

Regression Models with Heteroscedasticity using Bayesian Approach

Modelos de regresión heterocedásticos usando aproximación bayesiana

EDILBERTO CEPEDA CUERVO^{1,a}, JORGE ALBERTO ACHCAR^{2,b}

¹DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ, COLOMBIA

²DEPARTAMENTO DE MEDICINA SOCIAL, FACULDADE DE MEDICINA DE RIBEIRÃO PRETO, UNIVERSIDADE DE SÃO PAULO, SÃO PAULO, BRASIL

Abstract

In this paper, we compare the performance of two statistical approaches for the analysis of data obtained from the social research area. In the first approach, we use normal models with joint regression modelling for the mean and for the variance heterogeneity. In the second approach, we use hierarchical models. In the first case, individual and social variables are included in the regression modelling for the mean and for the variance, as explanatory variables, while in the second case, the variance at level 1 of the hierarchical model depends on the individuals (age of the individuals), and in the level 2 of the hierarchical model, the variance is assumed to change according to socioeconomic stratum. Applying these methodologies, we analyze a Colombian tallness data set to find differences that can be explained by socioeconomic conditions. We also present some theoretical and empirical results concerning the two models. From this comparative study, we conclude that it is better to jointly modelling the mean and variance heterogeneity in all cases. We also observe that the convergence of the Gibbs sampling chain used in the Markov Chain Monte Carlo method for the jointly modeling the mean and variance heterogeneity is quickly achieved.

Key words: Socioeconomic status, Variance heterogeneity, Bayesian methods, Bayesian hierarchical model.

Resumen

En este artículo, comparamos el desempeño de dos aproximaciones estadísticas para el análisis de datos obtenidos en el área de investigación social. En la primera, utilizamos modelos normales con modelación conjunta

^aProfesor asociado. E-mail: ecepedac@unal.edu.co

^bProfesor. E-mail: jorge@icmc.usp.br

de media y de heterogeneidad de varianza. En la segunda, utilizamos modelos jerárquicos. En el primer caso, se incluyen variables del individuo y de su entorno social en los modelos de media y varianza, como variables explicativas, mientras que, en el segundo, la variación en nivel 1 del modelo jerárquico depende de los individuos (edad de los individuos). En el nivel 2 del modelo jerárquico, se asume que la variación depende del estrato socioeconómico.

Aplicando estas metodologías, analizamos un conjunto de datos de talla de los colombianos, para encontrar diferencias que pueden explicarse por sus condiciones socioeconómicas. También presentamos resultados teóricos y empíricos relacionados con los dos modelos considerados. A partir de este estudio comparativo concluimos que, en todos los casos, es “mejor” la modelación conjunta de media y varianza. Además de una interpretación muy sencilla, observamos una rápida convergencia de las cadenas generadas con la metodología propuesta para el ajuste de estos modelos.

Palabras clave: metodología bayesiana, heterogeneidad de varianza, métodos bayesianos, estrato socioeconómico.

1. Introduction

Approximately 62% of Colombian children and teenagers lack of a satisfactory nutritional level and do not reach an optimal level of physical development. To get information about nutritional levels of this population, we model the mean of individual tallness in two groups of people, since the mean is one of the main nutritional level indicators. The first one includes children aging between 0 and 30 months old that belong to high socioeconomic level. The second one is a sample of people aging between six months and 20 years old that belong to lower, medium and higher socioeconomic strata. Tallness seems to be influenced by genetic factors (Chumlea et al. 1998), but the genetic homogeneity of the studied population seems to be apparent. Given that our main goal is to analyze this data set taking into account the variance heterogeneity, it is convenient to consider an analysis with explicit modeling of the variance, including age and socioeconomic stratum as explanatory variables. In this case, the variance can be modeled through an appropriate real function of the explanatory variables that takes into account the positivity of the variance (Aitkin 1987, Cepeda & Gamerman 2001). Thus, one possibility to analyze this data set is to use linear normal models with variance heterogeneity; other possibility is to consider hierarchical models (Bryk & Raudenbush 1992). The linear normal models with variance heterogeneity have two components: a linear function characterizing the mean response and a specification of the variance for each observation. In Section 2, we summarize the classical (Aitkin 1987) and the Bayesian methodologies (Cepeda & Gamerman 2001) used to fit these models. In the second model, a normal prior distribution is used and, given the orthogonality between mean and variance, a simple iterative process to draw samples from the posterior distribution can be performed. Hierarchical models are commonly used in educational and social statistics (Bryk & Raudenbush 1992, Raudenbush & Bryk 2002) to analyze this type of data. In this area, it is natural, for example, to consider that students are nested within classrooms and classrooms within schools;

individuals nested within socioeconomic stratum; groups may be nested in organizations (Steenbergen & Bradford 2002), and so forth. In this paper, we present, in Section 3, a general theoretical review of hierarchical models, since our intention is to use this model approach to analyze data taking into account the variance heterogeneity. For this purpose, we initially apply linear regression model with random coefficients (De Leeuw & Kreft 1986, Longford 1993). In Section 4, an application section, we initially compare standard linear normal regression models with regression models modelling the heterogeneity of the variances and standard normal regression models with regression models with random coefficients. Next, we analyze a people tallness data set, including age and socioeconomic stratum as explanatory variables. In this case, we include dummy variables and interaction terms between the dummy variables and one or more predictors in the modeling of the mean. However, since in general, it is not enough to model the contextual heterogeneity (Steenbergen & Bradford 2002), we consider a joint normal regression model for the mean and for the variance heterogeneity, where indicator and interaction factors are included in the mean and variance model. Here, we also analyze the data set using hierarchical models, including as many dummy variables as there are subgroups (socioeconomic stratum) in the second level to take into account the contextual differences. In all the applications, several explanatory models have been fitted and we include in all the cases, the estimates for the parameter of the model that produces the most satisfactory fit.

2. Modelling Variance Heterogeneity in Normal Regression

In this section, we summarize classical and Bayesian methodologies to get jointly maximum likelihood and posterior estimates, respectively, for the mean and variance parameters in normal regression models assuming variance heterogeneity.

2.1. Maximum Likelihood Estimation Using the Fisher Scoring Algorithm

Let $y_i, i = 1, \dots, n$, be the observed response on the i -th values $x_i = (x_{i1}, \dots, x_{ip})'$ and $z_i = (z_{i1}, \dots, z_{ir})'$ of the explanatory variables $X = (X_1, \dots, X_p)$ and $Z = (Z_1, \dots, Z_r)'$, respectively. Given the parameter vectors $\beta = (\beta_1, \dots, \beta_p)'$ and $\gamma = (\gamma_1, \dots, \gamma_r)'$, if the observations follow the model

$$y_i = x_i' \beta + \epsilon_i, \epsilon_i \sim N(0, \sigma_i^2), i = 1, \dots, n \quad (1)$$

where $\sigma_i^2 = g(z_i, \gamma)$ and g is an appropriate real function, the kernel of the likelihood function is given by

$$L(\beta, \gamma) = \prod_{i=1}^n \frac{1}{\sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} [y_i - x_i' \beta]^2 \right\}$$

Thus, given that the Fisher information matrix is a block diagonal matrix, an iterative alternate algorithm can be proposed to get maximum likelihood estimates for the parameters (Aitkin 1987). The summary of this algorithm, as it is presented in Cepeda & Gamerman (2001), is given in the next steps.

- a) Give the required initial values $\beta^{(0)}$ and $\gamma^{(0)}$ for the parameters.
- b) $\beta^{(k+1)}$ is obtained from $\beta^{(k+1)} = (X'W^{(k)}X)^{-1}X'W^{(k)}Y$, where $W^{(k)} = \text{diag}(w_i^{(k)})$, $w_i^{(k)} = 1/(\sigma_i^2)^{(k)}$ and $(\sigma_i^2)^{(k)} = \exp(z_i'\gamma^{(k)})$.
- c) $\gamma^{(k+1)}$ is obtained from $\gamma^{(k+1)} = (Z'WZ)^{-1}Z'W\tilde{Y}$, where $W = \frac{1}{2}I_n$, with I_n the $n \times n$ identity matrix, and \tilde{Y} is a n -dimensional vector with i -th component

$$\tilde{y}_i = \eta_i + \frac{1}{\sigma_i^2}(y_i - x_i\beta)^2 - 1$$

where $\eta_i = z_i'\gamma$.

- d) Steps (b) and (c) will be repeated iteratively until the pre-specified stopping criterion is satisfied.

2.2. Bayesian Methodology for Estimating Parameters

To implement a Bayesian approach to estimate the parameters of the model (1), we need to specify a prior distribution for the parameter of the model. For simplicity, we assign the prior distribution, $p(\beta, \gamma)$, given by

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim N \left[\begin{pmatrix} b \\ g \end{pmatrix}, \begin{pmatrix} B & C \\ C' & G \end{pmatrix} \right]$$

as in Cepeda & Gamerman (2001, 2005). Thus with the likelihood function $L(\beta, \gamma)$ given by a normal distribution, and using the Bayes theorem, we obtain the posterior distribution $\pi(\beta, \gamma | \text{data}) \propto L(\beta, \gamma)p(\beta, \gamma)$. Given that the posterior distribution $\pi(\beta, \gamma | \text{data})$ is intractable and it does not allow easily generating samples from it and taking into account that β and γ are orthogonal, we propose sampling these parameters using an iterative alternated process, that is, sampling β and γ from the conditional distributions $\pi(\beta | \gamma, \text{data})$ and $\pi(\gamma | \beta, \text{data})$, respectively. Since the full conditional distribution is given by

$$\pi(\beta, \text{data} | \gamma) = N(b^*, B^*) \quad (2)$$

where $b^* = B^*(B^{-1}b + X'\Sigma^{-1}Y)$ and $B^* = (B^{-1} + X'\Sigma^{-1}X)^{-1}$, with $\Sigma = \text{diag}(\sigma_i^2)$, samples of β can be generated from (2) and accepted with probability 1 (Geman and Geman, 1984). Since $\pi(\gamma | \beta, \text{data})$ are analytically intractable and it is not easy to generate samples from them, the following transition kernel (3) is proposed to get posterior samples of the parameters using the Metropolis-Hastings algorithm.

$$q(\gamma | \beta) = N(g^*, G^*) \quad (3)$$

where, $g^* = G^* \left(G^{-1}g + X'\Sigma^{-1}\tilde{Y} \right)$ and $G^* = \left(G^{-1} + Z'\Sigma^{-1}Z \right)^{-1}$, with $\Sigma = 2I_n$ and I_n the $n \times n$ identity matrix. The transition kernel q is obtained as the posterior distribution of γ , given by the combination of the conditional prior distribution $\gamma \mid \beta \sim N(g, G)$ with the working observational model $\tilde{y}_i \sim N(z'_i\beta, \sigma_i^2)$, where \tilde{y}_i is defined in item c), above. In this case, the quantity θ is updated in two blocks of parameters, β and λ . One of these blocks is updated in each iteration, as it is specified in the following algorithm:

- a) Begin the chain iteration counter in $j=1$ and set initial chain values $(\beta^{(0)}, \gamma^{(0)})$ for $(\beta, \gamma)'$.
- b) Move β to a new value ϕ generated from the proposed density (2).
- c) Update the γ vector to a new value ϕ generated from the proposed density (3).
- d) Calculate the acceptance probability of the movement, $\alpha(\gamma^{(j-1)}, \phi)$. If the movement is accepted, then $\gamma^{(j)} = \phi$. If it is not accepted, then $\gamma^{(j)} = \gamma^{(j-1)}$.
- e) Finally, update the counter from j to $j+1$ and return to b) until convergence.

3. Hierarchical Models

Hierarchical data are commonly studied in the social and behavioral sciences, since the variables of study often take place at different levels of aggregation. For example, in educational research, this approach is natural since the students are nested within classroom, classrooms within schools, schools within districts, and so on. In the two-level hierarchical models, given by N individuals nested within J groups, each one containing N_j individuals, if for simplicity we only consider one explanatory variable X , the first stage of the analysis is defined by

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2) \quad (4)$$

where Y_{ij} is the random quantity of interest associated to the i -th individual belonging to j -th group, $\beta_j = (\beta_{0j}, \beta_{1j})'$ is the parameter vector associated to the j -th strata, $j = 1, 2, \dots, J$, and ϵ_{ij} is the random error associated to the i -th subject belonging to j -th strata (Van Der Leeden 1998). In a second level, the regression coefficient behavior is explained by predicted variables of level 2, through the model

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + \eta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + \eta_{1j} \end{aligned} \quad (5)$$

where Z denotes the set of explanatory variables of level 2, and $\eta_{kj} \sim N(0, \sigma_k^2)$, $k = 0, 1$. In this model, Z is a context variable where its effect is assumed to be measured, for example, at the socioeconomic stratum, rather than at the individual

level. To complete the model, conjugate prior distributions are assumed for the hyperparameters.

The two level models are obtained by substitution of β_{0j} and β_{1j} , given by (5), into (4). Thus,

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{10}X_{ij} + \gamma_{11}Z_jX_{ij} + \eta_{1j}X_{ij} + \eta_{0j} + \epsilon_{ij} \quad (6)$$

In this model, given the error structure, $\eta_{1j}X_{ij} + \eta_{0j} + \epsilon_{ij}$, there is a heteroscedastic structure of the variance conditioned to the fixed part $\gamma_{00} + \gamma_{01}Z_j + \gamma_{10}X_{ij} + \gamma_{11}Z_jX_{ij}$. Thus, for each group

$$\text{Var}(Y_j) = \tilde{X}_j \Sigma_2 \tilde{X}_j' + \sigma_e^2 I_{N_j} \quad (7)$$

where $\tilde{X}_j = (1_{N_j}, X_j)$, with $X_j = (X_{1j}, \dots, X_{N_{jj}})'$. In (7) it is assumed that $e_{ij} \sim N(0, \sigma_e^2)$, that $(\eta_{0j}, \eta_{1j}) \sim N(0, \Sigma_2)$ and that the Level-1 random terms are distributed independently from the Level-2 random terms.

There are several ways to specify the level-2 model (Van Der Leeden 1998). If Z consists only of a vector of ones, the model specifies a random variation of the coefficient across the two units level. Such models are called Random Coefficient Models (De Leeuw & Kreft 1986, Prosser et al. 1991). In our applications, we consider models where the intercept and slope parameters are random, but there are other possibilities. For example, a simple model with random intercept and fixed slope can also be considered for use in practical work.

To implement a Bayesian approach to estimate the parameters of Hierarchical models, we include an appendix where we show the analytical processes to obtain the posterior distribution of interest.

4. Applications

In this section, we introduce some examples of growth and individual development studies for Colombian population. Studies about children's growth and development are very important in clinical research because they allow detection of factors which can affect children's health. As a first example, we consider a sample of babies between 0 and 30 months old. All the selected children in this example belong to a high socioeconomic stratum with a good welfare. In a second example, we consider a data set given by a group of 311 Colombian individuals aged from 6 months to 20 years old, sampled from different socioeconomic strata. For these examples, we consider as response variables of interest, the height measures for the individuals in the sample in order to concentrate on an appropriate specification of the time and socioeconomic stratum dependence.

4.1. Growth and Development of Babies

In this section, we analyze the growth and development of some groups of babies between 0 and 30 months old. The data set was selected by Pediatricians of

Bogota’s hospitals and by students of Los Andes University from the beginning of 2000 to the end of the year 2002. The data set shows some interesting characteristics: (a) the height increases with time but it does not have a linear behaviour as shown in figure (1). (b) Sample variance is not homogeneous; it seems to increase at an initially small time interval and then it decreases.

4.1.1. Applying Model 1

For this data, we assume the model $Y_i = \beta_0 + \beta_1\sqrt{t_i} + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma_i^2)$ and $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 t_i)$. Here Y_i is the tallness of the i -th babies at age t_i . The models were fitted by using a vague prior for the parameters. In all cases, we assigned a normal prior distributions $\beta_i \sim N(0, 10^5)$ and $\gamma_i \sim N(0, 10^5)$, $i = 1, 2$. The number 10^5 was chosen to impose large prior variances, but, as we have already checked in our analysis, increasing this value to larger orders of magnitude made no effective difference in the estimation process. Thus, the posterior means and standard deviations for the model parameters are given by: $\hat{\beta}_0 = 48.408(0.728)$, $\hat{\beta}_1 = 7.990(0.221)$, $\hat{\gamma}_0 = 2.303(0.2529)$, $\hat{\gamma}_1 = -0.038(0.022)$. Estimated values for the correlation parameters are presented in table 1. From Table 1, we see that $Corr(\beta_0, \beta_1)$ and $Corr(\gamma_0, \gamma_1)$ are significantly different from zero. Statistically, the correlation between β -s and γ -s are equal to zero. This result agrees with the diagonal form of the information matrix.

TABLE 1: Bayesian estimator for the parameter correlations.

	β_0	β_1	γ_0	γ_1
β_0	1	-0.904**	-0.021	-0.018
β_1		1	0.020	- 0.029
γ_0			1	-0.775**
γ_1				1

Figure 1 shows the plot for the posterior mean height of babies and the corresponding posterior 90% credibility interval. From this figure, we can see that the babies, all of whom have nutritional wealth, have tallness between international standards according to NCHS. Thus, there is no evidence of problems in the growth process. It is important to see that tallness is the most important parameter to validate the nutritional state and growing condition of babies. Figure 2 shows the behavior of the chain for the sample simulated for each parameter, where each one has small transient stage, indicating the speed convergence of simulation for the algorithm. The chain samples are given for the first 4500 iterations. The other results reported in this section are based on a sample of 4000 draws after a burn-in of 1000 draws to eliminate the effect of initial values. In Figure 3 we have the histograms for the posterior marginal distributions of parameters. These histogram seem to show that the posterior marginal distribution for all the parameters are approximately normal.

Finally, in this application we also considered normal bivariate prior distributions for the mean parameter $\beta = (\beta_0, \beta_1)$ and for the variance parameters $\gamma = (\gamma_0, \gamma_1)$.

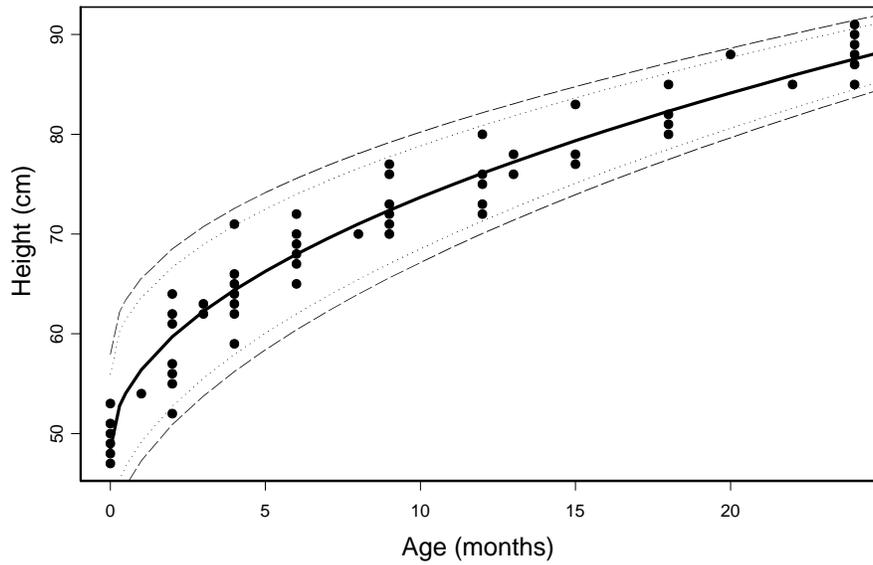


FIGURE 1: Predicted height for babies (the darker line represents the fitted posterior mean. The dotted lines represent the 90 and 95% prediction interval. The points correspond to the observations.)

4.1.2. Applying Normal Regression with Random Coefficient

In this case, we propose the model $Y_i = \beta_{0i} + \beta_{1i}\sqrt{t_i} + \epsilon_i$, where ϵ_i , $i = 1, 2, \dots, n$, are independent, identically distributed with normal distributions, that is $\epsilon_i \sim N(0, \sigma^2)$. Here Y_i is the tallness of the i -th individual at age t_i . Variation in the mean regression parameters is written in the second level as

$$\begin{aligned} \beta_{0i} &= \beta_0 + \eta_{0i} \\ \beta_{1i} &= \beta_1 + \eta_{1i} \end{aligned}$$

where $\eta_{ki} \sim N(0, \sigma_k^2)$, $k=0,1$. To complete the model, a conjugate prior is given for the hyperparameters: $\beta_k \sim N(0, 10^3)$, $\sigma_k^{-2} \sim \text{Gamma}(0.01, 0.01)$ and $\sigma^{-2} \sim \text{Gamma}(0.01, 0.01)$. Thus, the hierarchical model is given by

$$Y_i = \beta_0 + \beta_1\sqrt{t_i} + \eta_{0i} + \eta_{1i}\sqrt{t_i} + \epsilon_i$$

In this model, given the error structure, $r_i = \eta_{0i} + \eta_{1i}\sqrt{t_i} + \epsilon_i$, there is a heteroscedastic structure for the variance conditional on the level of the explanatory variable t . In this application we assume independence between observations, where

$$\text{Var}(r_i) = \sigma_0^2 + 2\sigma_{01}\sqrt{t_i} + \sigma_1^2 t_i + \sigma_e^2, \quad i = 1, 2, \dots, n$$

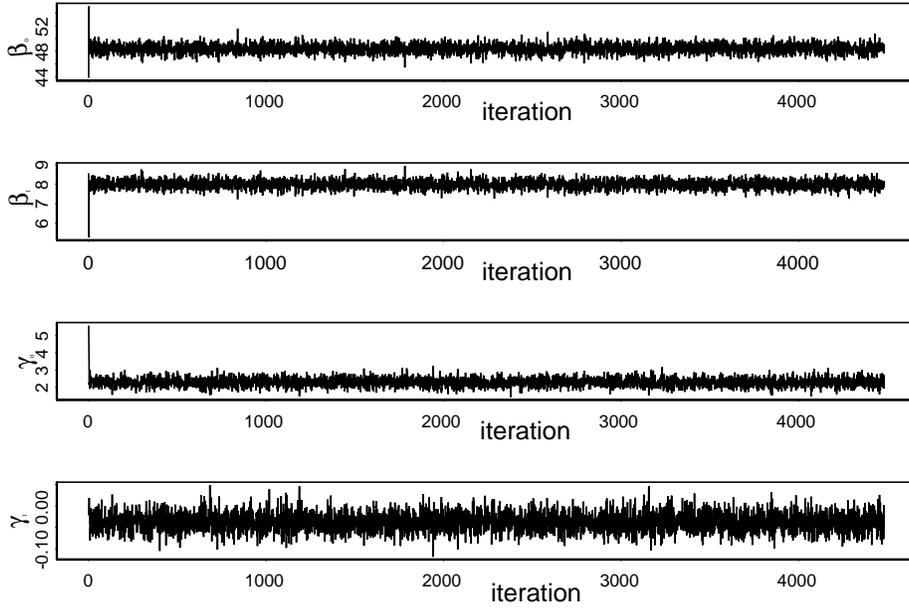


FIGURE 2: Behavior of the chain sample for parameters of the mean model β_i , and parameters of the variance model γ_i , $i=0,1$.

and the $corr(r_{i'}, r_i) = 0$, where $\sigma_{01} = Cov(\eta_{0i}, \eta_{1i})$, $\sigma_0^2 = Var(\eta_{0i})$ and $\sigma_1^2 = Var(\eta_{1i})$. It is clear that, as in the last model, the variance decreases with time. In this case, if $\sigma_{01} > 0$ the variance is increasing as function of time for $t > \sigma_{01}^2 \sigma_1^{-4}$ and increasing for all $t > 0$ if $\sigma_{01} > 0$.

In this analysis, the posterior means and standard deviations for the parameters models are: $\hat{\beta}_0 = 49.24(0.399)$, $\hat{\beta}_1 = 7.591(0.202)$. The estimates of other parameters are $\hat{\sigma}^{-2} = 0.942(0.097)$, $\hat{\sigma}_0^{-2} = 0.943(0.096)$, $\hat{\sigma}_{01} = -0.005(0.082)$ and $\hat{\sigma}_1^{-2} = 1.026(0.096)$.

TABLE 2: Model comparison.

Model	SSE	ln L	BIC
Variance heterogeneity	522.815	-174.022	5.003
Hierarchical	538.236	-191.447	5.597

In all cases, the obtained mean parameter estimates given by the two models agree. However, there are considerable differences in BIC values (Bayes Information Criterion) used for discrimination of models (Table 2). From the result in Table 2, we conclude that the model with joint modeling of the mean and variance heterogeneity is the best, since its BIC value is the smallest one.

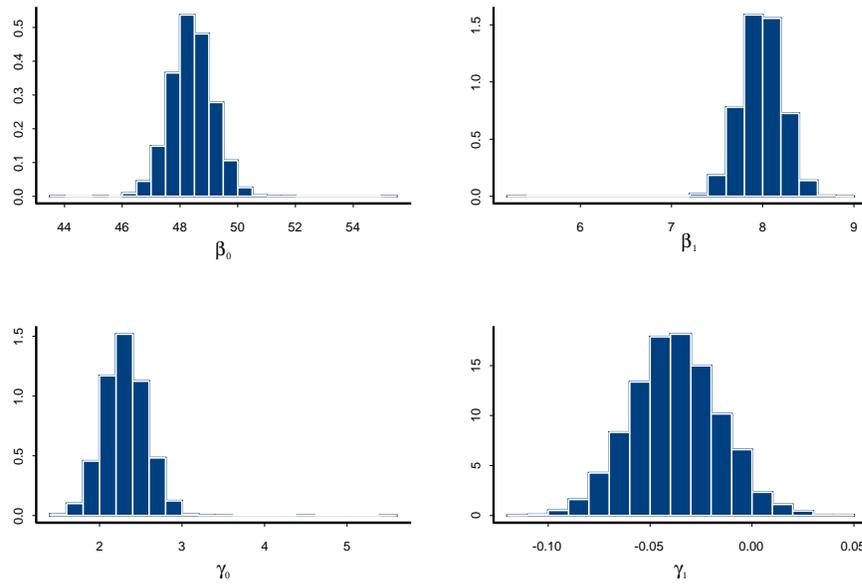


FIGURE 3: Histogram of the posterior marginal distributions. Parameter of the mean model β_i , and parameter of variance model γ_i , $i=0,1$.

4.2. Growth and Development of Population

In this section, we apply a Bayesian analysis to the growth and body development considering a sample of Bogota's individuals. We consider tallness as a response variable of interest to specify an appropriate dependence on age and socioeconomic level as possible factors associated with growing and body developmental processes (Adair et al. 2005).

The tallness of 311 person was measured and the age and socioeconomic stratum recorded. The individuals in the sample aged from 6 months to 20 years old and the socioeconomic strata were lower, medium and high, with approximately 100 individuals in each stratum. The data set shows some interesting characteristics exhibited in Figure 4: (a) the mean height increases in time but it does not have a linear behavior. (b) the means of tallness have a socioeconomic level dependence. (c) the sample variances are not homogeneous, increasing with the time and it seems to be different for each socioeconomic level.

4.2.1. A general approximation applying model 1

Taking into account these last points, we initially considered cubic polynomial models for the mean and variance. After a variable elimination process, a quadratic model $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$ for the mean and simple model $\log(\sigma_t^2) = \gamma_0 + \gamma_1 t$ for the

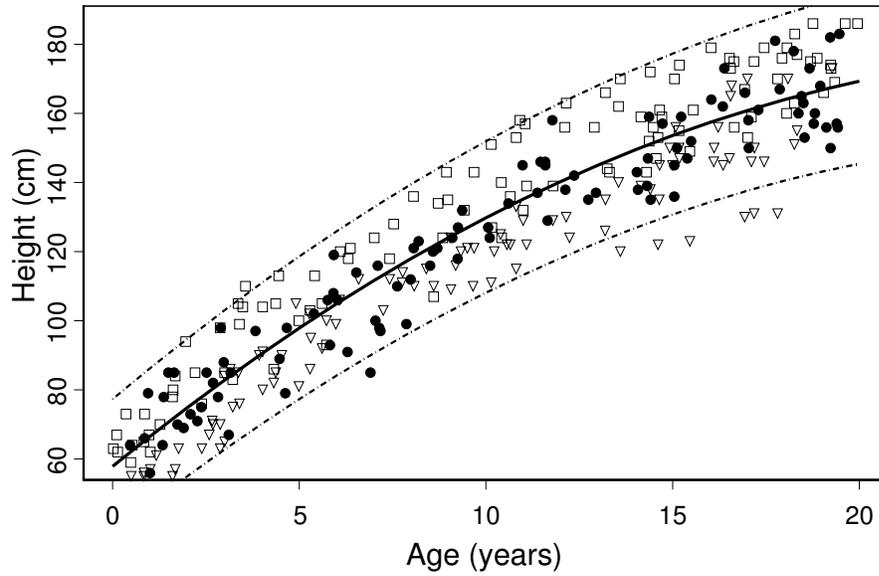


FIGURE 4: Plot for tallness versus age (the darkest line represents the fitted posterior mean. The dotted lines represent 95% prediction intervals. The triangle, point and square represent the observations corresponding to low, medium and high socioeconomic levels.)

variance seems to be appropriate for general data description, with t equal to age. These mean and variance models were fitted with vague priors distributions for each parameters. We further assumed prior independence among the parameters. The posterior means and standard deviations for the parameters model are given in Table 3.

TABLE 3: Posterior summaries for the mean and variance model.

Parameter	β_0	β_1	β_2	γ_0	γ_1
Mean	57,875	8,812	-0,162	4,573	0,024
S.d.	1,836	0,438	0,022	0,169	0,013

From Figure 4, we can establish some features of the growing process of this group of Bogota’s individuals. We also can compare the fitted model with the existing mean international curves for human growing and developing. We observe for this data set, the mean height curve is lower than the standard curve for human growth at all times, that is, there is some evidence of problems in the growth and body developing process. The differences could be an indication of malnutrition for some members of this group, an opinion that is shared by pediatricians and nutritional specialists.

4.2.2. General Approximation Using Normal Regression with Random Variables

For this data set, we also propose the model $Y_i = \beta_{0i} + \beta_{1i}t_i + \beta_{2i}t_i^2 + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma_e^2)$. In this model, Y_i is the tallness of the i -th individual at age t_i . Variation in the mean regression parameters is written in the second level as

$$\begin{aligned} \beta_{0i} &= \beta_0 + \eta_{0i} \\ \beta_{1i} &= \beta_1 + \eta_{1i} \\ \beta_{2i} &= \beta_2 + \eta_{2i} \end{aligned}$$

where $\eta_{ki} \sim N(0, \sigma_k^2)$, $k=0,1,2$. To complete the model, conjugate *Gamma*(0.00 019, 0.0001) prior distributions are assumed for the hyperparameters. Thus, the hierarchical model is given by

$$Y_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \eta_{0i} + \eta_{1i} t_i + \eta_{2i} t_i^2 + \epsilon_i \tag{8}$$

In this model, given the error structure, $r_i = \eta_{0i} + \eta_{1i}t_i + \eta_{2i}t_i^2 + \epsilon_i$, there is a heteroscedastic structure of the variance conditional on the level of the explanatory variable t . As in the last application, we assume independence between observations, that is,

$$Var(r_i) = \sigma_0^2 + \sigma_1^2 t_i^2 + \sigma_2^2 t_i^4 + 2\sigma_{01} t_i + 2\sigma_{02} t_i^2 + 2\sigma_{12} t_i^3 + \sigma_e^2, \quad i = 1, 2, \dots, n$$

where $\sigma_{i,i'} = Cov(\eta_i, \eta_{i'})$, and $\sigma_e^2 = Var(\epsilon_i)$.

The posterior means and standard deviations for the parameters models are: $\hat{\beta}_0 = 56.020(1.747)$, $\hat{\beta}_1 = 8.491(0.440)$, $\hat{\beta}_2 = -0.168(0.023)$. The Bayesian estimates obtained using the interactive procedure introduced in this paper for the other parameters are: $\hat{\sigma}_0^2 = 1.137(0.418)$, $\hat{\sigma}_1^2 = 0.066(0.0647)$, $\hat{\sigma}_2^2 = 7.115 \times 10^{-4}(2.909 \times 10^{-4})$ and $\hat{\sigma}_e^2 = 10.421$.

TABLE 4: Model comparison.

Model	SSE	ln L	BIC
Variance heterogeneity	42923.206	-1205.666	7.864
Hierarchical	42526.431	-2392.7662	15.535

As in the last example, we can see the Monte Carlo estimates for the posterior means for the parameters of interest assuming the two models agree. We also observe that there are considerable differences in the BIC values for the two models (Table 4). From the results of Table 4 we conclude that the model with joint modeling for the variance heterogeneity is better fitted by the data than the hierarchical model, since its BIC value is the smallest in comparison with the BIC value for the hierarchical model.

5. Growth and Development by Socioeconomic Stratum

In this section, we consider a second analysis for the tallness data. Here we are interested in determining if there are significant differences in the growing process depending on the socioeconomic levels, which are associated with nutritional status of the individuals, with a direct influence on their growing process.

5.1. Applying Model 1

In this analysis we considered indicator variables I_i , $i = 1, 2, 3$, for lower, medium and high socioeconomic stratum, respectively, and interaction variables $X_i = I_i t$, obtained by the product of t and I_i . Thus, taking into account the last description of the data, we propose the following model

$$Y_j = \beta_0 + \sum_{k=1}^2 \beta_k I_{kj} + \sum_{k=1}^3 (\lambda_k X_{kj} + \lambda_{k+3} X_{kj}^2) + e_{ij}$$

$$\sigma_j^2 = \exp \left(\beta'_0 + \sum_{k=1}^2 \beta'_k I_{kj} + \sum_{k=1}^3 (\lambda'_k X_{kj} + \lambda'_{k+3} X_{kj}^2) \right)$$

In the estimation process I_3 was eliminated from the model for the mean, and I_3 , X_1^2 , X_2^2 and X_3^2 from the model for the variance. With this new model, we obtain Monte Carlo estimates for the parameters in the model also assuming approximate non-informative priors for the parameters. For the mean model: $\hat{\beta}_0 = 61.735(1.712)$, $\hat{\beta}_1 = -17.137(2.452)$, $\hat{\lambda}_1 = 10.301(0.571)$, $\hat{\lambda}_2 = 7.551(1.741)$, $\hat{\lambda}_3 = 9.702(0.351)$, $\hat{\lambda}_4 = -0.251(0.032)$, $\hat{\lambda}_5 = -0.106(0.026)$, $\hat{\lambda}_6 = -0.199(0.027)$. For the variance models: $\hat{\beta}'_0 = 4.249(0.207)$, $\hat{\beta}'_2 = -0.939(0.317)$, $\hat{\lambda}'_1 = 0.092(0.022)$, $\hat{\lambda}'_2 = 0.017(0.018)$, $\hat{\lambda}'_3 = 0.0146(0, 018)$

Following the analysis, X_2 and X_3 were eliminated from the variance model. The parameter estimates of the resulting models are given in Tables 5 and 6.

TABLE 5: (a) Model for the mean.

Parameter	β_0	β_1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Mean	61.789	-17.234	10.299	7.565	9.656	-0.251	-0.107	-0.196
S. d.	1.803	2.035	0.580	0.489	0.497	0.032	0.025	0.026

TABLE 6: (b) Variance Model.

Parameter	β'_0	β'_1	λ'_1
Mean	4.412	-1.0981	0.092
S. d.	0.121	0,266	0,0216

From Figures 5 and 6, we observe that people from lower socioeconomic backgrounds present through time a significantly lower tallness than people from others

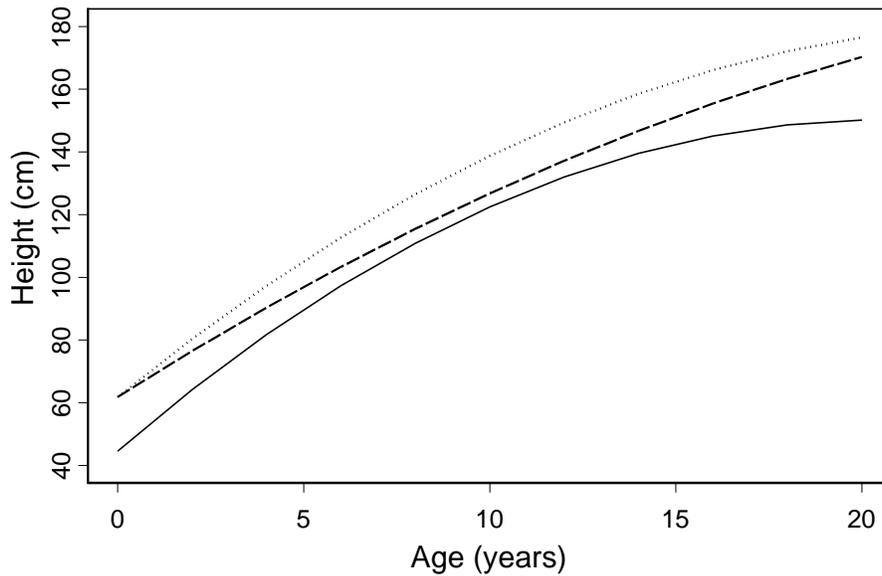


FIGURE 5: Mean of tallness of people by socioeconomic level. Continuous line for stratum 1, discontinuous line for stratum 2, and dotted line for stratum 3.

socioeconomic strata. People from medium socioeconomic stratum show a significantly lower tallness than people from high socioeconomic stratum, considering different ages. This fact is easy to understand, since in this case socioeconomic status has a special relevance on nutrition of people, and people from lower stratum probably have nutritional deficiencies (see Stein et al. 2004).

From the variance model, it can be seen that in the lower socioeconomic stratum, the variance behavior is given by $\hat{\sigma}_i^2 = 3.314 + 0.092X_i$. In the other strata, the variance is constant through time, $\hat{\sigma}_i = 4.412$. This behavior is shown in Figure 6.

This is a worrisome situation, since official reports from the Economic Commission for Latin America and the Caribbean (ECLAC) show the situation of poverty and indigence for Colombian children and teenagers: 45% can be considered poor and 17% indigents. If we add these numbers we obtain 62% of Colombians who do not have high life expectancy and whose nutritional supplies are not the proper ones, which results in a lack of optimal level of physical development.

5.2. Applying Hierarchical Models

In this section, the effect of socioeconomic stratum in the nutritional development of the people is evaluated through the tallness of the individuals. In our application, we have N individual belonging to low, medium and high socioeco-

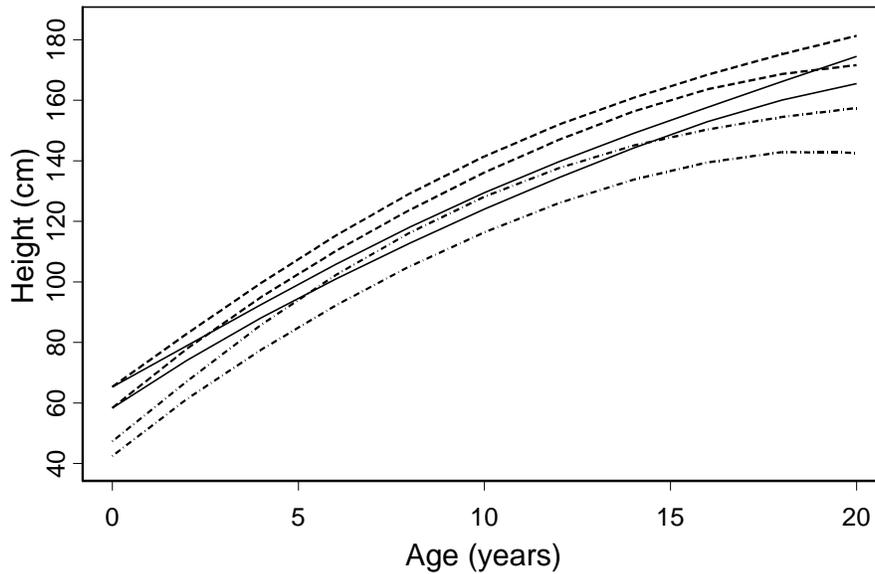


FIGURE 6: 95% prediction intervals for grown by stratum. Discontinuous line with point for stratum 1, continuous line for stratum 2, and discontinuous line for stratum 3.

nomic strata. The tallness and the age of each individual were determined and denoted Y_{ij} and t_{ij} , respectively, where the subscript ij indicate that the measures belong to the i -th individual belonging to j -th strata. Thus, as it is usual in multilevel analysis, in the first level, individuals are considered and a regression model (9) is defined for each group.

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + \beta_{2j}t_{ij}^2 + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2) \tag{9}$$

In this model, $\beta_j = (\beta_{0j}, \beta_{1j}, \beta_{2j})$ is the parameter vector associated to the j -th stratum, $j = 1, 2, 3$, and ϵ_{ij} is the random error associated in the i -th subject. In the second level of the model, the regression coefficient β_i is explained by predicted variables Z , through the model

$$\beta_{kj} = \gamma_{k0}Z_{0j} + \gamma_{k1}Z_{1j} + \gamma_{k2}Z_{2j} + \eta_{kj}, \quad k = 0, 1, 2 \tag{10}$$

where Z_{kj} , $k = 0, 1, 2$, is the set of explanatory variables of level 2, set of indicator variables of the stratum $k + 1$, and $\eta_{kj} \sim N(0, \sigma_k^2)$, $k = 0, 1, 2$. The model is completed with conjugate prior gamma for the hyperparameters.

Since in this application there is no prior information, we assume normal prior distribution with large variance, that is, approximate non informative prior. In this way, we assume normal prior distribution $\theta \sim N(0, 10^{-5}I_9)$, where θ is the

vector with components $\gamma_{k'k}$, $k'; k = 0, 1, 2$. In this application we assume the prior distributions $\sigma_e^{-2} \sim G(0.001, .0001)$ and $\sigma_k^{-2} \sim G(0.001, .0001)$, $k = 0, 1, 2$, for the hyperparameters. The estimates and standard deviations for the parameters of the model for the mean of the tallness in each one of the socioeconomic strata are given in Table 7. The estimate for the other parameters are: $\hat{\sigma}_e^2 = 4.527(3.549)$, $\hat{\sigma}_0^2 = 5.137(3.438)$, $\hat{\sigma}_1^2 = 0.1982(0.09646)$, $\hat{\sigma}_2^2 = 0.06458(0.004719)$.

TABLE 7: Posterior mean estimation on hierarchical models.

	γ_{k0}	γ_{k1}	γ_{k2}
$k = 0$	47.82 (2.371)	61.56 (2.272)	64.89 (2.061)
$k = 1$	10.25 (0.7447)	8.468 (0.606)	10.39 (0.593)
$k = 2$	-0.267 (0.045)	-0.171 (0.034)	-0.267(0.035)

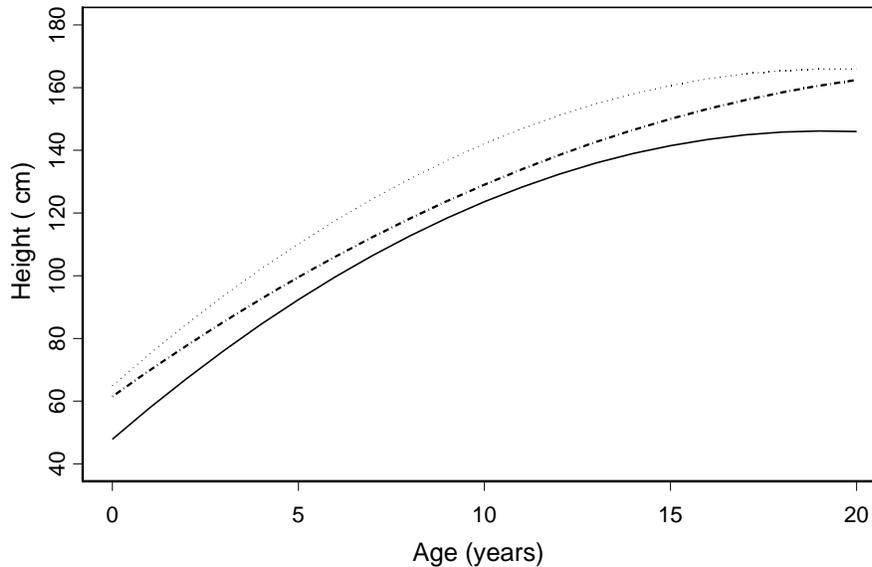


FIGURE 7: Mean tallness of people by socioeconomic level, hierarchical model. Continuous line for lower stratum, discontinuous line for medium stratum and dotted line for high stratum.

In Figure 7, we have the plot of the mean tallness against time for each one of the assumed stratum. We can see an agreement between average tallness behaviour showed in this figure and the behaviour showed in Figure 5, where the analysis was conducted using joint modelling of the mean and variance heterogeneity. In this example, we further observe considerable differences in the BIC values considering the different models as we can see in Table 8. From the results of this table, we again conclude that the model with joint modeling of the mean and of the variance heterogeneity is better fitted by the data than the hierarchical model.

TABLE 8: Model comparison.

Model	SSE	ln L	BIC
Variance heterogeneity	24842.694	-1113.897	387.928
Hierarchical	28610.283	-3397.753	1183.688

6. Concluding Remarks

From this comparative study for the two proposed models, we conclude that it is better to jointly model the mean and variance heterogeneity, as observed in all examples, in comparison to the use of hierarchical models. Additionally, convergence of the chain used to simulate samples for the posterior distribution of interest is quickly achieved and the model is easily interpretable. The proposed analysis can be convenient in social applications since these models perfectly capture any clustering by subgroups that may exist in the data. Other good reason for the modelling of the variance heterogeneity is related to the use of the explanatory variables in the regression model, since we can use the same explanatory variables assumed for the regression model for the mean or other different variables, for each one of the assumed groups. That is, this analysis also takes into consideration one of the most important goals of the multilevel statistical analysis, that is, it “substantively take into account for causal heterogeneity”. Although there is no statistical differences between intercepts in the growing curves for medium and high strata, apparently there is an advantage for high strata through time. We also observe that the joint modeling of mean and variance heterogeneity is simpler and easier to interpret.

[Recibido: marzo de 2009 — Aceptado: noviembre de 2009]

References

- Adair, L. S., Eckhardt, C. L., Gordon-Larsen, P. & Suchindran, C. (2005), ‘The Association Between Diet and Height in the Postinfancy Period Changes with Age and Socioeconomic Status in Filipino Youths’, *The Journal of Nutrition* **135**(9), 2192–2198).
- Aitkin, M. (1987), ‘Modelling Variance Heterogeneity in Normal Regression using Glim’, *Applied Statistics* **36**(4), 332–339.
- Bryk, A. & Raudenbush, S. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage publications, Inc, Newbury Park, United States.
- Cepeda, E. & Gamerman, D. (2001), ‘Bayesian Modeling of Variance Heterogeneity in Normal Regression Models’, *Brazilian Journal of Probability and Statistics* **14**(1), 207–221.
- Cepeda, E. & Gamerman, D. (2005), ‘Bayesian Methodology for Modeling Parameters in the two Parameter Exponential Family’, *Revista Estadística* **57**(168-169), 93–105.

- Chumlea, W. C., Guo, S. S., Wholihan, K., Cockram, D., Kuczmarski, R. J. & Johnson, C. L. (1998), 'Stature Prediction Equations for Elderly Non-Hispanic white, Non-Hispanic Black, and Mexican-American Persons Developed from NHANES III Data', *Journal of the American Dietetic Association* **98**(2), 137–142.
- De Leeuw, J. & Kreft, I. (1986), 'Random Coefficient Models for Multilevel Analysis', *Journal of Educational Statistics* **11**, 57–85.
- Longford, N. (1993), *Random Coefficient Models*, Oxford University Press, New York, United States.
- Prosser, R., Rasbash, J. & Goldstein, H. (1991), *ML3. Software for Three-Level Analysis. User's Guide for V. 2*, GB: Institute of Education, University of London, London, England.
- Raudenbush, S. & Bryk, A. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2 edn, Sage Publications, Inc., Thousand Oaks, United States.
- Steenbergen, M. & Bradford, S. (2002), 'Modeling Multilevel Data Structures', *American Journal of Political Science* **46**(1), 218–237.
- Stein, A. D., Barnhart, H. X., Wang, M., Hoshen, M. B., Ologoudou, K., Ramakrishnan, U., Grajeda, R., Ramírez, M. & Martorell, R. (2004), 'Comparison of Linear Growth Patterns in the first three Years of Life Across two Generations in Guatemala', *Pediatrics* **113**(3), 270–275.
- Van Der Leeden, R. (1998), 'Multilevel Analysis of Repeated Measures Data', *Quality & Quantity. Kluwer Academic Publishers. Netherlands* **32**, 15–29.

Appendix A.

Hierarchical Models

From the two level hierarchical model definition in Section 3, assuming independence between n_{0j} and n_{1j} , we have:

$$Y_{ij} \mid \beta_{0j}, \beta_{1j}, X_{ij}, \sigma^2 \sim N(\beta_{0j} + \beta_{1j}X_{ij}, \sigma^2)$$

for $i = 1, \dots, N_j$, $j = 1, \dots, J$, and

$$\beta_{0j} \mid \gamma_{00}, \gamma_{01}, Z_j, \tau_0 \sim N(\gamma_{00} + \gamma_{01}Z_j, \tau_0)$$

$$\beta_{1j} \mid \gamma_{10}, \gamma_{11}, Z_j, \tau_1 \sim N(\gamma_{10} + \gamma_{11}Z_j, \tau_1)$$

Thus the likelihood function is given by,

$$\begin{aligned}
 f(y | \beta_0, \beta_1, \sigma^2, X) &= \prod_{j=1}^J \prod_{i=1}^{N_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-1}{2\sigma^2} (y_{ij} - \beta_{0j} - \beta_{1j} X_{ij})^2 \right\} \\
 &\propto (\sigma^2)^{-\frac{1}{2} \sum_{j=1}^J N_j} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \beta_{0j} - \beta_{1j} X_{ij})^2 \right\}
 \end{aligned} \tag{11}$$

where $\beta_0 = (\beta_{01}, \dots, \beta_{0J})'$ and $\beta_1 = (\beta_{11}, \dots, \beta_{1J})'$.

To apply the Bayesian methodology, we need to specify a prior distribution function for θ . For simplicity, we assume independence between parameters, and assign the following prior distributions:

$$\begin{aligned}
 \sigma^2 &\sim IG(a, b) & \tau_0^2 &\sim IG(c_o, d_o) \\
 \tau_1^2 &\sim IG(c_1, d_1) & \gamma_{00} &\sim N(0, e_{00}) \\
 \gamma_{01} &\sim N(0, e_{01}) & \gamma_{10} &\sim N(0, e_{10}) \\
 \gamma_{11} &\sim N(0, e_{11})
 \end{aligned}$$

where $c_o, d_o, c_1, d_1, e_{00}, e_{01}, e_{10}, e_{11}$ are known constants.

With the assumed prior distribution and likelihood function given in (11), the posterior distribution for $\theta = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \tau_0, \tau_1, \beta_0, \beta_1)$ is given by

$$\begin{aligned}
 \pi(\theta | y, z, X) &\propto f(y | \beta_0, \beta_1, \sigma^2, X) \times \pi(\beta_0 | \gamma_{00}, \gamma_{01}, \tau_0, z) \pi(\beta_1 | \gamma_{10}, \gamma_{11}, \tau_1, z) \\
 &\quad \times \pi(\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma^2, \tau_0, \tau_1)
 \end{aligned}$$

where

$$\begin{aligned}
 \pi(\beta_0 | \gamma_{00}, \gamma_{01}, \tau_0, z) &\propto \prod_{j=1}^J \frac{1}{\sqrt{2\pi\tau_0}} \exp\left\{ \frac{-1}{2\tau_0} (\beta_{0j} - \gamma_{00} + \gamma_{01} Z_j)^2 \right\} \\
 &\propto (\tau_0)^{-\frac{J}{2}} \exp\left\{ \frac{-1}{2\tau_0} \sum_{j=1}^J \sum_{i=1}^{N_j} (\beta_{0j} - \gamma_{00} + \gamma_{01} Z_j)^2 \right\}
 \end{aligned}$$

$$\begin{aligned}
 \pi(\beta_1 | \gamma_{10}, \gamma_{11}, \tau_1, z) &\propto \prod_{j=1}^J \frac{1}{\sqrt{2\pi\tau_1}} \exp\left\{ \frac{-1}{2\tau_1} (\beta_{1j} - \gamma_{10} + \gamma_{11} Z_j)^2 \right\} \\
 &\propto (\tau_1)^{-\frac{J}{2}} \exp\left\{ \frac{-1}{2\tau_1} \sum_{j=1}^J \sum_{i=1}^{N_j} (\beta_{1j} - \gamma_{10} + \gamma_{11} Z_j)^2 \right\}
 \end{aligned}$$

$$\begin{aligned}
 \pi(\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma^2, \tau_0, \tau_1) &\propto \exp\left\{ -\frac{\gamma_{00}^2}{2e_{00}} \right\} \exp\left\{ -\frac{\gamma_{01}^2}{2e_{01}} \right\} \exp\left\{ -\frac{\gamma_{10}^2}{2e_{10}} \right\} \exp\left\{ -\frac{\gamma_{11}^2}{2e_{11}} \right\} \\
 &\quad \times (\tau_0)^{-(c_o+1)} e^{-d_o/\tau_0} (\tau_1)^{-(c_1+1)} e^{-d_1/\tau_1} (\sigma^2)^{-(a+1)} e^{-b/\sigma^2}
 \end{aligned}$$

Therefore, a joint posterior distribution for θ is given by:

$$\begin{aligned} \pi(\theta | y, z, X) &\propto (\sigma^2)^{-\frac{1}{2}} \sum_{j=1}^J N_j \exp\left\{\frac{-1}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \beta_{0j} - \beta_{1j} X_{ij})^2\right\} \\ &\times (\tau_0)^{-\frac{J}{2}} \exp\left\{\frac{-1}{2\tau_0} \sum_{j=1}^J \sum_{i=1}^{N_j} (\beta_{0j} - \gamma_{00} + \gamma_{01} Z_j)^2\right\} \\ &\times (\tau_1)^{-\frac{J}{2}} \exp\left\{\frac{-1}{2\tau_1} \sum_{j=1}^J \sum_{i=1}^{N_j} (\beta_{1j} - \gamma_{10} + \gamma_{11} Z_j)^2\right\} \\ &\times \exp\left\{-\frac{\gamma_{00}^2}{2e_{00}}\right\} \exp\left\{-\frac{\gamma_{01}^2}{2e_{01}}\right\} \exp\left\{-\frac{\gamma_{10}^2}{2e_{10}}\right\} \exp\left\{-\frac{\gamma_{11}^2}{2e_{11}}\right\} \\ &\times (\tau_0)^{-(c_0+1)} e^{-d_0/\tau_0} (\tau_1)^{-(c_1+1)} e^{-d_1/\tau_1} (\sigma^2)^{-(a+1)} e^{-b/\sigma^2} \end{aligned}$$

where $\sigma^2 > 0$, $\tau_0 > 0$, $\tau_1 > 0$, $\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \epsilon \mathbb{R}$ and $\beta_{1j}, \beta_{1j} \in \mathbb{R}; j = 1, \dots, J$. The parameter vector $\theta = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \sigma^2, \tau_0, \tau_1, \beta_{01}, \dots, \beta_{0J}, \beta_{11}, \dots, \beta_{1J})$ has 25+7 components.

Thus, the conditional posterior distribution is given by

$$\begin{aligned} \sigma^2 | \theta_{(\sigma^2)}, y, z, X &\sim IG\left\{a + \frac{1}{2} \sum_{j=1}^J N_j; b + \frac{1}{2} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \beta_{0j} - \beta_{1j} X_{ij})^2\right\} \\ \tau_0 | \theta_{(\tau_0)}, y, z, X &\sim IG\left\{c_0 + \frac{J}{2}; d_0 + \frac{1}{2} \sum_{j=1}^J (\beta_{0j} - \gamma_{00} - \gamma_{01} Z_j)^2\right\} \\ \tau_1 | \theta_{(\tau_1)}, y, z, X &\sim IG\left\{c_1 + \frac{J}{2}; d_1 + \frac{1}{2} \sum_{j=1}^J (\beta_{1j} - \gamma_{10} - \gamma_{11} Z_j)^2\right\} \\ \gamma_{00} | \theta_{(j_{00})}, y, z, X &\sim N\left(\frac{e_{00}^2 \sum_{j=1}^J \mu_{00j}}{\tau_0 + J e_{00}^2}; \frac{\tau_0 e_{00}^2}{\tau_0 + J e_{00}^2}\right) \\ \gamma_{10} | \theta_{(j_{10})}, y, z, X &\sim N\left(\frac{e_{10}^2 \sum_{j=1}^J \mu_{10j}}{\tau_1 + J e_{10}^2}; \frac{\tau_1 e_{10}^2}{\tau_1 + J e_{10}^2}\right) \\ \gamma_{01} | \theta_{(j_{01})}, y, z, X &\sim N\left(\frac{e_{01}^2 \sum_{j=1}^J z_j \mu_{01j}}{\tau_0 + e_{01}^2 \sum_{j=1}^J z_j^2}; \frac{\tau_0 e_{01}^2}{\tau_0 + e_{01}^2 \sum_{j=1}^J z_j^2}\right) \\ \gamma_{11} | \theta_{(j_{11})}, y, z, X &\sim N\left(\frac{e_{11}^2 \sum_{j=1}^J z_j \mu_{11j}}{\tau_1 + e_{11}^2 \sum_{j=1}^J z_j^2}; \frac{\tau_1 e_{11}^2}{\tau_1 + e_{11}^2 \sum_{j=1}^J z_j^2}\right) \end{aligned}$$

where $\mu_{00j} = \beta_{0j} - \gamma_{01} Z_j; j = 1, \dots, J$; $\mu_{10j} = \beta_{1j} - \gamma_{11} Z_j; j = 1, \dots, J$; $\mu_{01j} = \beta_{0j} - \gamma_{00}; j = 1, \dots, J$; $\mu_{11j} = \beta_{1j} - \gamma_{10}; j = 1, \dots, J$.

For β' s, given that

$$\begin{aligned} \pi(\beta_{0j} \mid \theta_{(\beta_{0j})}, y, z, X) &\propto \exp\left\{-\frac{1}{2\tau_0}(\beta_{0j} - \gamma_{00} - \gamma_{01}Z_j)^2\right\} \\ &\times \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{N_j} (y_{ij} - \beta_{0j} - \beta_{1j}X_{ij})^2\right\} \end{aligned}$$

then all posterior conditional distributions for β_{0j} are given by:

$$\beta_{0j} \mid \theta_{(\beta_{0j})}, y, z, X \sim N\left(\frac{\sigma^2(j_{00} + j_{01}Z_j) + \tau_0 \sum_{i=1}^{N_j} a_{0ij}}{\sigma^2 + \tau_0 N_j}; \frac{\sigma^2 \tau_0}{\sigma^2 + \tau_0 N_j}\right)$$

where $a_{0ij} = y_{ij} - \beta_{1j}X_{ij}; j = 1, \dots, J$.

In the same way, given that

$$\begin{aligned} \pi(\beta_{1j} \mid \theta_{(\beta_{1j})}, y, z, X) &\propto \exp\left\{-\frac{1}{2\tau_1}(\beta_{1j} - \gamma_{10} - \gamma_{11}Z_j)^2\right\} \\ &\times \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{N_j} (y_{ij} - \beta_{0j} - \beta_{1j}X_{ij})^2\right\} \end{aligned}$$

all posterior conditional distributions for β_{0j} are given by

$$\beta_{1j} \mid \theta_{(\beta_{1j})}, y, z, X \sim N\left(\frac{\sigma^2(j_{10} + j_{11}Z_j) + \tau_1 \sum_{i=1}^{N_j} a_{1ij}}{\sigma^2 + \tau_0 \sum_{i=1}^{N_j} X_{ij}}; \frac{\sigma^2 \tau_1}{\sigma^2 + \tau_1 N_j}\right)$$

where $a_{1ij} = y_{ij} - \beta_{0j}; j = 1, \dots, J$.