International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 9(Special Issue), 2009, pp 317-334

Estimating Binary Logit Regression Cohort Model for Analyzing Contraceptive Use Data of Repeated Surveys

Syeda Shahanara Huq
Department of Statistics
Jahangirnagar University
Dhaka-1342, Bangladesh

Makio Ishiguro
Institute of Statistical Mathematics
4-6-7 Minami-Azabu
Minato-ku, Tokyo 106, Japan

Md. Moshiur Rahman[1]
Department of Statistics
Jahangirnagar University
Dhaka-1342, Bangladesh

[Received January 4, 2007; Revised April 12, 2009; Accepted April 27, 2009]

Abstract

This study utilizes a Bayesian binary logit regression cohort model for explicating the contraceptive use dynamics of Bangladesh. Cohort analysis encounters an identification problem in age, period and cohort effects which appear difficult to decompose uniquely that has been got around through Bayesian smoothness priors approach. The marginal log-likelihood computation furnishes estimates of hyperparameters and maximizing penalized log-likelihood using a numerical optimization technique receives emphatic attention for prudent maximum a posteriori (MAP) estimates of the effect parameters of the model. The optimal model is selected appraising Akaike's Bayesian information criterion (ABIC) from a number of candidate models. Monte Carlo simulation studies are carried out to judge the performance of the method for estimating the cohort models. An application to contraceptive use data reveals that age, period, and cohort, and the instantaneous influence of important covariates along with the interaction effects consistently run toward different directions.

Keywords and Phrases: Bayesian Method, Smoothness Priors, Penalized Log-likelihood, ABIC, MAP Estimate, Numerical Optimization.

AMS Classification: Primary 62J02; Secondary 62J20.

[1]At present: *SAMHDA*, ICPSR, ISR, University of Michigan, Ann Arbor, MI 48106, USA.

1 Introduction

Cohort is an important concept in the study of social change. Data of cohort table are constituted from a set of repeated surveys classified by age and survey period. Cohort analysis is a method to investigate age, period (calender year), and cohort (birth time) effects and is widely utilized in the fields of demography, sociology, political science (Ryder, 1965; Glenn, 1977; Mason and Fienberg, 1985; Nakamura, 1986, 2002; Hayashi et al.,1992; Miller and Nakamura, 1996, 1997 and Sakamoto, 1999 among many). In this data, an exact equality condition exists among cohort, period, and age variables. So the partition of the variation in the response variable into the effects of age, period, and cohort in a linear functional form can not be unique. This is known as identification problem in cohort analysis, and modus operandi to deal with the problem have been independently discovered a number of times in different research fields (Schaie, 1965; Mason et al., 1973; Osmond and Gardner, 1982; Nakamura, 1986, 1996, 2002; and Robertson and Boyle, 1998). In statistical terms, cohort models are under-identified since there are more parameters to be estimated than there are degrees of freedom. Additional assumptions are, therefore, needed, and these assumptions should be hinged on some prior knowledge of the variables under study. Warshaw (1992) suggested that studying three dimensional graphs of a variable over time can provide information on which assumptions are appropriate for solving the identification problem. Nakamura (1986) proposed a Bayesian logit cohort model for grouped data and solved this problem by imposing smoothness constraints on the age, period and cohort effects. He assumed that the first order difference of the successive parameters for age, period, and cohort are close to zero. Akaike's (1980) Bayesian Information Criterion (ABIC) for model selection as well as estimation of the parameters of the prior distributions maintains the best balance between them.

Here an attempt has been made to analyze the binary cohort data accentuating on a new Bayesian model with a slightly modified computational procedure of Nakamura (1986). The layout design of the paper is as follows: Section 2 portrays the binary logit cohort regression model along with relevant notations and terminologies. A brief discussion on the identification problem of cohort analysis and making use of gradually changing parameter assumption in order to surmount the problem of decomposition of age, period, and cohort effects is an important focus of Section 3. The penalized log-likelihood of the Bayesian model and the maximum a posteriori (MAP) estimate of its parameters are also discussed here at large. Section 4 presents the numerical procedure for computation of the log-likelihood of the model and a technique for estimating starting values of the parameters of the log-likelihood computation. To check the effectiveness of the method, the results of simulation studies for two specific model cases are juxtaposed and analyzed in section 5. Reduction of high growth of population is a major concern of Bangladesh for which birth control by adopting contraception is the pivotal factor. Section 6 attempts to delineate the results of a real world application of proposed method to demographic data on individual level contraceptive use among women. Section 7 sums up with some concluding remarks.

2 Binary Logit Cohort Model

2.1 Development of the Model

Let a data set for several consecutive surveys be denoted by $\{(y_n, a_n, p_n, \boldsymbol{x}_n)|n = 1, 2, \ldots, N\}$ where N is the total sample size, y_n is a dichotomous response variable related to the respondent, y_n taking values 1 or 0, a_n is his/her age, p_n is the date of his/her interview or the survey year[2], and $\boldsymbol{x}_n = (x_{1n}, \ldots, x_{mn})^T$ is the vector of m covariates. Assuming that y_n takes value 1 with the probability $\Pr(y_n = 1) = \pi_n,\ 0 < \pi_n < 1$, a binary logit regression cohort model is defined by

$$\pi_n = \frac{\exp \eta(a_n, p_n, \boldsymbol{x}_n)}{1 + \exp \eta(a_n, p_n, \boldsymbol{x}_n)} \tag{2.1}$$

or equivalently

$$\log \frac{\pi_n}{1 - \pi_n} = \eta(a_n, p_n, \boldsymbol{x}_n) \tag{2.2}$$

with a structure $\eta(a, p, \boldsymbol{x}) = g_0(\boldsymbol{x}) + g_A(a, \boldsymbol{x}) + g_P(p, \boldsymbol{x}) + g_C(c, \boldsymbol{x})$ where $c\ (= p - a)$ is the birth time of a respondent (hereafter the suffix n is omitted). Subscripts A, P, and C indicate age, period, and cohort, respectively.

Supposing that the surveys have been conducted J times at Δ-year intervals and each survey was completed within a day of a certain year, the survey year p takes J different values $\{P_1, \ldots, P_J\}$ such that $P_{\tilde{p}} = P_1 + (\tilde{p} - 1)\Delta$, $\tilde{p} = 1, \ldots, J$ where P_1 is the earliest survey year. Thus age groups $A_{\tilde{a}}$ can be defined as intervals $A_{\tilde{a}} = [a^*_{\tilde{a}},\ a^*_{\tilde{a}+1})$, $a^*_{\tilde{a}} = a^*_1 + (\tilde{a} - 1)\Delta$, $\tilde{a} = 1, \ldots, I$ where a^*_1 is the age of the youngest respondents or suitably chosen lowest age and I is the number of age groups. Similarly, cohort groups $C_{\tilde{c}}$ are represented as intervals $C_{\tilde{c}} = (c^*_{\tilde{c}},\ c^*_{\tilde{c}+1}]$, $c^*_{\tilde{c}} = P_{\tilde{p}} - a^*_{\tilde{a}}$, $\tilde{c} = I - \tilde{a} + \tilde{p} = 1, \ldots, K$ where $K (= I + J - 1)$ is the number of cohort groups. Using the above notations, the functions g's are specified as beneath

$$\begin{aligned}
g_0(\boldsymbol{x}) &= \beta_0 + \sum_{j=1}^{m_0} \beta_{0j} x_j, \\
g_A(a, \boldsymbol{x}) &= \sum_{\tilde{a}}^{I} v_A(\tilde{a}, \boldsymbol{x}) \delta^A(a, \tilde{a}), \qquad \delta^A(a, \tilde{a}) = 1 \text{ if } a \in A_{\tilde{a}}, \text{ and 0 otherwise}, \\
g_P(p, \boldsymbol{x}) &= \sum_{\tilde{p}}^{J} v_P(\tilde{p}, \boldsymbol{x}) \delta^P(p, \tilde{p}), \qquad \delta^P(p, \tilde{p}) = 1 \text{ if } p = P_{\tilde{p}}, \text{ and 0 otherwise}, \\
g_C(c, \boldsymbol{x}) &= \sum_{\tilde{c}}^{K} v_C(\tilde{c}, \boldsymbol{x}) \delta^C(c, \tilde{c}), \qquad \delta^C(c, \tilde{c}) = 1 \text{ if } c \in C_{\tilde{c}}, \text{ and 0 otherwise},
\end{aligned}$$

[2] a_n and p_n are not necessarily measured in years but they could be in months or in days.

where functions $v_A(\tilde{a}, \boldsymbol{x})$, $v_P(\tilde{p}, \boldsymbol{x})$, and $v_C(\tilde{c}, \boldsymbol{x})$ are further expressed by $v_A(\tilde{a}, \boldsymbol{x}) = \beta^A_{\tilde{a}} + \sum_j^{m_A} \beta^A_{\tilde{a}j} x_j, v_P(\tilde{p}, \boldsymbol{x}) = \beta^P_{\tilde{p}} + \sum_j^{m_P} \beta^P_{\tilde{p}j} x_j$, and $v_C(\tilde{c}, \boldsymbol{x}) = \beta^C_{\tilde{c}} + \sum_j^{m_C} \beta^C_{\tilde{c}j} x_j$. β_0 is the grand mean; $\beta^A_{\tilde{a}}$, $\beta^P_{\tilde{p}}$, and $\beta^C_{\tilde{c}}$, the main effects due to age, period, and cohort, respectively. β_{0j} is the main effect of the j-th covariate, and β^A's, β^P's, β^C's with double subscripts are the interaction effects of age-by-covariate, period-by-covariate, and cohort-by-covariate, respectively. For example, $\beta^A_{\tilde{a}j}$ is the interaction effect between the $\tilde{a}$-th age group and the j-th covariate. $m_0 \leq m$ and m_A, m_P and m_C are suitable number of variables included in the model for covariates, age, period, and cohort respectively. Parameters β's are subject to the following zero-sum constraints: $\sum_{\tilde{a}=1}^{I} \beta^A_{\tilde{a}} = \sum_{\tilde{p}=1}^{J} \beta^P_{\tilde{p}} = \sum_{\tilde{c}=1}^{K} \beta^C_{\tilde{c}} = 0$, and $\sum_{\tilde{a}=1}^{I} \beta^A_{\tilde{a}j} = \sum_{\tilde{p}=1}^{J} \beta^P_{\tilde{p}j} = \sum_{\tilde{c}=1}^{K} \beta^C_{\tilde{c}j} = 0$. Model (2.2) named G$\{\boldsymbol{x}\}A\{\boldsymbol{x}\}P\{\boldsymbol{x}\}C\{\boldsymbol{x}\}$ becomes an ordinary age-period-cohort (GAPC) model if y_n does not depend on $\boldsymbol{x}_n$:

$$\log \frac{\pi_n}{1 - \pi_n} = \eta(a_n, p_n) = \beta_0 + \sum_{\tilde{a}}^{I} \beta^A_{\tilde{a}} \delta^A(a_n, \tilde{a}) + \sum_{\tilde{p}}^{J} \beta^P_{\tilde{p}} \delta^P(p_n, \tilde{p}) + \sum_{\tilde{c}}^{K} \beta^C_{\tilde{c}} \delta^C(c_n, \tilde{c}). \tag{2.3}$$

Considering the zero-sum constraints mentioned above, the main and interaction effects are expressed by vector notation as $\boldsymbol{\beta}_0 = (\beta_0, \beta_{01}, \ldots, \beta_{0m_0})^T$, $\boldsymbol{\beta}^A = (\beta^A_1, \ldots, \beta^A_{I-1})^T$, $\boldsymbol{\beta}^A_j = (\beta^A_{1j}, \ldots, \beta^A_{I-1,j})^T$, $j = 1, \ldots, m_A$; $\boldsymbol{\beta}^A_\star = ((\boldsymbol{\beta}^A_1)^T, \ldots, (\boldsymbol{\beta}^A_{m_A})^T)^T$, $\boldsymbol{\beta}^P = (\beta^P_1, \ldots, \beta^P_{J-1})^T, \boldsymbol{\beta}^P_j = (\beta^P_{1j}, \ldots, \beta^P_{J-1,j})^T$, $j = 1, \ldots, m_P; \boldsymbol{\beta}^P_\star = ((\boldsymbol{\beta}^P_1)^T, \ldots, (\boldsymbol{\beta}^P_{m_P})^T)^T$, $\boldsymbol{\beta}^C = (\beta^C_1, \ldots, \beta^C_{K-1})^T, \boldsymbol{\beta}^C_j = (\beta^C_{1j}, \ldots, \beta^C_{K-1,j})^T$, $j = 1, \ldots, m_C; \boldsymbol{\beta}^C_\star = ((\boldsymbol{\beta}^C_1)^T, \ldots, (\boldsymbol{\beta}^C_{m_C})^T)^T$, where ^{T} indicates the vector/matrix transpose. Further let $\boldsymbol{\beta} = (\boldsymbol{\beta}^T_0, \boldsymbol{\beta}^T_*)^T$, $\boldsymbol{\beta}_* = ((\boldsymbol{\beta}^A)^T, (\boldsymbol{\beta}^A_\star)^T, (\boldsymbol{\beta}^P)^T, (\boldsymbol{\beta}^P_\star)^T, (\boldsymbol{\beta}^C)^T, (\boldsymbol{\beta}^C_\star)^T)^T$. It is important to that $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_*$ are $(1 + m_0 + M)$ and M dimensional vectors, respectively, where $M = (I - 1) + m_A(I - 1) + (J - 1) + m_P(J - 1) + (K - 1) + m_C(K - 1)$.

Let $\boldsymbol{d}_n$ be a row vector of order $(1 + m_0 + M)$ for the n-th sample respondent,

$$\boldsymbol{d}_n = (1, \boldsymbol{x}^T_n, \boldsymbol{\delta}^A_n, \boldsymbol{x}^T_n \otimes \boldsymbol{\delta}^A_n, \boldsymbol{\delta}^P_n, \boldsymbol{x}^T_n \otimes \boldsymbol{\delta}^P_n, \boldsymbol{\delta}^C_n, \boldsymbol{x}^T_n \otimes \boldsymbol{\delta}^C_n),$$

which is the n-th row of the complete design matrix, $\boldsymbol{X}$. The operator $\otimes$ indicates the Kronecker product. The row vectors $\boldsymbol{\delta}^A_n$, $\boldsymbol{\delta}^P_n$, and $\boldsymbol{\delta}^C_n$ take the forms as $\boldsymbol{\delta}^A_n = (\delta^A(a_n, 1), \ldots, \delta^A(a_n, I - 1))$ if $a_n \notin A_I$; $\boldsymbol{\delta}^A_n = -\boldsymbol{1}^T_{I-1}$ if $a_n \in A_I$. $\boldsymbol{\delta}^P_n = (\delta^P(p_n, 1), \ldots, \delta^P(p_n, J - 1))$ if $p_n \neq P_J$; $\boldsymbol{\delta}^P_n = -\boldsymbol{1}^T_{J-1}$, otherwise. $\boldsymbol{\delta}^C_n = (\delta^C(c_n, 1), \ldots, \delta^C(c_n, K - 1))$ if $c_n \notin C_K$; $\boldsymbol{\delta}^C_n = -\boldsymbol{1}^T_{K-1}$, elsewhere. $\boldsymbol{1}_u$ is a u-dimensional unit vector, i.e., $\boldsymbol{1}_u = (\underbrace{1, \ldots, 1}_{u})^T$.

2.2 Identification Problem

For an arbitrary real number Δ, let $\gamma_0 = \beta_0 - \{I - (r_A - r_P + r_C)\}\Delta$, $\gamma^A_{\tilde{a}} = \beta^A_{\tilde{a}} + (\tilde{a} - r_A)\Delta$, $\gamma^P_{\tilde{p}} = \beta^P_{\tilde{p}} - (\tilde{p} - r_P)\Delta$, $\gamma^C_{\tilde{c}} = \beta^C_{\tilde{c}} + (\tilde{c} - r_C)\Delta$, where r_A, r_P, and r_C are suitably chosen so that γ's satisfy the zero sum constraints. Then

$$\begin{aligned}
\gamma_0 + \gamma^A_{\tilde{a}} + \gamma^P_{\tilde{p}} + \gamma^C_{\tilde{c}} &= \gamma_0 + \beta^A_{\tilde{a}} + \beta^P_{\tilde{p}} + \beta^C_{\tilde{c}} + \{\tilde{a} - \tilde{p} + \tilde{c} - (r_A - r_P + r_C)\}\Delta \\
&= \beta_0 + \beta^A_{\tilde{a}} + \beta^P_{\tilde{p}} + \beta^C_{\tilde{c}} \quad (\text{since} \quad I = \tilde{a} - \tilde{p} + \tilde{c}) \\
&= \eta(a_n, p_n),
\end{aligned}$$

which proves that the model (2.3) has infinite number of decompositions and similar reasons apply to model (2.2). This is known as identification problem in cohort analysis. Taking the sum of squares of the first-order differences of the effect parameters of each factor gives

$$\begin{aligned}
\sum_{\tilde{a}=1}^{I-1}(\gamma^A_{\tilde{a}} - \gamma^A_{\tilde{a}+1})^2 &= \sum_{\tilde{a}=1}^{I-1}(\beta^A_{\tilde{a}} - \beta^A_{\tilde{a}+1})^2 - 2(\beta^A_1 - \beta^A_I)\Delta + \Delta^2, \\
\sum_{\tilde{p}=1}^{J-1}(\gamma^P_{\tilde{p}} - \gamma^P_{\tilde{p}+1})^2 &= \sum_{\tilde{p}=1}^{J-1}(\beta^P_{\tilde{p}} - \beta^P_{\tilde{p}+1})^2 - 2(\beta^P_1 - \beta^P_J)\Delta + \Delta^2, \\
\sum_{\tilde{c}=1}^{K-1}(\gamma^C_{\tilde{c}} - \gamma^C_{\tilde{c}+1})^2 &= \sum_{\tilde{c}=1}^{K-1}(\beta^C_{\tilde{c}} - \beta^C_{\tilde{c}+1})^2 - 2(\beta^C_1 - \beta^C_K)\Delta + \Delta^2,
\end{aligned}$$

which suggests that minimizing these sums is the key to get a parsimonious decomposition and to overcome the identification problem.

2.3 Log-Likelihood Function and Smoothing Priors

Since y_n is a binary random variable, the log-likelihood of the model (2.1) is given by

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \boldsymbol{y}^T \log \boldsymbol{\pi} + (\mathbf{1} - \boldsymbol{y})^T \log(\mathbf{1} - \boldsymbol{\pi}) \tag{2.4}$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)^T$, and $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)^T$. The zero-sum constraints to parameters are not sufficient conditions for the uniqueness of ML estimates. Moreover, without further constraints the ML method will usually provide a rapidly fluctuating trajectory of age, period, and cohort effects.

To overcome the identification problem in the decomposition (2.2), it would be sensible to make the assumption that the successive parameters of the ages, periods, and birth cohorts change gradually (Nakamura, 2004). This can be expressed as minimizing the weighted sum of squares of the first-order differences of the parameters: $\frac{1}{\sigma^2_A}\sum_{\tilde{a}=1}^{I-1}(\beta^A_{\tilde{a}} - \beta^A_{\tilde{a}+1})^2 + \frac{1}{\sigma^2_P}\sum_{\tilde{p}=1}^{J-1}(\beta^P_{\tilde{p}} - \beta^P_{\tilde{p}+1})^2 + \frac{1}{\sigma^2_C}\sum_{\tilde{c}=1}^{K-1}(\beta^C_{\tilde{c}} - \beta^C_{\tilde{c}+1})^2 + \sum_{j=1}^{m_A}\frac{1}{\sigma^2_{Ax_j}}\sum_{\tilde{a}=1}^{I-1}(\beta^A_{\tilde{a}j} -$

$\beta^A_{\tilde{a}+1,j})^2 + \sum_{j=1}^{m_P} \frac{1}{\sigma^2_{Px_j}} \sum_{\tilde{p}=1}^{J-1} (\beta^P_{\tilde{p}j} - \beta^P_{\tilde{p}+1,j})^2 + \sum_{j=1}^{m_C} \frac{1}{\sigma^2_{Cx_j}} \sum_{\tilde{c}=1}^{K-1} (\beta^C_{\tilde{c}j} - \beta^C_{\tilde{c}+1,j})^2$, where σ^2's are called hyperparameters and this expression can compactly be expressed as $(\boldsymbol{X}_s\boldsymbol{\beta}_*)^T \boldsymbol{S}^{-1}(\boldsymbol{X}_s\boldsymbol{\beta}_*)$, which acts as a 'smoothness' penalty with respect to $\boldsymbol{\beta}$ for the log-likelihood given in (2.5) (Nakamura, 1986). The amount of smoothness of parameters is realized by the hyperparameters which act as the tradeoff among parameter effects by producing smoothed estimates (Lindley and Smith, 1972). $\boldsymbol{X}_s$ is an $M \times M$ matrix representing the first order differences of the parameters and $\boldsymbol{S}$ is an $M \times M$ diagonal matrix, $\boldsymbol{S} = \mathrm{diag}\big(\sigma^2_A \mathbf{1}^T_{I-1}, (\boldsymbol{I}_{m_A} \otimes \mathbf{1}_{I-1})\boldsymbol{\sigma}_A, \sigma^2_P \mathbf{1}^T_{J-1}, (\boldsymbol{I}_{m_P} \otimes \mathbf{1}_{J-1})\boldsymbol{\sigma}_P, \sigma^2_C \mathbf{1}^T_{K-1}, (\boldsymbol{I}_{m_C} \otimes \mathbf{1}_{K-1})\boldsymbol{\sigma}_C\big)$, where $\boldsymbol{\sigma}_A = (\sigma^2_{Ax_1}, \ldots, \sigma^2_{Ax_{m_A}})^T, \boldsymbol{\sigma}_P = (\sigma^2_{Px_1}, \ldots, \sigma^2_{Px_{m_P}})^T, \boldsymbol{\sigma}_C = (\sigma^2_{Cx_1}, \ldots, \sigma^2_{Cx_{m_C}})^T$. Here $\boldsymbol{I}_u$ is a $u \times u$ identity matrix.

Instead of (2.4), the following penalized log-likelihood is maximized

$$\ell_p(\boldsymbol{\beta}|\boldsymbol{\sigma}) = [\boldsymbol{y}^T \log \boldsymbol{\pi} + (\mathbf{1} - \boldsymbol{y})^T \log(\mathbf{1} - \boldsymbol{\pi})] - \frac{1}{2}(\boldsymbol{X}_s\boldsymbol{\beta}_*)^T \boldsymbol{S}^{-1}(\boldsymbol{X}_s\boldsymbol{\beta}_*) \tag{2.5}$$

with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\sigma} = (\sigma^2_A, \boldsymbol{\sigma}^T_A, \sigma^2_P, \boldsymbol{\sigma}^T_P, \sigma^2_C, \boldsymbol{\sigma}^T_C)^T$, a $(m_A + m_P + m_C + 3)$ dimensional vector of hyperparameters. The maximum estimate of $\ell_p(\boldsymbol{\beta}|\boldsymbol{\sigma})$ should be achieved with a trade-off between the log-likelihood of the model and the smoothness constraints (i.e., first order differences) to the parameters $\boldsymbol{\beta}_*$. The maximization of (2.5) can be interpreted as the computation of the mode of the Bayesian posterior distribution of $\boldsymbol{\beta}$ (Akaike, 1980; Jiang et al., 2001) when the prior distribution of $\boldsymbol{\beta}_*$ is assumed a multivariate Gaussian distribution defined by

$$\phi(\boldsymbol{\beta}_*|\boldsymbol{\sigma}) = (2\pi)^{-\frac{M}{2}} |\boldsymbol{X}^T_s \boldsymbol{S}^{-1} \boldsymbol{X}_s|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{X}_s\boldsymbol{\beta}_*)^T \boldsymbol{S}^{-1}(\boldsymbol{X}_s\boldsymbol{\beta}_*)\right\}. \tag{2.6}$$

To this end, estimating the parameters requires to take into account two conditions simultaneously — (i) to maximize the $\log L(\boldsymbol{\beta})$ that ensures the goodness of fit of a model to data, and (ii) to maximize $\phi(\boldsymbol{\beta}_*|\boldsymbol{\sigma})$ that should smooth the parameter estimates. The estimate of $\boldsymbol{\beta}$ is to be obtained by maximizing the penalized log-likelihood (2.5) for given $\boldsymbol{\sigma}$.

2.4 ABIC and MAP Estimate

As a criterion for model selection as well as for the determination of $\boldsymbol{\sigma}$, ABIC is defined (Akaike, 1980) by

$$\mathrm{ABIC} = -2\ \log L(\boldsymbol{\beta}_0, \boldsymbol{\sigma}) + 2\ (\dim \boldsymbol{\sigma} + m_0 + 1), \tag{2.7}$$

where dim indicates the dimension of a vector and the marginal likelihood function of $\boldsymbol{\sigma}$ is obtained as

$$L(\boldsymbol{\beta}_0, \boldsymbol{\sigma}) = \int L(\boldsymbol{\beta}) \cdot \phi(\boldsymbol{\beta}_*|\boldsymbol{\sigma}) \mathrm{d}\boldsymbol{\beta}_* = \int \exp\{\log L(\boldsymbol{\beta})\} \cdot \phi(\boldsymbol{\beta}_*|\boldsymbol{\sigma}) \mathrm{d}\boldsymbol{\beta}_*,$$

$$= c_0 \int \exp\Big\{\boldsymbol{y}^T \log \boldsymbol{\pi} + (\boldsymbol{1}-\boldsymbol{y})^T \log(\boldsymbol{1}-\boldsymbol{\pi}) - \frac{1}{2}(\boldsymbol{\beta}_*^T \boldsymbol{X}_s^T \boldsymbol{S}^{-1} \boldsymbol{X}_s \boldsymbol{\beta}_*)\Big\} \mathrm{d}\boldsymbol{\beta}_* \qquad (2.8)$$

where $c_0 = (2\pi)^{-\frac{M}{2}} |\boldsymbol{X}_s^T \boldsymbol{S}^{-1} \boldsymbol{X}_s|^{\frac{1}{2}}$. The exponent in (2.8) is nothing but the penalized log-likelihood defined in (2.5). Since the penalized log-likelihood is certainly based on non-Gaussian, the analytical solution of the integral in (2.8), unlike the linear models in Akaike (1980), is not possible. But the prior ϕ in (2.6) is Gaussian, so the Gaussian approximation of the posterior is plausible (Ishiguro and Sakamoto, 1983; for non-parametric regression, refer to Good and Gaskins, 1971; Silverman, 1982). An approximate computation of the integral is discussed in the next section. The estimate $\hat{\boldsymbol{\sigma}}$ of $\boldsymbol{\sigma}$ is obtained by maximizing (2.8) or equivalently by minimizing ABIC (2.7) and the maximum a posteriori (MAP) estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ for given $\hat{\boldsymbol{\sigma}}$ is then obtained by maximizing (2.5) with the numerical non-linear optimization procedure (Davidon, 1968; Ishiguro and Akaike, 1989; for MAP estimate, Green, 1997; and Jiang et al., 2001). ABIC will result a parsimonious model which maximizes the likelihood function while minimizing the influence of the somewhat arbitrary gradually changing parameter assumptions.

3 Numerical Procedure

To compute the maximum value of $\log L(\boldsymbol{\beta}_0, \boldsymbol{\sigma})$ numerically, we first maximize $\ell_p(\boldsymbol{\beta}|\boldsymbol{\sigma})$ with respect to $\boldsymbol{\beta}$ for given $\boldsymbol{\sigma}$. Assume that $\ell_p(\boldsymbol{\beta}|\boldsymbol{\sigma})$ is well approximated by the following quadratic form

$$\ell_p(\boldsymbol{\beta}|\boldsymbol{\sigma}) \approx \ell_p(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma}) - \frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T \boldsymbol{H}(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}),$$

where $\boldsymbol{H}(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma}) = -\dfrac{\boldsymbol{\partial}^2 \ell_p(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma})}{\boldsymbol{\partial\beta\partial\beta}^T}$ is the negative of the Hessian matrix of the penalized log-likelihood at $\hat{\boldsymbol{\beta}}$. After partitioning $\boldsymbol{H}(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma})$ for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_*$, the marginal log-likelihood (2.8) of the model for fixed $\boldsymbol{\sigma}$ is approximated to

$$\begin{aligned}\log L(\boldsymbol{\beta}_0, \boldsymbol{\sigma}) \quad &\approx \quad \log c_0 + \ell_p(\hat{\boldsymbol{\beta}}|\boldsymbol{\sigma}) - \frac{1}{2}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_0)^T \boldsymbol{H}_0(\hat{\boldsymbol{\beta}}_0|\boldsymbol{\sigma})(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_0) \\ &\quad + \log \int \exp\Big\{-\frac{1}{2}(\boldsymbol{\beta}_* - \hat{\boldsymbol{\beta}}_*)^T \boldsymbol{H}_*(\hat{\boldsymbol{\beta}}_*|\boldsymbol{\sigma})(\boldsymbol{\beta}_* - \hat{\boldsymbol{\beta}}_*)\Big\} \mathrm{d}\boldsymbol{\beta}_*.\end{aligned}$$

To obtain the minimum ABIC, $\log L(\boldsymbol{\beta}_0, \boldsymbol{\sigma})$ should be maximum at $\boldsymbol{\beta}_0 = \hat{\boldsymbol{\beta}}_0$ and becomes

$$\begin{aligned}\log L(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\sigma}) \approx \frac{1}{2}\log|\boldsymbol{X}_s^T \boldsymbol{S}^{-1} \boldsymbol{X}_s| - \frac{1}{2}\log|\hat{\boldsymbol{H}}_*(\hat{\boldsymbol{\beta}}_*|\boldsymbol{\sigma})| + \Big\{\boldsymbol{y}^T \log \hat{\boldsymbol{\pi}} \\ + (\boldsymbol{1}-\boldsymbol{y})^T \log(\boldsymbol{1}-\hat{\boldsymbol{\pi}}) - \frac{1}{2}(\boldsymbol{X}_s \hat{\boldsymbol{\beta}}_*)^T \boldsymbol{S}^{-1} (\boldsymbol{X}_s \hat{\boldsymbol{\beta}}_*)\Big\}. \qquad (3.1)\end{aligned}$$

This quantity is the trade-off between the estimated maximum log-likelihood (2.4) of the model and the other penalty for the 'smoothness' of the estimated parameters.

3.1 Calculation of Initial Estimates

The numerical computation of (2.5) of the binary regression cohort model requires initial values for the parameters and hyperparameters. The modus operandi is proposed here as follows: The estimate of $\boldsymbol{\beta}$ is obtained by minimizing the quantity

$$\begin{aligned}\mathcal{M}(\boldsymbol{\beta},\boldsymbol{\sigma}) &= \frac{1}{2\sigma^2}\sum_n \frac{1}{w^2(\delta)}(\mathring{y}_n - \boldsymbol{d}_n\boldsymbol{\beta})^2 + \frac{1}{2\sigma^2}(\boldsymbol{X}_s\boldsymbol{\beta}_*)^T\boldsymbol{S}^{-1}(\boldsymbol{X}_s\boldsymbol{\beta}_*),\\ &= \frac{1}{2\sigma^2}(\mathring{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{W}^{-1}(\mathring{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\beta}) + \frac{1}{2\sigma^2}(\boldsymbol{X}_s\boldsymbol{\beta}_*)^T\boldsymbol{S}^{-1}(\boldsymbol{X}_s\boldsymbol{\beta}_*), \qquad (3.2)\end{aligned}$$

where $\boldsymbol{X} = [\boldsymbol{X}_0, \boldsymbol{X}_*]$, $\boldsymbol{X}_0$ and $\boldsymbol{X}_*$ are the design matrices associated with $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_*$, respectively. $\mathring{\boldsymbol{y}}$ is an N-dimensional vector whose n-th element, $\mathring{y}_n = \log(y_n + \delta)/(1 - y_n + \delta)$, and $\boldsymbol{W} = w^2(\delta)\boldsymbol{I}_N$. $w^2(\delta)$ is the variance of the modified logit transformation for binary data, i.e., $w^2(\delta) = (\log(1+\delta)/\delta)^{-1} \simeq (\log\delta)^{-1}$. The weighted least squares estimate of $\boldsymbol{\beta}$ are obtained for fixed $\boldsymbol{\sigma}$ and δ. In the situation of binary response data, one may adopt $\delta = 0.5$, but here we choose $0 < \delta \ll 0.5$ because the approximation of (3.2) should be close to true one when δ tends to zero. Again let $\mathcal{L}_*(\boldsymbol{\beta})$ be the pseudo-likelihood of the model is given by $(2\pi\sigma^2)^{-\frac{N+M}{2}}|\boldsymbol{W}|^{-\frac{1}{2}} \times \exp\{-\frac{1}{2\sigma^2}(\mathring{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{W}^{-1}(\mathring{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\beta})\}$, then the estimate of σ^2 is obtained by maximizing $\mathcal{L}_*(\boldsymbol{\beta}_0,\boldsymbol{\sigma}) = \int \mathcal{L}_*(\boldsymbol{\beta})\cdot\phi(\boldsymbol{\beta}_*|\boldsymbol{\sigma})\mathrm{d}\boldsymbol{\beta}_*$ for given $\boldsymbol{\beta}$. Thus the marginal pseudo penalized log-likelihood is derived as

$$\log\mathcal{L}_*(\tilde{\boldsymbol{\beta}}_0,\boldsymbol{\sigma}) = -\frac{N}{2}(1+\log 2\pi\tilde{\sigma}^2) - \frac{1}{2}\log|\boldsymbol{W}| - \frac{1}{2}\log|\boldsymbol{W}_*| + \frac{1}{2}\log|\boldsymbol{V}|, \qquad (3.3)$$

where $\boldsymbol{W}_* = \boldsymbol{X}_*^T\boldsymbol{W}^{-1}\boldsymbol{X}_* + \boldsymbol{V}$ and $\boldsymbol{V} = \boldsymbol{X}_s^T\boldsymbol{S}^{-1}\boldsymbol{X}_s$. The estimate of $\boldsymbol{\sigma}$ are obtained by maximizing (3.3) using the numerical optimization procedure (Ishiguro and Akaike, 25). It is noted that the estimate $\tilde{\boldsymbol{\beta}}$ can be considered as a function of $\boldsymbol{\sigma}$ because $\tilde{\boldsymbol{\beta}}$ is computed for given $\boldsymbol{\sigma}$.

3.2 Estimation of Confidence Interval

The estimate $\hat{\boldsymbol{\beta}}_*$ in (3.1) is obtained by numerical optimization procedure. This is interpreted as the mode or MAP estimate of our Bayesian model. The Hessian matrix $\boldsymbol{H}_*(\hat{\boldsymbol{\beta}}_*|\boldsymbol{\sigma})$, the negative of the second derivatives of the penalized log-likelihood with respect to $\boldsymbol{\beta}_*$, calculated numerically at $\boldsymbol{\beta}_* = \hat{\boldsymbol{\beta}}_*$. According to the Bayesian interpretation, the inverse Hessian $\boldsymbol{H}_*^{-1}(\hat{\boldsymbol{\beta}}_*|\boldsymbol{\sigma}) = [h_{ij}]$, say, $(i = 1,\ldots,M; j = 1,\ldots,M)$ is considered as the variance-covariance matrix of the posterior distribution of the parameters. The standard error of the estimate, for example, $\hat{\beta}_{i*}$ [i-th element of $\boldsymbol{\beta}_*$] is given by $\sqrt{h_{ii}}$, the square root of the corresponding diagonal element of $\boldsymbol{H}_*^{-1}(\hat{\boldsymbol{\beta}}_*|\boldsymbol{\sigma})$. The reliability of the parameter estimate with 90% confidence can be inspected visually when $\hat{\beta}_{i*} \pm 1.65\sqrt{h_{ii}}$ are attached to the plot of the estimate $\hat{\beta}_{i*}$ $(i = 1, \ldots, M)$.

4 Simulation Study

Judging the performance of our proposed method by Monte Carlo simulation studies, a total of thirty five candidate models can be constructed practically when only a single covariate viz. x is considered and these are G, GA, GP, GC, GAP, GAC, GPC, GAPC,G{x}, G{x}A, G{x}P, G{x}C, G{x}AP, G{x}AC, G{x}PC, G{x}APC, G{x}A{x}, G{x}P{x}, G{x}C{x}, G{x}A{x}P, G{x}AP{x}, G{x}A{x}P{x}, G{x}A{x}C, G{x}AC{x}, G{x}A{x}C{x}, G{x}P{x}C, G{x}PC{x}, G{x}P{x}C{x}, G{x}A{x}PC, G{x}AP{x}C, G{x}APC{x}, G{x}A{x}P{x}C, G{x}A{x}PC{x}, G{x}AP{x}C{x}, G{x}A{x}P{x}C{x}. In the *first* simulation, data have been simulated based on the effects of age, period and cohort for a simple cohort model without covariate effects (i.e., GAPC model case), and in the *second* simulation, data with the effects of age, period, cohort and a single covariate on the response variable (i.e., G{x}A{x}P{x}C{x} model case) are generated. Data for five hypothetical repeated surveys have been generated. A total sample of size $N = 15000$ was taken with equal sizes for each of the surveys. The ages of women were randomly generated based on the empirical age distribution of currently married women of childbearing ages of Bangladesh (CPSs, 1983-1991; and DHSs, 1993-1997). The numbers for a dichotomous covariate according to the positive responses of probability 0.2 were also independently generated.

GAPC As True Model Case: In this case the data for the response variable are generated without the effect of any covariate. The dichotomy values of the response variable for N individuals were made according to the probability of positive response: $\Pr(y_n = 1) = \exp\eta(a_n, p_n, x_n)/\{1 + \exp\eta(a_n, p_n, x_n)\}$. The structure of the GAPC model is given by $\eta(a_n, p_n, x_n) = \mu + \sum_{\tilde{a}}^{I} \mu_{\tilde{a}}^{A}\delta^{A}(a_n, \tilde{a}) + \sum_{\tilde{p}}^{J} \mu_{\tilde{p}}^{P}\delta^{P}(p_n, \tilde{p}) + \sum_{\tilde{c}}^{K} \mu_{\tilde{c}}^{C}\delta^{C}(c_n, \tilde{c})$, which indicates that the true model has the age, period, and cohort effects but no covariate effect. Data have been generated with the model for a hypothetical assumed patterns of age, period and cohort effects. All of the practically possible candidate models are fitted to the data and their estimated minimum ABIC values are shown in Table 1 (first four columns) for top ten out performed models, where the fact stands out that ABIC has chosen GAPC model as the best model.

Table 1: ABIC Values of the Top Ten Models for Simulated Data: (a) When the GAPC Model is Assumed True, and (b) When the G{x}A{x}P{x}C{x} is assumed true.

[a]Models	dimension $\boldsymbol{\theta}$	dimension $\boldsymbol{\sigma}$	ABIC value	[b]Models	dimension $\boldsymbol{\theta}$	dimension $\boldsymbol{\sigma}$	ABIC value
GAPC	21	3	15454.9	G{x}A{x}P{x}C{x}	42	6	15925.6
G{x}AP{x}C	26	4	15455.1	G{x}A{x}PC{x}	38	5	15929.0
G{x}A{x}P{x}C	32	5	15458.4	G{x}APC{x}	32	4	15931.7
G{x}AP{x}C{x}	36	5	15460.3	G{x}A{x}P	18	3	15933.3
G{x}APC	22	3	15461.2	G{x}A{x}P{x}	22	4	15934.3
G{x}A{x}PC{x}	38	5	15462.9	G{x}AP{x}C	26	4	15935.4
G{x}A{x}P{x}C{x}	42	6	15464.7	G{x}A{x}C{x}	34	4	15936.4
GPC	15	2	15467.7	G{x}AP{x}C{x}	36	5	15937.5
G{x}PC	16	2	15468.0	G{x}A{x}P{x}C	32	5	15937.9
G{x}A{x}P	18	3	15469.0	G{x}AC{x}	28	3	15938.2

G{x}A{x}P{x}C{x} As True Model Case: In this simulation study, data are produced following the model structure $\eta(a_n, p_n, x_n) = \mu + \beta_0 x_n + \sum_{\tilde{a}}^{I}(\mu_{\tilde{a}}^{A} + \beta_{\tilde{a}}^{A} x_n)\delta^{A}(a_n, \tilde{a}) + \sum_{\tilde{p}}^{J}(\mu_{\tilde{p}}^{P} + \beta_{\tilde{p}}^{P} x_n)\delta^{P}(p_n, \tilde{p}) + \sum_{\tilde{c}}^{K}(\mu_{\tilde{c}}^{C} + \beta_{\tilde{c}}^{C} x_n)\delta^{C}(c_n, \tilde{c})$, where x is a single dichotomous covariate. The response variable evolves with a hypothetical assumed structures of the age, period, cohort and covariate effects. It is conspicuous from Table 1 (last four columns) that the true model appears as the optimal model with minimum ABIC value among the plausible models applied to the data.

5 An Application to Contraceptive Use Data

Contraceptive use dynamics data of Bangladesh stems from a series of six repeated retrospective surveys (CPSs 1983, 1985, 1989, and 1991; and DHSs 1993-94, 1996-97). In the previous section, Bayesian method for estimating cohort model has been evaluated by Monte Carlo simulation for data of equally spaced repeated surveys. But the survey years of real data are not equally distant and therefore are modified using spline basis function with appropriate order (Huq, 2002). Current contraceptive use status (CCUS) is considered as a response variable and seven explanatory variables such as children wanted in future (CWF), visit of family planning worker by six months (FPW), educational attainment of respondent (EDU), employment status by earning cash (EMP), place of residence (RES), religion (REL) and education of husband (EDH) (these variables' significance are recognized by voluminous research works, viz.

Table 2: Top ten candidate models arranged according to their minimum ABIC values.

Variables		[a]M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Constant		+	+	+	+	+	+	+	+	+	+
CWF		+	+	+	+	+	+	+	+	+	+
FPW		+	+	+	+	+	+	+	+	+	+
EDU		+	−	+	−	+	+	−	+	−	−
EMP		+	+	−	+	−	+	+	+	+	−
RES		+	+	−	+	+	+	+	−	+	+
REL		+	+	+	+	+	+	+	+	+	+
Age		+	+	+	+	+	+	+	+	+	+
Age	CWF	+	+	+	+	+	+	+	+	+	+
	FPW	−	−	+	−	−	+	−	+	−	−
	EDU	−	−	−	−	−	−	−	+	−	−
	EMP	−	−	−	−	−	−	−	−	−	−
	RES	+	−	−	+	+	+	+	−	+	+
	REL	+	+	+	+	+	+	+	+	+	+
Period		+	+	+	+	+	+	+	+	+	+
period	CWF	+	+	+	+	+	+	+	+	+	+
	FPW	+	+	+	+	+	+	+	+	+	+
	EDU	−	−	−	−	−	−	−	+	−	−
	EMP	−	−	−	+	−	−	+	−	−	−
	RES	+	+	+	+	+	+	+	−	+	+
	REL	−	−	−	+	−	+	+	+	−	+
Cohort		+	+	+	+	+	+	+	+	+	+
cohort	CWF	+	+	+	+	+	+	+	+	+	+
	FPW	+	+	+	+	+	+	+	+	+	+
	EDU	−	−	+	−	+	−	−	+	−	−
	EMP	−	−	−	−	−	−	+	−	−	−
	RES	−	+	−	+	−	+	−	−	+	+
	REL	+	+	+	+	+	+	+	+	+	+
dim of $\boldsymbol{\theta}$		94	97	102	117	103	117	117	116	103	109
dim of $\boldsymbol{\sigma}$		12	12	13	15	13	15	15	15	13	14
ABIC		48172.4	48176.1	48178.5	48181.2	48187.5	48188.8	48193.2	48199.0	48205.1	48208.4

[a]M's in column headings indicate model of type G$\{\boldsymbol{x}\}$A$\{\boldsymbol{x}\}$P$\{\boldsymbol{x}\}$C$\{\boldsymbol{x}\}$ and the number indicates rank. '+' and '−' indicates the inclusion and exclusion of the covariates in the models, respectively.

Bernhaart and Uddin, 1990; DeGraff, 1991; Ullah and Chakraborty, 1993; Kamal, 1994; and Islam and Mahmud, 1995) are primarily considered for age-period-cohort-covariate analysis.

For simplicity, dichotomy categorization of the variables are considered as follows: CCUS = 1 if respondent is currently using contraceptive, otherwise 0; CWF = 1 if she wants child in future, otherwise 0; FPW = 1 is she was visited by FPW, elsewhere 0; EDU = 1 for some education, 0 for no education; EMP = 1 if earns cash, otherwise 0; RES = 1 if she resides in rural, RES = 0 for urban woman; REL = 1 for Muslim woman and REL = 0 for other than Muslim; and EDH = 1 if her husband has some education, EDH = 0 if no education or she does not know. *Firstly* a best-fitting model GAPC has been chosen that has minimum ABIC value from all possible model combinations without any covariates. To reduce the exhaustive search of all the possible models in case of a large number of covariates, *secondly* four models of type GAPC (denoted by G$\{x\}$APC, G$\{x\}$A$\{x\}$PC, G$\{x\}$AP$\{x\}$C, G$\{x\}$APC$\{x\}$) for each single variable are fitted and their minimum ABIC values are reckoned. From the results, husband's education has found to be the lowest impact on the contraceptive practice for three of four categories of models and therefore is not considered for further model investigations. *Finally*, based on the performances for each of the covariates on the response variable, several dozens of candidate models for main and interaction effects for a number of covariates with age, period and cohort effects are fitted to the data.

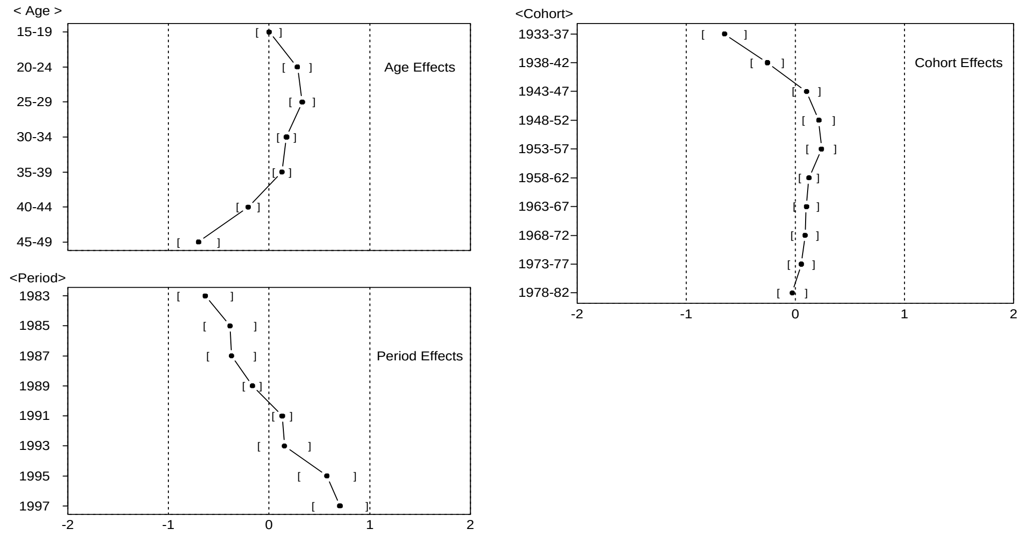

Figure 1: Estimated age, period, and cohort effects (represent by bullets) for the minimum ABIC model and the square brackets indicate the estimated 90% confidence intervals of the corresponding effect parameter.

Table 2 reveals the summaries of top ten models arranged in ascending order with their minimum ABIC values. Model M1 (denoted by G$\{\boldsymbol{v}_G\}$A$\{\boldsymbol{v}_A\}$P$\{\boldsymbol{v}_P\}$C$\{\boldsymbol{v}_C\}$ where $\boldsymbol{v}_G$=(CWF, FPW, EDU, EMP, RES, REL), $\boldsymbol{v}_A$=(CWF, RES, REL), $\boldsymbol{v}_P$=(CWF, FPW, RES), and $\boldsymbol{v}_C$=(CWF, PFW, REL)) is found to be the optimal model which has the smallest minimum ABIC value. Inclusion of main and interaction terms in a model are so flexible by considering the stepwise approach for selecting a best model among all possible candidate models. The best minimum-ABIC model (model M1) presented in Table 2 includes six of seven main effect of covariates and nine interaction terms between age, period and cohort with different combinations of four variables of CWF,FPW, RES and REL out of seven variables. Other nine models labelled M2 to M10 with various combination of main and interaction effects of covariates are also presented in this table. The estimate of grand mean is found to be -0.726 (32.6%) and the estimates of main effect of covariates in $\boldsymbol{v}_G$ are -0.76, 0.41, 0.22, 0.27, -0.29 and -0.21 respectively which are harmonious with the prior expectations amongst demographers, and the findings of many of the variables effect agreed with them. Effects of CWF, RES and REL on currently using contraceptive method are found negative and while the factors FPW, EDU and EMP are positively associated with the contraceptive prevalence rate. Negative effect of CWF interprets that woman who yet to complete family size is inversely related to contraceptive prevalence.

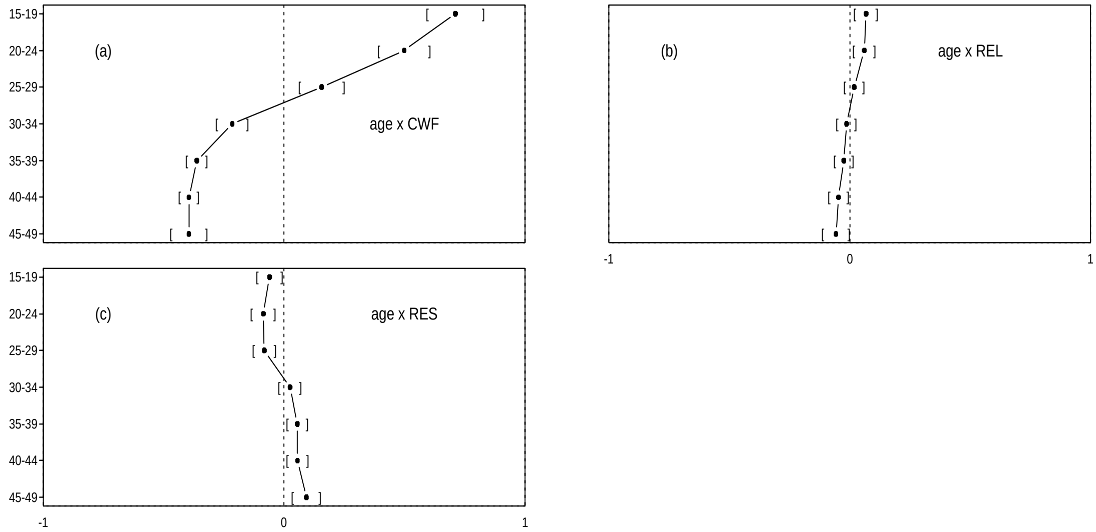

Figure 2: Estimated interaction effects (indicated by bullets) between the covariates and age of the minimum-ABIC model and the square brackets indicate the estimated 90% confidence intervals of the corresponding interation-effect parameter.

Figure 1 displaying age, period, and cohort changes over time. Movement to the right signifies an increase in contraceptive use while movement to the left indicates a decrease. It is obvious that contraceptive prevalence is higher among women of 20-30 years of age and the trajectory of age effects gradually declines as age increases. Increased time period clearly leads to increased contraceptive prevalence. The rising pattern of time effects is quite linear, though there are slight variations around it. That is, period effects indicate an increasing level of contraceptive use during the years surveys are conducted. Older cohorts of women have been found to possess less inclination in contraceptive use as many of them became out of risk in getting pregnancy. Women of middle cohorts exhibit more modernized attitude with regard to contraception than the younger cohorts. The premise behind this may be higher unmet need for contraception and/or higher discontinuation (those who abandon of use) among younger cohorts. Both the age and cohort effects show non-linear patterns but the period effects depict a sharp linear steep (rising) pattern.

Figure 2 shows the interaction effects between age, period, and cohort and covariates with their 90% confidence limits. This joint plot of the estimated effects with confidence intervals is used to assess the fit of the Bayesian models. The pattern for age-CWF interaction effects in panel (a) implies that women of age ranges from 15-35

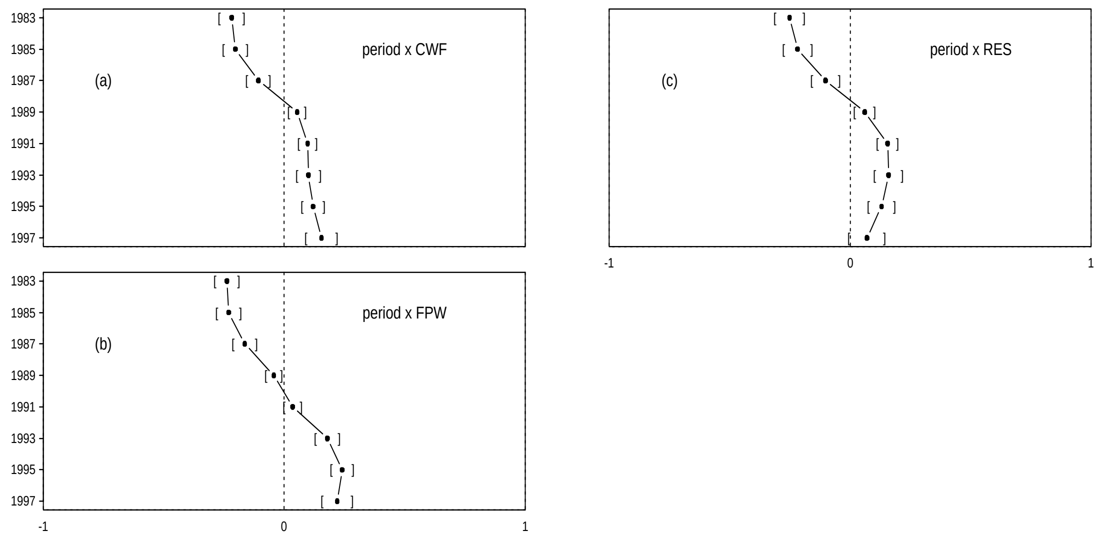

Figure 3: Estimated interaction effects between the covariate and period of the minimum ABIC model and the square brackets indicate the estimated 90% confidence intervals of the corresponding interaction-effect parameter.

who want additional children are practising contraceptive method more frequently to avoid conception than the 35-49 years of age, i.e., increased age of those women who want child in future leads to decreased contraceptive prevalence and the prevalence rate is hovering about women of age 35 and higher. Young woman may wants to give space for the next birth and they have higher prevalence in the use of contraception. Interaction effects between age and place of residence and age and religion shown in panel (b) and (c) respectively appear to be weakly consistent in their direction, i.e., increased age with women living in urban area follows a slightly increased contraceptive use and increased age of Muslim women leads to slowly decreased contraceptive use rate. As can be seen that the main effect of religion has negative influence on contraceptive use for religious regulations and/or prohibitions are widely perceived and respected by the Muslim women of the country which reflects in their older ages.

In Figure 3 the 80s shows a steady increased rate of contraception (panel (a)) but the rate after 1990 plummets to a stagnant state among women who desire child in future. Among women who desired additional children it was observed in Figure 4 that, younger to middle cohorts show a consistent decreased contraceptive use, while effects in older cohorts are inconsistent (panel (a)). Recent cohorts may want to give space for the next birth by practising contraceptives. Younger cohorts of Muslim eligible respondents showed a higher prevalence of contraception (panel (b)).

Interactions between time period and FPW in panel (b) of Figure 3 leads to increasing pattern of contraceptive prevalence during the middle survey periods though the direction changes at 1995, i.e., stable effects observed to both the ends and in

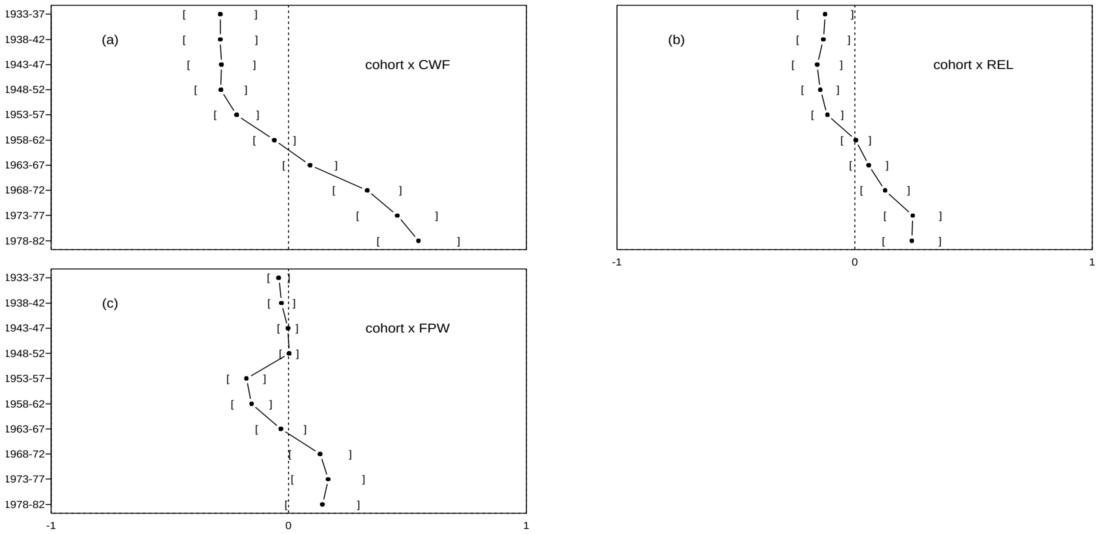

Figure 4: Estimated interaction effects between the covariate and cohort of the minimum-ABIC model and the square brackets indicate the estimated 90% confidence intervals of the corresponding parameter.

panel (c) of Figure 4, the joint effects of family planning assistance with cohorts prior to 1950s appear to be weak but slightly inconsistent, which may suggest that family planning women workers assistance did not reach to these cohorts of women. While the contraceptive prevalence dropped significantly in early 1950s and afterwards there is a sharp increase but constant effect observed from late-1960s to younger cohorts. The reason might be due to the withdrawals of governmental policy to nationwide continuing family planning services since late 90s. This fact may be reflected in panels (b) of Figure 3 and (c) of Figure 4.

6 Summary

Contraceptive use dynamics in Bangladesh has been investigated by using Bayesian binary logit regression cohort model which not only helps to examine the effects of age, period, and cohort but also to explore the instantaneous influence of covariates along with the interaction effects of age-by-covariate, period-by-covariate, and cohort-by-covariate. This model confronted an identification problem in that the age, period, and cohort effects cannot be partitioned uniquely and the problem is solved by a Bayesian method with gradually-changing-parameter assumption. Numerical optimization technique is utilized to maximize the likelihood for finding the estimates of the parameters. Effectiveness of the method have been checked by two simulation studies. The fit of candidate models are evaluated by the model selection statistic ABIC, and the best-fitting model that yields with the smallest ABIC value. The reliability of estimates of the parameters and the solutions of the identification problem have also been inspected visually by the joint display of the effects with their confidence bounds.

A real application of the models to the contraceptive use dynamics discloses many factors that have significant influence on fertility control in Bangladesh. Results suggest that period effects are somewhat strong and consistent. There appears to be a clear indication of increased changes in contraceptive use patterns. The patterns of contraceptive prevalence by age is typically U-shaped; relatively old or young women of childbearing ages have lower contraceptive prevalence rate than others because biologically older women's decline their fecundability but younger's may yet reach to a desired family size or a desired sex composition of children. The older cohorts have the lowest tendency in the contraceptive use practice, the highest prevalence observed in the mid 50s while from 1960s cohort effects are decreasing steadily but relatively stable. While taking interactions of CWF, RES and REL with age into account, the trajectory of these effects do not show U-shaped but have different effect patterns.

The findings reveal that women in younger cohorts are conscious enough about the expected number of offsprings compared to older cohorts; higher rate of contraception use grows apace as a sequel of changing attitude of Muslim women cohorts which is also boosted up by the field workers' visit; rural women of the periods between 80s and 90s are found not very receptive about contraception; field workers' visit increased its efficiency towards higher rate of contraception use during the historical periods. More importantly, application of sophisticated model brings into light that age, period, and cohort effects and many of the interaction effects with socioeconomic and demographic variables tend to be present and run in different directions, which might be suggested that this method can be plausible for analyzing data of repeated surveys.

References

[1] Akaike, H. (1980). Likelihood and the Bayes procedure. *Bayesian Statistics*; J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Eds.; University press, Valencia, Spain, 143-166.

[2] Bangladesh Contraceptive Prevalence Survey (CPSs) (1983, 1985, 1989, and 1991). National Instit7ute of Population Research and Training (NIPORT), Mitra and Associates, Dhaka.

[3] Bangladesh Demographic and Health Survey (DHSs) (1993-94, 1996-97), NIPORT, Mitra and Associates, Dhaka and Macro International Inc., USA.

[4] Bernhaart, M. H. and Uddin, M. M. (1990). "Islam and Family Planning Acceptance in Bangladesh ". *Studies in Family Planning*, *21*, 287-292.

[5] Davidon, W. C. (1968). "Variance Algorithm for Minimization". *Computer Journal*, *10*, 406-410.

[6] DeGraff, D. S. (1991). "Increasing Contraceptive Use in Bangladesh: The Role of Demand and Supply Factors". *Demography*, *28(1)*.

[7] Glenn, N. D. (1977). *Cohort Analysis*, Beverly Hills, California, Sage Publishers.

[8] Green, P. J. (1997). "Penalized Likelihood". *Encyclopedia of Statistical Sciences*, update volume (invited entry).

[9] Good, I. J. and Gaskins, R. A. (1971). "Nonparametric Roughness Penalties for Probability Densities". *Biometrika*, *58*, 255-277.

[10] Hayashi, C., Suzuki, T. and Sasaki, M. (1992). *Data Analysis for Comparative Social Research: International Perspectives*, Elsevier Science Pub., The Netherlands.

[11] Huq, S. S. (2002). "A Bayesian Binary Regression Cohort Model for Analyzing the Determinants of Family Planning in Bangladesh: With Special Focus on Unmet Contraceptive Need". *Unpublished Ph.D. Dissertation*, The Graduate University for Advanced Studies, Tokyo.

[12] Ishiguro, M. and Akaike, H. (1989). "DALL: Davidon's Algorithm for Log-likelihood Maximization - A Fortran subroutine for statistical model builders". *Computer Science monographs No. 25*, The Institute of Statistical Mathematics, Tokyo.

[13] Islam, M. M. and Mahmud, M. (1995). "Contraception Among Adolescents in Bangladesh". *Asia-Pacific Population Journal*, *10(1)*.

[14] Jiang, J., Jia, H. and Chen, H. (2001). "Maximum Posterior Estimation of Random Effects in Generalized Linear Models". *Statistica Sinica*, *11*, 97-120.

[15] Kamal, N. (1994). "Role of Government Family Planning Workers and Health Centers as Determinants of Contraceptives Use in Bangladesh". *Asia Pacific Population Journal*, *9(4)*, 59-65.

[16] Lindley, D. V. and Smith, A. F. M. (1972). "Bayes' Estimates for the Linear Model". *Journal of the Royal Statistical Society (Series B)*, *34*, 1-41.

[17] Mason, K. O., Mason, W. M., Winsborough, H. H. and Poole, W. K. (1973). "Some Methodological Issues in Cohort Analysis of Archival Data". *American Sociological Review*, *38*: 242-58.

[18] Mason, W. M. and Fienberg, S. E. (1985). *Cohort Analysis in Social Research*, Springer-Verlag, New York.

[19] Miller, A. S. and Nakamura, T. (1996). "On the Stability of Church Attendance Patterns During a Time of Demographic Changes: 1965-1988". *Journal for the Scientific Study of Religion*, *35(3)*, 275-284.

[20] Miller, A. S. and Nakamura, T. (1997). "Trends in American Public Opinion: A Cohort Analysis of Shifting Attitudes from 1972-1990". *Behaviormetrika, 24(2)*, 179-191.

[21] Nakamura, T. (1986). "Bayesian Cohort Models for General Cohort Table Analyses". *Annals of the Institute of Statistical Mathematics, 38(2)*, part B, 353-370.

[22] Nakamura, T. (1996). Re-examination of the Identifiability Problem in Cohort Analysis- A Bayesian Cohort Model. *Proceedings of the 24th Annual Meeting of the Behaviometric Society of Japan*, 208-209 (in Japanese).

[23] Nakamura, T. (2002). Cohort Analysis of Data Obtained Using a Multiple Choice Question. *Measurement and Multivariate Analysis*, S. Nishisato, Y. Baba, H. Bozdogan and K. Kanefuji (Eds); Springer, Tokyo, 241-248.

[24] Nakamura, T. (2004). "A Bayesian Logit Age-Period-Cohort Model". *Research Memorandum No. 922*, Institute of Statistical Mathematics, Tokyo.

[25] Osmond, C. and Gardner, M. J. (1982). "Age, Period and Cohort Models Applied to Cancer Mortality Rates". *Statistics in Medicine, 1*, 245-259.

[26] Robertson, C. and Boyle, P. (1998). "Age-Period-Cohort Analysis of Chronic Disease Rates. I: Modelling Approach". *Statistics in Medicine, 17*, 1305-1323.

[27] Ryder, N. B. (1965). "The Cohort as a Concept in the Study of Social Change", *American Sociological Review, 30*, 843-861.

[28] Sakamoto, Y. (1999). Statistical Research of the Japanese National Character: General Trends in Attitude over the Latter Half of the Twentieth Century. *World Association for Public Opinion Research 52nd Conference*, Paris, Sept. 3-5.

[29] Schaie, W. K. (1965). "A general model for the Study of Developmental Problems". *Psychological Bulletin, 64*, 92-107.

[30] Silverman, B. W. (1982). "On the Estimation of a Probability Desnsity Function by the Maximum Penalized Likelihood Method". *Annals of Statistics, 10*, 795-810.

[31] Ullah, M. S. and Chakraborty, N. (1993). "Factors Affecting the Use of Contraception in Bangladesh: A Multivariate Analysis". *Asia-Pacific Population Journal. 8(3)*, 19-30.

[32] Warshaw, M. (1992). The Use of Conditional Probability Graphs to Examine Age-Period-Cohort Data. *1992 Proceedings of the Section on Statistical Graphics.* Alexandria, VA: American Statistical Association, 55-60.