AUSTRIAN JOURNAL OF STATISTICS
Volume 38 (2009), Number 2, 83–99

A Semiparametric Sequential Ordinal Model with Applications to Analyse First Birth Intervals

Lawrence Kazembe
University of Malawi, Zomba, Malawi

Abstract: A semiparametric sequential ordinal model is proposed to analyze socio-demographic and spatial determinants of first birth intervals after marriage. Random effects are introduced to capture spatially structured and unstructured latent covariates. The structured effects are modelled by assuming conditional autoregressive priors, and for the unstructured effects we use an exchangeable Gaussian prior, while the smooth effects of continuous covariates are modelled by penalized splines. Inference is based on the mixed model approach. The model is applied to data from a cross-sectional survey. Compared to a spatial parametric predictor, the spatial semiparametric model better fits the data.

Zusammenfassung: Ein semiparametrisches, sequentielles, ordinales Modell wird zur Analyse sozialdemografischer und räumlicher Faktoren für Intervalle von Erstgeburten nach Verehelichung vorgeschlagen. Zufällige Effekte werden eingesetzt um räumlich strukturierte und unstrukturierte latente Kovariablen zu erfassen. Die strukturierten Effekte werden modelliert indem konditionale autoregressive Priors angenommen werden, und für die unstrukturierten Effekte verwenden wir einen austauschbaren Gauss-Prior, während die glatten Effekte der stetigen Kovariablen durch pönalisierter Splines modelliert sind. Die Inferenz basiert auf dem Ansatz Gemischter Modelle. Das Modell wird auf Daten aus einer Querschnittserhebung angewandt. Verglichen mit einem räumlichen parametrischen Prädiktor passt das räumliche semiparametrische Modell besser zu den Daten.

Keywords: Mixed-Model based Inference, Spatial Modelling, Malawi.

1 Introduction

Modelling fertility data is of great interest in population economics study (Henry, 1973; Lloyd, 2005). Several indicators are used to measure fertility patterns, among which is first birth intervals (FBI) after marriage (Lloyd, 2005). The timing of first birth is strongly correlated with the pace of subsequent fertility and, often, rapid first birth leads to rapid transition to higher parities and higher fertility. It may also suggest social and cultural changes to fertility, values of family formation and parenthood. In many societies, especially in developing countries, birth carries multivalent social implications. For example, child bearing contributes significantly to the woman's identity in society, proves her fertility and reduces the anxiety surrounding family continuance (Lloyd, 2005).

The aim of this article is to develop a statistical model to analyze FBI, and investigate how global and local spatial effects on FBI can be successfully assessed, adjusting for a variety of socio-demographic factors. Most of the previous studies of FBI have employed the discrete-time duration model because the duration times are reported in months and are discrete in nature (Feng and Quanhe, 1996; Zhenzhen, 2000). In this paper, we propose working with ordinal representation of the length of the waiting interval as an alternative approach to modelling waiting time data. The ordinal responses arise by categorizing the continuous outcomes (i.e., the interval in months) by adjacent intervals along the continuous scale. The observed response can be regarded as the result of a sequential process in which each time point (response category) can be reached successively.

The sequential ordinal model, as described by Albert and Chib (1997, 2001), can be used to analyze such categorical responses that occur in sequential order. The sequential ordinal model, also referred to as the continuation ratio model, is equivalent to the most commonly used cumulative ordinal model where the distribution function is the extreme value distribution (Läärä and Matthews, 1985; Albert and Chib, 2001). For various extensions and comparisons among these models, see the overview by Liu and Agresti (2005). Tutz (2003) showed that the sequential ordinal model belongs to the multivariate exponential family, and the generalized linear model framework applies. Several reasons justify the choice of sequential ordinal models to analyze event history data. Firstly, the sequential ordinal model compared to other duration models (e.g. the classical proportional hazards model) avoids the estimation bias introduced by long-term survivors. Secondly, the sequential ordinal model can be used to model non-proportional and non-monotonic hazard functions, and the effect of time-varying covariates can be allowed (Albert and Chib, 2001; Tutz, 2003).

Applications of the sequential ordinal model in the analysis of event history demographic data, to our knowledge, are few. Such a use, however, is common in several other fields. In epidemiological studies, for instance, Knorr-Held, Raber, and Becker (2002) applied both cumulative and sequential ordinal models to map disease-specific cancer incidence data. Albert and Chib (2001) developed a sequential ordinal model to analyze length of hospital stay data. In another study, Tutz (2005) developed an isotonic sequential ordinal model to analyze repeated ordinal measurements. Similar applications have appeared in educational and economic studies. Albert and Chib (1997) applied sequential ordinal models to analyze education attainment, creating an ordinal response by categorizing duration (i.e., the number of years) of schooling. Omori (2003) compared the proportional hazard model and the sequential ordinal model to estimate Japanese diffusion index data.

In the following, we extend the sequential ordinal model of Albert and Chib (1997, 2001) by modelling FBI with a flexible geoadditive predictor (Fahrmeir, Kneib, and Lang, 2004) that incorporates random effects to account for spatial correlation and heterogeneity and allows nonlinear effects of continuous covariates and the usual fixed effects. For example, social norms associated with FBI can exhibit spatial effects (Entwisle, Casterline, and Sayed, 1989) and may be useful to quantify in order to formulate socio-economic policies. Furthermore, continuous variables such as age at marriage and year of marriage are estimated using categorical dummies or quadratic components, but this assumption may be too restrictive and such factors may exhibit nonlinear effects (Zhang and Steele, 2004). Inference follows the mixed model approach (Fahrmeir et al., 2004).

The rest of this paper is structured as follows. Section 2 describes the model and the estimation procedure. Section 3 outlines the data and the analysis plan. In Section 4, we give the results. We conclude in Section 5 with a discussion.

2 The Model

We consider the common situation of a cross-sectional regression analysis. Let y_i be a response variable with J ordered categories. In additional, we have a vector $w_i = (w_{i1}, \dots, w_{ip})'$ of p covariates. The observations (y_i, w_i) are assumed independent. The basic idea is to cast the model in terms of conditional transition probabilities $\Pr(y_i = j|y_i \geq j)$, $j = 1, \dots, J-1$. In our example, this is the characteristic of the ith woman who has first birth in interval j, which occurs only after passing levels $1, 2, \dots, j-1$ and only bears at level j or higher. The probability of having birth at interval j, conditional on the event that the jth interval is reached is given by,

$$\Pr(y_i = j|y_i \geq j, w_i) = F(\theta_j - w_i'\alpha)\,, \qquad j = 1, \dots, J-1\,, \tag{1}$$

where $\theta = (\theta_1, \dots, \theta_{J-1})$ are the cutpoints, one of which is normalized to 0 to ensure identifiability. F is a strictly monotone function, $w_i'\alpha$ is the effect of covariates associated with the response. If F is chosen to be a logistic distribution function we obtain a sequential logit (Tutz, 2003; Liu and Agresti, 2005),

$$\Pr(y_i = j|y_i \geq j, w_i) = \frac{\exp(\theta_j - w_i'\alpha)}{1 + \exp(\theta_j - w_i'\alpha)}\,, \tag{2}$$

or equivalently in logit form

$$\eta_{ij} = \log \frac{\Pr(y_i = j|w_i)}{\Pr(y_i > j|w_i)} = \theta_j - w_i'\alpha\,, \qquad j = 1, \dots, J-1\,. \tag{3}$$

where η_{ij} is a predictor.

Model (1) can also be formulated in terms of latent variables expressing the propensity of a woman to reach category j before bearing her first child. Corresponding to the jth category (of time to first birth), define latent variable $\{z_{ij}\}$, where $z_{ij} = \theta_j - w_i'\alpha + \epsilon_{ij} = \eta_{ij} + \epsilon_{ij}$, where ϵ_{ij} is an error variable. We observe $y_i = 1$ if $z_{i1} \leq \theta_1$, and we observe $y_i = 2$ if the first latent variable $z_{i1} > \theta_1$ and the second latent variable $z_{i2} \leq \theta_2$. In general we have

$$y_i = \begin{cases} 1 & \text{if } z_{i1} \leq \theta_1 \\ 2 & \text{if } z_{i1} > \theta_1, z_{i2} \leq \theta_2 \\ \vdots & \\ J-1 & \text{if } z_{i1} > \theta_1, z_{i2} > \theta_2, \dots, z_{iJ-1} \leq \theta_{J-1} \\ J & \text{if } z_{i1} > \theta_1, z_{i2} > \theta_2, \dots, z_{iJ-1} > \theta_{J-1}, z_{iJ} \leq \theta_J\,. \end{cases} \tag{4}$$

The latent variable representation can be simplified by incorporating the cutpoints $\{\theta_j\}$ into the mean function and fixing one of the cutpoints, $\theta_J = 0$. It can be shown that this categorization implies that $P(y_i = j|y_i \geq j, w_i)$ is as specified in equation (1). In the discrete time survival context, the outcome variable $y_i = j$ corresponds to an event in a pre-specified time interval $[a_{j-1}, a_j)$. Thus the sequential ordinal model for $\Pr(y_i = j|y_i \geq j)$ provides a discrete version of hazard regression (Tutz, 1991).

An alternative to the sequential model is the cumulative model given by

$$\Pr(y_i \leq j|w_i) = F(\theta_j - w_i'\alpha)\,, \qquad j = 1, \dots, J-1\,. \tag{5}$$

If F is the logistic distribution function one obtains the proportional odds model

$$\log \frac{\Pr(y_i \leq j|w_i)}{\Pr(y_i > j|w_i)} = \theta_j - w_i'\alpha \,. \tag{6}$$

When the logit link is replaced by the complimentary log-log link, the resulting model

$$\log \left[-\log \frac{\Pr(y_i \leq j|w_i)}{\Pr(y_i > j|w_i)} \right] = \theta_j - w_i'\alpha \tag{7}$$

is equivalent to the proportional hazards model. Läärä and Matthews (1985) showed that the cumulative model and sequential model are identical when the complementary log-log link is used.

Modelling of heterogeneity and spatially structured variation may be obtained by introducing random effects. Similarly, nonlinear effects are introduced in the model through smoothing functions. The predictor (3) is expanded to include all possible explanatory variables like fixed, nonlinear and spatial covariates, giving a semi-parametric predictor (Tutz, 2003),

$$\eta_{ij} = \theta_j - w_i'\alpha + \sum_{k=1}^{q} f_k(x_{ik}) + f_s(s_i) \,, \tag{8}$$

where α are fixed effects corresponding to $w_i = (w_{i1}, \ldots, w_{ip})'$, f_k, $k = 1, \ldots, q$ are unknown smooth functions of continuous covariates $x_i = (x_{i1}, \ldots, x_{iq})'$ that enter nonlinearly, and $f_s(s_i)$ is the spatial component of the model that captures random effects of area s_i, $s \in \{1, \ldots, S\}$, where woman i lives. The component $f_s(s_i)$ is split further into spatially structured and unstructured random effects, $f_{str}(s_i)$ and $f_{unstr}(s_i)$ respectively, to capture any residual between-and-within district variation in FBI that is not explained by components of the model.

To obtain a mixed model formulation of the predictor in generic form, we introduce some matrix notation. Let $\eta = (\eta_{1j}, \ldots, \eta_{nj})'$ denote the predictor, $f_k = (f_k(x_{1j}), \ldots, f_k(x_{nj}))'$ the effects of covariate x_j, $j = 1, \ldots, p$, $f_{str} = (f_{str}(s_1), \ldots, f_{str}(s_n))'$ the spatial effects, and $f_{unstr} = (f_{unstr}(s_1), \ldots, f_{unstr}(s_n))'$ the uncorrelated random effects. Then f_k, f_{str}, and f_{unstr} can always be expressed as the matrix product of an appropriately defined design matrix X and a (possible high-dimensional) vector of regression coefficients β, such that $f_h = X_h\beta_h$. Further, define $\gamma = (\theta, \alpha')'$ as the overall fixed regression coefficients (including the threshold parameters), and

$$V = \begin{pmatrix} 1 & & -w' \\ & \ddots & \vdots \\ & 1 & -w' \end{pmatrix}$$

the corresponding design matrix constructed from the covariates w_i and thresholds θ. Then, after reindexing, we can rewrite the predictor (8) in generic matrix notation as

$$\eta = V\gamma + X_1\beta_1 + \cdots + X_L\beta_L \,, \tag{9}$$

where $V\gamma$ represents fixed effects (including the threshold parameters) while each of the term $X_h\beta_h$ represents a nonparametric, spatial or random effect.

2.1 Prior Assumptions

Specification of the model (9) is completed by assigning appropriate prior distributions for the regression coefficients. In the empirical Bayesian approach, the parameters γ are considered fixed effects, while $\beta_1, \ldots, \beta_L$ are random effects. In the Bayesian framework we assign diffuse priors for the fixed effects i.e. $p(\gamma) \propto$ const, and informative priors for the random effects.

By assuming the effects of continuous covariates vary smoothly over their codomain, their priors can be modelled through P-splines (Eilers and Marx, 1996). The approach assumes the unknown function f_k can be approximated by a polynomial spline of degree l with equally spaced knots ($x_{k,min} = \zeta_{k0} < \zeta_{k1} < \cdots < \zeta_{k,r-1} < \zeta_{kr} = x_{k,max}$) within the domain of x_k. For each covariate x_k, an l degree P-spline approximation of f_k is defined as

$$f_k(x_k) = \sum_{m=1}^{n} \beta_{km} B_{km}(x_k)\,. \tag{10}$$

The P-spline is a linear combination of $n = r + l$ B-spline basis functions B_{km}. The estimation of f_k is reduced to the estimation of $\beta_k = (\beta_{k1}, \ldots, \beta_{km})'$. Here, a cubic P-spline in combination with second order random walk priors for β_{km} is employed

$$\beta_{km} = 2\beta_{k,m-1} + \beta_{k,m-2} + u_{km} \tag{11}$$

with $u_{km} \sim N(0, \tau_k^2)$ for $m > 2$ with β_{m1} and β_{m2} assigned diffuse priors.

For the spatial component, we distinguish the spatially structured and unstructured effects, defined in matrix form as $f_{str} = X_{str}\beta_{str}$ and $f_{unstr} = X_{unstr}\beta_{unstr}$, respectively. The spatially structured component is modelled by assuming a conditional autoregressive (CAR) prior (Besag, York, and Mollie, 1991). The CAR prior define areas as neighbors if they share a common boundary and assume that the effect of area s is conditionally Gaussian, with the mean of the effects of neighboring areas as expectation and a variance that is inverse proportional to the number of neighbors of areas s. Hence the most commonly used spatial smoothness prior is,

$$\beta_{str}(s)|\beta_{str}(r)\,, s \neq r\,, \tau_{str}^2 \sim N\left(\frac{1}{N_s}\sum_{r\in\delta_s}\beta_{str}(r), \frac{\tau_{str}^2}{N_s}\right)\,, \qquad s \in \{1, \ldots, S\}\,, \tag{12}$$

where N_s is the number of adjacent sites, and $r \in \delta_s$ denotes that r is a neighbor of area s. The parameter τ_{str}^2 quantifies the amount of spatial variation present in the data and control the smoothness. For the unstructured heterogeneity we introduce additional i.i.d. Gaussian priors with

$$\beta_{unstr}(s) \sim N(0, \tau_{unstr}^2)\,, \qquad s \in \{1, \ldots, S\}\,. \tag{13}$$

2.2 Mixed Model based Inference

Inference for the semiparametric sequential ordinal models is based on the empirical Bayesian approach, also called the mixed model methodology (Fahrmeir et al., 2004; Brezger, Kneib, and Lang, 2005). This is achieved by recasting the predictor model (9) as a generalized linear mixed model (GLMM) after appropriate reparametrization. This provides the key for simultaneous estimation of the function evaluations f_h and the variance parameters τ_h^2 in the empirical Bayes approach. To rewrite model (9) as mixed model, we assume that β_h has dimension d_h and the corresponding penalty matrix has rank $r_h < d_h = \dim(\beta_h)$. Each parameter vector β_h is partitioned into a penalized (β_h^{pen}) and unpenalized (β_h^{unp}) part yielding a variance component model (Fahrmeir et al., 2004; Brezger et al., 2005),

$$\beta_h = \Psi_h^{unp}\beta_h^{unp} + \Psi_h^{pen}\beta_h^{pen} \tag{14}$$

for some well defined $d_h \times (d_h - r_h)$ matrix Ψ_h^{unp} and a $d_h \times r_h$ matrix Ψ_h^{pen}. The following priors are assumed. For the penalized part, an i.i.d. Gaussian prior is suitable, while for the unpenalized part we assume a flat prior, this is

$$p(\beta_h^{pen}) \sim N(0, \tau_k^2 I_{r_h})\,, \qquad p(\beta_h^{unp}) \propto \text{const.} \tag{15}$$

Applying decomposition (14) to all the components of predictor (9) yields

$$\eta = X^{unp}\beta^{unp} + X^{pen}\beta^{pen}\,. \tag{16}$$

We have obtained in (16) a GLMM with fixed effects β^{unp} and random effects β^{pen}. The posterior, in terms of the GLMM representation, is given by

$$p(\beta^{unp}, \beta^{pen}|y) \propto L(y, \beta^{unp}, \beta^{pen}) \prod_{h=1}^{L} p(\beta_h^{pen}|\tau_h^2)\,, \tag{17}$$

where $L(\cdot)$ denotes the likelihood which is the product of individual likelihood contributions and $p(\beta_h^{pen}|\tau_h^2)$ is as defined above.

Estimation of regression coefficients and variance parameters is carried out using iteratively weighted least squares and approximate restricted maximum likelihood. At the first iteration the default (starting) values are assumed for the penalized, unpenalized and variance parameters. Then updates for $\bar{\beta}_h^{unp}$ and $\bar{\beta}_h^{pen}$ are obtained in the first step by solving a system of linear equations given estimates for the variance parameters. In the second step updates of the variance parameters are obtained by maximizing the approximate restricted log-likelihood. The restricted log-likelihood is maximized through a Fisher scoring technique. The two steps above are iterated until convergence. Fahrmeir et al. (2004) derived numerically efficient formulae that allow for handling large data sets.

3 Applications

3.1 Data

For applications of the methodology we consider data from the 2000 Malawi Demographic and Health Survey (MDHS). The 2000 MDHS interviewed a representative sample of more than 13000 eligible women aged between 15 and 49 years (National Statistical Office and ORC Macro 2001, 2000). A two-stage stratified sampling design was implemented to collect the data. The data were realized through a questionnaire that included

Table 1: Categorization of marriage to first birth intervals (in months).

Marriage to first birth interval (months)	Response	Frequency	Percent
8-10	1	1375	16.5
11-12	2	1354	16.2
13-16	3	1644	19.7
17-21	4	1277	15.3
22-31	5	1333	16.0
≥32	6	1358	16.3
Total		8342	100

questions on marriage and reproductive histories, of which detailed dates of birth of all women and their children were collected.

We analyze the time between marriage and first birth. We consider data from women who indicated births after post-marital conceptions. Those classified as premarital births (who gave a date of first birth that preceded date of marriage), and those classified premarital conceptions (who married within 7 months of their first birth) are excluded from further analysis, yielding 8342 (79.6%) ever-married women of postmarital conceptions.

The response variable is constructed by categorising the intervals between marriage and first births (in months) into six segments as shown in Table 1. This type of segmentation is consistent with previous studies of FBI (Feng and Quanhe, 1996), as well as guided by exploratory analysis of the empirical hazard function. In general, the interval of 8-11 months tried to capture the behavior of contemporary women or rural counterparts who are pressed to have an heir soon after marriage in order to consolidate their marriage. For the interval 12-24 months, it is argued that within this period an average woman would have had a birth. Beyond this, it is a deliberate attempt to delay birth.

Several covariates, grouped into two broad categories: community and bio-demographic covariates, are included in the analyses. These have been found important in previous studies of first birth intervals (Zhenzhen, 2000). Community factors allow for socio-cultural and/or socio-economic factors both at small and large scale, and include region, place of residence (rural/urban), education of the woman and ethnicity. Bio-demographic characteristics include age at marriage, year of marriage and age differences between spouses. Descriptive statistics of the variables are shown in Table 2. The DHS data also has information on district of residence, which permit inclusion of spatial correlation effects to capture residual or unobserved factors that may influence the pattern of the response.

3.2 Analysis

We fit the following three models, for both the sequential and cumulative models. The cumulative model is estimated for comparison purposes. The first model, M1, is purely spatial

$$\text{M1: } \eta_{ij} = \theta_j - f_{str}(district_i) - f_{unstr}(district_i)\,.$$

In this model we introduce spatial smoothness priors to capture spatial correlations at district level. This is achieved by assuming CAR priors (12). Further, the model permits

Table 2: Summary of covariates used in the model. Given are the counts in each birth interval category.

	Birth interval segments						Total
Covariate	1	2	3	4	5	6	(n)
Region							
North	301	219	324	183	194	199	1420
Centre	544	578	630	463	458	419	3092
South	530	557	690	631	681	741	3830
Residence							
Rural	279	223	243	231	205	199	1380
Urban	1096	1131	1401	1046	1128	1160	6962
Woman's education							
None	338	395	463	357	430	537	2520
Primary	899	896	1094	849	836	791	5365
Secondary & higher	138	63	87	71	67	31	457
Ethnicity							
Chewa	415	463	491	340	354	320	2383
Lomwe	223	219	259	256	248	287	1492
Yao	144	183	210	173	207	268	1185
Ngoni	157	126	188	150	136	114	871
Others	436	363	496	358	388	370	2411
Spouses age diff.							
Wife older	23	26	32	21	18	130	150
Husband older (≤ 5)	658	684	807	615	594	615	3973
Husband older (> 5)	567	524	628	546	597	556	3418
Year of Marriage							
1966-1975	41	59	58	57	70	128	413
1976-1985	226	256	323	280	317	411	1813
1986-1995	740	686	893	687	733	767	4506
1996-2000	368	353	370	253	213	53	1610
Age at first marriage							
<15 yr	149	129	197	185	236	416	1312
15-17 yr	566	646	749	605	632	558	3756
18-19 yr	409	372	436	290	290	210	2007
20-24 yr	242	187	230	179	156	147	1141
≥25 yr	9	20	32	18	19	28	126

unstructured heterogeneity. This model investigates whether there is substantial spatial variation in the first birth intervals, and if the answer is yes can this variation be explained by community and bio-demographic factors.

The second model, M2, is a spatial parametric model which adjusts for covariates, i.e.

$$\text{M2: } \eta_{ij} = \theta_j - w_i'\alpha - f_{str}(district_i) - f_{unstr}(district_i)\,.$$

With this model, we assess how much of the spatial variation is attenuated by the inclusion of fixed effects of all considerable covariates. Here the effect of *age at marriage* (age) and *year of marriage* (cohort) are estimated as fixed effects, categorized as in Table 2.

In the last model, M3, we fit a spatial semi-parametric model with age at marriage and marriage cohort assumed nonlinear and the rest of the variables assumed fixed

$$\text{M3: } \eta_{ij} = \theta_j - w_i'\alpha - f_1(age_i) - f_2(cohort_i) - f_{str}(district_i) - f_{unstr}(district_i)\,.$$

For the nonlinear effects we use a second-order random walk prior (11). Model M3 investigates the bias of fitting restrictive linear model, M2.

We compare the fitted models using Akaike Information criterion (AIC) or Bayesian Information Criterion (BIC). These are defined as sum of the the log-likelihood and the degrees of freedom (df). The log-likelihood measures the goodness of fit whereas the df measures model complexity. The smaller the AIC or BIC, the better the model. Implementation of these models were carried out in `BayesX` (Brezger et al., 2005). In `BayesX`, regression coefficients are estimated iteratively. For each model fitted, convergence is achieved when the change in regression parameters is 0.0001 and terminated at 400 iterations if convergence is not achieved. However at under 25 iterations all models converged.

4 Results

4.1 Model Selection

In Table 3, model selection values are given for the two types of ordinal models (sequential and cumulative) with different specifications of the covariates. The results show that the sequential logit models have smaller AIC and BIC values than the cumulative logit models. In model M1, the AIC and BIC criterion have a slight preference for the sequential model, with differences of $\Delta\text{AIC} = 5.7$ and $\Delta\text{BIC} = 3.4$ in AIC and BIC, respectively. In model M2 the differences in AIC and BIC values between the sequential model and the proportional odds model are large ($\Delta\text{AIC} = 86.9$ and $\Delta\text{BIC} = 75.9$). The proportional odds model fits the data worst between the two. Considering model M3, again as evidenced by the large differences in AIC and BIC values ($\Delta\text{AIC} = 103.8$ and $\Delta\text{BIC} = 93.6$), the sequential logit models fit the data much better than the proportional odds models. A look at the maps based on model M1 and M3 (results not shown), reveals that the estimated spatial effects are fairly similar, with slightly more pronounced pattern in the cumulative model. This may be caused by the order restrictions. In summary, based on the AIC and BIC alone, the sequential model is chosen.

4.2 Model Estimates

Now turning to the sequential model, Figure 1a shows the structured spatial variation in FBI estimated from the model without covariates (M1). The estimates ranged between -0.24 and $+0.27$. Dark gray indicates areas with increased chance of early first birth, while areas with white to light gray denote those with lower or delayed first birth. The figure displays considerable spatial autocorrelation in the underlying hazard towards first births. The 80% credible intervals (Figure 1b), show areas of significant positive and negative effects. The variance component for spatially structured effects is estimated at 0.056. The unstructured geographical effects (Figure 2a), with estimates ranging from

Table 3: Model comparison values based on AIC and on BIC for the three models, together with the marginal log-likelihood (LL). Also given are variance components τ^2_{str} for the spatial effects.

Model	Description	-2LL	df	AIC	BIC	τ^2_{str}
Cumulative						
M1	Spatial random effects (RE) only	29570.3	25.5	29621.4	29801.0	0.099
M2	Fixed + RE	29048.2	40.1	29128.2	29409.2	0.027
M3	Fixed + Nonlinear + RE	28877.2	50.3	28977.7	29331.0	0.014
Sequential						
M1	Spatial random effects (RE) only	29564.0	25.9	29615.7	29797.6	0.056
M2	Fixed + RE	28961.0	40.1	29041.3	29323.3	0.012
M3	Fixed + Nonlinear + RE	28770.5	51.7	28873.9	29237.4	0.006

-0.22 to $+0.21$, have noticeable influence on the model as confirmed by the corresponding confidence intervals map (Figure 2b).

We continue the analysis by including community and bio-demographic characteristics in model M2. The improved fit of the model is evidenced by the values of log-likelihood, AIC or BIC (Table 3). The purely spatial model (M1) is the least complex (df $= 25.9$) and fitted poorly (AIC = $29615.7, \text{BIC} = 29797.6$ in M1), when compared to model M2 (AIC $= 29041.3$ and BIC $= 29323.3$). Accounting for these risk factors, in model M2, eliminated considerable regional variation as evidenced by the reduction in the variance components for the structured spatial effects ($\sigma^2_{str} = 0.012$). This suggests that community and bio-demographic factors partly explain geographical differences in FBI. The estimates of the covariates are given in Table 4.

We fit the last model (M3) by assuming nonlinear smoothing functions for the continuous covariates: age at marriage and year of marriage. Values of LL, AIC and BIC for the model are again given in Table 3. There is a notable improvement in model fit compared to the spatial parametric model (M2). The adjusted spatial residual effects are given in Figure 3. The estimated smooth geographical effects (Figure 3a), with values ranging from -0.029 to +0.018, are very weak. Indeed, none of the effects are significant (Figure 3b). The spatial variance is again reduced to 0.006. The unstructured geographical effects (Figure not shown) are estimated between -0.25 and +0.19. In general, the number areas of statistically significant effects are slightly reduced, but the overall variability remains the same.

Table 4 shows estimates of covariates obtained from model M3. Results from the cumulative model are also given for comparison purposes. Included in the table are estimates of the threshold parameters, $\theta_1, \dots, \theta_5$ for first five categories, with the last category (≥ 32 months) assigned as reference. The threshold parameters are interpreted as follows. Higher values of the threshold i.e., ($\theta > 0$) correspond to early first birth and lower values ($\theta < 0$) correspond to delayed first birth. For example, lower (higher) values of θ signify a shift to the right (left) side on the latent scale, which implies a decreased (increased) probability for that category. Generally, estimates for threshold parameters increase from θ_1 to θ_5, which indicates that the probability of having birth increases with increasing time in

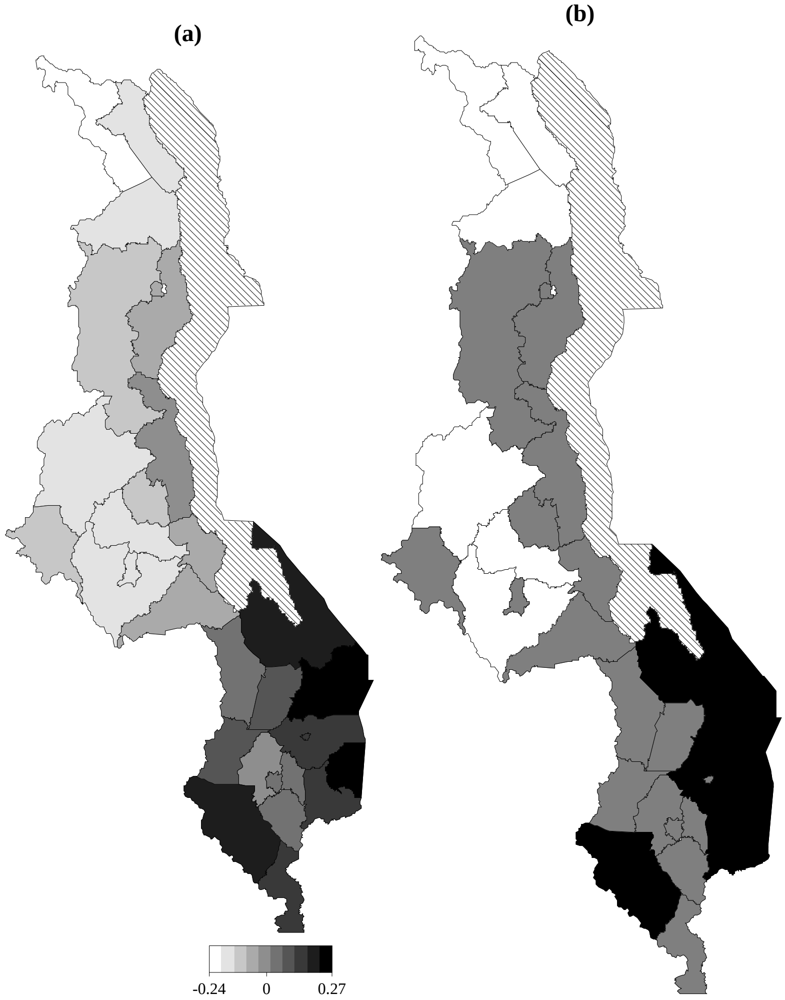

Figure 1: (a) Structured spatial effects, at district level, of first birth intervals (Model M1). Shown are the posterior modes. (b): Corresponding posterior probabilities at 80% nominal level, white denotes regions with strictly negative credible intervals, black denotes regions with strictly positive credible intervals, and gray depicts regions of nonsignificant effects.

marriage. Note that when compared, the threshold parameter estimates for the cumulative and sequential model are similar for θ_1 only. The difference comes in because in cumulative model, interest is to estimate cumulative probabilities, while in the sequential model, the aim is to estimate conditional probabilities. Thus these two model equal in definition at the first threshold only. Clearly, based on the cumulative model, the likelihood of first birth increases with increasing time in marriage. Fixed effects associated with FBI are region, education, ethinicity, marriage cohort and age at first marriage (Table 4). However, marriage cohort and age of marriage are better estimated as nonlinear effects (Figure 4). Indeed, considering the values of log-likelihood, AIC or BIC (Table 3), the model with nonlinear effects (M3) is better than the two (M1 and M2).

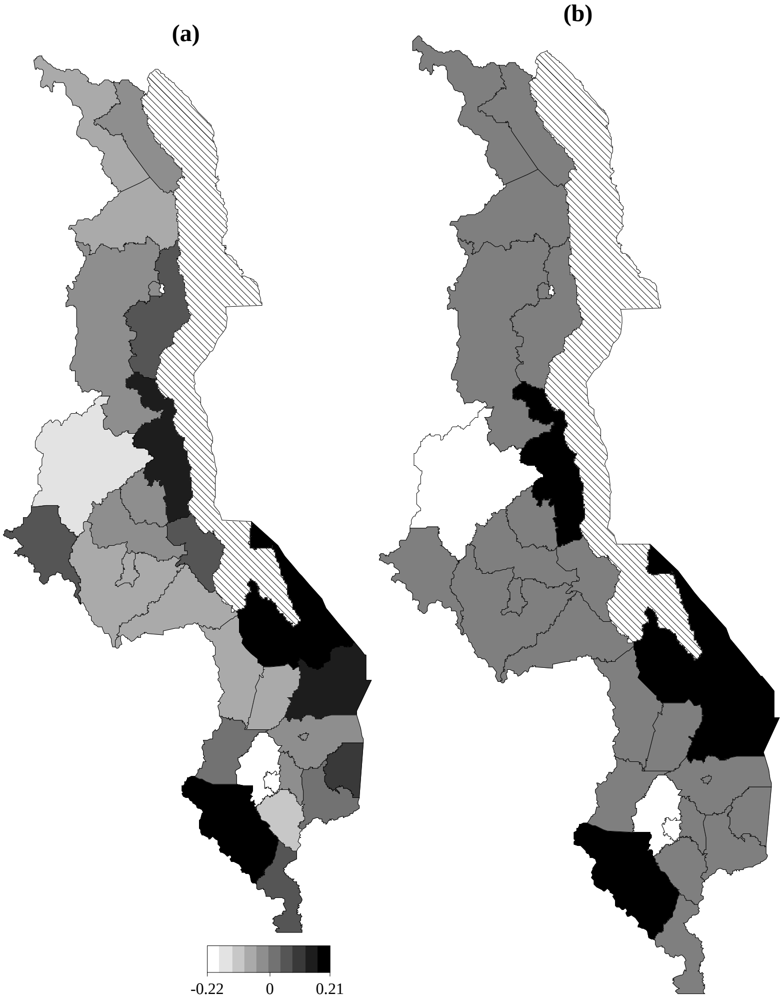

Figure 2: (a) Unstructured spatial effects, at district level, of first birth intervals (Model M1). Shown are the posterior modes. (b): Corresponding posterior probabilities at 80% nominal level, white denotes regions with strictly negative credible intervals, black denotes regions with strictly positive credible intervals, and gray depicts regions of non-significant effects.

5 Discussion and Conclusion

We have proposed a sequential ordinal model to analyze small-scale geographical variability in first birth intervals. The model assumed a semiparametric predictor, which facilitates smoothing of spatial effects and nonlinear effects of continuous covariates, while estimating other fixed effects in a single framework. A recently developed mixed model approach is used for inference (Fahrmeir et al., 2004; Tutz, 2003). In the following, we discuss the approach adopted, the results obtained, and limitations which appeal for further future studies.

The semiparametric model has been shown to provide flexible models in situations where the set of covariates consists of categorical and continuous variables. Here we note that the relationship between age at marriage and year of marriage with FBI are

Table 4: Estimates of fixed effects for the three models fitted. Given are the posterior modes and standard errors in brackets.

	Sequential Models			Cumulative Models		
Covariate	M1	M2	M3	M1	M2	M3
Threshold						
θ_1	-1.64 (0.04)	-1.74 (0.06)	-2.07 (0.14)	-1.67 (0.05)	-1.75 (0.08)	-2.23 (0.19)
θ_2	-1.43 (0.04)	-1.51 (0.06)	-1.82 (0.15)	-0.75 (0.05)	-0.81 (0.07)	-1.28 (0.19)
θ_3	-0.88 (0.04)	-0.93 (0.07)	-1.22 (0.15)	0.08 (0.05)	0.05 (0.07)	-0.41 (0.19)
θ_4	-0.73 (0.05)	-0.73 (0.07)	-1.02 (0.15)	0.74 (0.05)	0.74 (0.07)	0.29 (0.19)
θ_5	0.02 (0.05)	0.09 (0.07)	-0.19 (0.15)	1.65 (0.05)	1.69 (0.08)	1.26 (0.19)
θ_6	0	0	0	0	0	0
Region						
Northern		0	0		0	0
Central		-0.04 (0.06)	-0.04 (0.05)		-0.04 (0.08)	-0.04 (0.08)
Southern		0.17 (0.07)	0.16 (0.06)		0.28 (0.11)	0.26 (0.09)
Residence						
Rural		0	0		0	0
Urban		-0.01 (0.02)	-0.02 (0.02)		-0.01 (0.03)	-0.01 (0.03)
Ethnicity						
Chewa		-0.07 (0.03)	-0.07 (0.04)		-0.08 (0.05)	-0.07 (0.05)
Lomwe		0.05 (0.04)	0.04 (0.04)		0.07 (0.05)	0.06 (0.05)
Yao		0.04 (0.04)	0.05 (0.04)		0.03 (0.05)	0.04 (0.05)
Ngoni		-0.02 (0.04)	-0.02 (0.04)		-0.04 (0.06)	-0.03 (0.05)
Others		0	0		0	0
Woman's Education						
None		0	0		0	0
Primary		0.04 (0.03)	0.04 (0.03)		0.09 (0.04)	0.09 (0.04)
Secondary & higher		-0.13 (0.05)	-0.12 (0.05)		-0.20 (0.06)	-0.19 (0.06)
Spouses age diff.						
Wife older		0	0		0	0
Husband older (≤ 5)		-0.02 (0.03)	-0.002 (0.03)		0.002 (0.04)	0.02 (0.05)
Husband older (> 5)		-0.05 (0.03)	-0.05 (0.03)		-0.05 (0.05)	-0.05 (0.05)
Year of Marriage						
1966-1975		0.34 (0.05)			0.42 (0.07)	
1976-1985		0.18 (0.03)			0.21 (0.04)	
1986-1995		0.01 (0.03)			-0.01 (0.03)	
1996-2000		0			0	
Age at first marriage						
<15 years		0			0	
15-17 years		-0.06 (0.03)			-0.07 (0.04)	
18-19 years		-0.26 (0.04)			-0.34 (0.05)	
20-24 years		-0.19 (0.04)			-0.23 (0.06)	
≥25 years		0.13 (0.09)			0.19 (0.13)	

nonlinear. Such behavior, apart from improving model prediction, does emphasize that demographic relations are not as simplistic as often depicted. Adequate and appropriate statistical modelling is useful for answering substantial questions in applied research.

The analysis in this article is based on the mixed model approach. It provides a valuable alternative to estimation via MCMC simulation techniques. The fully Bayesian inference has received considerable coverage (Knorr-Held et al., 2002; Omori, 2003; Albert

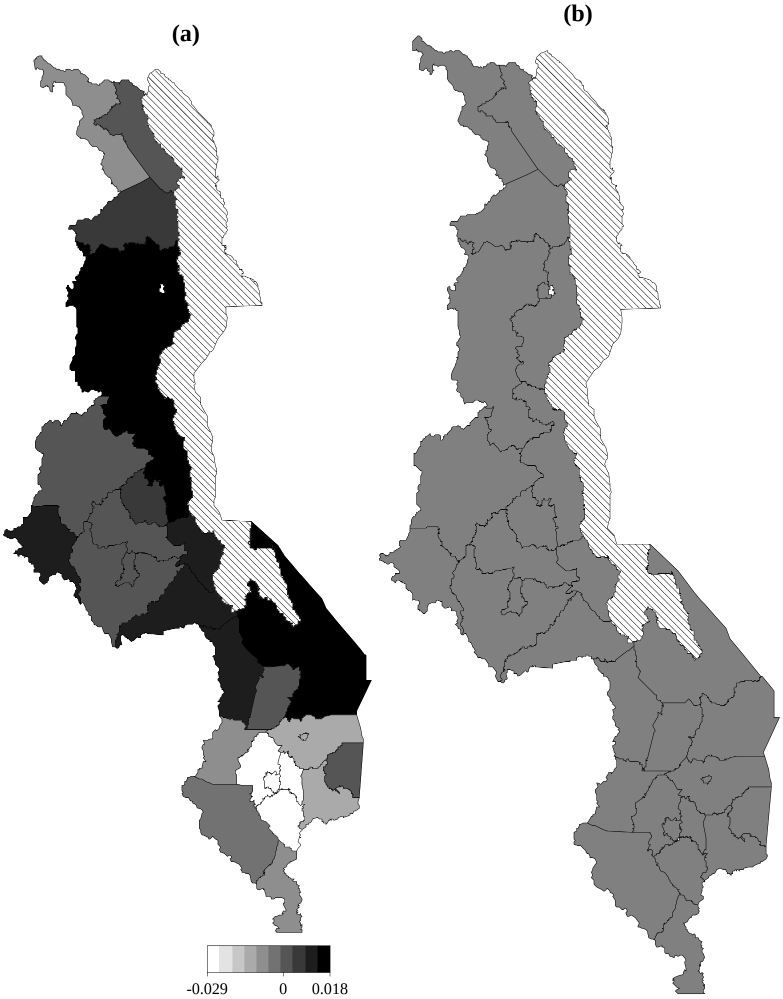

Figure 3: (a) Structured spatial effects, at district level, of first birth intervals (Model M3). Shown are the posterior modes. (b): Corresponding posterior probabilities at 80% nominal level, white denotes regions with strictly negative credible intervals, black denotes regions with strictly positive credible intervals, and gray depicts regions of nonsignificant effects.

and Chib, 2001). The mixed-model based approach as implemented in *BayesX* has been explained in detail in Fahrmeir et al. (2004), and this has been closely adopted here.

Many of the previous studies concerning FBI considered discrete-time duration models. Here we assessed the ordinal representation using the sequential model. Although the performance of the sequential ordinal model against proportional hazard model should have been evaluated, such comparisons have been reported elsewhere. Omori (2003) compared a proportional hazard model with sequential probit model and found that the estimates from the two models were consistently similar. More generally, the sequential ordinal model can estimate non-proportional and non-monotone hazard functions. An immediate extension to the model is to consider category-specific effects. We plan to explore the applicability of such models in future studies.

The choice between cumulative and sequential ordinal models merits further discus-

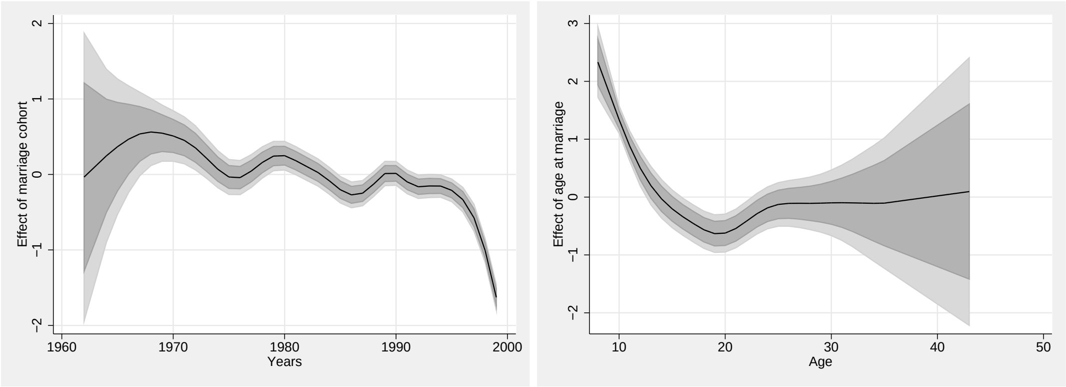

Figure 4: Non-linear effects of *year of marriage* (left) and *age at first marriage* (right). Shown are the posterior modes within 80% and 95% credible bands.

sion. Läärä and Matthews (1985), Tutz (1991), and Liu and Agresti (2005) provide a more detailed review of their properties. A more intuitive choice between the cumulative and sequential models can be based on the goals of the analysis. The sequential model is recommended when the underlying outcome is irreversible, and where the process is step-wise. This indeed is the case in our example data on waiting time till first birth after marriage. Our final results indicate that the sequential model has a much better fit than the cumulative model. Inclusion of spatial random effects and nonlinear effects leads to a further best fitting model. Knorr-Held et al. (2002) pointed out that the sequential model is better compared to the cumulative model. They further stated four arguments in favor of the sequential model. In general, the cumulative and sequential models are equivalent when the distribution function F is the extreme value distribution.

Overall, we found considerable spatial variability in FBI even after controlling for socio-demographic covariates. These spatial effects are surrogates of factors not captured by the survey instruments. Understanding the geographical variability of fertility behavior is an increasingly important research problem (Borgoni and Billari, 2003). However, this has often been done implicitly and at gross scale using categorical variables to measure geographic effects (Gould, Herrchen, and Pham, 1998). In our approach, we explicitly introduced spatial effects and modelled them using CAR priors.

In summary, the primary objective of this article was to illustrate a novel application of a recently developed structured additive regression model to analyze demographic data. The approach is data driven. The results emphasize that adequate statistical modelling and analysis is of importance in understanding complex relations that may exist in social and demographic processes.

Acknowledgements

I like to thank the anonymous referee for his careful scrutiny of the original manuscript and thus contributed tremendous to the readability of the final version of the manuscript. I acknowledge the permission granted by Measure DHS to use the 2000 Malawi Demographic Survey Data.

References

Albert, J., and Chib, S. (1997). *Bayesian methods for cumulative, sequential and two-step ordinal data regression models* (Tech. Rep.). Department of Mathematics and Statistics, Bowling Green State University. (Preprint Series)

Albert, J., and Chib, S. (2001). Sequential ordinal modelling with applications to survival data. *Biometrics*, *57*, 829-836.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, *43*, 1-59.

Borgoni, R., and Billari, F. C. (2003). Bayesian spatial analysis of demographic survey data: An application to contraceptive use at first sexual intercourse. *Demographic Research*, *8*, 3.

Brezger, A., Kneib, T., and Lang, S. (2005). BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, *14*, 11.

Eilers, P. H. C., and Marx, B. D. (1996). Flexible smoothing using B-splines and penalties. *Statistical Science*, *11*, 89-121.

Entwisle, B., Casterline, J. B., and Sayed, H. A. A. (1989). Villages as contexts for contraceptive behaviour in rural Egypt. *American Sociological Review*, *54*, 1019-1034.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, *14*, 715-745.

Feng, W., and Quanhe, Y. (1996). Age at marriage and the first birth interval: the emerging change in sexual behaviour among young couples in china. *Population and Development Review*, *22*, 299-320.

Gould, J. B., Herrchen, B., and Pham, T. (1998). Small-area analysis: targeting high-risk areas for adolescent pregnancy prevention programs. *Family Planning Perspective*, *30*, 173–176.

Henry, L. (1973). *Human Fertility: The Basic Components*. Chicago: The University of Chicago.

Knorr-Held, L., Raber, G., and Becker, N. (2002). Disease mapping of stage-specific cancer incidence data. *Biometrics*, *25*, 492-501.

Liu, I., and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, *14*, 1-73.

Lloyd, C. B. (2005). *Growing up Global: The Changing Transitions to Adulthood in Developing Countries*. New York: National Academy Press.

Läärä, E., and Matthews, J. N. (1985). The equivalence of two models for ordinal data. *Biometrika*, *72*, 206-207.

National Statistical Office and ORC Macro 2001. (2000). Malawi demographic and health survey 2000 [Computer software manual]. Zomba, Malawi: NSO.

Omori, Y. (2003). Discrete duration model having autoregressive random effects with application to Japanese diffusion index. *Journal of the Japanese Statistical Society*, *33*, 1-22.

Tutz, G. (1991). Sequential models in ordinal regression. *Computational Statistics and Data Analysis*, *11*, 275-295.

Tutz, G. (2003). Generalized semiparametrically structured ordinal models. *Biometrics*, *59*, 263-273.

Tutz, G. (2005). Modelling of repeated ordered measurements by isotonic sequential regression. *Statistical Modelling*, *5*, 269-287.

Zhang, W., and Steele, F. (2004). A semiparametric multilevel survival model. *Journal of the Royal Statistical Society, Series C*, *53*, 387–404.

Zhenzhen, Z. (2000). Social-demographic influence on first birth interval in China, 1980-1992. *Journal of Biosocial Sciences*, *32*, 315-327.

Author's Address:

Lawrence Kazembe
Applied Statistical and Epidemiology Research Unit
Mathematical Sciences Department
Chancellor College
University of Malawi, PO Box 280
Zomba, Malawi

E-mail: `lkazembe@yahoo.com`