We then have from Equation (3.20) and Equation (3.21)

$$\begin{aligned} M(u) \;\approx\; & \int (2\pi)^{-\frac{1}{2}mK}(\Pi_{k=1}^K \omega_k)^{\frac{1}{2}m}(\det M)^{-\frac{K}{2}} \cdot \exp\big[-\frac{1}{2}\sum_k \omega_k \boldsymbol{x}_k^T M^{-1}\boldsymbol{x}_k + uX^2\big] \\ \times\; & \underbrace{\Big\{1 + 2u\sum_k \frac{\omega_k}{n_k}(1 - \frac{\omega_k}{\sum_{i=1}^K \omega_i})\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k + u^2 \sum_k \frac{\omega_k^2}{n_k}\big[\boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k\big]^2\Big\}}_{:\quad g(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K)} d\boldsymbol{x}_1 \ldots \boldsymbol{x}_K, \end{aligned} \tag{3.22}$$

where $X^2 = \sum_{k=1}^K \omega_k \boldsymbol{z}_k^T M^{-1}\boldsymbol{z}_k$, $\boldsymbol{z}_k = \boldsymbol{x}_k - \overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}} = \sum_k \omega_k \boldsymbol{x}_k / (\sum_i \omega_i)$.

Since it is difficult to integrate, we use again the Taylor's series of $g(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K)$ at the point $(\mathbf{0}, \ldots, \mathbf{0})$, then

$$\begin{aligned} g(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K) &= g(\mathbf{0}, \ldots, \mathbf{0}) + \sum_k \boldsymbol{z}_k^T \boldsymbol{\delta}_k \\ &= (I + \sum_k \boldsymbol{z}_k^T \boldsymbol{\delta}_k)\Big(g(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K)\Big) \\ &= \exp(\sum_k \boldsymbol{z}_k^T \boldsymbol{\delta}_k)\Big(g(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K)\Big), \end{aligned} \tag{3.23}$$

where $I = I|_{\boldsymbol{z}=\mathbf{0}}$ is an identity operator and $\boldsymbol{\delta}_k = \frac{\partial}{\partial \boldsymbol{z}_k}|_{\boldsymbol{z}=\mathbf{0}}$ denotes differentiation with

respect to $\boldsymbol{z}_k$ with $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ setting to zero after all differentiation have been carried out.

By plugging Equation (3.23) in Equation (3.22), we have

$$M(u) \approx \Psi\Big(g(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K)\Big), \tag{3.24}$$

where

$$\begin{aligned} \Psi &= \int (2\pi)^{-\frac{1}{2}mK} (\Pi_{k=1}^{K} \omega_k)^{\frac{1}{2}m} (\det M)^{-\frac{K}{2}} \cdot \exp\Big\{ -\frac{1}{2} \sum_k \omega_k \boldsymbol{x}_k^T M^{-1} \boldsymbol{x}_k \\ &+ u \sum_{k=1}^{K} \omega_k (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^T M^{-1} (\boldsymbol{x}_k - \overline{\boldsymbol{x}}) + \sum_k (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^T \boldsymbol{\delta}_k) \Big\} d\boldsymbol{x}_1 \ldots \boldsymbol{x}_K, \end{aligned} \tag{3.25}$$

where $\overline{\boldsymbol{x}} = \sum_k \omega_k \boldsymbol{x}_k / (\sum_i \omega_i)$.

On integration,

$$\Psi = (1-2u)^{-\frac{1}{2}m(K-1)} \exp\Big\{ \frac{1}{2} (1-2u)^{-1} \big[\sum_k \frac{\boldsymbol{\delta}_k^T M \boldsymbol{\delta}_k}{\omega_k} - \frac{(\sum_i \boldsymbol{\delta}_i)^T M (\sum_i \boldsymbol{\delta}_i)}{\sum_i \omega_i} \big] \Big\}. \tag{3.26}$$

Next, we keep only a few terms after applying the operator Φ to the function $g(\cdot)$,

$$\Psi_{\boldsymbol{z}_k^T M^{-1} \boldsymbol{z}_k} = (1-2u)^{-\frac{1}{2}m(K-1)} \Big[(1-2u)^{-1} C_1 (\frac{1}{\omega_k} - \frac{1}{\sum_i \omega_i}) \Big], \tag{3.27}$$

$$\Psi_{[\boldsymbol{z}_k^T M^{-1} \boldsymbol{z}_k]^2} = (1-2u)^{-\frac{1}{2}m(K-1)} \Big[3(1-2u)^{-2} C_2 (\frac{1}{\omega_k} - \frac{1}{\sum_i \omega_i})^2 \Big], \tag{3.28}$$

where

$$C_1 = \sum_{1 \le i \le j \le m} M_{ij} \cdot (M^{-1})_{ij}, \tag{3.29}$$

$$C_2 = \sum_{i=1}^{m} M_{ii}^2 (M^{-1})_{ii}^2 + \frac{1}{3} \sum_{1 \le i < j \le m} M_{ij}^2 (M^{-1})_{ij}^2. \tag{3.30}$$

From Equations (3.24) and (3.26),

$$M(u) = (1-2u)^{-\frac{1}{2}m(K-1)}\Big\{1+\big[2C_1u(1-2u)^{-1}+3C_2u^2(1-2u)^{-2}\big]\big[\sum_k \frac{1}{n_k}\big(1-\frac{\omega_k}{\sum_i\omega_i}\big)^2\big]\Big\}, \tag{3.31}$$

then to the order $1/n_k$, the corresponding cumulant-generating function of T^2 is

$$K(u) \approx -\frac{1}{2}m(K-1)\log(1-2u)+\big[2C_1u(1-2u)^{-1}+3C_2u^2(1-2u)^{-2}\big]\big[\sum_k \frac{1}{n_k}\big(1-\frac{\omega_k}{\sum_i\omega_i}\big)^2\big]. \tag{3.32}$$

We may invert $M(u)$ in Equation (3.31) term by term to produce a distribution function, which is more complicated. Welch used one of the Pearson curves followed by the variance ratio F distribution by matching its moment generating function to $M(u)$. The following induction is done by Welch, I list the result for completion.

Let $F := \frac{\chi_1^2/f_1}{\chi_1^2/f_2}$, where χ_1^2 and χ_2^2 are distributed independently as χ^2 with degrees of freedom f_1 and f_2, respectively. Then the moment generating function $M_F(u)$ of F in Welch's paper, to the order of $1/f_2$, is given by

$$M_F(u) = (1-2u/f_1)^{-\frac{1}{2}f_1}\Big[1+\frac{2u}{f_2}(1-\frac{2u}{f_1})^{-1}+\frac{f_1+2}{f_1f_2}u^2(1-\frac{2u}{f_1})^{-2}\Big]. \tag{3.33}$$

Next, let $f_1 = m(K-1)$ and $G = \big[f_1 + A/f_2\big]F$, by Equation (3.33)

$$\begin{aligned} M_G(u) &= (1-2u)^{-\frac{1}{2}m(K-1)}\Big[1+\frac{A+2m(K-1)u(1-2u)^{-1}}{f_2} \\ &+ \frac{\big(m(K-1)+1\big)^2-1}{f_2}u^2(1-2u)^{-2}\Big], \end{aligned} \tag{3.34}$$

then,

$$\begin{aligned} K_G(u) &= -\frac{1}{2}m(K-1)\log(1-2u) + \frac{A + 2m(K-1)u(1-2u)^{-1}}{f_2} \\ &+ \frac{\big(m(K-1)+1\big)^2 - 1}{f_2} u^2(1-2u)^{-2}. \end{aligned} \tag{3.35}$$

By comparing $K(u)$ in Equation (3.32) with $K_G(u)$ in Equation (3.35), they are equal if

$$\frac{A + 2m(K-1)u(1-2u)^{-1}}{f_2} = 2C_1 \sum_k \frac{1}{n_k}\big(1 - \frac{\omega_k}{\sum_i \omega_i}\big)^2, \tag{3.36}$$

$$\frac{\big(m(K-1)+1\big)^2 - 1}{f_2} = 3C_2 \sum_k \frac{1}{n_k}\big(1 - \frac{\omega_k}{\sum_i \omega_i}\big)^2, \tag{3.37}$$

i.e.

$$\frac{A}{f_2} = \Big[2C_1 - \frac{6C_2}{m(K-1)+2}\Big] \sum_k \frac{1}{n_k}\big(1 - \frac{\omega_k}{\sum_i \omega_i}\big)^2, \tag{3.38}$$

$$\frac{1}{f_2} = \frac{3C_2}{\big(m(K-1)+1\big)^2 - 1} \sum_k \frac{1}{n_k}\big(1 - \frac{\omega_k}{\sum_i \omega_i}\big)^2. \tag{3.39}$$

Therefore, to order $1/n_k$ it appears that T^2 is distributed as $G = \big[f_1 + A/f_2\big]F$, where $f_1 = m(K-1)$, and f_2 and A/f_2 are given by Equations (3.38) and (3.39).

If we define a statistic as follows

$$F^* := \frac{T^2}{f_1 + A/f_2} = \frac{\Big[\sum_{k=1}^{K} \mathrm{w}_k (\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p)^T M^{-1} (\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p)\Big]/f_1}{1 + \frac{1}{f_2}\Big[\frac{2C_1(f_1+2)}{3C_2} - 2\Big]}, \tag{3.40}$$

where

$$f_1 = m(K-1), \tag{3.41}$$

$$f_2 = \left[\frac{3C_2}{\big(f_1+1\big)^2-1}\sum_k \frac{1}{n_k}\big(1-\frac{\omega_k}{\sum_i \omega_i}\big)^2\right]^{-1}, \tag{3.42}$$

then, when $\boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^2 = \cdots = \boldsymbol{\alpha}^K$, we have approximately

$$Pr(F^* > F_P) = P, \tag{3.43}$$

where F_P is the tabled value F distribution exceeded with probability P for degrees of freedom f_1 and f_2.

It will be seen from Equation (3.40) that F^* involves the unknown quantity $\omega_k = 1/\sigma_k^2$ in f_2. We still have a probability statement Equation (3.43) by substituting ω_k with w_k to the order $1/n_k$, because ω_k enter only into expressions of this order. The approximate test procedure will be

- Calculate the statistic

$$F^* = \frac{\Big[\sum_{k=1}^K \mathrm{w}_k(\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p)^T M^{-1}(\tilde{\boldsymbol{\alpha}}_p^k - \hat{\boldsymbol{\alpha}}_p)\Big]/f_1}{1+\frac{1}{f_2}\Big[\frac{2C_1(f_1+2)}{3C_2}-2\Big]}, \tag{3.44}$$

  where

$$f_1 = m(K-1); \qquad f_2 = \left[\frac{3C_2}{\big(f_1+1\big)^2-1}\sum_k \frac{1}{n_k}\big(1-\frac{\mathrm{w}_k}{\sum_i \mathrm{w}_i}\big)^2\right]^{-1}, \tag{3.45}$$

$$C_1 = \sum_{1\le i\le j\le m} M_{ij}\cdot(M^{-1})_{ij}; \quad C_2 = \sum_{i=1}^m M_{ii}^2(M^{-1})_{ii}^2 + \frac{1}{3}\sum_{1\le i<j\le m} M_{ij}^2(M^{-1})_{ij}^2,$$

M is the covariance matrix of age effect excluding the variance component.

- Refer F^* to a table of F distribution with degrees of freedom f_1 and f_2.

3.4 Summary

The extended F test in terms of vectors was derived due to the special form of variance-covariance matrix of age effects as $\sigma^2 M$, in which the variation is from the variance component only and the matrix part, M, is fixed. The other assumption of the test is to assume the same dimension of APC table across different populations. Under this case, we can have the same orthogonal decomposition of the matrix part, then the weighted average of age effects can be constructed from all vectors of age effects. Not only can we test the equality of age trends across multiple populations, but also we can test the period effects using the same test derived above due to the symmetry between the rows (or age groups) and columns (periods) of an APC table.

CHAPTER 4

Simulation and Application

In this chapter, the US lung and bronchus cancer incidence data and heart disease mortality data among different racial groups and gender were considered, which were retrieved from the **Surveillance Epidemiology and End Results** (SEER) database. These data sets were explored to illustrate statistical models and approaches in a descriptive and analytic way.

4.1 Simulation

In this section, I did the simulation of hypothesis testing by the F test, which was introduced in Chapter 3. Only three populations were being considered. For convenience, period and cohort effects for the remaining two populations were given by multiplying a coefficient or taking a reverse order to that given in the first population, which were shown in Tables 4.1 and 4.2 under the null hypothesis $H_0 : \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3$ and the alternative hypothesis $H_1 : \boldsymbol{\alpha}_i \neq \boldsymbol{\alpha}_j$, for $i \neq j$, respectively. After specifying each significant level, the proportions of the test being rejected were calculated for testing on the null hypothesis of identical age curves across all populations.

4.1.1 Parameters Specification of APC Model for Different Population

For each population, age curves were specified to be equal to the same values as in row set 1 with a=9 in Figure 2.7, and was given by $\boldsymbol{\alpha}$=(-0.224, -0.775, -0.989, -0.775, -0.224,0.427, 0.901, 0.993, 0.665)^T. For the first population, the period and cohort effects were specified the same as before, which were $\beta_j = sin(j)$ for $j = 1, \ldots, p$, and $\gamma_k = cos(k) + sin(k \cdot 10^9)$ for $k = 1, \ldots, a + p - 1$, respectively. The intercept was specified to be $\mu = -8$ for all populations. An example of APC trend for each population specified from Table 4.2 when $p = 10$ is given in Figure 4.1.

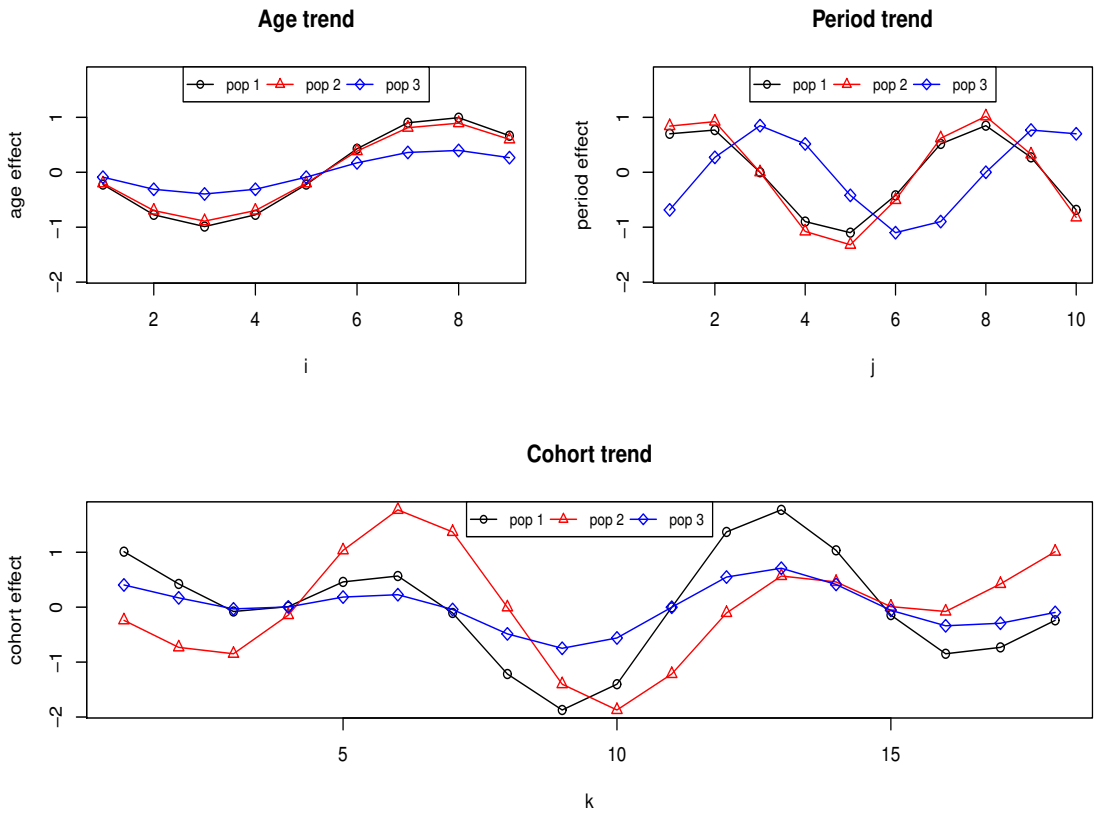

Figure 4.1: **The APC trend for each population specified for simulation when p=10 under** H_1. The i, j, and k are the indices for age, period, and cohort effects, respectively, for $i = 1, \ldots, 9$, $j = 1, \ldots, 10$, and $k = 1, \ldots, 18$.

Table 4.1: Parameter setting under H_0

	Parameters			
Pop(k)	Intercept(μ)	Age($\boldsymbol{\alpha}$)	Period ($\boldsymbol{\beta}$)	Cohort($\boldsymbol{\gamma}$)
1	-8	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$
2	-8	$\boldsymbol{\alpha}$	$1.2 * \boldsymbol{\beta}$	rev($\boldsymbol{\gamma}$)
3	-8	$\boldsymbol{\alpha}$	rev($\boldsymbol{\beta}$)	$0.4 * \boldsymbol{\gamma}$

Table 4.2: Parameter setting under H_1

	Parameters			
Pop(k)	Intercept(μ)	Age($\boldsymbol{\alpha}$)	Period ($\boldsymbol{\beta}$)	Cohort($\boldsymbol{\gamma}$)
1	-8	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$
2	-8	$0.9 * \boldsymbol{\alpha}$	$1.2 * \boldsymbol{\beta}$	rev($\boldsymbol{\gamma}$)
3	-8	$0.4 * \boldsymbol{\alpha}$	rev($\boldsymbol{\beta}$)	$0.4 * \boldsymbol{\gamma}$

4.1.2 Data Generation for Each Population

For each set of specified parameters $(\mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_{a+p-1})$ with given $a = 9$, $p = 10$, 20, and 30, in each iteration among N (N=1000) times, the responses for kth population were generated by a linear model for normal variables

$$y_{ij}^k = \mu + \alpha_i^k + \beta_j^k + \gamma_{a-i+j}^k + \epsilon_{ij}^k \quad \text{for } k = 1, 2, 3,$$

where the error term ϵ_{ij}^k follows a normal distribution $N(0, \sigma_k^2)$, the variance component, σ_k^2, were specified from a signal-noise ratio of 3, which are given in Table 4.3.

Table 4.3: The variance component under H_0 and H_1

	variance $\sigma_k^2(H_0)$			variance $\sigma_k^2(H_1)$		
p	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
p=10	0.680	0.775	0.377	0.680	0.741	0.236
p=20	0.608	0.807	0.400	0.608	0.771	0.258
p=30	0.625	0.809	0.416	0.625	0.775	0.269

4.1.3 Simulation Result

Table 4.4 displays the proportion of tests being rejected among 1000 iterations under the null hypothesis after specifying the significant level of 0.01, 0.03, 0.05, and 0.1 for $p = 10$, 20, and 30, respectively. The rejection rate are very close to the given significance level. Specially, we obtained higher value of rejection rate than significance level for 0.05 and 0.01. Table 4.5 displays the power of the test under the alternative hypothesis, all values are close to one, which means that the extended F test is powerful.

Table 4.4: The proportion of tests being rejected under H_0

	Type I error			
p	$\alpha = 0.01$	$\alpha = 0.03$	$\alpha = 0.05$	$\alpha = 0.1$
p=10	0.008	0.036	0.057	0.124
p=20	0.013	0.025	0.054	0.108
p=30	0.009	0.029	0.054	0.123

Table 4.5: The proportion of tests being rejected under H_1

	Power			
p	$\alpha = 0.01$	$\alpha = 0.03$	$\alpha = 0.05$	$\alpha = 0.1$
p=10	0.59	0.741	0.81	0.893
p=20	0.975	0.991	0.998	1
p=30	1	1	1	1

4.2 Application

4.2.1 Data Source

The **Surveillance Epidemiology and End Results** (SEER) Program collects information through cancer registries in various states on newly diagnosed invasive cancers in the United States. The SEER data include cancer incidence, mortality, and population data associated by age, year of diagnosis, gender, race, incidence site, geographic areas, and so on. We retrieved from the SEER database incidence rate data of lung and bronchus cancer and mortality rate data of heart disease by age, year of diagnosis, gender, and race from 1973 to 2012 and 1973 to 2010, respectively. Here three major racial groups of white, black, and other (American Indian/Alaska and Asian/Pacific Islander combined) and two gender groups (male and female) were considered. More information are on the website http://seer.cancer.gov/.

4.2.2 Data

The Lung and Bronchus Cancer Incidence Data

Tables 4.6, 4.7, 4.8, 4.9, 4.10, and 4.11 display the incidence rate (per 100,000 person-years) with frequency of lung and bronchus cancer grouped in five-year intervals of age and year. These data correspond to people in U.S. among different racial groups—white, black, and, other racial and different gender—male and female, respectively. These tables have 12 rows representing age groups and 8 columns representing period groups with 5 year intervals for each. Age groups are from 30-34, 35-39, to 80-84, and 85+, and period groups are from 1973-1977, 1978-1982, to 2008-2012.

The Heart Disease Mortality Data

Tables 4.12, 4.13, 4.14, 4.15, 4.16, and 4.17 are the mortality rate (per 100,000 person-years) with frequency of heart disease associated with three racial groups (white, black, and other) and two genders (male and female). These tables have 14 rows representing age groups and 8 columns representing period groups with 5 year intervals for each, except for the last column, which covers only three years (2008-2010), in which age groups are from 20-24, 25-29, to 80-84, and 85+, and period groups are from 1973-1977, 1978-1982, to 2008-2012.

Table 4.6: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US white males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	2.3	1.5	1.8	1.5	1.2	1.1	1.1	1.2
	67	56	74	64	54	44	40	45
35-39	7.7	7.3	5.7	4.7	4	3.5	2.9	2.6
	176	214	203	189	177	150	109	91
40-44	22.1	20.1	18.7	13.3	11.5	11	9.3	6.7
	479	474	535	476	463	469	381	250
45-49	59	55.1	46	39.7	29.9	26.4	22.6	19.1
	1346	1212	1057	1116	1041	1021	929	761
50-54	110.1	117	104.9	92.1	72	56.5	52.3	44.7
	2558	2708	2212	2036	1955	1913	1945	1785
55-59	183.1	197.7	203.4	173.5	145.3	121.7	100.2	85.5
	3757	4435	4325	3404	3044	3115	3192	3056
60-64	287	302.1	305.1	307.5	259.4	237	188.7	146.7
	4841	5787	6033	5883	4660	4482	4408	4379
65-69	392.3	429.9	428.2	430	410	359.8	311.2	271.7
	5014	6402	7015	7358	6849	5624	5224	5808
70-74	485.5	524.6	539	529.2	510.8	498	440.7	386.8
	4445	5601	6481	7048	7313	7043	5960	5728
75-79	518.4	568.6	591.2	595.7	574.7	552.4	559.3	501.6
	3173	3976	4736	5513	6096	6372	6454	5664
80-84	446.6	521.7	566.8	602.6	562.5	553.5	551	552.5
	1707	2119	2543	3144	3562	4122	4594	4769
85+	345.7	410.9	441.4	474.6	472.9	477.8	474.2	464.5
	810	1172	1351	1608	1933	2370	2846	3401

Table 4.7: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US white females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	1.7	1.5	1.4	1.1	1.6	1.3	1.6	1.5
	48	54	56	45	65	49	54	51
35-39	5.8	6.2	4.6	3.6	4.6	4.4	2.8	2.7
	133	181	164	143	196	182	104	90
40-44	15.4	16.9	14.5	11.4	9.5	11.3	10.2	7.8
	337	403	418	410	379	477	410	283
45-49	30.7	33.8	34	31.1	27.1	22.8	25.7	21.6
	722	759	790	876	944	882	1052	849
50-54	46.9	61.1	65.9	69.8	57.7	49	44.7	44.8
	1151	1472	1427	1573	1596	1682	1680	1804
55-59	72.2	91.7	109.6	120.5	109	103.7	79.1	69.7
	1585	2222	2485	2472	2360	2719	2593	2556
60-64	88.8	125.7	155.2	173.3	175.4	171.5	155.9	126.2
	1662	2679	3416	3648	3397	3446	3826	3951
65-69	96.9	146.1	188.1	224.6	252.2	253.4	244.8	220.4
	1540	2679	3727	4617	4930	4496	4551	5112
70-74	91.8	142.3	203.6	260.8	296.4	318.6	338.2	310.4
	1176	2115	3334	4577	5452	5580	5478	5332
75-79	83.1	120.2	180.6	254.1	301.1	330.5	379.9	386.6
	823	1372	2313	3596	4671	5387	5841	5534
80-84	70.8	101.2	144.3	206	268.9	298.4	349.8	378.2
	498	804	1267	2045	3024	3665	4588	4756
85+	69.8	82.2	104.9	131.9	164.4	203.1	248.4	262.1
	354	558	843	1205	1762	2447	3287	3891

Table 4.8: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US black males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	3.6	4.6	3.2	2.4	2.1	1.7	2	2.3
	10	19	16	14	13	11	12	14
35-39	17.3	16.5	12.5	15.4	10.4	6	3.7	2.9
	37	49	49	77	62	38	23	18
40-44	52.6	51.9	42.3	39.2	35.6	23.6	13	9.1
	108	125	127	160	179	139	81	57
45-49	130.3	111.6	118.4	104.4	65	57.8	50.4	30
	261	238	284	306	257	285	288	183
50-54	211	244	243.6	199	155.2	127.1	106.2	88.7
	414	507	499	444	440	501	508	500
55-59	325.7	385.3	377.3	329.1	288.5	242.6	199.6	156.9
	532	755	732	615	601	649	738	715
60-64	397.8	465.2	537.1	500.9	419.2	373.9	302.2	257.5
	524	721	887	857	732	721	738	873
65-69	494.1	611.8	667.7	645	573.8	505.9	479.6	386.9
	492	766	916	956	872	761	810	848
70-74	560.4	689.4	747.3	745.3	761	644.1	539.5	486.8
	356	559	696	771	846	754	674	699
75-79	440.7	523.7	662	672.7	712.4	738.4	618.9	629.5
	165	261	393	461	559	647	559	601
80-84	495.7	501.1	709.3	706.1	726.2	659.6	680.3	671.1
	85	116	208	249	301	321	387	407
85+	420.6	447.1	454	459.9	514.9	535.2	569.4	503.5
	45	68	91	111	143	169	207	240

Table 4.9: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US black females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	1.2	1.5	1.9	3.1	1.8	1.5	2	1.1
	4	7	11	20	13	11	14	8
35-39	7.2	6.2	4.4	7.6	8.1	6.3	3.9	2.5
	18	21	20	44	55	45	28	18
40-44	26	19.5	27.3	18.4	17.7	19	14	8.3
	61	54	95	87	103	127	99	59
45-49	53.1	52.5	51.7	42.5	39.9	35	43.1	32.1
	121	130	143	142	183	201	281	223
50-54	59.8	83	86.3	91.6	81.9	65.1	67.4	65.9
	132	201	206	239	270	302	378	424
55-59	69.2	97.6	139.2	132.7	113.7	137.4	102.9	97.3
	127	220	318	302	287	435	459	531
60-64	79.7	127.2	165.2	202.2	193	176.3	172.4	146.2
	121	239	340	435	418	421	518	624
65-69	79.6	120.6	164.7	256.8	245.6	276	244.4	241.9
	95	193	295	504	490	556	541	684
70-74	102.2	128.8	186.1	253.8	267.4	308.8	296.7	298.6
	83	144	252	384	440	530	536	597
75-79	79.6	115	147.7	199.8	245.2	319.1	349.7	345.5
	43	88	141	228	310	448	504	525
80-84	75.1	102.8	125.5	196	195.3	300	334.4	349.2
	22	42	67	131	161	278	357	394
85+	94.7	91.3	103.7	116.6	170.4	196	247.5	268.4
	21	31	47	66	121	165	232	313

Table 4.10: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US other males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	1.5	1.7	1.3	1.3	1.1	1.3	2.3	1.6
	3	5	5	6	6	8	16	12
35-39	6.4	6.7	4	5.2	3.2	2.8	3.9	2.6
	10	15	13	22	17	17	25	19
40-44	21.4	21.6	14.6	9.8	12	8.3	8.2	5.3
	31	39	37	35	56	46	51	36
45-49	52.2	36	34.9	32.8	24.5	19.6	20	13.1
	74	57	68	85	92	95	115	85
50-54	67.7	84.5	68.3	62	52.8	37.8	37.5	35.1
	92	130	116	129	153	155	189	208
55-59	107.5	121	132.3	128.6	99.9	89.5	81.6	62.9
	124	178	216	231	220	259	339	321
60-64	148.6	174.4	194.4	206.9	179.3	166.7	134.9	120
	152	208	295	357	341	382	391	489
65-69	201.8	235.6	258.4	273.8	282.2	234.5	228.9	205.5
	168	234	315	425	488	442	515	582
70-74	236	274.3	304.6	342.5	396.6	339.6	355.1	282.1
	134	208	281	383	576	546	625	595
75-79	333.3	339.8	329.9	374	381.2	434.9	451.7	385.9
	110	162	209	296	372	548	624	593
80-84	329.7	390.7	320.1	366.5	365.5	446.2	425.5	480.9
	55	88	103	168	212	322	405	520
85+	484.8	379.6	382	314.4	406.2	431.2	433.2	459.4
	63	54	75	88	161	226	290	423

Table 4.11: Lung and bronchus cancer incidence rate (per 10^5 person-year) and frequency among US other females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-12
30-34	0.5	1.5	0.5	1	1.6	1.6	1.1	1.3
	1	5	2	5	9	10	8	10
35-39	5.3	3.7	2.8	3.3	2.2	2.4	2.6	2.6
	9	9	10	15	12	15	18	21
40-44	12.6	8.1	12.3	8.3	7.8	6.5	7.2	6.6
	21	16	34	33	40	39	49	49
45-49	16.8	16.3	13.2	14.3	14.7	13.8	15.8	13
	27	30	28	41	62	75	100	93
50-54	37.2	28.4	31.7	38.3	33.5	23.7	26.2	28.6
	56	53	64	90	108	109	150	190
55-59	45	43.2	50.2	64.9	51.4	48.5	48.7	42.3
	52	70	99	142	130	159	231	250
60-64	53.2	63.6	73.1	72.1	85	81.5	80	69.6
	47	81	131	159	202	219	271	337
65-69	91.1	70.6	91.7	111.4	116.8	129.6	134.4	119
	60	69	127	207	261	314	368	407
70-74	87.5	127.4	109.9	125.7	157.2	150.3	180.8	167.4
	43	84	106	167	282	328	424	442
75-79	127.8	148.1	152.6	164.7	192.5	165.2	208.5	214.9
	44	69	95	146	234	271	414	463
80-84	112.2	186.8	100.5	115	207.4	197.2	218.3	244.6
	24	54	38	57	148	200	303	419
85+	166.6	137	145.1	194.3	196.1	212.2	215.8	214.6
	25	31	45	81	110	171	245	347

Table 4.12: Heart disease mortality rate (per 10^5 person-year) and frequency among US white males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	3.3	3.4	3.4	3.2	3.2	3.3	3.8	3.5
	1384	1546	1501	1296	1208	1281	1580	901
25-29	6.0	5.9	5.9	5.7	5.6	5.7	6.6	6.5
	2239	2483	2713	2509	2305	2198	2564	1625
30-34	15.1	13.3	13.3	12.0	12.0	11.2	12.1	12.2
	4648	5049	5623	5519	5518	4740	4755	2823
35-39	43.8	37.3	32.8	27.0	26.8	24.7	23.9	22.3
	11009	11351	12281	11366	12507	11370	10057	5392
40-44	112.5	91.2	77.2	61.6	56.5	52.7	50.4	44.8
	27363	23062	23397	23126	23834	24379	22971	11309
45-49	232.9	193.1	156.3	126.5	112.0	98.4	92.3	85.3
	59168	46042	38805	37552	41382	41053	42057	23452
50-54	408.3	346.5	293.8	227.2	202.0	167.8	153.7	143.2
	104886	86495	68062	55053	59436	61775	63123	38195
55-59	674.3	575.5	492.2	398.0	334.5	276.1	232.6	215.4
	154283	139687	116579	88088	78636	78854	83338	50549
60-64	1069.2	906.0	788.2	645.3	557.6	441.4	357.9	313.1
	210498	192963	177274	142450	117277	99307	97756	62261
65-69	1597.9	1396.3	1211.4	1000.3	855.9	689.4	526.3	449.7
	254454	244792	229496	200747	172292	133628	109935	66084
70-74	2351.7	2119.5	1910.7	1560.8	1328.4	1093.4	830.7	691.2
	268615	274822	272603	244261	228656	190299	141755	74926
75-79	3537.4	3181.3	2903.3	2427.9	2097.6	1770.9	1379.7	1148.1
	264621	267014	279293	266125	259523	242771	194836	96499
80-84	5295.5	4897.4	4544.0	3938.4	3491.7	3012.0	2435.3	2035.6
	235494	231140	241192	244450	256095	256179	236260	125147

Table 4.13: Heart disease mortality rate (per 10^5 person-year) and frequency among US white females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	1.9	2.0	2.0	2.0	2.0	2.0	2.0	1.5
	802	881	855	753	720	715	785	369
25-29	2.9	2.9	3.1	3.0	3.2	3.1	3.1	2.8
	1082	1201	1377	1284	1246	1140	1149	677
30-34	5.7	5.0	4.9	4.9	5.4	5.3	5.2	5.3
	1786	1909	2061	2213	2442	2171	1992	1190
35-39	12.3	10.5	9.0	8.2	9.4	10.1	9.8	9.5
	3182	3236	3384	3425	4317	4583	4036	2242
40-44	27.6	23.5	20.2	16.2	16.7	17.6	18.6	17.5
	6864	6082	6229	6114	7041	8097	8453	4360
45-49	53.6	48.0	41.7	34.2	32.2	29.8	31.8	31.4
	14295	11873	10636	10372	12071	12577	14593	8658
50-54	102.3	91.3	84.5	69.5	63.0	55.1	50.5	49.9
	28279	24266	20561	17523	19143	20763	21224	13569
55-59	193.7	172.9	158.3	136.9	119.9	103.1	84.6	75.6
	48822	46578	41001	32466	29852	30889	31536	18499
60-64	361.0	325.0	298.5	250.6	225.6	187.5	150.8	128.6
	81556	79409	76768	62298	52118	45708	44174	27125
65-69	634.6	574.3	523.7	435.3	381.9	328.7	251.4	210.0
	126165	125321	121457	106275	90779	72781	58721	34104
70-74	1145.3	1025.4	935.9	778.3	675.6	572.7	451.6	366.5
	181821	184213	183071	161908	150024	124024	92213	46400
75-79	2074.6	1822.2	1650.9	1390.9	1209.9	1041.2	828.7	688.5
	246923	245729	253012	234605	219454	201782	156186	74061
80-84	3611.1	3323.6	3029.2	2597.6	2295.4	2010.7	1627.6	1339.5
	291199	297707	308851	304824	301806	286073	249606	122835

Table 4.14: Heart disease mortality rate (per 10^5 person-year) and frequency among US black males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	9.6	9.2	8.7	9.3	10.1	9.0	9.0	9.4
	531	601	599	611	662	625	700	465
25-29	20.8	19.2	17.7	16.6	16.5	16.2	17.6	15.4
	890	1052	1139	1097	1054	1028	1157	669
30-34	45.0	39.6	40.8	34.6	30.8	28.4	30.5	28.8
	1544	1752	2249	2210	2076	1863	1948	1133
35-39	101.6	89.8	81.2	72.2	60.9	53.6	53.2	50.2
	2839	2982	3557	4001	3998	3692	3511	1987
40-44	191.8	180.0	166.4	138.3	121.7	104.0	92.2	83.1
	5250	5175	5609	6104	6775	6852	6369	3341
45-49	359.4	325.3	281.6	264.0	232.1	192.4	171.3	144.6
	9321	8490	8005	8645	10030	10616	11209	6019
50-54	580.1	533.2	493.9	426.0	392.6	323.1	294.4	244.2
	14411	13398	12425	11480	12699	14091	16145	9371
55-59	831.6	817.2	734.2	682.8	599.8	525.0	439.4	384.2
	17823	19056	17401	15886	15370	16217	18521	11737
60-64	1226.1	1155.7	1152.5	1012.1	900.9	763.4	645.5	548.4
	22774	22459	23708	21340	19653	18509	18712	12150
65-69	1629.2	1550.6	1530.7	1423.2	1212.1	1056.0	868.1	772.7
	25389	25682	26019	25639	23174	20744	19103	11711
70-74	2435.2	2183.2	2220.1	1983.5	1846.2	1555.7	1250.1	1044.3
	25449	25657	27354	25545	25559	23224	20431	11245
75-79	3085.4	2927.0	2880.6	2720.5	2416.6	2215.5	1842.5	1560.9
	20542	22541	24475	24339	23231	23736	20924	11291
80-84	4248.3	4373.4	4451.6	3965.4	3641.5	3247.9	2778.2	2335.4
	14402	16539	19333	19506	19531	19353	19127	10417

Table 4.15: Heart disease mortality rate (per 10^5 person-year) and frequency among US black females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	6.9	6.1	5.8	5.6	5.9	5.5	4.9	5.0
	421	434	424	383	404	402	389	250
25-29	10.7	10.5	10.3	9.7	10.2	9.7	9.8	9.6
	531	658	737	705	720	682	713	452
30-34	22.1	18.7	18.2	17.2	17.1	16.9	16.8	15.5
	898	970	1157	1252	1301	1238	1205	680
35-39	48.8	38.6	34.3	32.0	32.9	29.7	27.5	24.4
	1664	1544	1774	2048	2451	2307	2047	1091
40-44	95.1	81.0	70.6	64.5	60.9	58.5	51.2	46.5
	3125	2815	2829	3329	3921	4368	3996	2090
45-49	174.4	149.4	127.0	118.3	109.0	99.8	88.5	78.7
	5415	4738	4354	4615	5547	6352	6555	3685
50-54	288.5	252.9	232.6	205.9	185.5	158.6	144.3	126.6
	8600	7841	7156	6757	7199	8116	9160	5528
55-59	446.8	406.5	388.7	344.3	307.1	264.5	213.3	186.3
	11461	11593	11393	10062	9798	9948	10700	6736
60-64	714.5	647.0	629.1	563.1	485.8	416.5	331.5	278.7
	16305	15886	16784	15613	13665	12803	12021	7578
65-69	1026.2	931.8	912.4	850.7	729.6	639.6	492.3	400.9
	20513	20749	21123	21103	18855	16916	14280	7951
70-74	1724.8	1455.1	1425.4	1245.4	1171.8	990.2	760.6	617.3
	24182	24176	26407	24333	24675	22147	17922	9353
75-79	2245.6	2133.6	2056.3	1877.5	1671.8	1558.3	1218.3	998.1
	21533	25188	28128	28619	27093	27711	22897	11579
80-84	3198.3	3334.3	3377.8	2969.4	2695.7	2418.0	2030.5	1660.0
	17267	21299	26687	28271	29519	28718	26937	14088

Table 4.16: Heart disease mortality rate (per 10^5 person-year) and frequency among US other males

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	2.9	4.0	3.2	2.8	3.0	2.7	2.3	2.7
	28	53	57	62	81	85	87	68
25-29	7.2	5.8	6.2	4.0	4.6	3.3	4.8	5.1
	62	74	111	92	124	109	185	132
30-34	10.4	10.7	10.1	7.6	8.7	7.5	8.9	6.9
	77	129	180	174	238	242	344	170
35-39	28.0	25.3	19.7	15.7	16.6	16.0	13.9	14.7
	165	235	288	319	430	496	499	364
40-44	57.9	48.9	40.8	32.9	32.3	30.0	29.0	25.5
	305	372	465	556	724	846	968	563
45-49	112.1	93.9	78.1	65.3	65.9	53.8	51.0	47.0
	498	564	690	812	1194	1302	1541	977
50-54	216.5	173.8	147.6	121.3	112.1	91.1	81.9	74.7
	825	872	1010	1171	1519	1813	2109	1376
55-59	301.5	276.9	250.6	220.6	192.0	156.7	132.5	114.6
	938	1207	1409	1626	1957	2204	2724	1730
60-64	499.6	418.2	386.5	351.3	308.1	249.7	199.5	178.7
	1324	1414	1882	2143	2433	2698	2889	2078
65-69	728.1	659.4	588.3	547.3	503.4	397.7	321.6	273.8
	1636	1829	2161	2750	3128	3184	3506	2230
70-74	1184.5	1080.7	985.3	844.5	780.0	650.6	477.9	407.1
	1871	2268	2673	2934	3789	3916	3685	2384
75-79	1716.6	1712.3	1588.4	1402.7	1237.3	1157.4	819.9	677.5
	1608	2311	2891	3242	3708	4905	4365	2606
80-84	2249.9	2567.3	2723.1	2449.1	2131.1	1802.4	1481.2	1265.9
	1169	1643	2444	3098	3637	4201	4909	3015

Table 4.17: Heart disease mortality rate (per 10^5 person-year) and frequency among US other females

	Period							
Age	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07	2008-10
20-24	2.6	1.8	1.8	1.1	1.5	1.2	1.5	1.4
	25	24	30	24	39	37	55	35
25-29	3.5	3.5	1.7	2.2	1.7	1.8	1.9	1.4
	33	50	33	51	47	62	76	37
30-34	5.1	5.0	2.9	2.7	3.0	2.5	2.5	2.3
	42	68	57	66	85	83	104	62
35-39	10.9	7.4	6.4	6.0	6.0	4.7	4.4	3.4
	72	76	105	133	166	155	169	92
40-44	20.0	15.6	11.1	8.7	9.8	7.9	8.7	7.5
	116	125	137	166	247	243	315	180
45-49	31.9	29.3	25.2	18.2	16.5	13.7	14.3	12.5
	157	197	231	245	339	376	482	288
50-54	56.0	48.1	48.4	41.2	37.0	29.4	24.4	20.5
	236	288	378	425	550	665	722	430
55-59	117.6	93.8	97.2	85.0	75.3	57.2	43.1	37.7
	377	462	671	739	852	900	1032	671
60-64	184.1	172.4	159.2	146.9	135.1	108.0	89.1	67.9
	468	661	927	1141	1300	1317	1482	935
65-69	278.1	285.2	273.2	262.8	235.0	202.7	159.8	120.4
	558	870	1216	1648	1934	2037	2033	1152
70-74	537.0	475.6	470.3	459.1	415.3	358.0	274.8	224.2
	800	1005	1496	2002	2575	2899	2748	1596
75-79	941.7	797.0	795.9	846.6	765.1	667.7	518.5	438.6
	982	1168	1641	2433	3039	3823	3908	2326
80-84	1401.0	1446.3	1522.9	1511.9	1506.8	1272.9	1053.3	869.6
	896	1258	1736	2310	3434	4302	5156	3188

4.2.3 Data Visualization

In the descriptive analysis of APC model, the rate is usually plotted against the age, period, or cohort to show how these three effects impact on the incidence or mortality rates of cancers for males and females. In this section, the lung and bronchus cancer incidence rate data was used. For the notation in the figures, the left end point of each age group was used to represent each age group, e.g., age 30 represented the age group 30-34, and the midpoint was used to denote each period group, like period 1975 representing the period group of 1973-1977.

Figures 4.2 and 4.3 present the age trend of the incidence rate of lung cancer by period in the original scale and the logarithmic scale, respectively. From Figure 4.2, the incidence rate showed rapid increased after the age of 45 for both males and females, then decreased later for old age. Also, the incidence rate for males along the age increased faster than females for all races, but there was not much difference of trends between males and females after taking the logarithmic transformation of the incidence rate from Figure 4.3. There were higher incidence rates along the age in males over early period than later period for all races. But it had the opposite effect for females. More females in the later period were diagnosed with lung cancer than early period. By comparing the age trend of incidence rate in original scale with the logarithmic scale, the curves were more parallel in the logarithmic scale. This indicated that the incidence rate was additive across different periods after taking the logarithmic transformation.

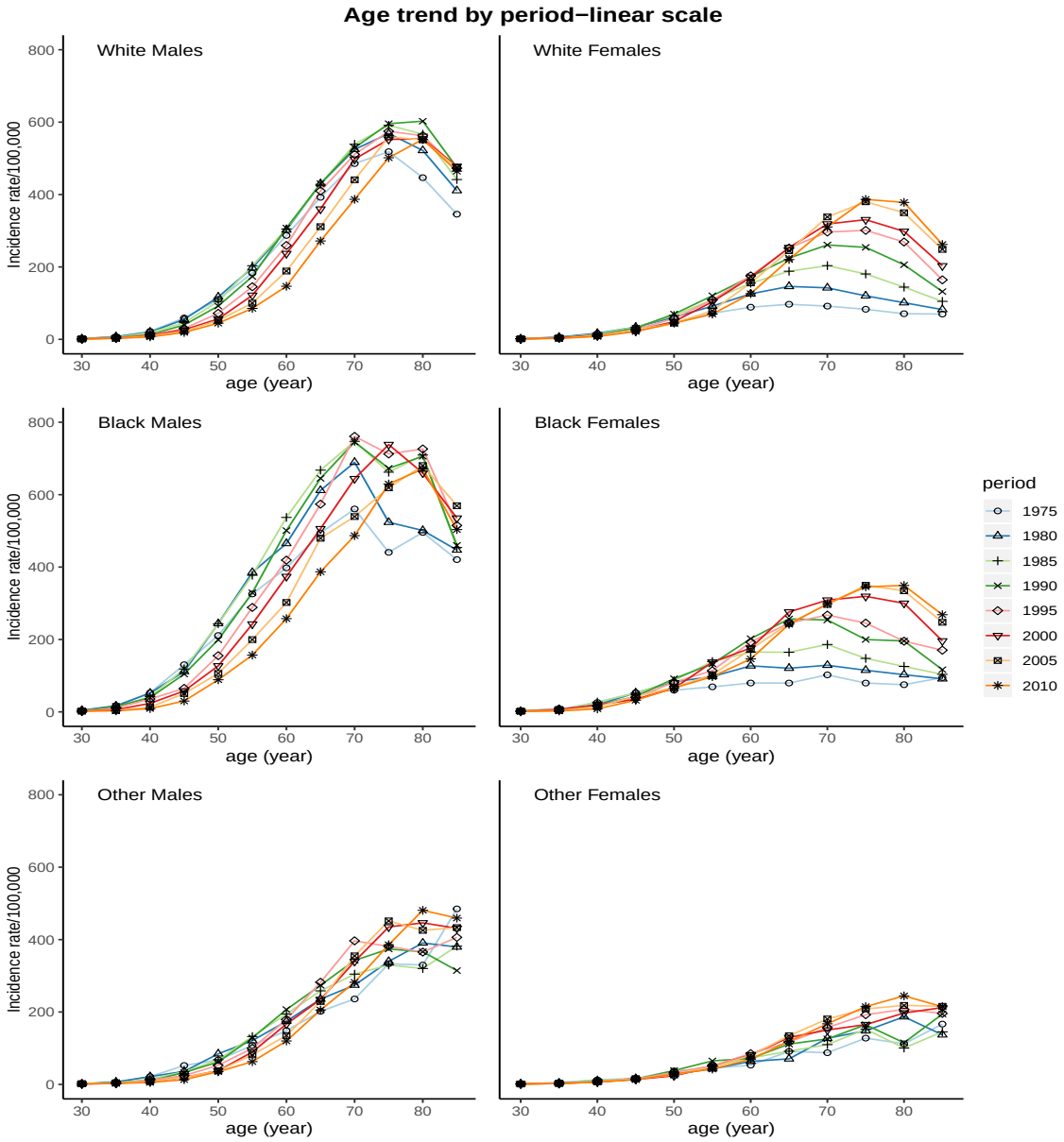

Figure 4.2: **Plot of lung cancer incidence rate among different race and gender against age by period**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012.

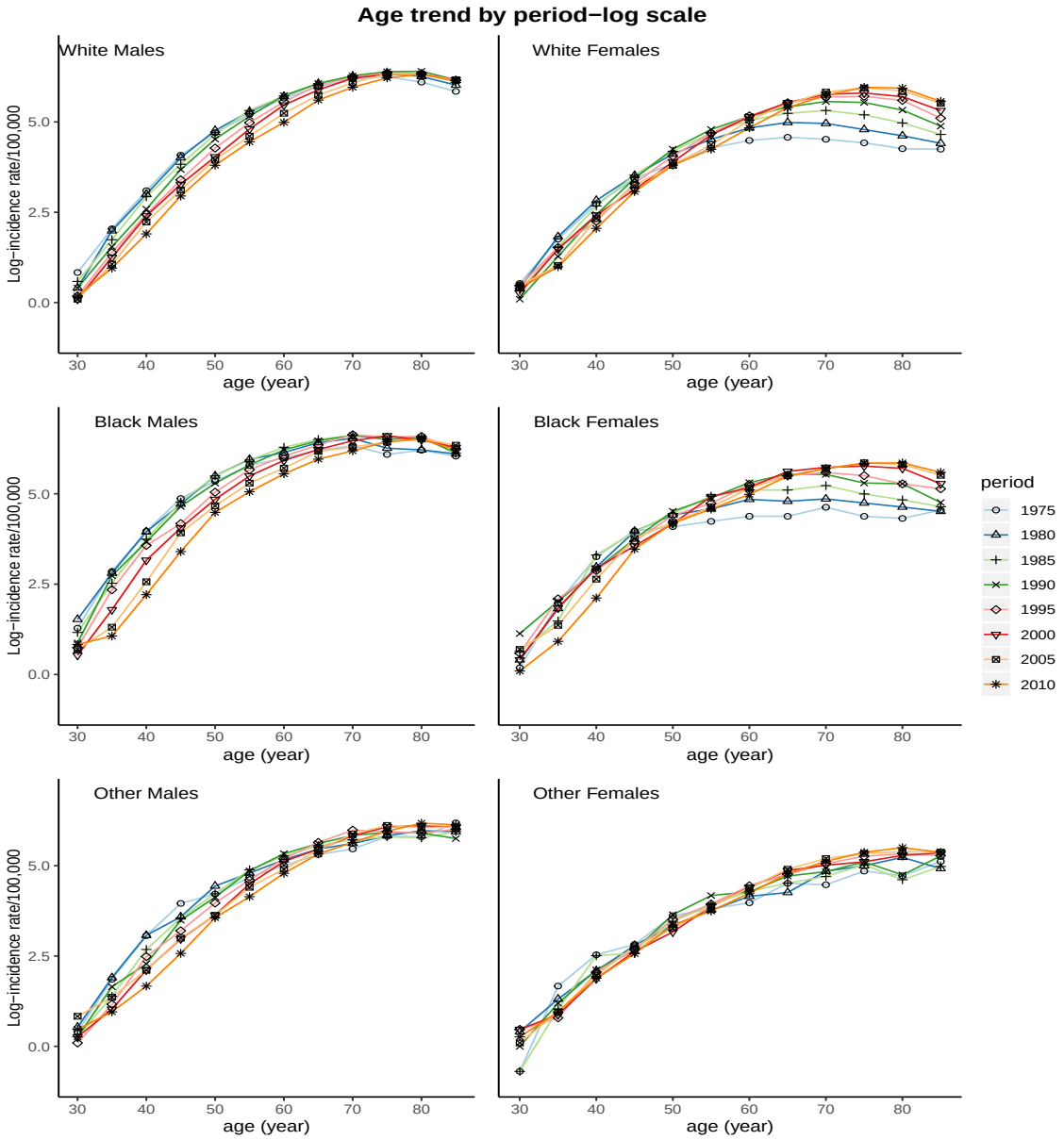

Figure 4.3: **Plot of lung cancer incidence rate in log scale among different race and gender against age by period**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012.

4.2.4 Hypothesis Testing on Equality of Age Trends Across Multiple Populations

In this section, we examined the age, period, and cohort effect on lung cancer incidence and heart disease mortality rate by plotting their trends with 95% confidence interval using intrinsic estimator. Then, we tested the equivalence of age and period trends among three racial groups for both males and females. Also tested the equality of age and period trends between males and females for each racial group.

The lung and bronchus cancer incidence rate data in US

Figures 4.4 and 4.5 show the trends of age, period, and cohort effects with results testing on the age and period trends among three racial groups (white, black, and other) for males and females. We rejected the null hypothesis on equivalence of age and period trends among races for both males and females based on the p-value of zero, which means that at least one of the age trends among three racial groups for both male and female differed, the same conclusion for period trend was reached.

The age trends increased faster from age 30s to peaks in old ages and then leveled off for different gender and race, indicating a concave pattern. The period effects on lung cancer incidence showed slow increasing trends for both gender and race except for black males. Females had a steeper increase than males, especially for white and black races as seen in Figures 4.6 and 4.7. The cohort effects showed an inverse-U shape that increased in older cohorts and decreased for the subsequent younger cohorts with tail leveled up a little bit, in which the decrease may due to the smoking cessation. Comparing the cohort trends between males and females, earlier cohort effect for male was higher than females,

and the pattern preceded that of females by two decades [29].

Figures 4.6, 4.7 and 4.8 show the trends of age, period, and cohort effects with results testing the age and period trends between males and females within the same racial group for white, black, and other, respectively. The age effects were higher before age 50 then lower after for females than males for all racial groups, but a clear difference in age trend between male and female was found in white people from Figure 4.6. However, period trend had the opposite effect in which it had a lower value before 1990 then higher value after 1990 for females than males. The smallest difference in age and period trends between males and females can be found in other racial groups because it had the largest p-value from Figure 4.8.

The heart disease mortality rate data in US

Figures 4.9 and 4.10 show the trends of age, period, and cohort effects with 95% confidence interval and testing statistic on the age and period trends among three racial groups (white, black, and other) for males and females, respectively. We rejected the null hypothesis on equivalence of age and period trends among races for both males and females based on the p-value of zero. Even though it looks like the period trend for these three racial groups overlapped, we still rejected the null hypothesis to conclude that at least one period trend differ because the 95% confidence interval was very narrow not like the lung cancer incidence data, which had a wider confidence interval.

The age trends increased faster from earlier adulthood to age of 45, then increased lower in older ages for males, while there was a linear increase for white and black females. The period effects on heart disease mortality showed similar kinds of flat trends for both gender and race. The trends of cohort effects decreased slowly along the cohort with a

relatively flat tail.

Figures 4.11, 4.12 and 4.13 show the trends of age, period, and cohort effects with testing on the age and period trends between males and females within the same racial group for white, black, and other, respectively. We rejected the null hypothesis of equal age trend between male and female for all racial groups. For period trend between females and males, we rejected the null hypothesis only for white and black, but we failed to reject the null hypothesis of equal period trend between males and females for other racial group, because the p-value was 0.35174 from Figure 4.13.

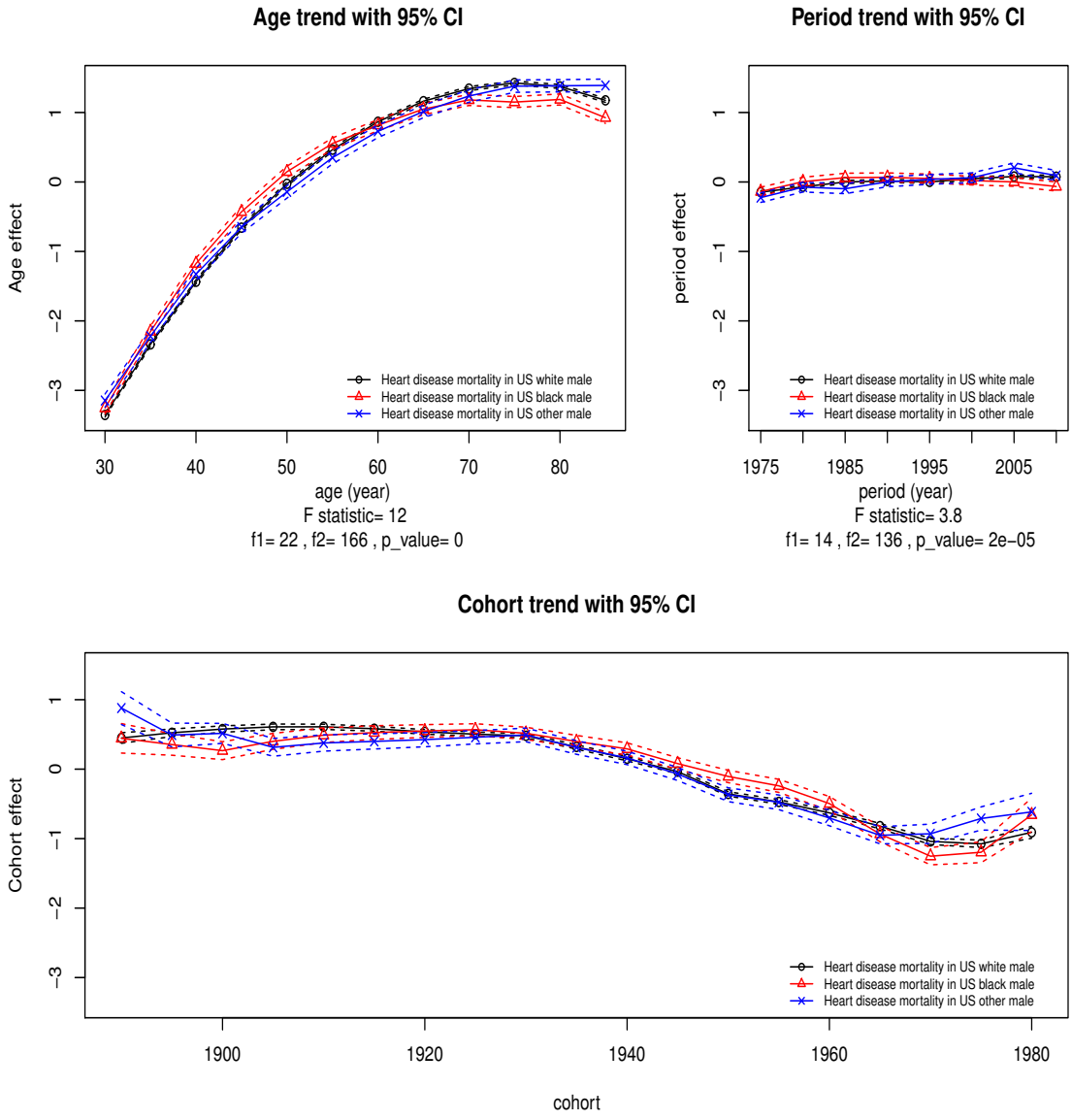

Figure 4.4: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed lung cancer incidence rate in US males among different races**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012, and cohort represent birth cohort groups of 9-years interval from 1884-1892, 1889-1897, to 1974-1982.

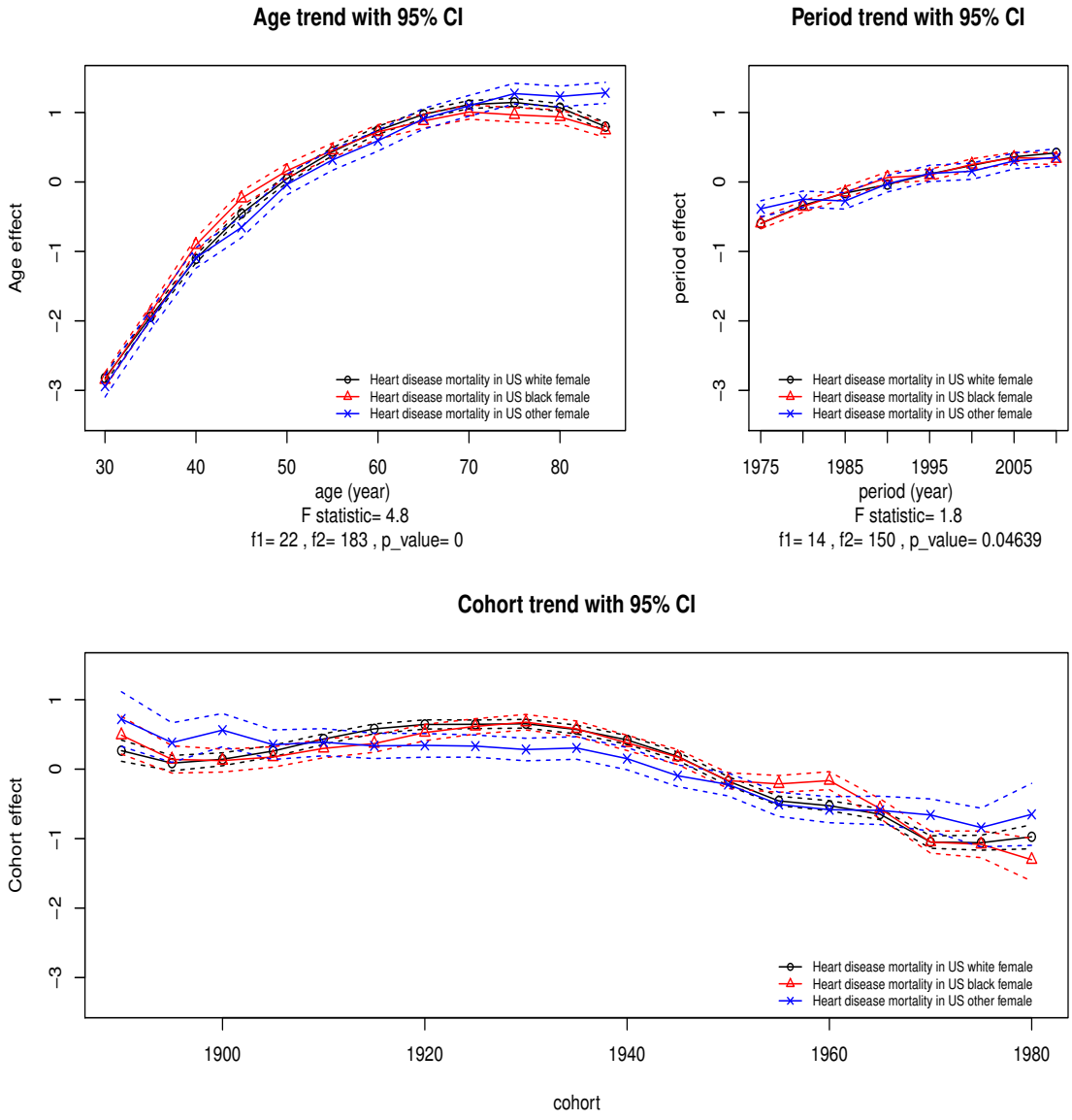

Figure 4.5: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed lung cancer incidence rate in US females among different races**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012, and cohort represent birth cohort groups of 9-years interval from 1884-1892, 1889-1897, to 1974-1982.

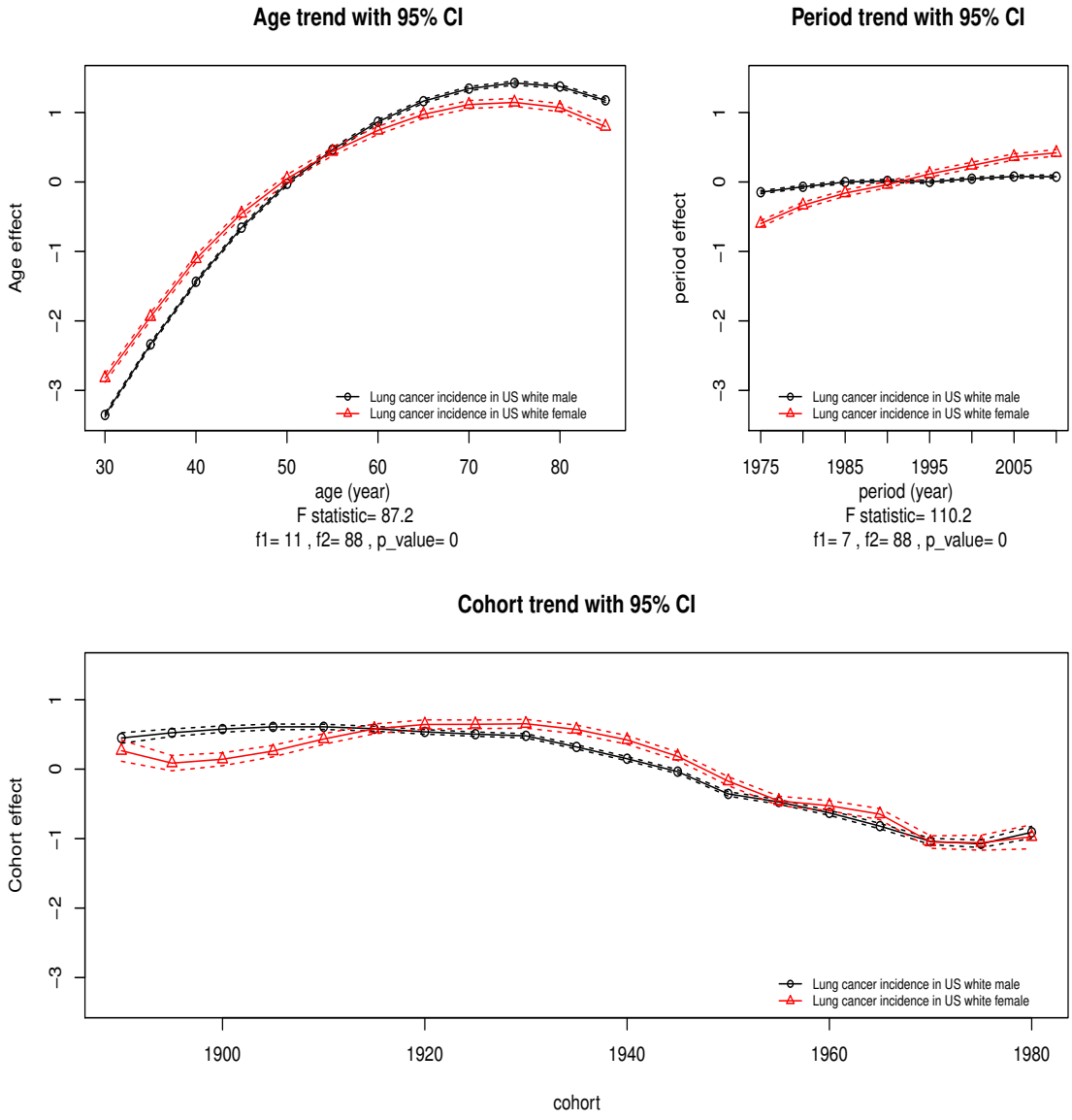

Figure 4.6: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed lung cancer incidence rate in US white males and females**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012, and cohort represent birth cohort groups of 9-years interval from 1884-1892, 1889-1897, to 1974-1982.

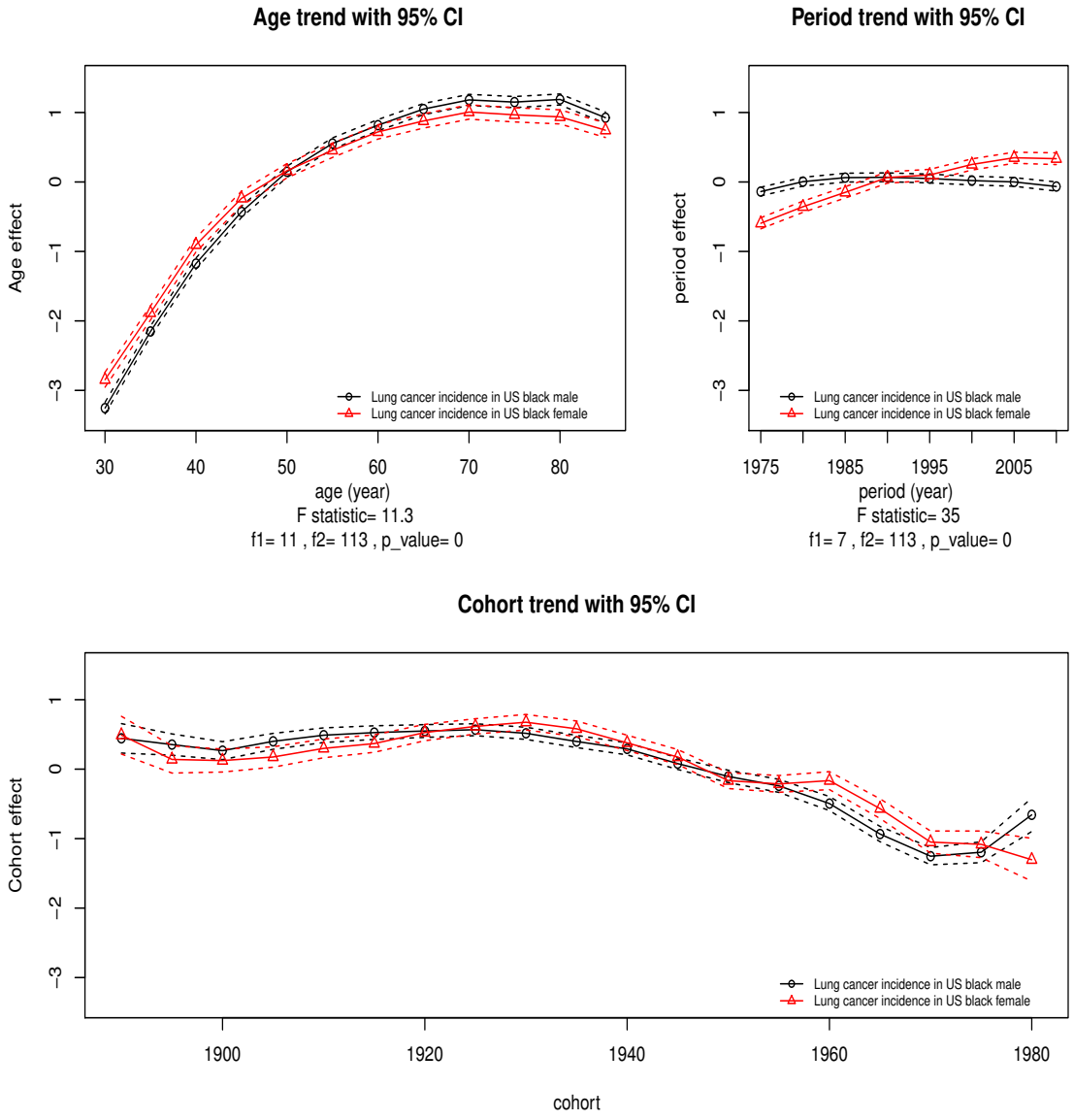

Figure 4.7: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed lung cancer incidence rate in US black males and females**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012, and cohort represent birth cohort groups of 9-years interval from 1884-1892, 1889-1897, to 1974-1982.

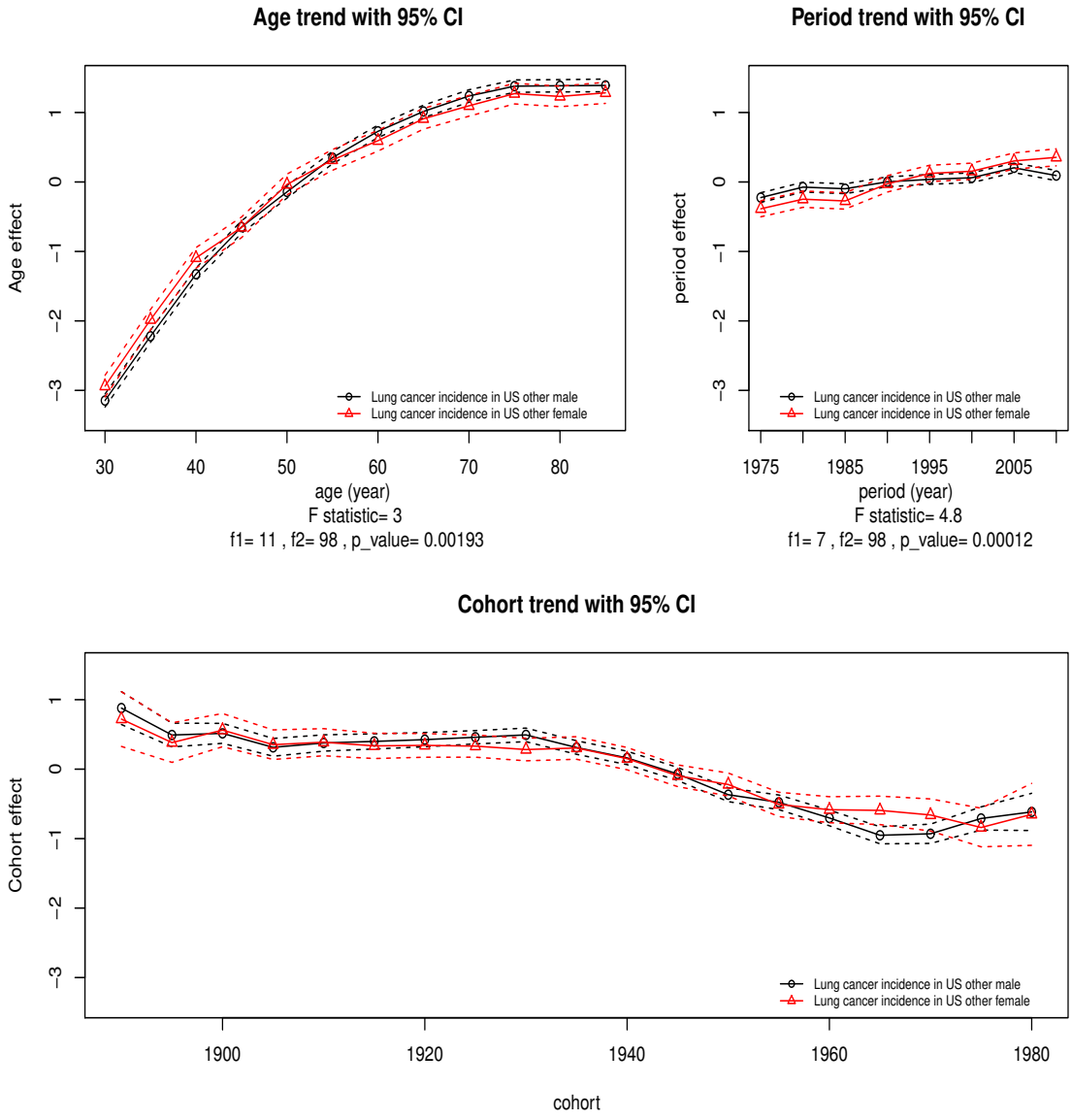

Figure 4.8: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed lung cancer incidence rate in US other males and females**. Age represent age groups of 5-years interval from 30-34, 35-39, to 85+, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2012, and cohort represent birth cohort groups of 9-years interval from 1884-1892, 1889-1897, to 1974-1982.

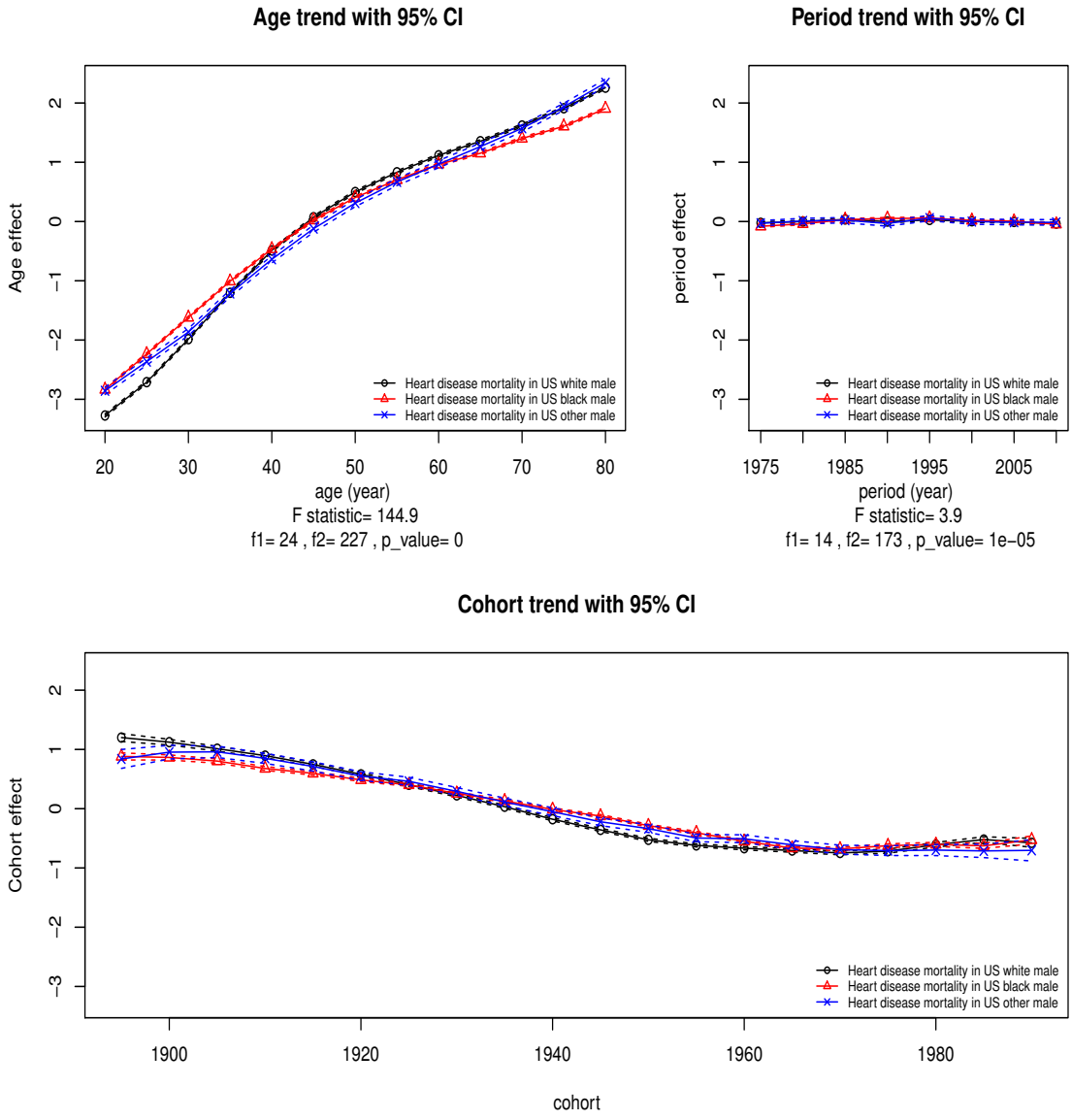

Figure 4.9: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed heart disease mortality in US males among different races**. Age represent age groups of 5-years interval from 20-24, 25-29, to 80-84, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2010, and cohort represent birth cohort groups of 9-years interval from 1889-1897, 1894-1902, to 1984-1992.

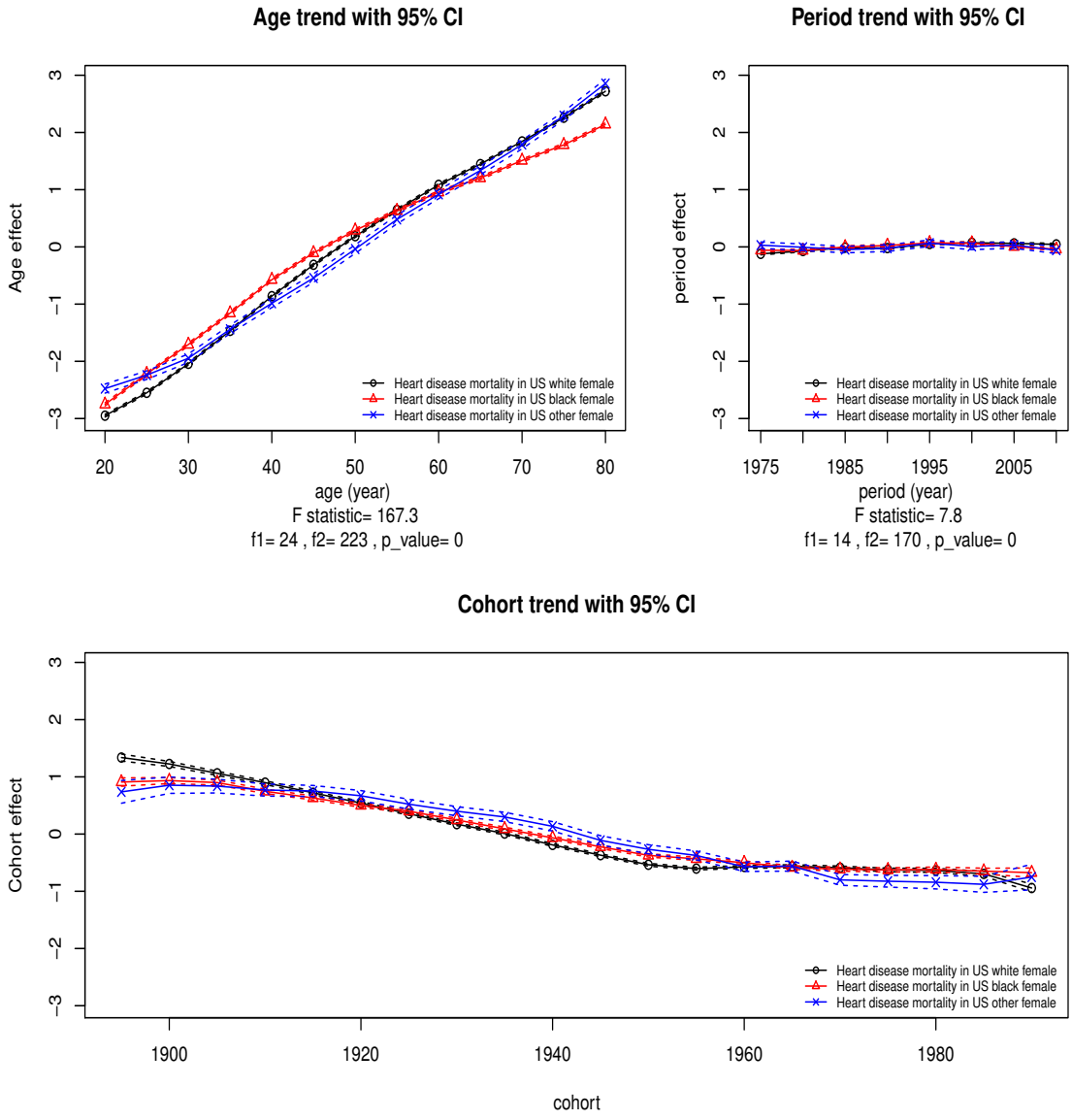

Figure 4.10: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed heart disease mortality rate in US females among different races**. Age represent age groups of 5-years interval from 20-24, 25-29, to 80-84, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2010, and cohort represent birth cohort groups of 9-years interval from 1889-1897, 1894-1902, to 1984-1992.

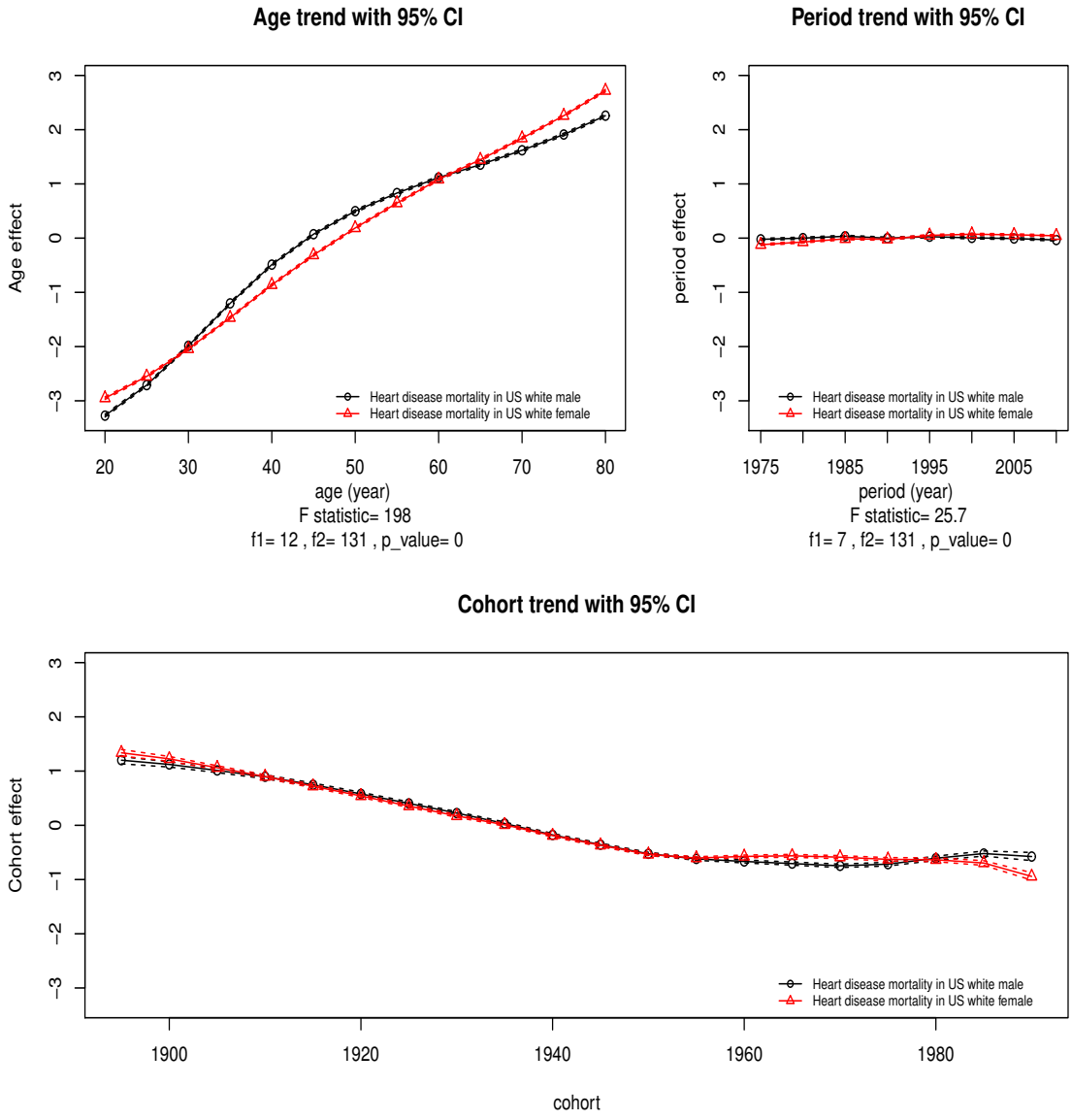

Figure 4.11: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed heart disease mortality rate in US white males and females**. Age represent age groups of 5-years interval from 20-24, 25-29, to 80-84, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2010, and cohort represent birth cohort groups of 9-years interval from 1889-1897, 1894-1902, to 1984-1992.

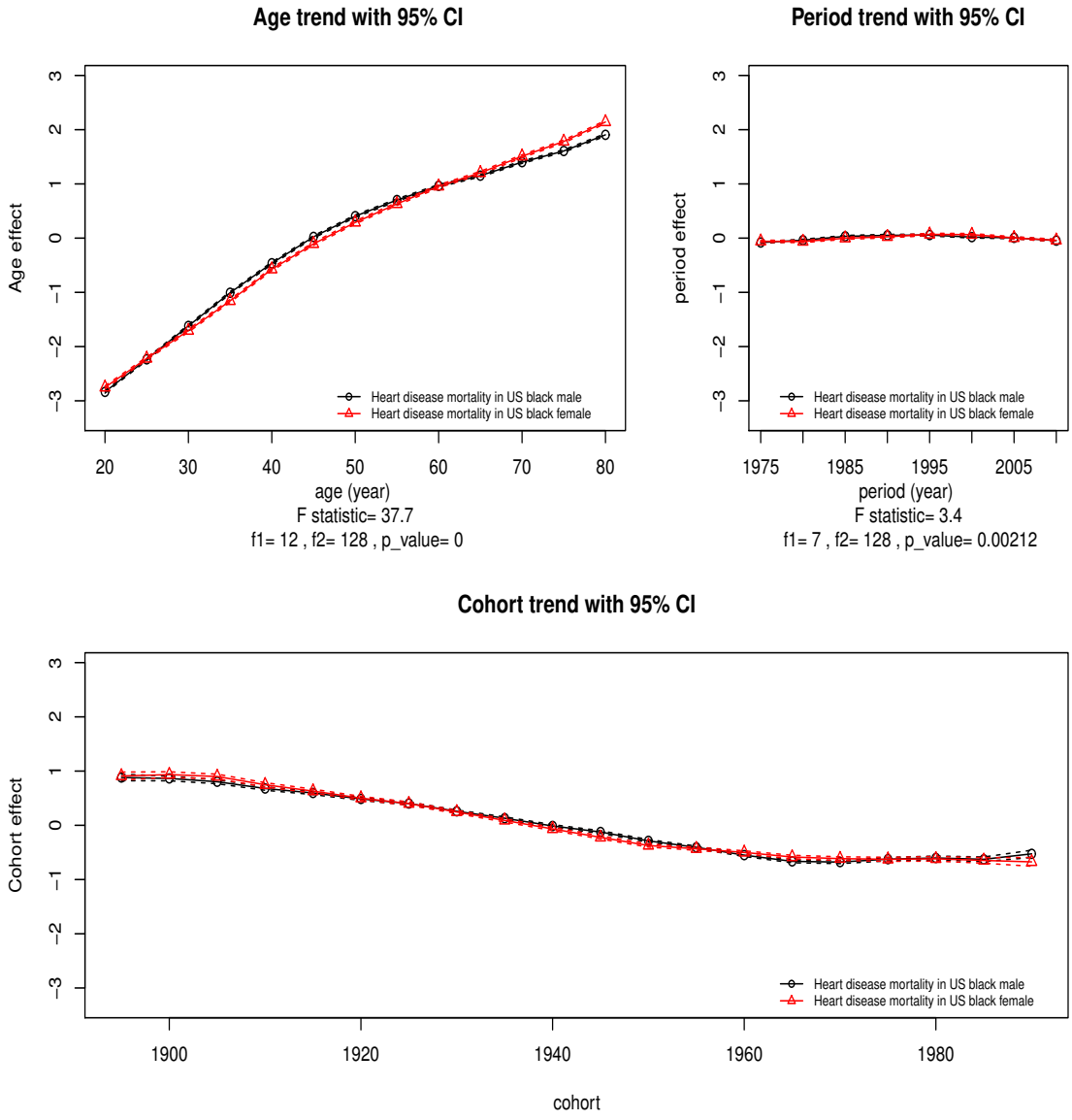

Figure 4.12: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed heart disease mortality rate in US black males and females**. Age represent age groups of 5-years interval from 20-24, 25-29, to 80-84, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2010, and cohort represent birth cohort groups of 9-years interval from 1889-1897, 1894-1902, to 1984-1992.

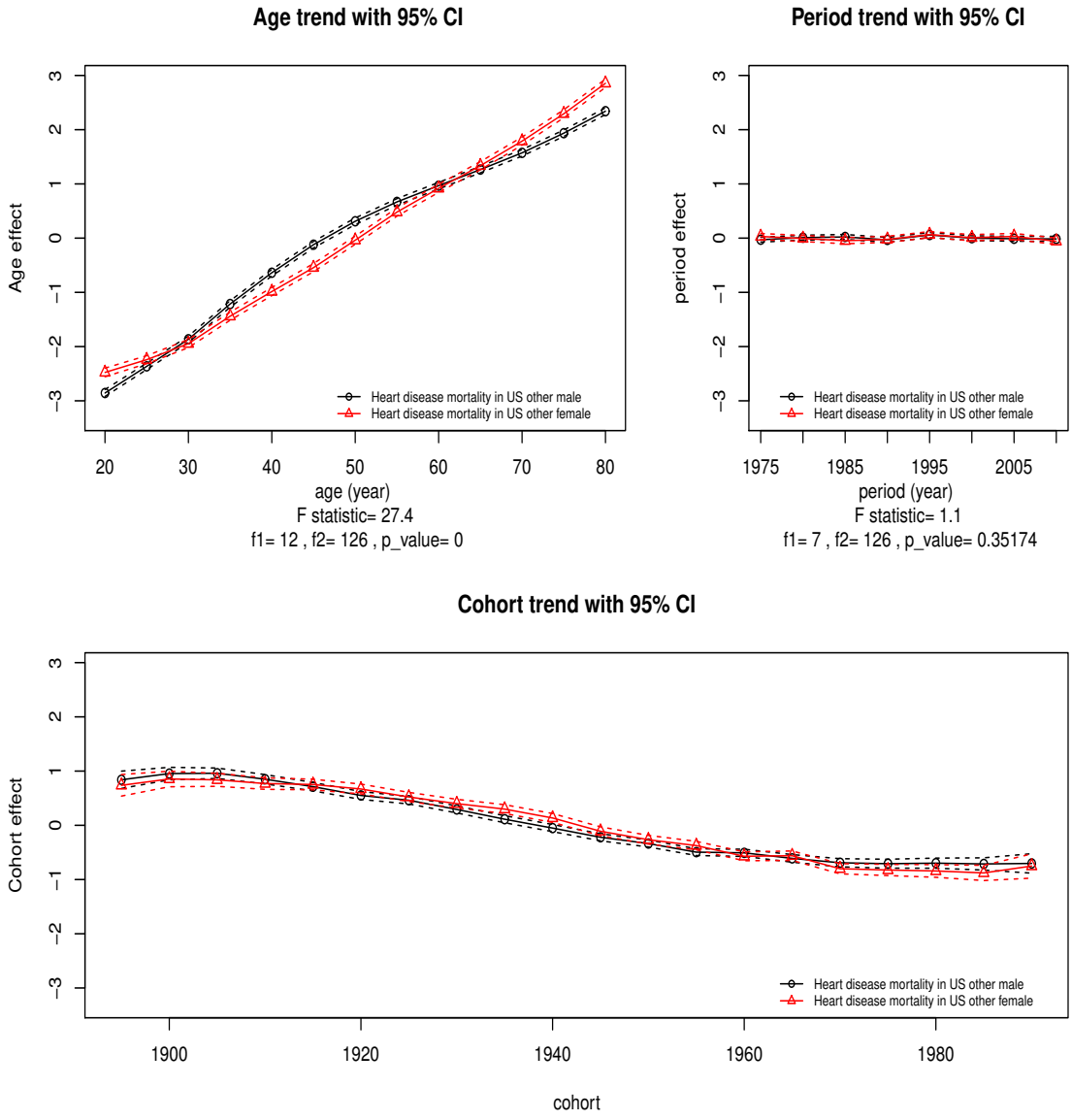

Figure 4.13: **Plot of age, period, and cohort trends by intrinsic estimator of log-transformed heart disease mortality rate in US other males and females**. Age represent age groups of 5-years interval from 20-24, 25-29, to 80-84, period represent period groups of 5-years interval from 1973-1977, 1978-1982, to 2008-2010, and cohort represent birth cohort groups of 9-years interval from 1889-1897, 1894-1902, to 1984-1992.

CHAPTER 5

Conclusion and Outlook

5.1 Conclusion

The age, period, and cohort (APC) models have served as a statistical tool over the past four decades in APC analysis, since it was proposed by Mason (1973). Due to the linear dependency among these three variables: $period - age = cohort$, the models suffered with the identification problem in APC analysis. In the context, many conventional approaches introduced in previous chapter of introduction have been proposed to address the identification problem, which often have limitations and yield biased estimates. In recent development of estimation, the intrinsic estimator proposed by Fu (2000) has very good

properties, such as unbiasedness, estimablity and later on being proved to be consistent [11]. In this dissertation, the parameter estimation of APC models in data analysis and simulation study are all based on the intrinsic estimator method.

In this dissertation, I addressed two issues of APC models in parameter estimation in Chapter 3. One is the selection of side condition on age, period, and cohort effects, even though we can get a unique intrinsic estimator by principal component analysis (PCA) method. Based on the statistical theory of efficient estimates, variances of parameter estimates by different side condition were compared by analytic analysis and simulation study in one-way and two-way ANOVA models first, because the APC model is a special case of two-way ANOVA model. We found that the centralization yielded smaller variance of estimates than setting reference levels. For APC models, we recommend to use the centralization for age and period duo to the balanced data design, and the centralization is still preferred on cohort effects by the simulation study. The other is to derive a more accurate variance estimate for period and cohort effects when fitted with a generalized linear model for an exponential family distributed response variable except for Gaussian case, in which the model default variance estimation by the PCA method was incorrect. The analytic expression of the variance-covariance matrix of period and cohort effects was derived with the Delta method, and was compared with the PCA method by the simulation study, in which the Delta method yielded smaller variance of estimates.

Chapter 4 derived a new statistic (the extended F test) to test on the equality of age trends across multiple populations when fitted with a linear APC model under the assumption of heteroscedasticity, especially with small number of columns of APC table. The formula of this test was given by extending the Welch's F test[47] in the univariate case to our vector case with theoretical justification. Also, the test was effective and powerfully

confirmed by the simulation study. Later on, the extended F test was applied to the lung cancer incidence rate and heart disease mortality rate data, which showed that there was significant difference among age trends across different race and gender.

5.2 Outlook

The selection of side conditions on age, period, and cohort effects in APC models and variance estimation on period and cohort effects have been fully discussed with theoretical justification and simulation. However, there is a limitation of the extended F test—it assumes that the dimension of table for each population has to be the same, i.e. same number of rows and same number of columns. The case that all population have same number of rows but different number of columns still needs further consideration. Tests on other exponential family distributions can be done following the likelihood ratio test approach. Overall, further investigation is needed to the testing procedure on trends for APC models.

Bibliography

[1] P. B. Baltes. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 11(3):145–171, 1968.

[2] W. F. Christensen and A. C. Rencher. A comparison of type I error rates and power levels for seven solutions to the multivariate behrens-fisher problem. *Communications in Statistics-Simulation and Computation*, 26(4):1251–1273, 1997.

[3] D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. II: age–period–cohort models. *Statistics in Medicine*, 6(4):469–481, 1987.

[4] R. Dorfman. A note on the delta-method for finding variance formulae. *The Biometric Bulletin*, 1(129-137):92, 1938.

[5] B. Everitt and A. Skrondal. *The Cambridge Dictionary of Statistics*, volume 106. Cambridge University Press Cambridge, 2002.

[6] S. E. Fienberg and W. M. Mason. Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology*, 10:1–67, 1979.

[7] R. A. Fisher. Statistical methods for research workers. In *Breakthroughs in Statistics*, pages 66–70. Springer, 1992.

[8] W. H. Frost. The age selection of mortality from tuberculosis in successive decades. *American Journal of Epidemiology*, 30(3):91–96, 1939.

[9] W. Fu. Ridge estimator in singular design with application to age-period-cohort analysis of disease rates. *Communications in Statistics-Theory and Methods*, 29(2):263–278, 2000.

[10] 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.

[11] 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.

[12] N. D. Glenn. Distinguishing age, period, and cohort effects. In *Handbook of the Life Course*, pages 465–476. Springer, 2003.

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

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

[15] T. R. Holford. Analysing the temporal effects of age, period, and cohort. *Statistical Methods in Medical Research*, 1(3):317–337, 1992.

[16] T. R. Holford, K. A. Cronin, A. B. Mariotto, and E. J. Feuer. Chapter 4: Changing patterns in breast cancer incidence trends. *Journal of the National Cancer Institute. Monographs*, 2006(36):19–25, 2006.

[17] H. Hotelling. The generalization of Student's ratio. *The Annals of Mathematical Statistics*, 2(3):360–378, 1931.

[18] G. S. James. The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 38(3/4):324–329, 1951.

[19] G. S. James. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, 41(1/2):19–43, 1954.

[20] R. I. Jennrich and P. Sampson. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18(1):11–17, 1976.

[21] S. Johansen. The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, 67(1):85–92, 1980.

[22] M. G. Kendall, A. Stuart, and J. K. Ord. *Kendall's Advanced Theory of Statistics*, volume 1. Oxford University Press, 1987.

[23] K. M. Keyes, R. Nicholson, J. Kinley, S. Raposo, M. B. Stein, E. M. Goldner, and J. Sareen. Age, period, and cohort effects in psychological distress in the united states and canada. *American Journal of Epidemiology*, 179(10):1216–1227, 2014.

[24] D. Knoke and M. Hout. Social and demographic factors in american political party affiliations. *American Sociological Review*, 39(5):700–713, 1974.

[25] L. L. Kupper, J. M. Janis, A. Karmous, and B. G. Greenberg. Statistical age-period-cohort analysis: a review and critique. *Journal of Chronic Diseases*, 38(10):811–830, 1985.

[26] L. L. Kupper, J. M. Janis, I. A. Salama, C. N. Yoshizawa, B. G. Greenberg, and H. Winsborough. 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):201–217, 1983.

[27] P. W. Lavori, G. L. Klerman, M. B. Keller, T. Reich, J. Rice, and J. Endicott. Age-period-cohort analysis of secular trends in onset of major depression: findings in siblings of patients with major affective disorder. *Journal of Psychiatric Research*, 21(1):23–35, 1987.

[28] N. T. Longford. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4):817–827, 1987.

[29] A. D. Lopez. *Adult Mortality in Developed Countries: from Description to Explanation*. Clarendon Press; Oxford University Press Oxford: New York, 1995.

[30] L. Luo. Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, 50(6):1945–1967, 2013.

[31] L. Luo, J. Hodges, C. Winship, and D. Powers. The sensitivity of the intrinsic estimator to coding schemes: comment on Yang, Schulhofer-Wohl, Fu, and Land. *American Journal of Sociology*, 122(3):930–961, 2016.

[32] K. O. Mason, W. M. Mason, H. H. Winsborough, and W. K. Poole. Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 38(2):242–258, 1973.

[33] P. McCullagh and J. Nelder. *Generalized Linear Models, Second Edition*. Taylor & Francis, 1989.

[34] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 1972.

[35] C. Osmond and M. Gardner. Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine*, 1(3):245–259, 1982.

[36] J. Rice. *Mathematical Statistics and Data Analysis*. Nelson Education, 2006.

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

[38] C. Robertson and P. Boyle. Age-period-cohort analysis of chronic disease rates I: modelling approach. *Statistics in Medicine*, 17(12):1305–1323, 1998.

[39] 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.

[40] W. L. Rodgers. Estimable functions of age, period, and cohort effects. *American Sociological Review*, 47(6):774–787, 1982.

[41] A. Sen and M. Srivastava. *Regression Analysis: Theory, Methods, and Applications*. Springer Science & Business Media, 2012.

[42] R. G. Stevens, S. H. Moolgavkar, and J. A. Lee. Temporal trends in breast cancer. *American Journal of Epidemiology*, 115(5):759–777, 1982.

[43] R. E. Tarone and K. C. Chu. Implications of birth cohort patterns in interpreting trends in breast cancer rates. *Journal of the National Cancer Institute*, 84(18):1402–1410, 1992.

[44] J. M. Ver Hoef. Who invented the delta method? *The American Statistician*, 66(2):124–127, 2012.

[45] J. Wardle, K. Robb, S. Vernon, and J. Waller. Screening for prevention and early diagnosis of cancer. *American Psychologist*, 70(2):119–133, 2015.

[46] B. L. Welch. The generalization of Student's problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947.

[47] B. L. Welch. On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4):330–336, 1951.

[48] H. H. Winsborough. *Age, Period, Cohort, and Education Effects on Earnings by Race*. University of Wisconsin, Institute for Research on Poverty, 1975.

[49] Y. Yang and K. C. Land. A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, 36(1):75–97, 2006.

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

[51] Y. Yao. An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika*, 52(1/2):139–147, 1965.