CIMAT

CENTRO DE INVESTIGACION EN MATEMÁTICAS, A.C.

BAYESIAN ESTIMATION IN A MEASUREMENT ERROR MODEL IN NUTRIENT DENSITY STUDY APPLIED TO A SAMPLE OF MEXICAN POPULATION OF CHILDREN AND ADOLESCENTS, ENSANUT 2012

T E S I S

QUE PARA OBTENER EL GRADO DE
Maestra en Ciencias con Especialidad en Probabilidad y Estadística

PRESENTA:
Silvia Xochitl Ibañez Vara

DIRECTOR DE TESIS:
Dra. Alicia L. Carriquiry
Iowa State University

Guanajuato, Guanajuato Marzo 2014.

Dedico esta tesis a:

Mis papás Juan C. Ibañez y Silvia Vara, gracias a su esfuerzo y apoyo alcanzo esta meta.

Mi hermana Sofía por creer en mí y darme ánimos.

Mis abuelos, primos y tíos.

Eduardo Trujillo, gracias por todas tus atenciones y paciencia durante mi estancia en Ames, Iowa.

Oscar Santiago, Orimar Sauri y Arturo Jaramillo por compartir conmigo su entender y largas horas de estudios.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología que a través del Centro de Investigación en Matemáticas me brindó los recursos para estudiar un posgrado.

A la Dra. Alicia Carriquiry por haber aceptado la dirección de este trabajo y por las atenciones que tuvo para conmigo durante mi estancia en Iowa State University.

Al Dr. Joaquín Ortega por el apoyarme antes, durante y después del desarrollo de la esta tesis.

Al Dr. Rogelio Ramos y Dr. Andrés Christen por el tiempo dedicado a la lectura de este trabajo.

Silvia Xochitl Ibanez Vara
Guanajuato, Gto.
Marz0, 2014.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 2012 the *Instituto Nacional de Salud Pública* in Mexico (INSP) fielded a nationwide health and nutrition survey called ENSANUT 2012 [1]. The goal of the survey is to collect information to monitor and evaluate trends in health and nutrition of the Mexican population. Details about the design and objectives of ENSANUT 2012 can be found in [12].

The ENSANUT 2012 included a questionnaire called a 24-Hour Dietary Recall (R24Hr-2012) that aims to capture intake of food energy (calories), nutrients and non nutrient components from food and beverages that were consumed during the 24-hour period prior to the interview by each participant. From a public health perspective, however, what is of interest is the *usual* or *regular* nutrient intake which we envision to be the *long-run average* intake of the nutrient by the individual. Usual or regular intake is unobservable, unless daily intakes for a large sample of persons are collected over a long period of time. This is impractical, both from a cost and from a respondent burden perspective. Therefore, we typically have only one or two observations of daily intake per person. From a statistical viewpoint, this is equivalent to observing noisy measurements of the variables of interest.

The focus of this thesis is to describe the association between usual intake of sugary drinks and usual intake of nutrients in a sample of Mexican children and adolescents.

The difficulty here is that neither usual intake of the nutrients nor usual intake of sugary drinks are observable as discussed above. Given independent replicates of daily intake, however we can formulate a measurement error model in order to estimate the associations of interest.

In general, measurement error models are useful in cases where one or more of the variables in the model are not observable. When both the response and the predictor(s) are not observable, we sometimes refer to the regression problem as an errors in variables problem. Unless one is willing to make strong and typically unverifiable assumptions, having independent replicate observations from which to obtain information about the distribution of the measurement errors is critically important. In the case of ENSANUT, a replicate 24-hr recall was collected from a random sub-sample of about 10% of survey participants. While this replicate permits in principle estimating the distribution of the measurement errors in both the response and the predictor, we simplify this estimation problem by assuming normality of the measurement errors.

The Bayesian approach to estimation, which we adopt for our analyses, allows incorporation of prior information given what we know about the contextual framework, without which it may be difficult to estimate parameters in the absence of richer data at the level of the individual. This fact permits introducing more complexity (and thus, flexibility) into the measurement error models. For example, by formulating our model in a hierarchical manner, we can accommodate heterogeneity of measurement error variances across individuals. However, the challenge of the Bayesian approach is computational; we need

[1] The previous surveys are ENSANUT 2001 and ENN 1999.

to program a Markov chain Monte Carlo (MCMC) algorithm in order to approximate the posterior distributions of interest.

The package rjags, available in R, is able to implement a Gibbs sampler algorithm but only for models that are not very complex. In this last case it is necessary to program the algorithm from scratch.

Although measurement error models are not a new topic in Statistics, few articles address estimation in these models from a Bayesian point of view [1]. For this reason, presenting an application of measurement error models to ENSANUT data using a Bayesian approach is a valuable contribution. The models we propose to describe the association between usual intakes of sugary drinks and of nutrients have attractive attributes:

- They use all the information (replicates) available at the level of the individual.
- They account for both the between individual and the within individual variances in intakes.
- They are flexible enough to allow realistic assumptions about the measurement error variances in intake.

These characteristics are present in an heteroscedastic measurement error model with replicates using a Gibbs sampler algorithm programed in rjags package to estimate the parameters.

The results that we compute in this thesis are not to be generalized to the entire population of Mexican children and adolescents. This is because the ENSANUT is not a simple random sample of the Mexican population. To obtain results that are generalizable to the entire Mexican population we would need to carry out a weighted analysis, using the survey weights that are provided in the database. Because ENSANUT is far from self-weighting, we expect that results obtained using weights will differ significantly from those we present here. Thus, a limitation of our application is that results apply exclusively to the sample of children and adolescents used in the analyses. The next step is to carry out a weighted analyses using the same methodology, so that results can be more widely useful.

1.1 Structure of the Thesis

The thesis is composed of three chapters and a conclusions section.
In Chapter 2: we explain how the dataset used in our analyses was built by using in the application of the models described in Chapter 3 using the data provided by INSP. It will be described, in a general way, the origin of the data and the structure of ENSANUT 2012 data base. It will also be explained which are the important variables according to the purpose of this thesis and the variables that compose the data for analysis.

Chapter 3: we propose two variants of the measurement error models with replicates. The concept of measurement error model will be introduced, then the models that will be applied will be described and the methodology to estimate their parameters will be explained.

Chapter 4: here the parameters of models in the previous chapter are estimated using simulated data with similar characteristics to the data analyzed in Chapter 5. This chapter has the objective to analyze the performance of the credible intervals and compare them to the

models proposed.

Chapter 5: we present an application of the models introduced in Chapter 2. We explain how to fit the models in Section 3.1.1 and 3.1.2, and the results will be shown. The analysis of fitting the models will be compared in order to choose the best option to explain the association between consumption of sugary drinks and nutrient density.

Finally, we include a short conclusions section and an appendix.

Chapter 2

Data Description

The aim of this chapter is to describe the intake data provided by the Instituto Nacional de Salud Pública, Cuernavaca, México. First the structure of the ENSANUT 2012 data base will be described. Secondly the important variables according to the purpose of this thesis will be explained. Finally a description of the variables that compose the data used in the application will be given.

2.1 ENSANUT R24Hr-2012 Data Base

In 2012 the Instituto Nacional de Salud Pública (INSP) implemented the Encuesta Nacional de Salud y Nutrición (ENSANUT) 2012ENSANUT is a nation-wide survey designed to be representative at the level of regions. Food intake information is collected using a questionnaire called 24-Hour Dietary Recall (R24Hr-2012) which is meant to capture individual total intake of foods, beverages and supplements during a 24-hour period. Using extensive food composition databases, consumption of food and beverages is transformed into its components: energy (calories), nutrients and non-nutrient components from food and beverages that were consumed during a 24-hour period prior to the interview.

Neither the handicapped nor children younger than 15 years were interviewed directly; instead, information was collected from the person responsible to feed the family. Children and adolescents were asked to confirm the information provided by the care-taker.[1]

The INSP obtained a replicated, independent recall for a subsample of those who participated in the first interview; thus R24Hr-2012 information is organized in two data sets; one for the first recall and the other for the second. The first recall data set contains information on about 10,886 individuals from which a second data set of 981 individuals were interviewed again. Both sets were built in the same way; each row in data base corresponds to one beverage or food consumed by the individual with certain folio number, both data bases include a total of 214 variables grouped into six types (See Table 2.1). These types of variables are described.

- Folio: it identifies the individual interviewed.
- Demographic: Socioeconomic characteristics.
- Anthropometric: Physical characteristics.
- Intake: How often, where, how etc. the interviewed person consumes the product in question.
- Characteristics of food: it includes codes for the type of food, the quantity consumed, the ingredients etc.

[1] More informations about R24Hr-2012 consult [10].

Variable Class	Variable Type	Number of Variables
Interviewee	Folio	1
	Demographic	22
	Anthropometric	5
Habits	Intake	21
Foods	Characteristics of food	35
	Decomposition	130
	Total	214

Table 2.1: R24Hr-2012 Data Base Variables.

Age Group	First Recall	Second Recall
Preschool	2,655	231
School	2,783	261
Adolescents	2,138	197
Total	7,576	689

Table 2.2: Number of interviewees per Age Group.

- Composition: Nutrient and no nutrient content in each food.

Here only a subset of the variables will be considered, since the focus of this thesis is on the association between nutrient intake and sugary drinks.

2.1.1 Target Population

The R24Hr-2012 data sets contain the variable called *GrupoPOB* that classifies the information in four groups of different intervals of ages:

- *Preschool*: from 1 to <5 years old
- *School*: from 5 to <12 years old
- *Adolescents*: from 12 to $\leq$ 19 years old
- *Adults*: at least 20 years old

The target population in our analyses includes Preschool, School and Adolescents. This study will not consider Adults in our work since we assume that children and young adults are more likely to consume sugary drinks in greater quantities. Therefore from now on, the description will include only the groups of interest. Table 2.2 shows the number of individuals interviewed per group of age in the first and second recall.

2.1.2 Sugary Drinks and Nutrients

The codification of R24Hr-2012 classified all non-alcoholic beverages into 22 different types of non alcoholic drinks. Among the sugary drinks consumed in Mexico, we select bever-

Calium	Iron	Magnesium	Phosphorous	Potasium
Zinc	Folate	Thiamin	Riboflavin	Niacin
Vitamin B6	Vitamin C	Vitamin A	Vitamin E	Vitamin D

Table 2.3: Nutrients Analyzed

ages with highest sugar content and highest consumption among children and adolescents. The sugary drinks we included in our analyses are:

- Industrialized flavored water.
- Sodas.
- Artificial fruit/vegetables drinks.

From now on when we refer to a sugary drinks, we mean the three beverages mentioned above. In terms of nutrients, see Table 2.3 :

- Total of each nutrient consumed 24 hrs. prior to recall j per individual i.
- Total of Kcal consumed 24 hrs. prior to recall j per individual i.
- Total of Kcal consumed 24 hrs. prior to recall j through sugary drinks per individual i.

Here $j = 1, 2$ and $i = 1, \dots n$. These totals will be indispensable in order to obtain the final data base as will be shown next.

2.1.3 Final Analyticial Data Base

Now a new set of variables will be computed by taking into account the nutrient and Kcal ingested by individuals. First we define the concept of Nutrient Density :

$$\text{Nutrient Density} = \frac{\text{Nutrient content}}{\text{Total Kcal content}} \times 100.$$

Thinking of diet quality in terms of nutrient density rather than nutrient amounts permits comparing individuals whose total intake differs due to age, body size, etc. Nutrient density, as the name suggests, is an indicator of the amount of a nutrient that is present in 100 kcal.
Once the totals of nutrients and calories ingested per each person were computed, the next variables can be calculated :

- l_{ij} : Total amount of nutrient l consumed by individual i 24 hrs. prior to recall j. Where l could be whatever nutrient in the table 2.3.
- $Kcal_{ij}$: Total kcal consumed by individual i 24 hrs. prior to recall j.
- $KcalSB_{ij}$: Total kcal that contain the sugary drinks consumed by individual i 24 hrs. prior to recall j.

Recall		First	Second
Total of Individuals		3,571	226
Columns	Folio	1	1
	GrupoPOB	1	1
	Body Mass Index	1	1
	Age	1	1
	X_{ij}	1	1
	$Y_{l_{ij}}$	15	15
Total of Columns		20	20

Table 2.4: Final Data Base per Recall.

Finally, we define two additional variables:

$$Y_{l_{ij}} = \frac{l_{ij}}{Kcal_{ij}} \tag{2.1}$$

$$X_{ij} = \frac{KcalSB_{ij}}{Kcal_{ij}}, \tag{2.2}$$

where $Y_{l_{ij}}$ is the density of nutrient l in the diet of participant i on day j of the survey and X_{ij} is the proportion of total calories in the diet that are contributed by sugary drinks for that person on that day.
It is worth mentioning that a considerable portion of the data set contains missing data or zero value in variables $Kcal_{ij}$ and $KcalSB_{ij}$. This loss of information could be produced by error in the cleaning data bases processes. Since to compute densities we need to have positive energy consumption in the day, all participants whose energy intake was missing or was recorded as zero were deleted. After filtering the information according to the type of beverages consumed and after eliminating the individuals with zero or NA in variables $Kcal_{ij}$ and $KcalSB_{ij}$, the data set of the first recall $j = 1$ contains information of n=3,571 individuals and the second recall $j = 2$, only contains 226 observations. The final data set will contain draws of equations (5.3) and (5.4) and the following variables from the original data set: Folio, Age and Body Mass Index. In table 2.4 the structure of the data sets that will be used in the next chapter are shown.

Chapter 3

Methodology

Two variants of the measurement error models with replicates will be described in this chapter. First the concept of measurement error model will be introduced, then the models that will be used in the application will be described and at the end the methodology to estimate the parameters of the models will be explained. A Bayesian approach will be used for which the distributions of unknown parameters and latent variables will be assigned.

3.1 Measurement Error Models

Let a classical linear regression model

$$y_i = x_i\beta + e_i \tag{3.1}$$

where x_i is a covariate and e_i is a random variable. When x_i is not observed directly but instead of observing x_i one observes X_i

$$X_i = x_i + u_i \tag{3.2}$$

where u_i is a random variable, then equations (3.1) and (3.2) comprise a measurement error model. If x_i is fixed, the model is functional, while a model with random x_i is called structural.

When the response variable y_i is not directly observed either, it means that one observes Y_i

$$Y_i = y_i + w_i \tag{3.3}$$

with w_i a random variable, then equations(3.1), (3.2) and (3.3) comprise a model with an error in the equation as described by [6].
Two structural models will be used in this thesis. The first model will have an error in the equation, it will be added a **z** observable variable and u_i and w_i will be assumed as heteroscedastic errors. The second model is a modification of a measurement error model assuming u_i as heteroscedastic errors as well. Both models assume that replicates are available to estimate the parameters of the measurement error distribution. In the following sections these models will be described in detail.

3.1.1 Model 1: Model with an Error in the Equation with Replicates

Consider an experiment where the response y_i , $i = 1, \ldots, n$ is independent among individuals i. Let x_i, and z_i be explanatory variables for the individual i. It will be assumed that there is a linear relation among y_i , x_i and z_i of the next form:

$$y_i = \beta_0 + x_i\beta_x + z_i\beta_z + q_i \quad , \quad q_i|\sigma_q^2 \sim (0, \sigma_q^2), \tag{3.4}$$

where β_x and β_z are the regression coefficients.

It will be denoted as $\beta = (\beta_{\mathbf{0}}, \beta_{\mathbf{x}}, \beta_{\mathbf{z}})$ the vector that contains all the coefficients. In the model (3.4) the variables x_i and y_i, are not directly observed, instead of observing them the experiment was replicated j times for some individual i and the observed variables Y_{ij} and X_{ij} will give information about $\mathbf{x}_i$ and y_i in the next way:

$$Y_{ij} = y_i + w_{ij} \quad , \quad w_{ij} \overset{ind}{\sim} (0, \sigma^2_{w_i}) \tag{3.5}$$

$$X_{ij} = x_i + u_{ij} \quad , \quad u_{ij} \overset{ind}{\sim} (0, \sigma^2_{u_i}). \tag{3.6}$$

The variables y_i and x_i are called latent variables , q_i, w_{ij} and u_{ij} are the measurement errors , these are random variables independent to the respective latent variable and among them. It will be assumed these errors have mean zero and variance $\sigma^2_{w_i}$ and $\sigma^2_{u_i}$. [1]

Notice that the three equations comprise a model with an error in the equation, if y_i in (3.4) is replaced by (3.5) then we obtain:

$$Y_{ij} = \beta_0 + \mathbf{x}_i\beta_x + \mathbf{z}_i\beta_z + q_i + w_{ij} \tag{3.7}$$

$$X_{ij} = x_i + u_{ij}. \tag{3.8}$$

It will be reasonable to assume a normal distribution for the measurement errors w_{ij} and u_{ij}, perhaps after an appropriate transformation of the data. Thus, we assume

$$\begin{aligned} q_i|\sigma^2_q &\overset{ind}{\sim} N\big(0, \sigma^2_q\big) \\ w_{ij}|\sigma^2_{w_i} &\overset{ind}{\sim} N\big(0, \sigma^2_{w_i}\big) \\ u_{ij}|\sigma^2_{u_i} &\overset{ind}{\sim} N\big(0, \sigma^2_{u_i}\big). \end{aligned} \tag{3.9}$$

The sum of errors $e_{ij} = q_i + w_{ij}$ in (3.7) is known as error in the equation . Then, replacing

$$e_{ij} = q_i + w_{ij}$$

in the equation (3.7) it is obtained,

$$Y_{ij} = \beta_0 + \mathbf{x}_i\beta_x + \mathbf{z}_i\beta_z + e_{ij} \tag{3.10}$$

$$X_{ij} = x_i + u_{ij}. \tag{3.11}$$

[1] In the homoscedastic case $\sigma^2_{w_i}$ and $\sigma^2_{u_i}$ are the same for all the individuals.

Consistent with a Bayesian approach it will be assumed that x_i has a conditional normal prior distribution with mean $\alpha_0 + \alpha_1 z_i$ and variance σ_x^2, then

$$x_i|\alpha_0, \alpha_1, \sigma_x^2 \sim N\big(\alpha_0 + \alpha_1 z_i, \sigma_x^2\big). \tag{3.12}$$

Using (3.9) and (3.10), it is obtained

$$\begin{aligned} Y_{ij}|\beta_0, \beta_x, \beta_z, x_i, y_i, \sigma_{w_i}^2, \sigma_{u_i}^2 &\sim N\big(\beta_0 + \mathbf{x}_i\beta_x + z_i\beta_z, \sigma_{u_i}^2 + \sigma_{w_i}^2\big) \\ X_{ij}|x_i, \alpha_0, \alpha_1, \sigma_{u_i}^2 &\sim N\big(x_i, \sigma_{u_i}^2\big), \\ y_i|\beta_0, \beta_x, \beta_z, \mathbf{x}_i, \sigma_q^2 &\sim N\big(\beta_0 + x_i\beta_x + z_i\beta_z, \sigma_q^2\big). \end{aligned} \tag{3.13}$$

In the model with an error in the equation with replicates that we have introduced, parameters $\beta_0, \beta_x, \beta_z, \sigma_q^2, \sigma_{w_i}^2, \sigma_{u_i}^2, \alpha_0, \alpha_1$ are unknown and the variables y_i and x_i are no observable directly. This makes estimation of model parameters challenging. The Gibbs sampler algorithm is an option to obtain simulations of y_i and $\mathbf{x}_i$ from their respective posterior distributions in order to estimate the model parameters. In Section 3.2 we describe how to construct this algorithm.

3.1.2 Model 2: Modified Measurement Error Model with Replicates

Consider now an experiment where the responses Y_{ij} , $i = 1, \ldots, n$ are independent between individuals i. Let x_i a explanatory variable for the individual i. It will be assumed that there is a linear relation between Y_{ij} and x_i in the next form:

$$Y_{ij} = \beta_0 + x_i\beta_x + w_{ij} \quad , \quad w_{ij}|\sigma_w^2 \sim (0, \sigma_w^2) \tag{3.14}$$

$$X_{ij} = x_i + u_{ij} \quad , \quad u_{ij}|\sigma_{u_i}^2 \sim (0, \sigma_{u_i}^2), \tag{3.15}$$

where the errors u_{ij} and w_{ij} were defined in (3.7). According to a Bayesian approach, x_i is assumed to have a conditional normal prior distribution with parameters:

$$x_i|\mu_x, \sigma_x^2 \sim N\big(\mu_x, \sigma_x^2\big) \tag{3.16}$$

From (3.14) and (3.15) it is obtained

$$\begin{aligned} Y_{ij}|\beta_x, x_i, \sigma_w^2 &\sim N\big(\beta_0 + x_i\beta_x, \sigma_w^2\big) \\ X_{ij}|x_i, \sigma_{u_i}^2 &\sim N\big(x_i, \sigma_{u_i}^2\big). \end{aligned} \tag{3.17}$$

In this model, unlike the previous model, all we have to estimate is x_i and the parameters $\beta_0, \beta_x, \sigma_w^2, \mu_x, \sigma_u^2, \sigma_x^2$. As in the previous case, we design a Gibbs Sampler to draw values from the posterior distributions of the parameters. The method will be explained in next section.

3.2 Estimation via Gibbs Sampler

The Gibbs sampler works iteratively, and it is useful in many mutidimensional problems where the objective is to estimate the posterior distribution of

$$\theta = \big(\theta_1, \ldots, \theta_d\big).$$

Each iteration of the algorithm cycles through the subvectors of θ, drawing each subset conditional on the value of all others. There are d steps in the iteration t. At each iteration t, an ordering of the components of θ is chosen and, in turn, each θ_j^t is sampled from the conditional distribution given all the other components of θ:

$$\mathbf{p}\big(\theta_j|\theta_{-j}^{t-1}, y\big), \tag{3.18}$$

where

$$\theta_{-j}^{t-1} = \big(\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \ldots, \theta_d^{t-1}\big).$$

Thus, each subvector θ_j is updated conditional on the latest values of the other components of θ, which are the iteration t values for the components already updated and the iteration $t-1$ values for the others [7]. The generic algorithm Gibbs is given below:

1. For any j not selected before, $\mathbf{p}(\theta_j^t|\theta^t$ without $\theta_j^t, y)$.

2. Simulate from the conditional distribution $\mathbf{p}(\theta_j^t|\theta^t$ without $\theta_j^t, y)$ and obtain θ_j^{t+1}.

3. Update $\theta^t = (\theta_1^t, \ldots, \theta_j^{t+1}, \theta_{j+1}^t, \ldots, \theta_n^t, y)$, with $t = 1, \ldots, n$

4. Repeat.

At the end of the iteration t all the vector θ must to be updated. On other hand, at the beginning of the algorithm initial values for θ have to be specified to start the iterations.

In Section 3.2.1 and 3.2.2 the conditional distributions that allow to simulate from (3.18) will be given for the respective model described above.

3.2.1 Gibbs Sampler for Model 1

According to the model in section 3.1.1 and supposing that $\beta_0, \beta_x, \beta_z, \sigma_q^2, \sigma_{w_i}^2, \sigma_{u_i}^2, \alpha_0, \alpha_1$ are given, Y_{ij} , $\mathbf{X}_{ij}$ and $\mathbf{x}_i$ are conditionally independent [1], then the likelihood for one individual with m_i replicates is

$$\begin{aligned}
\mathbf{p}(\mathbf{Y}_i, \mathbf{X}_i, x_i, y_i|z_i, \beta, \alpha, \sigma_q^2, \sigma_x^2, \sigma_{w_i}^2, \sigma_{u_i}^2) \propto & \prod_{j=1}^{m_i} \mathbf{p}(Y_{ij}|\beta, x_i, y_i, \sigma_{w_i}^2) \\
& \times \prod_{j=1}^{m} \mathbf{p}(X_{ij}|x_i, \alpha, \sigma_{u_i}^2) \\
& \times \mathbf{p}(x_i|\alpha, \sigma_x^2) \\
& \times \mathbf{p}(y_i|z_i, \beta, x_i, \sigma_q^2),
\end{aligned} \tag{3.19}$$

where

$$\beta = (\beta_0, \beta_x, \beta_z)$$

$$\alpha = (\alpha_0, \alpha_1)$$

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})$$

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{im_i}).$$

Then the target (or posterior) distribution is:

$$\begin{aligned}\mathbf{p}(\beta, \alpha, \sigma_q^2, \sigma_x^2, \sigma_{w_i}^2, \sigma_{u_i}^2 | \mathbf{Y}, \mathbf{X}, \mathbf{x}, \mathbf{y}, \mathbf{z}) \propto & \prod_{i=1}^{n} \mathbf{p}(\mathbf{Y}_i, \mathbf{X}_i, x_i, y_i | z_i, \beta, \alpha, \sigma_q^2, \sigma_{w_i}^2, \sigma_{u_i}^2) \\ & \times \mathbf{p}(\beta)\mathbf{p}(\alpha)\mathbf{p}(\sigma_q^2)\mathbf{p}(\sigma_x^2)\mathbf{p}(\sigma_{w_i}^2)\mathbf{p}(\sigma_{u_i}^2)\end{aligned} \tag{3.20}$$

Under the assumptions that (3.12) and (3.13) and replacing (3.19) in (3.20) the target distribution is expressed by,

$$\begin{aligned}\mathbf{p}(\beta, \alpha, \sigma_q^2, \sigma_{w_i}^2, \sigma_{u_i}^2 | \mathbf{Y}, \mathbf{X}, \mathbf{x}, \mathbf{y}, \mathbf{z}) \propto & \prod_{i=1}^{n} \Bigg[\big(\sigma_{w_i}\sigma_{u_i}\big)^{-m_i} \big(\sigma_x \sigma_q\big)^{-1} \\ & \times exp\Big\{ -\frac{1}{2\sigma_q^2}(y_i - C_i^t\beta)^2 - \frac{1}{2\sigma_x^2}(x_i - D_i^t\alpha)^2 \Big\} \\ & \times exp\Big\{ \frac{-\sum_{j=1}^{m_i}(X_{ij} - x_i)^2}{2\sigma_{u_i}^2} - \frac{\sum_{j=1}^{m_i}(Y_{ij} - y_i)^2}{2\sigma_{w_i}^2} \Big\} \Bigg] \\ & \times \mathbf{p}(\sigma_q^2)\mathbf{p}(\sigma_x^2)\mathbf{p}(\sigma_{w_i}^2)\mathbf{p}(\sigma_{u_i}^2).\end{aligned} \tag{3.21}$$

Here $C_i = (1, z_i, x_i)$ and $D_i = (1, z_i)$.

Following an application in Ruppert(2006) [9] the priors that will be used are, [2]

[2] IG meaning Inverse Gamma.

$$\begin{aligned}
\mathbf{p}(\beta) &\sim N_3\big(0, \sigma_\beta^2 \mathbf{I}\big) \\
\mathbf{p}(\alpha) &\sim N_2\big(0, \sigma_\alpha^2 \mathbf{I}\big) \\
\mathbf{p}(\sigma_q^2) &\sim IG(\gamma_q, \delta_q) \\
\mathbf{p}(\sigma_{w_i}^2) &\overset{iid}{\sim} IG(\gamma_w, \delta_w) \\
\mathbf{p}(\sigma_{u_i}^2) &\overset{iid}{\sim} IG(\gamma_u, \delta_u) \\
\mathbf{p}(\sigma_x^2) &\sim IG(\gamma_x, \delta_x).
\end{aligned} \tag{3.22}$$

The hyperparameters $\sigma_\beta^2, \sigma_\alpha^2, \gamma_*, \delta_*$ are constants. These constants will be specified later. Thus the full conditionals for each parameter taking into account m_i replicates per individual i is given by, [3]

$$\begin{aligned}
\mathbf{p}(x_i|others) &\propto exp\Bigg\{ -\frac{\beta_x^2 x_i - 2x_i(\beta_0 + \beta_z z_i - y_i)}{2\sigma_q^2} - \frac{x_i^2 - 2x_i(\alpha_0 - \alpha_1 z_i)}{2\sigma_x^2} - \frac{\sum_{j=1}^{m_i}\big(X_{ij} - x_i\big)^2}{\sigma_{u_i}^2} \Bigg\} \\
&= exp\Bigg\{ -\frac{1}{2}\bigg(\frac{\beta_x^2}{\sigma_q^2} + \frac{1}{\sigma_x^2} + \frac{m_i}{\sigma_{u_i}^2}\bigg)x_i^2 - 2x_i\bigg(\frac{\beta_0 + \beta_z z_i - y_i}{\beta_z} + \frac{\alpha_0 + \alpha_1 z_i}{\sigma_x^2} + \frac{m_i \bar{X}_{ij}}{\sigma_{u_i}^2}\bigg)\Bigg\}
\end{aligned}$$

$$\therefore \mathbf{p}(x_i|others) \sim N\Bigg(\bigg[\frac{\beta_x^2}{\sigma_q^2} + \frac{1}{\sigma_x^2} + \frac{m_i}{\sigma_{u_i}}\bigg]^{-1}\bigg(\frac{\beta_0 + \beta_z z_i - y_i}{\beta_x} + \frac{\alpha_0 + \alpha_1 z_i}{\sigma_x^2} + \frac{m_i \bar{X}_{ij}}{\sigma_{u_i}^2}\bigg), \bigg[\frac{\beta_x^2}{\sigma_q^2} + \frac{1}{\sigma_x^2} + \frac{m_i}{\sigma_{u_i}}\bigg]^{-1}\Bigg). \tag{3.23}$$

$$\begin{aligned}
\mathbf{p}(y_i|others) &\propto exp\Bigg\{ -\frac{1}{2\sigma_q^2}\big(y_i - \beta_0 - \beta_x x_i - \beta_z z_i\big)^2 - \frac{\sum_{j=1}^{m_i}\big(Y_{ij} - y_i\big)^2}{2\sigma_{w_i}^2} \Bigg\} \\
&= exp\Bigg\{ \frac{y_i^2 - 2y_i(\beta_0 + \beta_x x_i + \beta_z z_i)}{2\sigma_q^2} - \frac{m_i y_i^2 - 2m_i \bar{Y}_{ij} y_i}{2\sigma_{w_i}^2} \Bigg\} \\
&= exp\Bigg\{ y_i^2\bigg(\frac{1}{\sigma_q^2} + \frac{m_i}{\sigma_{w_i}^2}\bigg) - 2y_i\bigg(\frac{\beta_0 + \beta_x x_i + \beta_z z_i}{\sigma_q^2} + \frac{m_i \bar{Y}_{ij}}{\sigma_{w_i}^2}\bigg)\Bigg\}
\end{aligned}$$

$$\therefore \mathbf{p}(y_i|others) \sim N\Bigg(\bigg[\frac{1}{2\sigma_q 2} + \frac{m_i}{\sigma_{w_i}^2}\bigg]^{-1}\bigg(\frac{\beta_0 + \beta_x x_i + \beta_z z_i}{\sigma_q^2}\bigg), \bigg[\frac{1}{2\sigma_q^2} + \frac{m_i}{\sigma_{w_i}^2}\bigg]^{-1}\Bigg) \tag{3.24}$$

[3] Rule: if some p-dimensional parameter θ,

$$\mathbf{p}(\theta|others) \propto exp\big\{ -(\theta^t A\theta - 2b\theta)/2\big\}$$

where the constant of proportionality is independent of θ, then $\mathbf{p}(\theta|others)$ is $N(A^{-1}b, A^{-1})$.

$$\mathbf{p}(\beta|others) \propto exp\Big\{ -\frac{1}{2\sigma_q^2}\sum_{i=1}^{n}(y_i - C_i^t\beta)^t(y_i - C_i^t\beta) + \frac{\beta^t\beta}{2\sigma_\beta^2}\Big\}$$

$$\text{let} \quad C = [C_1^t, \ldots, C_n^t] \text{a matrix}$$

$$\therefore \mathbf{p}(\beta|others) \sim N\Bigg(\Big[C^tC\sigma_q^{-2} + I_3\sigma_\beta^{-2}\Big]^{-1} C^t y\sigma_q^{-2}, \Big[C^tC\sigma_q^{-2} + I_3\sigma_\beta^{-2}\Big]^{-1}\Bigg) \tag{3.25}$$

$$\mathbf{p}(\alpha|others) \propto exp\Big\{ -\frac{1}{2\sigma_x^2}(\mathbf{x} - D\alpha)^t(\mathbf{x} - D\alpha) - \frac{1}{2\sigma_\alpha^2}\alpha^t\alpha\Big\}$$

$$\text{where} \quad D = [\mathbf{1_{(n\times 1)}} \quad \mathbf{z}] \text{ is a matrix}$$

$$\therefore \mathbf{p}(\alpha|others) \sim N\Bigg(\Big[D^tD + \delta_\alpha I_3\Big]^{-1} D^t\mathbf{x}, \Big[D^tD + \delta_\alpha I_3\Big]^{-1}\Bigg) \tag{3.26}$$

$$\mathbf{p}(\sigma_q^2|others) \propto \big(\sigma_q^2\big)^{-n/2} exp\Big\{ -\frac{1}{2\sigma_q^2}\sum_{i=1}^{n}\big(y_i - \beta_0 - \beta_x x_i - \beta_z z\big)^2 - \frac{\delta_q}{\sigma_q^2}\Big\}\big(\sigma_q^2\big)^{-\gamma_q+1}$$

$$= \big(\sigma_q^2\big)^{-\left(\gamma_q + n/2 - 1\right)} exp\Big\{ -\frac{\delta_q + \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_x x_i + \beta_z z_i))^2}{\sigma_q^2}\Big\}$$

$$\therefore \mathbf{p}(\sigma_q^2|others) \sim IG\Bigg(\gamma_q + n/2, \delta_q + \sum_{i=1}^{n}\big(y_i - (\beta_0 + \beta_x x_i + \beta_z z_i)\big)^2\Bigg) \tag{3.27}$$

$$\mathbf{p}(\sigma_x^2|others) \sim exp\Big\{ -\frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i - \alpha_0 - \alpha_1 z_i)^2 - \frac{\delta_x}{\sigma_x^2}\Big\}\big(\sigma_x^2\big)^{-(\sum_{i=1}^{n} m_i/2 - \gamma_x - 1)}$$

$$\therefore \mathbf{p}(\sigma_x^2|others) \sim IG\Bigg(\gamma_x + \sum_{i=1}^{n} m_i/2, \delta_x + \sum_{i=1}^{n}\big(x_i - (\alpha_0 + \alpha_x z_i)\big)^2\Bigg) \tag{3.28}$$

$$\mathbf{p}(\sigma_{u_i}^2|others) \propto \big(\sigma_{u_i^2}\big)^{-m_i/2} exp\Big\{ -\frac{1}{2\sigma_{u_i}^2}\sum_{j=1}^{m_i}(X_{ij} - x_i)^2 - \frac{\delta_u}{2\sigma_u^2}\Big\}\big(\sigma_{u_i}^2\big)^{\gamma_u - 1}$$

$$\therefore \mathbf{p}(\sigma_{u_i}^2|others) \sim IG\Bigg(\gamma_u + m_i/2, \delta_u + \frac{\sum_{j=1}^{m_i}\big(X_{ij} - x_i\big)^2}{2}\Bigg) \tag{3.29}$$

$$\mathbf{p}(\sigma^2_{w_i}|others) \propto \left(\sigma_{w_i^2}\right)^{-m_i/2} exp\Bigg\{ -\frac{1}{2\sigma^2_{w_i}} \sum_{j=1}^{m_i} (Y_{ij} - y_i)^2 - \frac{\delta_w}{2\sigma^2_w} \Bigg\} \left(\sigma^2_{w_i}\right)^{\gamma_w - 1}$$

$$\therefore \mathbf{p}(\sigma^2_{w_i}|others) \sim IG\Bigg(\gamma_w + m_i/2, \delta_w + \frac{\sum_{j=1}^{m_i} \left(Y_{ij} - y_i\right)^2}{2}\Bigg) \tag{3.30}$$

Once we have derived the full conditionals of the parameters $\beta, \sigma^2_q, \sigma^2_x, \sigma^2_{w_i}, \sigma^2_{u_i}, \alpha$ and non observable variables x_i,y_i, a Gibbs sampler can be carried out as was shown at the beginning of this section (3.2). Let $\mathbf{p}(.|\theta^t$ without $\theta_.)$ a distribution expressed in (3.23) to (3.30),

$$\begin{aligned} \theta &= (x_i, y_i, \beta_0, \beta_x, \beta_z, \sigma^2_q, \sigma^2_x, \sigma^2_{w_i}, \sigma^2_{u_i}, \alpha_0, \alpha_1). \\ &= (\theta_1, \ldots, \theta_{2n+9}). \end{aligned} \tag{3.31}$$

Remember that the model is heteroscedastic in the errors u_{ij} and w_{ij}, therefore $2n + 9$ values will be obtained in each step, thus $\theta^t = (\theta^t_1, \ldots, \theta^t_{2n+9})$ is the estimation of the entries of the vector θ in the step t. The initial value θ_0 will be specified in the application.

3.2.2 Gibbs Sampler for Model 2

Now we list the full conditionals corresponding to Model (3.1.2), where the parameters are $\beta_x, \sigma^2_w, \sigma^2_{u_i}, \mu_x, \sigma^2_x$ and a non observable variable x_i have to be estimated. The likelihood for one individual with m_i replicates is [1]:

$$\begin{aligned} \mathbf{p}(\mathbf{Y}_i, \mathbf{X}_i, x_i|\beta_0, \beta_x, \sigma^2_x, \mu_x, \sigma^2_w, \sigma^2_{u_i}) \propto & \prod_{j=1}^{m_i} \mathbf{p}(Y_{ij}|\beta_x, x_i, \sigma^2_w) \\ & \times \prod_{j=1}^{m_i} \mathbf{p}(X_{ij}|x_i, \beta_0, \beta_x, \sigma^2_{u_i}) \\ & \times \mathbf{p}(x_i|\mu_x, \sigma^2_x), \end{aligned} \tag{3.32}$$

where

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})$$

$$\begin{aligned} \mathbf{X}_i &= (X_{i1}, \ldots, X_{im_i}). \\ \beta &= (\beta_0, \beta_x) \end{aligned}$$

The target distribution is:

$$\begin{aligned} \mathbf{p}(\beta_0, \beta_x, \mu_x, \sigma^2_w, \sigma^2_{u_i}, \sigma^2_x|\mathbf{Y}, \mathbf{X}, \mathbf{x}) \propto & \prod_{i=1}^{n} \Big[\mathbf{p}(\mathbf{Y}_i, \mathbf{X}_i, x_i|\beta_0, \beta_x, \mu_x, \sigma^2_{u_i}, \sigma^2_w, \sigma^2_x)\Big] \\ & \times \mathbf{p}(\beta_0, \beta_x)\mathbf{p}(\mu_x)\mathbf{p}(\sigma^2_w)\mathbf{p}(\sigma^2_u)\mathbf{p}(\sigma^2_x). \end{aligned} \tag{3.33}$$

Under the suppositions in (3.16) and (3.17) and replacing them in (3.33), the expression of the target distribution is:

$$\begin{aligned}\mathbf{p}(\beta_0,\beta_x,\mu_x,\sigma_w^2,\sigma_{u_i}^2,\sigma_x^2|\mathbf{Y},\mathbf{X},\mathbf{x}) \propto \prod_{i=1}^{n}\Bigg[&\big(\sigma_x^2\big)^{-1/2}\big(\sigma_w^2\sigma_{u_i}^2\big)^{\sum_{i=1}^n m_i/2}\\ &\times exp\Bigg\{-\frac{\sum_{j=1}^{m_i}(Y_{ij}-\beta_0-x_i\beta_x)^2}{2\sigma_w^2}-\frac{\sum_{j=1}^{m_i}(X_{ij}-x_i)^2}{2\sigma_{u_i}^2}-\frac{(x_i-\mu_x)^2}{2\sigma_x^2}\Bigg\}\Bigg]\\ &\times \mathbf{p}(\beta_0,\beta_x)\mathbf{p}(\mu_x)\mathbf{p}(\sigma_w^2)\mathbf{p}(\sigma_{u_i}^2)\mathbf{p}(\sigma_x^2). \end{aligned} \tag{3.34}$$

The priors proposed are the following:

$$\begin{aligned}\mathbf{p}(\beta) &\sim N\big(0,\sigma_\beta^2 I\big)\\ \mathbf{p}(\mu_x) &\sim N\big(\mu,\sigma_{\mu_x}^2\big)\\ \mathbf{p}(\sigma_w^2) &\sim IG\big(\gamma_w,\delta_w\big)\\ \mathbf{p}(\sigma_{u_i}^2) &\overset{iid}{\sim} IG\big(\gamma_u,\delta_u\big)\\ \mathbf{p}(\sigma_x^2) &\sim IG\big(\gamma_x,\delta_x\big)\end{aligned} \tag{3.35}$$

The parameters $\sigma_\beta^2,\mu,\sigma_{\mu_x}^2,\gamma_*,\delta_*$ are constants. From (3.34) and (3.35) the full conditionals are obtained.

$$\begin{aligned}\mathbf{p}(x_i|others) &\propto exp\Bigg\{-\frac{\sum_j^{m_i}(Y_{ij}-\beta_0-x_i\beta_x)^2}{2\sigma_w^2}-\frac{\sum_j^{m_i}(X_{ij}-x_i)^2}{2\sigma_{u_i}^2}-\frac{(x_i-\mu_x)^2}{2\sigma_x^2}\Bigg\}\\ \therefore \mathbf{p}(x_i|others) &\sim N\Bigg(\Bigg[m_i\Bigg(\frac{\beta_x^2}{\sigma_w^2}+\sigma_{u_i}^{-2}\Bigg)+\sigma_x^2\Bigg]^{-1}\Bigg(-\frac{\beta_x\bar{Y}_{ij}+\bar{X}_{ij}}{\sigma_w^2}+\frac{\mu_x}{\sigma_x^2}\Bigg),\Bigg[m_i\Bigg(\frac{\beta_x^2}{\sigma_w^2}+\sigma_{u_i}^{-2}\Bigg)+\sigma_x^2\Bigg]^{-1}\Bigg)\end{aligned} \tag{3.36}$$

Let $\beta=(\beta_0,\beta_x)$, $\mathbf{X}=[\mathbf{1}_{(\mathbf{n}\times\mathbf{1})}\quad \mathbf{x}]$ be a matrix, $\mathbf{m}$ is a vector with the number of repetitions per individual m_i, $M=\sum_i^n m_i$ and $\mathbf{Y}_j$ is a vector that contain all the observations in replicate j, then

$$\mathbf{p}(\beta|others)\propto exp\Bigg\{-\frac{\sum_{j=1}^{m_i}\Big(Y_{ij}-\mathbf{X}\beta\Big)^t\Big(Y_{ij}-\mathbf{X}\beta\Big)}{2\sigma_w^2}-\frac{\beta^t\beta}{2\sigma_\beta^2}\Bigg\}$$

$$\therefore \mathbf{p}(\beta|others)\sim N\Bigg(\Bigg[I\sigma_\beta^{-2}+\frac{M\mathbf{X}^t\mathbf{X}}{\sigma_w^2}\Bigg]^{-1}\frac{\mathbf{X}^t(\mathbf{m}\otimes\bar{Y}_j)}{\sigma_w^2},\Bigg[I\sigma_\beta^{-2}+\frac{M\mathbf{X}^t\mathbf{X}}{\sigma_w^2}\Bigg]^{-1}\Bigg) \tag{3.37}$$

$$\mathbf{p}(\mu_x|others) \propto exp\Bigg\{ -\mu_x^2\left(\frac{1}{2\sigma_x^2}+\frac{1}{2\sigma_{\mu_x}^2}\right) - \mu_x\left(\frac{\bar{x}}{\sigma_x^2}-\frac{\mu}{\sigma_{\mu_x}^2}\right)\Bigg\}$$

$$\therefore \mathbf{p}(\mu_x|others) \sim N\left(\left(\frac{1}{2\sigma_x^2}+\frac{1}{2\sigma_{\mu_x}^2}\right)^{-1}\left(\frac{\bar{x}}{\sigma_x^2}-\frac{\mu}{\sigma_{\mu_x}^2}\right), \left(\frac{1}{2\sigma_x^2}+\frac{1}{2\sigma_{\mu_x}^2}\right)^{-1}\right) \tag{3.38}$$

$$\mathbf{p}(\sigma_w^2|others) \propto \left(\sigma_w^2\right)^{n\sum_{i=1}^n m_i/2+\gamma_w+1} exp\Bigg\{ -\frac{\sum_i^n\sum_{j=1}^n(Y_{ij}-\beta_0-\mathbf{x}_i\beta_x)^2}{2}+\delta_w\Bigg)/\sigma_w^2\Bigg\}$$

$$\therefore \mathbf{p}(\sigma_w^2|others) \sim IG\left(\frac{n\sum_{i=1}^n m_i}{2}, \frac{\sum_{i=1}^n\sum_{j=1}^{m_i}(Y_{ij}-\beta_0-x_i\beta_x)^2}{2}+\delta_w\right) \tag{3.39}$$

$$\mathbf{p}(\sigma_{u_i}^2|others) \propto \left(\sigma_{u_i}^2\right)^{-\sum_{i=1}^n m_i/2-\gamma_u-1} exp\Bigg\{ -\frac{1}{\sigma_{\sigma_{u_i}^2}}\left(\frac{\sum_{i=1}^n(X_{ij}-x_i)^2}{2}+\delta_u\right)\Bigg\}$$

$$\therefore \mathbf{p}(\sigma_{u_i}^2|others) \sim IG\left(\frac{\sum_{i=1}^n m_i}{2}+\gamma_u, \frac{\sum_{i=1}^n(X_{ij}-x_i)^2}{2}+\delta_u\right) \tag{3.40}$$

$$\mathbf{p}(\sigma_x^2|others) \propto \left(\sigma_x^2\right)^{-\gamma_x-n/2-1} exp\Bigg\{ -\frac{1}{\sigma_x^2}\left(\frac{\sum_{i=1}^n(x_i-\mu_x)^2}{2}+\delta_x\right)+\delta_x\Bigg\}$$

$$\therefore \mathbf{p}(\sigma_x^2|others) \sim IG\left(\gamma_u+n/2, \frac{\sum_{i=1}^n(\mu_x-x_i)^2}{2}+\delta_x\right) \tag{3.41}$$

The full conditionals given above will be used in a Gibbs sampler algorithm in order to estimate the parameters.

How to apply the models given and how to use the Gibbs sampler to made estimations will be shown in the next chapter.

Chapter 4

Simulation Study

The aim we pursue in this chapter is to assess the performance of the algorithm we use to estimate the parameters in the models proposed in Chapter 2.

The vector of regression coefficients β in the models is of interest since the coefficients quantify the strength of the association between the response variable and consumption of sugary drinks. One of the difficulties we encountered in our analysis (as discussed in Chapter 3) that nutrient and sugary drinks intakes are highly variable, both within persons (between days) and also between persons. For this reason we try to simulate data with these characteristics to observe the quality of the credible intervals.

4.1 Simulated Data

Data were simulated according to model (3.13), (3.17) with the priors given in (3.2.1) and (3.35) they were used to asses performance of credible intervals.

Three data set with different values for the variance were simulated per each model. In Table 4.1 and 4.2 are the respective values used. The main interest of the simulation is to observe the behavior of the credible intervals under high variability in the parameters σ_{u_i} and σ_{w_i} because the real data to be analyzed in Chapter 5 have these characteristics.

Value of σ_{u_i} are drawn from an Inverse Gamma(2.5,6) distribution. The values for σ_{w_i} were simulated from three distributions with different mean and variance: Inverse Gamma(3,30), Inverse Gamma(2.5,6) and Inverse Gamma(3,1). These parameters are similar to the estimation obtained by moments method using the real data.

Parameter	Simulation 1	Simulation 2	Simulation 3
N. Observations	1000	1000	1000
N. Chains	100	100	100
α_0	5	5	5
α_1	1.5	1.5	1.5
β_0	70	70	70
β_x	-0.5	-0.5	-0.5
β_z	-0.1	-0.1	-0.1
$\bar{\sigma_{u_i}}$	4.45	4.45	4.08
$\bar{\sigma_{w_i}}$	14.68	5.05	0.17
σ_q	21	21	1
σ_x	5	5	5

Table 4.1: Parameters for simulation for Model 1.

[1]

[1]Mean of σ_{*_i} is denoted by $\bar{\sigma_{*_i}}$.

Parameter	Simulation 1	Simulation 2	Simulation 3
N. Observations	1000	1000	1000
N. Chains	100	100	100
β_0	70	70	70
β_x	-0.5	-0.5	-0.5
μ_x	5	5	5
$\sigma_{u_i}^-$	5.39	4.22	2.89
σ_x	5	5	5
σ_w	25	0.69	1

Table 4.2: Parameters of simulation for Model 2.

4.2 Expected credible intervals

100 chains of size 100,000 per simulation were simulating by using a burn-in of 10,000. After that, the expected quantile 2.5% and 97.5% were computed.

Parameter	$\sigma_{u_i}^-$ =4.45 $\sigma_{w_i}^-$ =14.85	$\sigma_{u_i}^-$ =4.45 $\sigma_{w_i}^-$ =5.05	$\sigma_{u_i}^-$ =4.33 $\sigma_{w_i}^-$ =0.17, σ_q =1
α_0 =5	(4.79 , 5.99)	(4.22 , 5.63)	(4.56 , 5.67)
α_1 =1.5	(0.99 , 1.83)	(0.93 , 1.89)	(1.04 , 1.82)
β_0 =70	(66.56 , 72.63)	(66.24 , 70.57)	(68.95 , 69.36)
β_x =-0.5	(-0.87 , -0.27)	(-0.68 , 0.08)	(-0.53 , -0.31)
β_z =-0.1	(-0.091 , 3.74)	(-1.88 , 0.88)	(-0.46 , -0.22)
σ_{u_i}	(1.92 , 2.09)	(2.02 , 2.19)	(1.71 , 3.16)
σ_{w_i}	(1.78 , 1.89)	(3.28 , 3.73)	(0.51 , 0.53)
σ_q =21	(21.44 , 30.15)	(20.32 , 22.21)	(0.66 , 1.86)
σ_x =5	(5.16 , 6.83)	(4.92 , 6.72)	(5.72 , 6.35)

Table 4.3: Expected quantiles of Model 1.

Parameter	$\sigma_{u_i}^-$ =5.39 σ_w =25	$\sigma_{u_i}^-$ =4.22 σ_w =0.69	$\sigma_{u_i}^-$ =4.33 σ_w =1
β_0 =7	(6.17 , 7.87)	(2.64 , 14.48)	(6.74 , 7.67)
β_1 =-0.5	(-0.43 , -0.02)	(-2.87 , -0.52)	(-1.17 , 1.47)
μ_x =5	(5.15 , 5.45)	(3.98 , 4.37)	(4.9 , 5.27)
$\sigma_{u_i}^-$	(2.18 , 2.37)	(3.26 , 3.45)	(2.84 , 3)
σ_x =5	(7.16 , 7.95)	(0.16 , 0.19)	(0.16 , 0.21)
σ_w	(24.25 , 26.36)	(0.6 , 0.81)	(0.9 , 1.45)

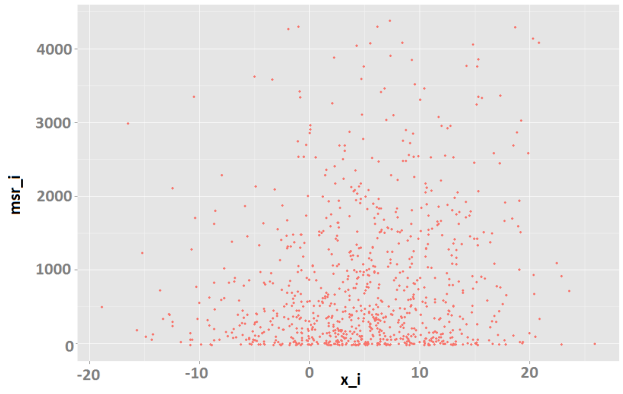
Table 4.4: Expected quantiles of Model 2.

From the results shown in Table 4.3, we observe that the means of the $\sigma's$ computed from the data are not contained in the intervals between the 2.5% and the 97.5% expected quantiles for $\sigma_{u_i}^-$ and $\sigma_{w_i}^-$. This, however, does not appear to impact the estimation of the regression coefficients. Further, the residuals computed from the model are plotted in Figure 4.2 [2] where it can be observed that there is not a pattern but there are high values of residuals. This behavior is expected because when the model is fitted it is supposed that the response and covariate are not observable and data have considerable variability. Taking all of this into account, we find that the expected credible sets are sufficiently narrow.

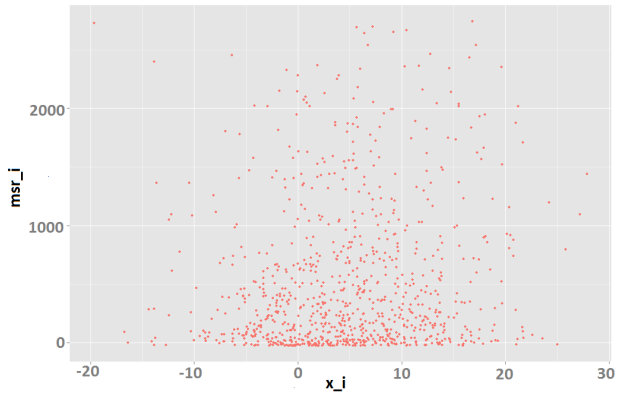
In Model 2, we observe a similar behavior. While the expected quantiles 2.5% and 97.5% for $\sigma_{u_i}^-$ and σ_w do not bracket the true parameter values, for the other parameters they do. Comparing the residuals based on simulations from model 2 in Figure 4.2, we notice that the residuals of model 2 are large in cases of high variance. When the variances σ_{w_i} are low, Model 2 produces smaller residuals than Model 1.

The heteroscedastic measurement errors in the variables make it challenging to estimate the regression parameters. It is worth mentioning that we also simulated the homoscedastic case to determine whether a simpler model might also be applicable to the type of data collected in ENSANUT. We failed to reach converge, but since the real data display clear heterogenity of variances, we did not pursue this issue any furrther.

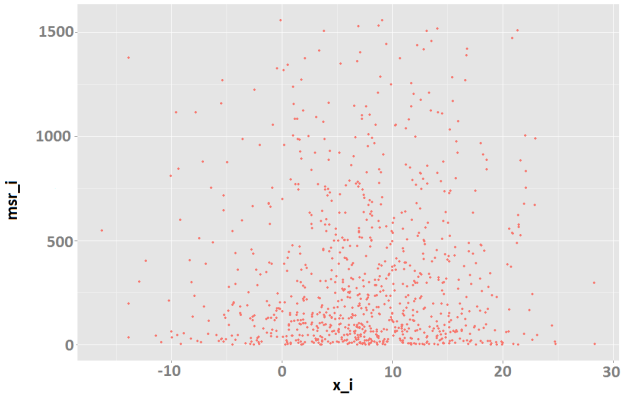
[2] Details about computation of msr are in Appendix B.1.



(a)

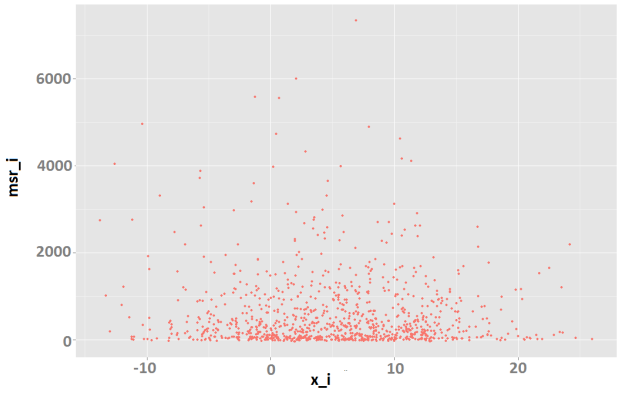


(b)

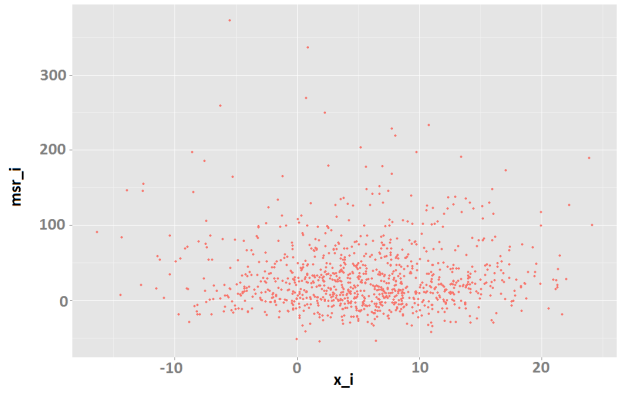


(c)

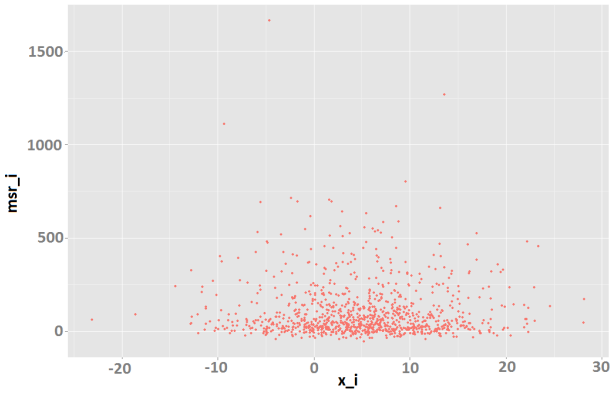
Figure 4.1: Residuals msr_i of simulation. In (a), (b), (c) are the residuals of simulations 1,2 and 3 for Model 1.



(a)



(b)



(c)

Figure 4.2: Residuals msr_i of simulation. In (a), (b), (c) are the residuals of simulations 1,2 and 3 for Model 2.

Chapter 5

Application

The models introduced in the last chapter will be fitted to consumption data collected by the 2013 ENSANUT. We first briefly describe the problem. Then we explain how to use the models in Section 3.1.1 and 3.1.2 and present some results. Finally we revisit the substantive problem and interpret our findings in terms of the association of sugary drinks intake and nutrient density in the diet of Mexican children and adolescents.

5.1 Question of interest

This application is focused on quantifying the association between intake of sugary drinks and usual consumption of certain nutrients in a sample of Mexican children and adolescents. This population is divided into three age groups: Preschool, School and Adolescents, as was mentioned in Section (2.1.1).

One way to describe the association between consumption of sugary drinks and nutrient density of the diet is through a linear relation between the regular density of the nutrient in question and the percentage of the kcal provided by the sugary drinks consumed regularly. More precisely, if we let

$$y_i = \frac{\text{Regular intake of nutrient } N}{\text{Regular intake of Kcal}}$$

$$x_i = \frac{\text{Regular intake of Kcal provided by sugary drinks}}{\text{Regular intake of Kcal}}, \tag{5.1}$$

then a linear model with error in the equation is given by

$$y_i = \beta_0 + \beta_x x_i + q_i \quad , \quad q_i \overset{ind}{\sim} (\mu, \sigma^2) \tag{5.2}$$

The linear relation in (5.2) permits quantifying, at an individual level, what happen with the usual (or regular) density of the nutrients of interest [1] if the proportion of the energy consumed by the individual provided by sugary drinks increases o decreases. Using proportion of kilo-calories as the metric for consumption of sugary drinks makes sense because in addition to added sugar, these beverages contribute close to nothing else to the diet. The difficulty in this problems is that regular intake of nutrients is not observable [5], however information from one or two 24 Hours Dietary Recalls can be used in order to estimate the regular intake of the nutrients of interest. Given the data described in Section 2.1 and supposing that we are interested in analyzing a specific nutrient N, we recall the definitions from Section 2.1.3:

[1] Nutrients of interest are in table 2.3

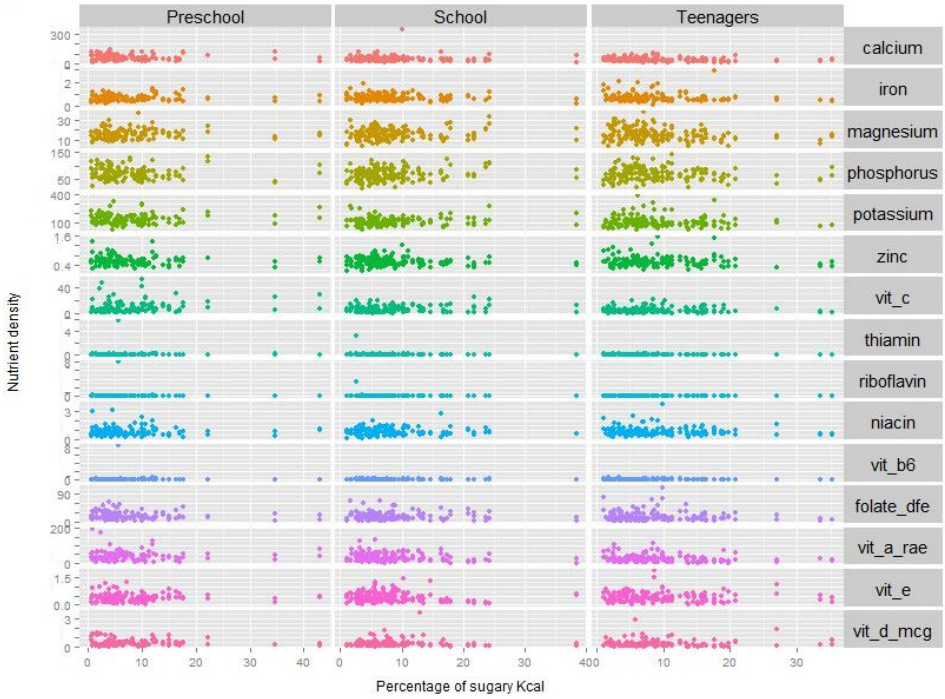


Figure 5.1: Linear relation between X and Y.

- l_{ij} : Total amount of a nutrient consumed by individual i on day j. Here l can represent any of the nutrients listed in the Table 2.3.
- $Kcal_{ij}$: Total kcal consumed by individual i on day j.
- $KcalSB_{ij}$: Total kcal contributed by sugary drinks consumed by individual i on day j.

Thus instead of observing x_i and y_i, we observe noisy measurements:

$$Y_{ij} = \frac{l_{ij}}{Kcal_{ij}} \tag{5.3}$$

$$X_{ij} = \frac{KcalSB_{ij}}{Kcal_{ij}} \tag{5.4}$$

It is important to notice that due to the measure error, the linear relation between Y_{ij} and X_{ij} cannot be observed, as it is shown in Figure 5.1, however the linearity assumption between x_i and y_i can still be assumed.

Given that the predictor and the response are not observed, and given that there are replicate measurements available on a sub-sample of individuals, we can fit a measurement error model to determine the relation between x_i and y_i.

The two measurement error models we fitted in this chapter assume heteroscedastic error in the measurement of x_i and y_i. This is a good characteristic of the model since it permits assuming that the within-individual variation of energy and nutrient consumption is different across persons, something which is known to actually happen.

The difference between the two models is two-fold; in the first model we assume that the response depends on x, an non observable variable, and z an observable variable. We use body mass index (BMI) as the observable covariate in the model . The second model assumes that measurement error of y only depends on w_{ij}, the error of the observation Y_{ij}.

The objective of this chapter is to fit both and compare them in order to select the a model for each nutrient. Before fitting the models to the ENSANUT data, we carried out a simulation study (see Chapter 4.4).

5.2 Model 1

Recall the model in Section (3.1.1) where

$$\begin{aligned} y_i &= \beta_0 + x_i\beta_x + z_i\beta_z + q_i, \quad q_i|\sigma_q^2 \overset{iid}{\sim} N(0,\sigma_q^2), \\ Y_{ij} &= y_i + w_i \quad , \quad w_{ij}|\sigma_{w_i}^2 \overset{iid}{\sim} N(0,\sigma_{w_i}^2) \\ X_{ij} &= x_i + u_{ij} \quad , \quad u_{ij}|\sigma_{u_i}^2 \overset{iid}{\sim} N(0,\sigma_{u_i}^2) \end{aligned} \tag{5.5}$$

where z is an observable variable and y_i, x_i are latent variables. To apply a model with an error in the equation the variables *BMI*, $Y_{N_{ij}}$ and X_{ij} will be taken from the final data base [2], as z, Y_{ij}, X_{ij} respectively.

Given the full conditional distributions and the priors distributions in (3.2.1) for the model parameters, it is necessary to specify the values of the hyper-parameters σ_β^2, σ_α^2, γ_q, δ_q, γ_w, δ_w,γ_u, δ_u, γ_x, δ_x to implement the Gibbs sampler algorithm.

In the case of α and β it will be used a non informative prior so

$$\sigma_\beta^2 = \sigma_\alpha^2 = 100.$$

The errors q_i and w_i are not observable and will be supposed to have small values because the largest source of variation is day-to-day variability in individual intake more than in the error of the regression. In order to ensure that the posterior distribution of σ_q and σ_w will have most of it mass on small values, we propose the following prior parameter values:

$$\gamma_q = \gamma_w = 3$$

$$\delta_q = \delta_w = 1.$$

The errors u_{ij} and w_{ij} are estimable from the replicate observations, therefore we can compute sample estimates of σ_u^2 and σ_w^2 in the following way:

$$\tilde{\sigma}_{u_i}^2 = \text{sample variance}\big(X_{i1}, X_{i2}\big)$$

$$\tilde{\sigma}_u^2 = (\tilde{\sigma}_{u_1}, \ldots, \tilde{\sigma}_{u_n})$$

$$\tilde{\sigma}_{w_i}^2 = \text{sample variance}\big(Y_{i1}, Y_{i2}\big)$$

$$\tilde{\sigma}_w^2 = (\tilde{\sigma}_{w_1}, \ldots, \tilde{\sigma}_{w_n}).$$

[2] See table 2.4

Variable	Nutrient Density	Adolescents		School		Preschool	
		γ_*	δ_*	γ_*	δ_*	γ_*	δ_*
X_{ij}	Sugar drinks Kcal	2.59	7.92	2.53	5.89	2.46	6.04
	Calcium	2.61	20.62	2.34	22.60	2.59	25.85
	Iron	2.32	0.26	2.53	0.24	2.51	0.25
	Magnaesium	2.56	6.19	2.59	6.25	2.56	4.93
	Phosphorus	2.63	27.21	2.63	27.69	2.67	23.44
	Potassium	2.48	49.45	2.51	43.40	2.56	53.61
	Zinc	2.51	0.19	2.57	0.18	2.51	0.18
	Vit C	2.49	4.23	2.54	5.16	2.39	6.64
Y_{ij}	Thiamin	2.53	0.03	2.04	0.04	2.02	0.22
	Riboflavin	2.32	0.03	2.04	0.04	2.02	0.24
	Niacin	2.38	0.39	2.43	0.39	2.43	0.42
	Vit B	2.31	0.05	2.43	0.06	2.04	0.16
	Folate	2.43	9.60	2.50	11.89	2.55	13.38
	Vit A	2.48	18.57	2.38	20.64	2.42	24.25
	Vit E	2.58	0.28	2.60	0.26	2.53	0.25
	Vit D	2.34	0.25	2.31	0.26	2.44	0.28

Table 5.1: Values for hyperparameters of Inv. Gamma(γ_*,δ_*) for the distribution of the error u_{ij} and.w_{ij} per group of age.

Assuming that σ_u^2 and σ_w^2 are Inverse Gamma distributed, using (5.2) γ_w,δ_w,γ_u,δ_u can be estimated by method of moments for each group of age:

$$\gamma_* = \frac{\left(\sum_{i=1}^n \tilde{\sigma}_{*_i}^2\right)^2}{n\sum_{i=1}^n (\tilde{\sigma}_{*_i}^2)^2} + 2$$

$$\delta_* = \frac{\sum_{i=1}^n \tilde{\sigma}_{*_i}^2}{n}\Big(\tilde{\alpha} - 1\Big) \tag{5.6}$$

We wish to fit the model separately for the three age groups defined earlier. The corresponding values of the hyperparameters are shown in table 5.1.

From the table, we observe that for some nutrients the values of the hyperparameters computed from the sample data are high. This will result in high values of the mean of σ_w^2 for these nutrients. This is particularly true for the Pre-school group. For the Adolescents there appears to be more variability in the intake percentage of kcal of sugary drinks. These facts could affect the residuals.

Before implementing the Gibbs sampler algorithm using real data, a simulation was made in order to asses the performance of the estimation, details about that are in Appendix 4.2.

The Gibbs sampler algorithm given in Section 3.2.1 was programed in R using the package rjags [11]. For each nutrient and for group of age were simulated two chains with

100,000 iterations and burn-in of 10,000.

The convergence of the chains and the fit of the models using residual analysis will be assessed and the results will be exposed in following sections.

5.2.1 Convergence

Two convergence test were used. First, we used the Geweke test [13] using proportions 10% and 50% for chains of size 90,000. In Table 5.2.1 the values of the statistic $Z - score$ are given for chains of each parameter, nutrient and group of age.

The Geweke test considers as evidence against the hypothesis of convergence an absolute value of $Z - score$ greater than 2.5. Under this criterion, one of the two chains of some of the parameters for a few nutrients show evidence against convergence. These instances are underlined in Table 5.2.1.

Shaded results correspond to α_0 and α_1 of Adolescents group and σ_x of Preschool in zinc. σ_u in iron of Preschool group. In thiamin β_0 of School, σ_w and σ_q of Preschool group. In riboflavin σ_u and σ_w of in School, and in folate σ_x of Preschool group.

The graphics of the chains for which the null hypothesis of convergence is rejected with Geweke test are shown in Figures 5.2 to 5.11 where both chains per parameter are compared. It can be observed that both chains have almost the same density estimation and also look like a convergent chain, therefore these chains will be assumed as convergent.

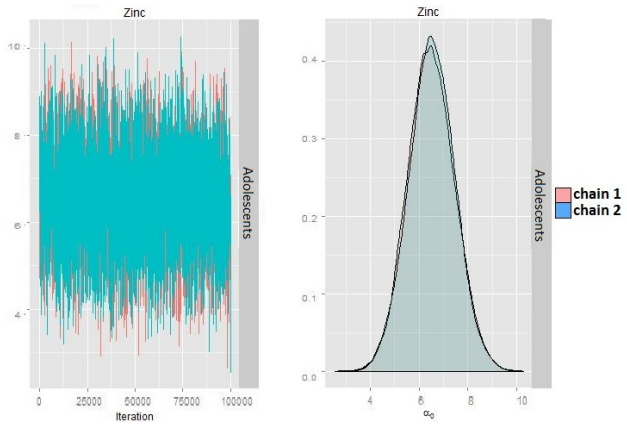


Figure 5.2: Geweke test refused convergence of chain from α_0 Adolescents group, Zinc.

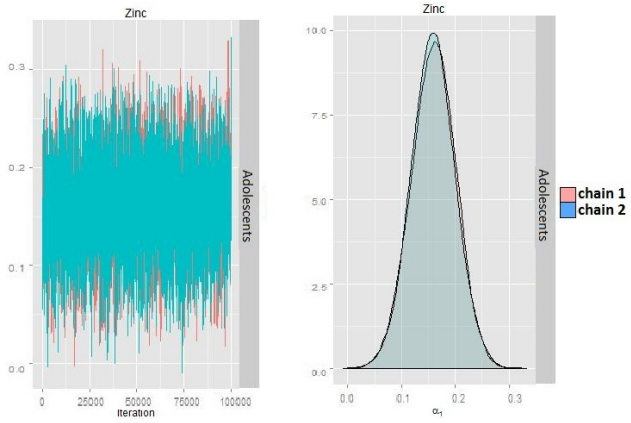


Figure 5.3: Geweke test refused convergence of chain from α_1 Adolescents group, Zinc.

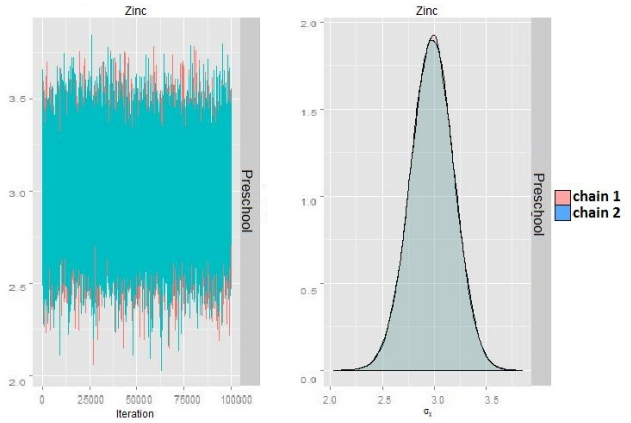


Figure 5.4: Geweke test refused convergence of chain from σ_x Preschool group, Zinc.

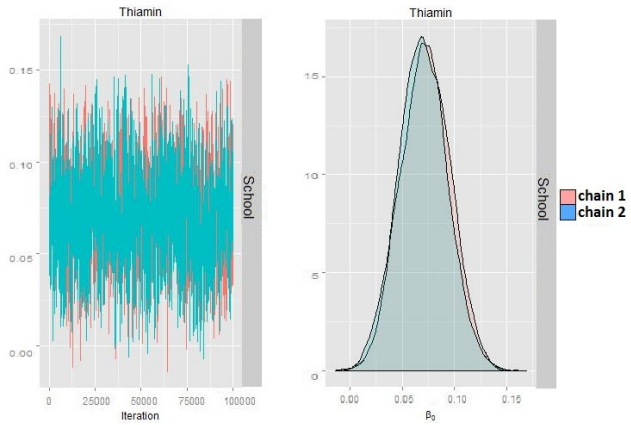


Figure 5.5: Geweke test refused convergence of chain from β_0 School group, Thiamin.

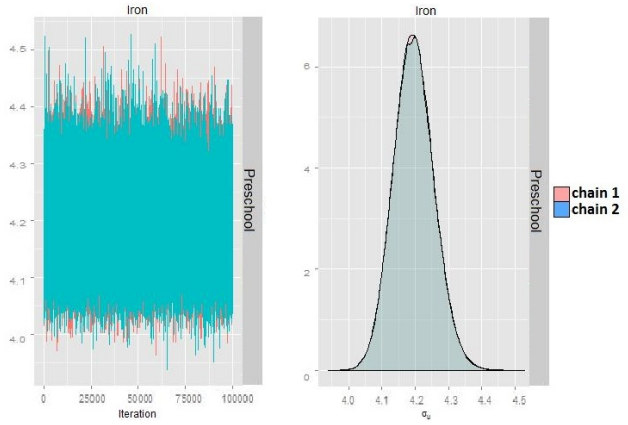


Figure 5.6: Geweke test refused convergence of chain from σ_{u_i} Preschool group, Iron.

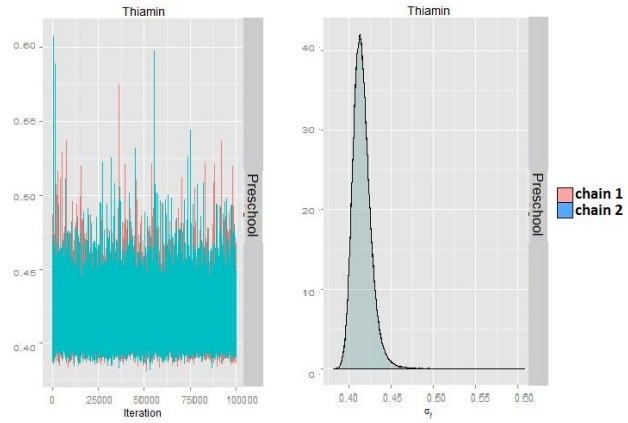


Figure 5.7: Geweke test refused convergence of chain from σ_{w_i} Preschool group, Thiamin.

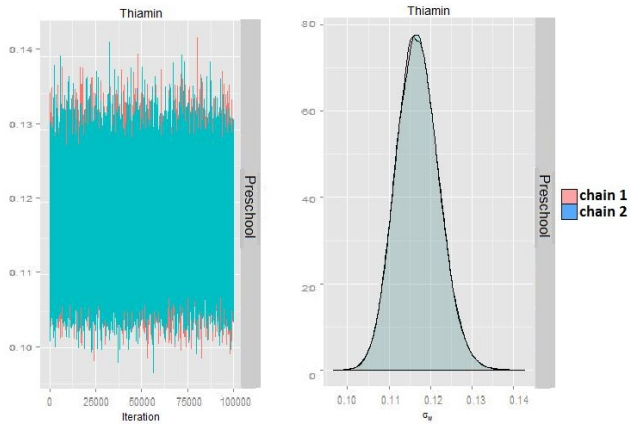


Figure 5.8: Geweke test refused convergence of chain from σ_q Preschool group, Thiamin.

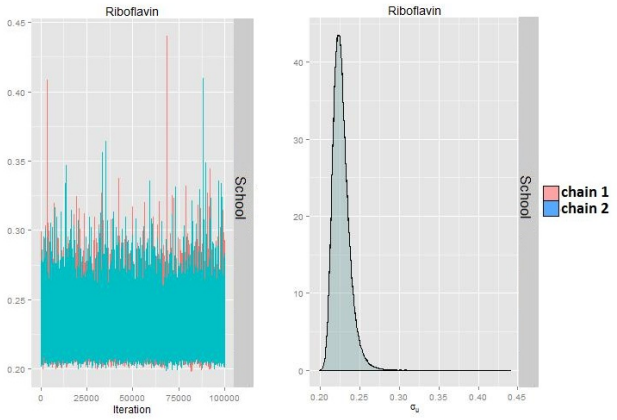


Figure 5.9: Geweke test refused convergence of chain from $\sigma_{u_i}^{-}$ School group, Riboflavin.

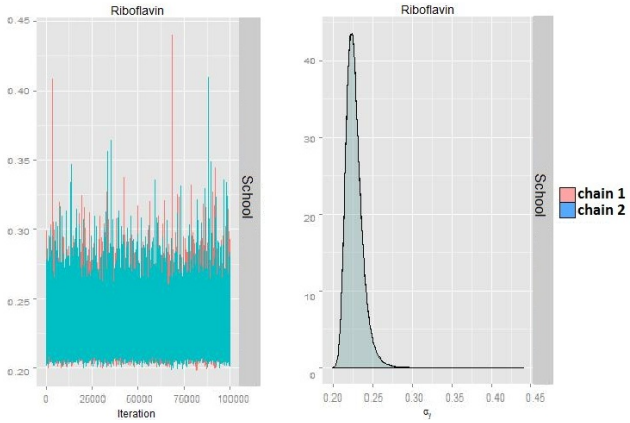


Figure 5.10: Geweke test refused convergence of chain from σ_{w_i} School group, Riboflavin.

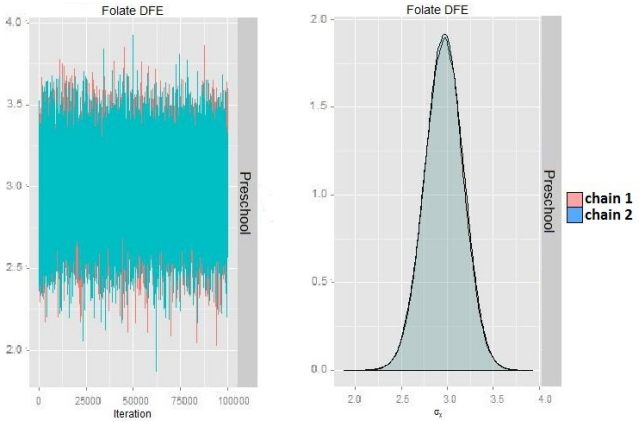


Figure 5.11: Geweke test refused convergence of chain from σ_x Preschool group, Folate.

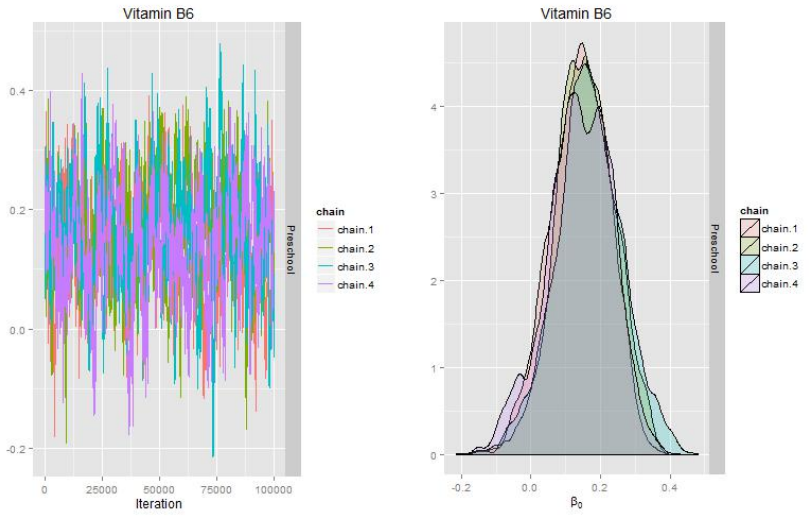
The second test used was the Gelman-Rubin test. Here the criterion to reject the null hypothesis of convergence is an $R - statistic$ greater than 1.06. The value of the $R-$statistic indicates the potential reduction in the Monte Carlo variance that can be achieved by continuing the iterations to infinity. A value of 1.06 indicates that we cannot decrease the variance more than 6% even if we were to continue iterating indefinitely. For this test, the null hypothesis was rejected for chains of parameters β_0 and β_z of Preschool group of vitamin B, and for the same parameters and same group of vitamin E the hypothesis was rejected too. These results are shaded in Table 5.3.

Given this evidence against convergence another two chains for Preschool group for vitamin B and E were simulated, now with 100,000 iterations and burn-in of 100,000, both test were done again, the result are in Table 5.2. Gelman-Rubin and Geweke test were not reject convergence in this last simulation, however the posterior densities were not different than in the first one.

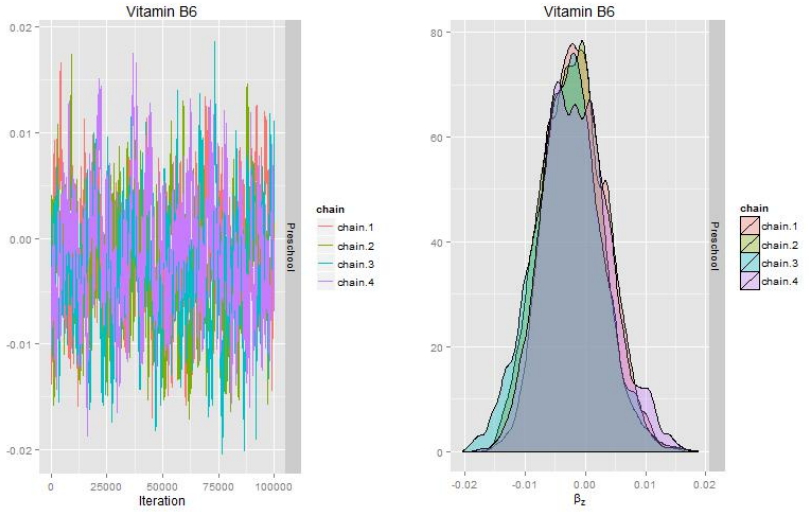
Figures 5.13 and 5.13 contain four chains. The first two correspond to the first simulation where Gelman-Rubin test concluded to reject the null hypothesis while Geweke did not. Chain three and four correspond to the last simulation, for these chains Geweke test and Gelman-Rubin did not reject convergence.

Parameter	Nutrient	Gelma-Rubin	Geweke		Nutrient	Gelma-Rubin	Geweke	
			Chain 1	Chain 2			Chain 1	Chain 2
α_0		1.00	-3.11	0.16		1.01	-1.82	-0.34
α_1		1.00	3.08	-0.17		1.01	1.76	0.32
β_0		1.03	1.16	0.86		1.01	-0.69	1.85
β_x		1.00	-1.47	-1.28		1.00	0.31	-0.99
β_z	Vit B	1.04	-1.07	-0.81	Vit E	1.01	0.73	-1.81
σ_u		1.00	-0.71	-0.82		1.00	0.49	0.53
σ_y		1.00	-0.71	-0.82		1.00	0.49	0.53
σ_q		1.00	-0.23	-0.24		1.00	-0.28	0.35
σ_x		1.00	-0.15	-0.29		1.00	-2.07	-1.11

Table 5.2: Test results for Preschool group, Vitamin B and E after 100,000 simulations with burn in of 100,000

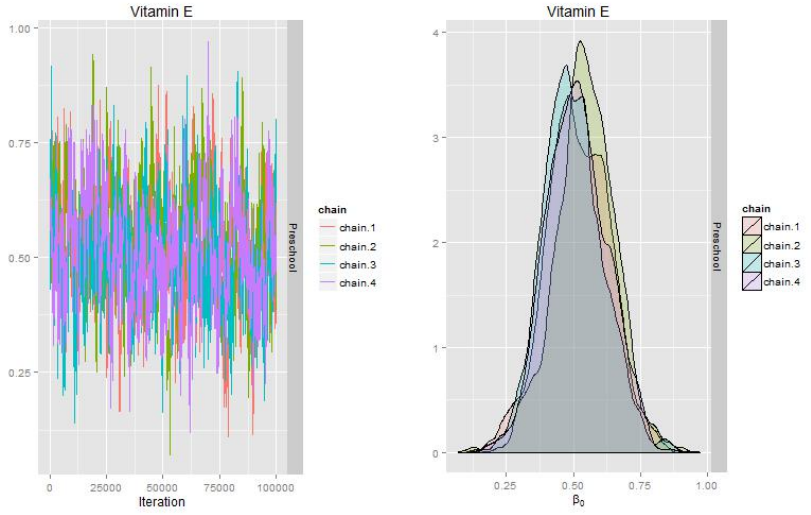


(a)

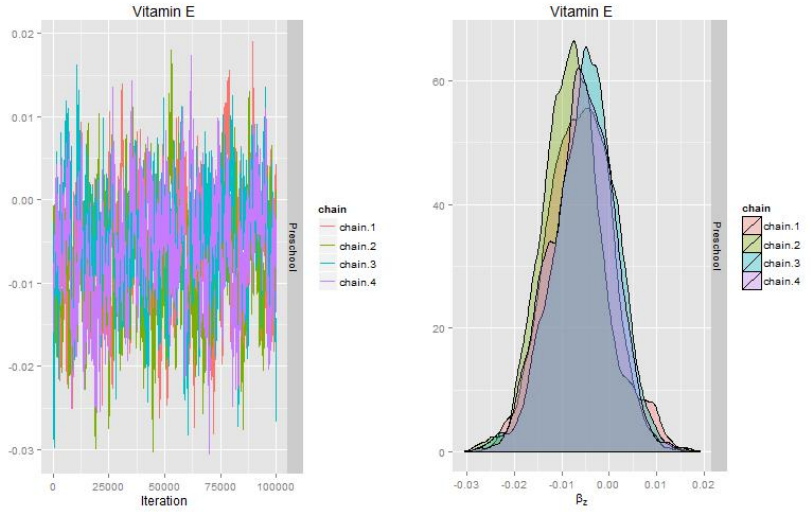


(b)

Figure 5.12: Posterior of (a) β_0 Preschool group, Vitamin B .(b) β_z Preschool group, Vitamin B.



(a)



(b)

Figure 5.13: Posterior of (a) β_0 Preschool group, Vitamin E. (b) β_z Preschool group, Vitamin E.

		Adolescents		School		Preschool	
Nutrient	Parameter	Chain 1	Chain 2	Chain 1	Chain 2	Chain 1	Chain 2
Calcium	α_0	-1.4	-0.8	0.34	0.49	-0.46	0.11
	α_1	1.37	0.79	-0.38	-0.54	0.48	-0.09
	β_0	0.32	-0.86	-1.06	-0.21	0.34	-0.42
	β_x	0.07	-0.11	-0.11	0.04	-0.45	0.19
	β_z	-0.33	0.93	1.37	0.26	-0.31	0.45
	σ_u	1.64	0.28	-0.52	-1.25	0.72	0.08
	σ_w	-0.32	0.03	0.44	0.56	0.2	1.2
	σ_q	1.78	-1.63	0.32	-0.23	-0.85	-1.17
	σ_x	0.38	0.29	-0.51	-0.77	0.25	0.32
Iron	α_0	0.22	0.16	-2.2	0.74	2.19	1.7
	α_1	-0.16	-0.2	2.16	-0.74	-2.15	-1.72
	β_0	-0.64	0.14	-0.96	-1.66	-0.77	0.42
	β_x	1.97	-0.24	-0.24	1.25	1.86	0.53
	β_z	0.25	-0.12	1.34	1.47	0.57	-0.54
	σ_u	0.13	-0.35	-0.04	-0.13	-0.8	2.41
	σ_w	-0.88	-1.09	0.3	-1.91	0.46	0.92
	σ_q	1	-0.35	-0.71	1.25	0.15	-0.49
	σ_x	1.84	-1.74	0.37	0.87	1.89	-0.07
Magnesium	α_0	-0.61	0.79	-0.78	-0.35	-1.54	-0.74
	α_1	0.62	-0.76	0.8	0.41	1.52	0.74
	β_0	2.36	-0.78	-0.95	-0.11	-1.02	-0.51
	β_x	-1.92	-0.73	1.46	0.96	-1.56	-0.64
	β_z	-1.87	1	0.53	-0.39	1.48	0.73
	σ_u	-0.33	-1.55	1.79	0.53	-2.07	0.73
	σ_w	-0.03	0.23	0.6	-0.54	-0.74	0.36
	σ_q	0.25	1.06	-0.78	-0.48	1.14	-1.19
	σ_x	0.56	-1.55	1.58	0.67	-1.62	0.45
Phosphorus	α_0	-1.05	0.53	1.03	0.38	2.01	-0.48
	α_1	1.03	-0.5	-1.07	-0.34	-2	0.47
	β_0	0.97	-0.15	0.22	-0.59	-0.55	-0.36
	β_x	1.06	-0.14	0.13	0.63	1.32	0.3
	β_z	-1.37	0.18	-0.35	0.44	0.45	0.39
	σ_u	0.56	-1.07	-0.85	-1.57	0.75	-0.15
	σ_w	0.04	2.39	-0.57	-1.03	-0.44	-0.29
	σ_q	-1.93	-0.42	0.42	-0.17	1.58	0.8
	σ_x	0.58	1.21	-1.08	0.54	1.11	-0.29
Potassium	α_0	-1.51	-0.96	-0.7	0.97	-0.96	0.95
	α_1	1.53	0.91	0.7	-0.97	0.97	-0.97
	β_0	-1.84	-0.73	0.26	-0.31	1.44	0.21
	β_x	1.35	-0.07	-0.06	0.83	-1.36	0.29
	β_z	1.61	0.84	-0.25	-0.03	-1.49	-0.29
	σ_u	0.87	0.29	0.2	-0.94	-0.35	1.25
	σ_w	-0.62	-2.65	-1.25	-2.48	0.74	-0.99
	σ_q	0.49	2.09	1.17	2.06	-0.84	0.94
	σ_x	0.31	-1.27	0.09	0.35	1.77	-1.47

		Adolescents		School		Preschool	
Nutrient	Parameter	Chain 1	Chain 2	Chain 1	Chain 2	Chain 1	Chain 2
Zinc	α_0	3.06	-0.01	-0.01	1.26	-0.66	1.15
	α_1	-3.11	0.05	-0.02	-1.29	0.65	-1.12
	β_0	1.87	0.38	0.23	2.08	0.08	-0.34
	β_x	-0.76	-0.57	-0.3	-1.02	0.45	2.39
	β_z	-1.82	-0.3	-0.19	-2.14	-0.12	0.18
	σ_u	0.21	-0.14	-0.94	0.23	0.22	0.51
	σ_w	0.34	-1.35	1.76	-0.55	1.54	-0.55
	σ_q	0.02	0.46	-0.46	0.15	0.01	-1.12
	σ_x	-0.43	1.03	-0.53	-0.23	0.98	2.83
Vit C	α_0	1.17	0.88	-0.74	-0.5	1.22	2.17
	α_1	-1.27	-0.84	0.72	0.63	-1.25	-2.16
	β_0	1.66	-1.08	0.41	1.82	-1.27	-0.12
	β_x	-0.14	-0.42	-0.09	-2.4	0.58	-0.79
	β_z	-1.7	1.35	-0.55	-1.01	1.36	0.43
	σ_u	1.07	1.73	-1.54	2.54	0.59	2.05
	σ_w	-1.56	-0.48	1.14	0.1	0.25	0.23
	σ_q	1.88	0.88	-0.53	1.73	-0.67	0.37
	σ_x	0.93	0.73	-1.04	2.69	-0.72	0.89
Thiamin	α_0	-1.37	0.36	-0.79	-0.07	0.96	-0.7
	α_1	1.31	-0.39	0.78	0.12	-0.96	0.72
	β_0	-0.82	0.09	-2.81	-0.01	-0.02	-1.25
	β_x	-0.5	0.65	0.78	0.22	-0.12	1.1
	β_z	1	-0.2	2.91	-0.06	0.03	1.36
	σ_u	-1.07	-3.27	-0.85	-1.72	1.17	-1.7
	σ_w	-0.68	-0.65	-0.67	1.02	3.02	-0.22
	σ_q	-0.93	1.13	2.17	0.68	0.69	-3.09
	σ_x	0.11	-1.82	1.17	0.9	-0.99	-0.3
Riboflavin	α_0	0.19	-0.54	-0.56	1.91	2.5	-2.12
	α_1	-0.24	0.53	0.53	-1.84	-2.51	2.11
	β_0	-1.09	0.11	-0.04	-1.39	-0.79	-1.89
	β_x	1.46	0.41	0.87	1.12	0.17	-0.4
	β_z	0.83	-0.19	-0.16	1.26	0.8	1.89
	σ_u	-0.48	0.71	-0.3	-0.09	0.64	0.91
	σ_w	0.37	-0.19	2.94	1.34	-0.54	0.63
	σ_q	1.23	-0.41	0.98	-1.81	-0.05	0.36
	σ_x	0.64	-0.69	-1.05	0.76	-1.13	-0.78
Niacin	α_0	-0.93	-0.08	-0.63	0.18	-0.72	-1.05
	α_1	0.86	0.09	0.69	-0.18	0.72	1.04
	β_0	0.41	-1.69	1.21	-0.87	0.34	0.98
	β_x	0.16	0.25	-0.4	1.59	-0.53	-1.06
	β_z	-0.43	1.67	-1.3	0.57	-0.3	-0.95
	σ_u	1.43	-1.27	-0.88	1.42	-0.18	-2.59
	σ_w	0.39	-1.09	-0.8	1	-0.42	0.71
	σ_q	-0.16	-0.28	0.41	-0.31	-1.78	0.03
	σ_x	-1.77	-0.56	-0.92	0.73	-0.08	0.44

		Adolescents		School		Preschool	
Nutrient	Parameter	Chain 1	Chain 2	Chain 1	Chain 2	Chain 1	Chain 2
Vit B	α_0	1.75	1.64	-0.96	0.8	-2.2	1.28
	α_1	-1.73	-1.52	1.02	-0.84	2.24	-1.26
	β_0	0.41	2.2	-0.71	1.43	0.65	-0.08
	β_x	0.87	-0.27	-0.36	-0.02	0.32	-0.21
	β_z	-0.63	-2.35	0.91	-1.57	-0.68	0.08
	σ_u	-2.33	-1.22	1.02	-1.68	0.2	-0.15
	σ_w	-0.16	1.79	0.07	0	-0.25	0.93
	σ_q	0.68	-0.65	-0.87	0.11	0.52	-0.09
	σ_x	0.95	-1.35	0.68	0.24	0.89	-0.6
Folate	α_0	0.02	-0.44	-0.23	-1.07	2.25	-0.69
	α_1	0	0.39	0.16	0.98	-2.28	0.68
	β_0	0.06	-1.82	0.52	-0.3	0.14	0.52
	β_x	0.43	1.23	-0.88	1.11	0.55	-0.3
	β_z	-0.17	1.76	-0.29	-0.14	-0.27	-0.55
	σ_u	0.55	-0.34	-0.66	-0.11	-0.61	1.18
	σ_w	0.22	0.29	-0.03	0.92	0.98	0.21
	σ_q	0.15	-0.26	-0.71	-0.33	-0.63	-0.76
	σ_x	1.3	-0.9	-2.53	0.13	3.92	-0.86
Vit A	α_0	-0.52	-0.79	0.87	-0.04	-0.41	1.11
	α_1	0.53	0.82	-0.9	0.08	0.41	-1.13
	β_0	1.1	0.94	0.74	0.93	0.08	-0.59
	β_x	-1.19	0.3	-0.96	0.1	0.95	-0.32
	β_z	-0.81	-1.07	-0.54	-1.24	-0.3	0.81
	σ_u	-1.16	-0.11	-0.72	-0.54	-0.8	0.23
	σ_w	0.53	-1.45	-1.44	-2.25	0.68	0.26
	σ_q	-0.66	-0.28	0.29	0.57	-0.82	-0.17
	σ_x	-1.37	1.2	-1.43	0.37	-0.57	1.64
Vit E	α_0	-0.33	-0.94	-0.06	-0.06	-0.06	-0.2
	α_1	0.36	0.97	0.04	0.06	0.06	0.2
	β_0	0.42	-0.82	-0.19	-0.26	0.26	0.62
	β_x	-1.17	0.72	0.02	0.55	-0.08	0.11
	β_z	-0.11	0.7	0.26	0.21	-0.28	-0.62
	σ_u	1.85	1.35	0.32	-2.54	-0.73	-1.33
	σ_w	0.32	0.12	0.8	-1.32	-1.22	0.95
	σ_q	0.63	0.64	-0.8	0.2	-0.68	-2.13
	σ_x	-0.55	1.1	-0.95	-1.15	0.26	0.25
Vit D	α_0	0.66	0.44	1.71	1.01	1.69	1.29
	α_1	-0.66	-0.43	-1.66	-1.07	-1.71	-1.29
	β_0	-0.21	0.8	-0.98	-1.3	0.28	-0.81
	β_x	-0.91	-0.49	1.13	0.56	-0.47	1.33
	β_z	0.37	-0.78	0.68	1.31	-0.24	0.78
	σ_u	-0.82	-1.32	-0.22	-1.42	-0.18	1.78
	σ_w	-0.12	1.43	0.03	-0.55	0.86	0.12
	σ_q	-0.02	-0.36	-0.47	1.52	0.36	-2.47
	σ_x	-0.09	-0.25	1.58	-2.33	-0.55	0.52

Parameter	Nutrient	Adolescents	School	Preschool	Nutrient	Adolescents	School	Preschool
α_0		1	1.01	1		1	1	1
α_1		1	1.01	1		1	1	1
β_0		1	1	1.01		1	1.01	1
β_x		1	1	1		1	1	1
β_z	Calcium	1	1	1.01	Phosphorus	1	1.01	1.01
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1	1		1	1	1.01
α_1		1	1	1		1	1	1.01
β_0		1	1.01	1.03		1	1	1
β_x		1	1	1		1	1	1
β_z	Iron	1.01	1	1.02	Potassium	1	1	1
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1.01
σ_q		1	1	1		1	1	1.01
σ_x		1	1	1		1	1	1
α_0		1	1	1		1	1.01	1
α_1		1	1	1.01		1	1.01	1
β_0		1	1	1		1.01	1	1.02
β_x		1	1	1		1	1	1
β_z	Magnesium	1	1	1	Zinc	1.01	1	1.02
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1	1.01		1	1	1
α_1		1	1	1.01		1	1	1
β_0		1	1	1		1.01	1.02	1.02
β_x		1	1	1		1	1.01	1.01
β_z	Vit C	1	1	1	Niacin	1.01	1.01	1.02
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1	1.01		1	1	1.02
α_1		1	1	1.01		1	1	1.02
β_0		1.01	1.02	1.02		1.02	1	1.13
β_x		1	1	1		1.01	1	1
β_z	Thiamin	1.01	1.02	1.02	Vit B	1.01	1	1.15
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1.01	1.04		1	1.01	1
α_1		1	1.02	1.04		1	1.01	1
β_0		1.01	1.01	1		1	1	1
β_x		1	1	1		1	1	1
β_z	Riboflavin	1.01	1	1	Folate	1	1	1
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1	1.02		1.01	1	1.01
α_1		1	1	1.01		1.01	1	1.01
β_0		1	1	1		1	1.03	1
β_x		1	1.01	1		1	1	1
β_z	Vit A	1	1	1	Vit D	1	1.02	1
σ_u		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
σ_q		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
α_0		1	1	1				
α_1		1	1	1				
β_0		1	1.01	1.08				
β_x		1	1	1				
β_z	Vit E	1	1.01	1.08				
σ_u		1	1	1				
σ_w		1	1	1				
σ_q		1	1	1				
σ_x		1	1	1				

Table 5.3: Model 1: Gelman-Rubin test results.

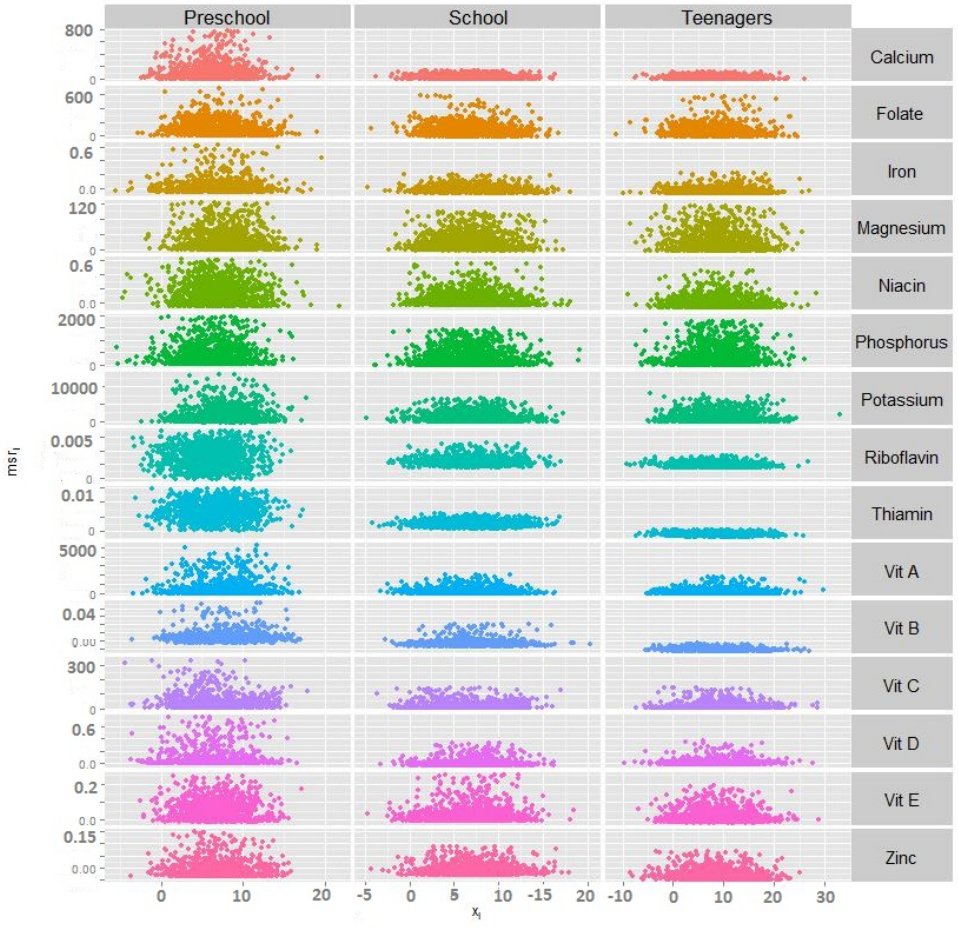


Figure 5.14: Model 1: msr_i vs $\hat{x}_i$.

5.2.2 Residuals

In order to detect non linearity in the regression, lack of the homogeneity of the error variances and non normality of the errors, the graph of simulation of x_i versus modified square residuals (msr_i) [2] will be analyzed.

$$msr_i = r_i^2 - \hat{\sigma}_{q_i}^2 - \hat{\beta}_x^2 \hat{\sigma}_{u_i}^2 + 2\hat{\beta}_x^2 \hat{\sigma}_x^2. \tag{5.7}$$

The way to compute msr_i is described in appendix B.1.

The plot of the residuals is shown in Figure 5.14. It can be seen that there is no pattern in behavior of the msr_i and high values of the residuals can be observed, mainly for Preschool group, for nutrients calcium, magnesium, phosphorus, potassium, folate and vitamin A and C. This is a consequence of high variability in the sample and the measurement error. Given the analysis of convergence and residuals and comparing the behavior of residuals with the results of simulations, the fit of the model 5.2 for all the nutrients and groups with low variability in data can be accepted.

Once that the results of the estimations were assessed, the credible intervals are given in Appendix A.1.

5.3 Model 2

Here we consider the model in 3.1.2

$$
\begin{aligned}
Y_{ij} = \beta_0 + x_i\beta_x + w_{ij}, \quad & w_{ij}|\sigma_w^2 \overset{iid}{\sim} (0, \sigma_w^2) \\
X_{ij} = x_i + u_{ij}, \quad & u_{ij}|\sigma_{u_i}^2 \overset{iid}{\sim} (0, \sigma_u^2) \\
& x_i|\mu_x, \sigma_x^2 \overset{iid}{\sim} N\big(\mu_x, \sigma_x^2\big) \\
& Y_{ij}|\beta_x, x_i, \sigma_w^2 \overset{iid}{\sim} N\big(\beta_0 + x_i\beta_x, \sigma_w^2\big) \\
& X_{ij}|x_i, \sigma_{u_i}^2 \overset{iid}{\sim} N\big(x_i, \sigma_u^2\big) \qquad (5.8)
\end{aligned}
$$

The variables X_{ij} and Y_{ij} are as defined in 5.2 and the parameters that we need to estimate are μ_x, σ_u^2, σ_x^2, σ_w^2, with main interest in β_0 and β_x. Initial values are $\sigma_\beta^2 = 100$, $\gamma_q = \gamma_w = 3$, $\delta_q = \delta_w = 1$. The values for the hyperparameters corresponding to the Inverse Gamma distribution of the errors u_{ij} and w_{ij} are in Table 5.1.

A Gibbs sampler algorithm programed in rjags with 100,000 iterations was used for estimating the parameters of this model. The credible intervals for β_0 and β_x are in table A.2 in Appendix A.2. Next the analysis of the convergence of the chains and residual will be made.

5.3.1 Convergence

Analogously to 5.2.1, the convergence of the chains was analyzed by using the Geweke and the Gelman-Rubin tests.

Using the Gelman-Rubin test, the hypothesis of convergence was not rejected for any chain as can be observed in Table 5.6. The Geweke test however rejected convergence in the Preschool group for parameters β_0 and β_x in magnesium and folate. In zinc, for σ_u and σ_x convergence was also rejected. These results are shaded in Table 5.5. Even in these cases, the posterior density estimated from these chains were practically non-distinguishable from the densities estimated from the chains that had presumably not converged to their stationary distribution. The chains in question can be observed in Figures 5.15 to 5.20.

The 95% credible intervals for all the parameters were computed and are shown in Table A.2. The interpretation of these intervals will be addressed later.

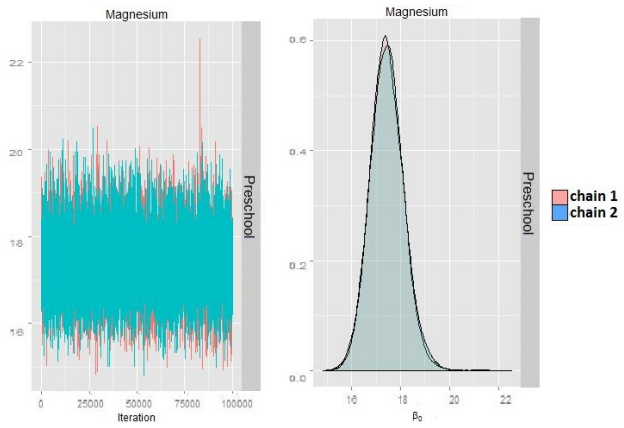


Figure 5.15: Geweke test refuse convergence of chain from β_0 Preschool group, Magnesium.

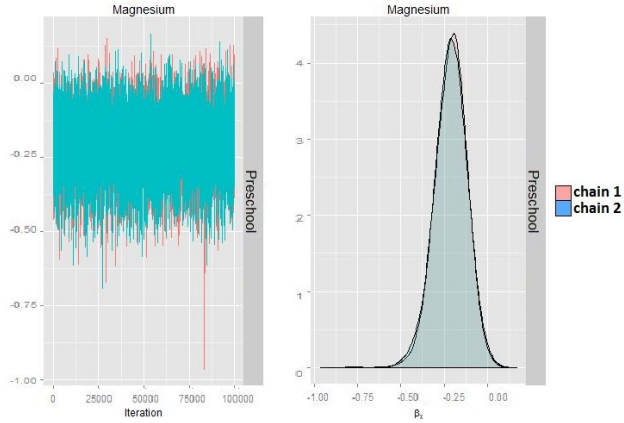


Figure 5.16: Geweke test refuse convergence of chain from β_x Preschool group, Magnesium.

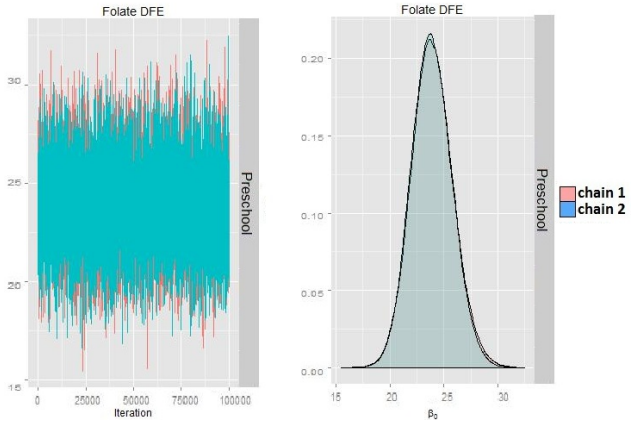


Figure 5.17: Geweke test refuse convergence of chain from β_0 Preschool group, Folate.

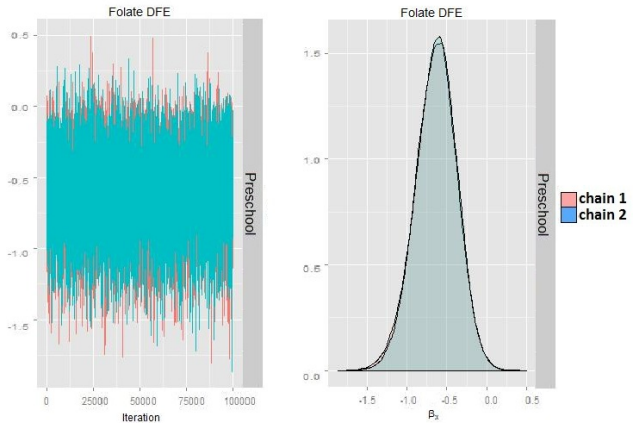


Figure 5.18: Geweke test refuse convergence of chain from β_x Preschool group, Folate.

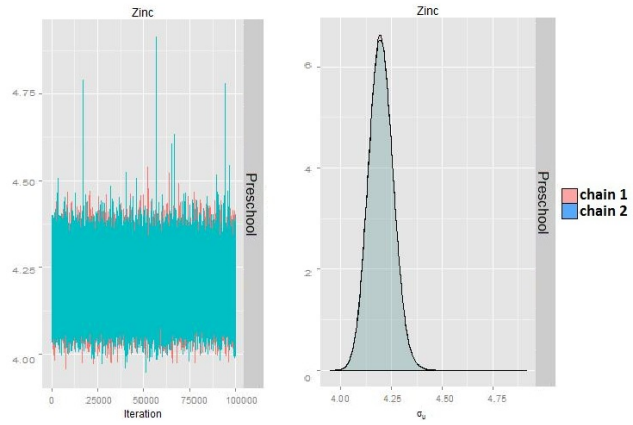


Figure 5.19: Geweke test refuse convergence of chain from σ_u Preschool group, Zinc.

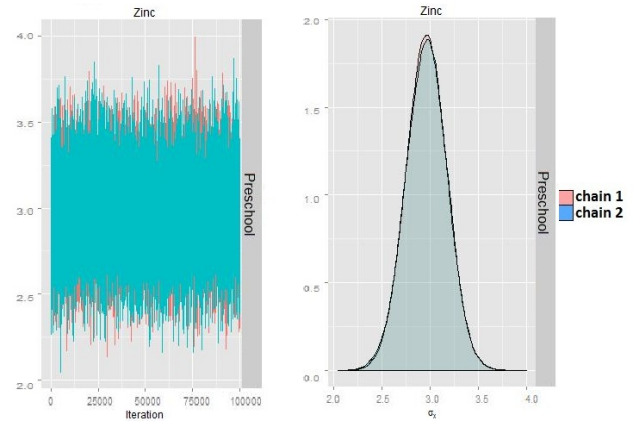


Figure 5.20: Geweke test refuse convergence of chain from σ_x Preschool group, Zinc.

		Adolescents		School		Preschool	
Nutrient	Parameter	Chain 1	Chain 2	Chain 1	Chain 2	Chain 1	Chain 2
Calcium	β_0	-0.65	0.9	-1.14	-0.74	1.04	-0.83
	β_x	0.54	-0.93	1.14	0.76	-1.06	0.85
	μ_x	-1.55	-0.21	-0.13	1	-1.24	0.44
	σ_u	1.73	0.04	0.61	0.25	1.49	-1.06
	σ_x	-1.44	0.01	-0.59	0.02	-1.53	0.74
	σ_w	0.02	-1.63	-1.18	1.33	1.25	-0.26
Iron	β_0	-1.82	0.2	-1.65	-1.67	-1.99	-0.85
	β_x	1.85	-0.28	1.59	1.67	2.07	0.84
	μ_x	1.48	0.89	-0.17	0.61	0.35	0
	σ_u	-1.69	0.23	-0.08	-1.14	-1.84	-0.35
	σ_x	1.38	-0.59	-0.04	1.01	2.55	0.46
	σ_w	0.73	0.11	1.57	1.82	2.55	0.43
Magnesium	β_0	-1.05	0.56	0.76	1.43	3.14	0.49
	β_x	1.04	-0.57	-0.76	-1.39	-3.1	-0.52
	μ_x	-0.22	-0.95	-0.46	0.72	-0.27	0.01
	σ_u	-1.42	1.55	-0.08	-0.75	0.78	1.59
	σ_x	0.4	-1.58	-0.64	0.56	-1.55	-0.97
	σ_w	-0.78	1.23	0.54	-0.07	-1.92	-0.36
Phosphorus	β_0	2.3	-0.91	-0.18	1	-0.77	0.51
	β_x	-2.26	0.8	0.15	-0.97	0.78	-0.49
	μ_x	-0.36	-0.2	0.54	0.54	-0.78	0.19
	σ_u	0.75	-0.07	0.56	0.28	0.55	0.11
	σ_x	-0.99	-0.73	0.49	0.03	0.44	-0.19
	σ_w	-1.62	0.47	-0.28	1.31	1.5	-0.56
Potassium	β_0	0.96	0.27	-1.18	-0.39	-1.19	-1.15
	β_x	-1	-0.38	1.18	0.35	1.2	1.16
	μ_x	-0.64	0.07	1.47	-0.88	0.01	-0.94
	σ_u	-0.43	-0.12	-1	-0.4	0.47	1.58
	σ_x	-0.33	-0.35	0.91	0.25	-0.94	-1.47
	σ_w	0.6	1.32	0.78	-1.04	-1.42	-0.5
Zinc	β_0	-1.08	0.66	0.94	1.31	-0.49	-0.55
	β_x	1.03	-0.7	-0.93	-1.3	0.5	0.59
	μ_x	1.06	-0.18	-1.41	-0.19	0.3	2.53
	σ_u	-0.42	0.91	0.86	0.36	-0.44	-3.03
	σ_x	1.78	-1.02	-0.52	-0.19	-0.12	2.81
	σ_w	1.11	-1.29	-0.25	-1.47	0.25	0.01
Vit C	β_0	-0.98	0.33	1.28	0.54	0.72	-0.67
	β_x	1.05	-0.41	-1.26	-0.47	-0.76	0.66
	μ_x	-0.55	-0.11	0.55	-0.63	1.46	-0.91
	σ_u	0.7	-0.41	-1.2	0.35	-1.46	0.48
	σ_x	-0.45	0.71	0.39	-0.46	0.75	-1.17
	σ_w	-0.06	0.43	1.38	1.17	0.76	-0.36

Table 5.4: Model 2: Geweke test results.

		Adolescents		School		Preschool	
Nutrient	Parameter	Chain 1	Chain 2	Chain 1	Chain 2	Chain 1	Chain 2
	β_0	0.27	0.8	-0.16	1.22	1.49	-0.43
	β_x	-0.25	-0.67	0.16	-1.28	-1.44	0.4
	μ_x	-1.7	0.86	0.19	0.05	0.14	-0.41
	σ_u	1.4	-0.28	-1.55	-1.15	-0.66	0.28
Thiamin	σ_x	-1.49	1.38	1.53	0.7	1.58	-0.42
	σ_w	-1.85	-0.05	0.24	-0.3	-1.16	0.36
	β_0	-1.26	-1.3	0.84	-0.47	-0.06	2.4
	β_x	1.11	1.35	-0.84	0.48	0.05	-2.38
	μ_x	-0.59	-1.72	0.54	-0.15	-0.24	-0.75
	σ_u	0.15	0.35	-2.44	-0.52	0.3	0.61
Riboflavin	σ_x	-0.22	-0.71	1.25	0.8	-0.89	-0.02
	σ_w	0.55	-0.65	1.93	1.57	0.73	0.07
	β_0	-1.68	2.43	0.37	-0.34	-0.34	-1.15
	β_x	1.7	-2.42	-0.44	0.37	0.38	1.15
	μ_x	0.25	-1.01	-0.72	0.45	0.77	-1.12
	σ_u	0.05	1.48	1.21	-0.7	-0.54	-1.34
Niacin	σ_x	0.6	-1.21	-0.99	0.6	0.47	0.45
	σ_w	-0.25	-0.84	0.44	-0.91	0.27	1.45
	β_0	0.26	0.32	-0.24	1.36	1.91	-1.63
	β_x	-0.15	-0.2	0.3	-1.35	-1.86	1.68
	μ_x	0.91	0.66	-0.37	-0.01	0.02	-0.44
	σ_u	-0.34	0.4	1.81	0.08	0.71	-0.16
Vit B	σ_x	0.41	0.01	-0.77	0.52	-0.13	-0.63
	σ_w	0.93	1.19	0.31	-0.93	1.28	1.94
	β_0	1.07	0.17	0.6	1.03	3.27	0.08
	β_x	-1.16	-0.15	-0.6	-1.05	-3.29	-0.08
	μ_x	1.22	0.06	-0.74	-1.04	-0.79	0.32
	σ_u	-1.57	1.05	-0.38	1.08	1.22	0.17
Folate	σ_x	1.52	0.18	0.21	-0.88	-1.06	-0.13
	σ_w	-0.63	-0.89	-0.87	-0.72	-2.49	-0.5
	β_0	0.16	-0.07	-0.79	2.05	-0.43	0.51
	β_x	-0.15	-0.04	0.81	-2.01	0.42	-0.54
	μ_x	0.02	0.16	-0.69	0.92	1.39	-1.65
	σ_u	-1.14	1.04	0.75	0.26	-0.96	2.12
Vit A	σ_x	0.37	0.41	-0.19	-0.41	0.78	-2.58
	σ_w	-0.82	-0.03	-0.77	0.14	-1.21	0.2
	β_0	-1.99	2.52	1.2	-0.32	-0.03	-0.45
	β_x	2.04	-2.42	-1.18	0.31	-0.02	0.39
	μ_x	0.67	-0.87	0.69	-0.21	-0.91	-0.66
	σ_u	-0.35	0.66	0.66	-0.2	0.19	0.7
Vit E	σ_x	0.25	-1.88	0.02	0.81	-0.87	-0.84
	σ_w	0.15	-0.1	-0.4	0.26	-0.03	-0.45
	β_0	0.72	-0.98	-0.36	0.55	-0.12	-1.76
	β_x	-0.82	1.04	0.37	-0.59	0.13	1.68
	μ_x	-0.26	0.23	1.98	0.89	-0.69	-2.02
	σ_u	-0.15	0.17	-0.93	0.24	1.56	1.05
Vit D	σ_x	-1.14	0.01	1.95	0.15	-0.83	-1.16
	σ_w	-1.42	0.83	0.76	-0.59	0.43	2

Table 5.5: Model 2: Geweke test results.

Parameter	Nutrient	Adolescents	School	Preschool	Nutrient	Adolescents	School	Preschool
β_0	Calcium	1	1	1.01	Thiamin	1	1	1
β_x		1	1	1.01		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Iron	1	1	1	Riboflavin	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Magnesium	1	1	1	Niacin	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Phosphorus	1	1	1	Vit B	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Potassium	1	1	1	Folate	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Zinc	1	1	1	Vit A	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Vit C	1	1	1	Vit E	1	1	1
β_x		1	1	1		1	1	1
μ_x		1	1	1		1	1	1
σ_u		1	1	1		1	1	1
σ_x		1	1	1		1	1	1
σ_w		1	1	1		1	1	1
β_0	Vit D	1	1	1				
β_x		1	1	1				
μ_x		1	1	1				
σ_u		1	1	1				
σ_x		1	1	1				
σ_w		1	1	1				

Table 5.6: Model 2: Gelman-Rubin test results.

5.3.2 Residuals

Similar to Section 5.2.2, predictions for the posterior predictive distribution of msr_i expressed as:

$$msr_i = r_i^2 - (\sigma_w^2 + \beta_x^2 \sigma_{u_i}^2).$$

In Appendix B.2 we describe the msr_i for this model. The plot of the residuals is in Figure 5.21, and is similar to the plot of residuals that was drawn for Model 1. There is no pattern in the behavior of the msr_i. We do, however, observe some high values of the residuals for the nutrients calcium, magnesium, phosphorus, potassium, folate and vitamin A and C. This is not true for iron, zinc, thiamin, riboflavin, niacin, vitamin B, E and D. Although the behavior of both models are similar the residual values of Model 1 appear to be smaller than the residuals of Model 2.

The chains of Model 2 show better a behavior of convergence, but results of residuals are almost the same. The conclusion is an acceptable fit of Model 2 if the variability is low. In the cases with high variability in data, the fit could not be so good. The credible intervals for all the parameters in this model are in Appendix A.2.

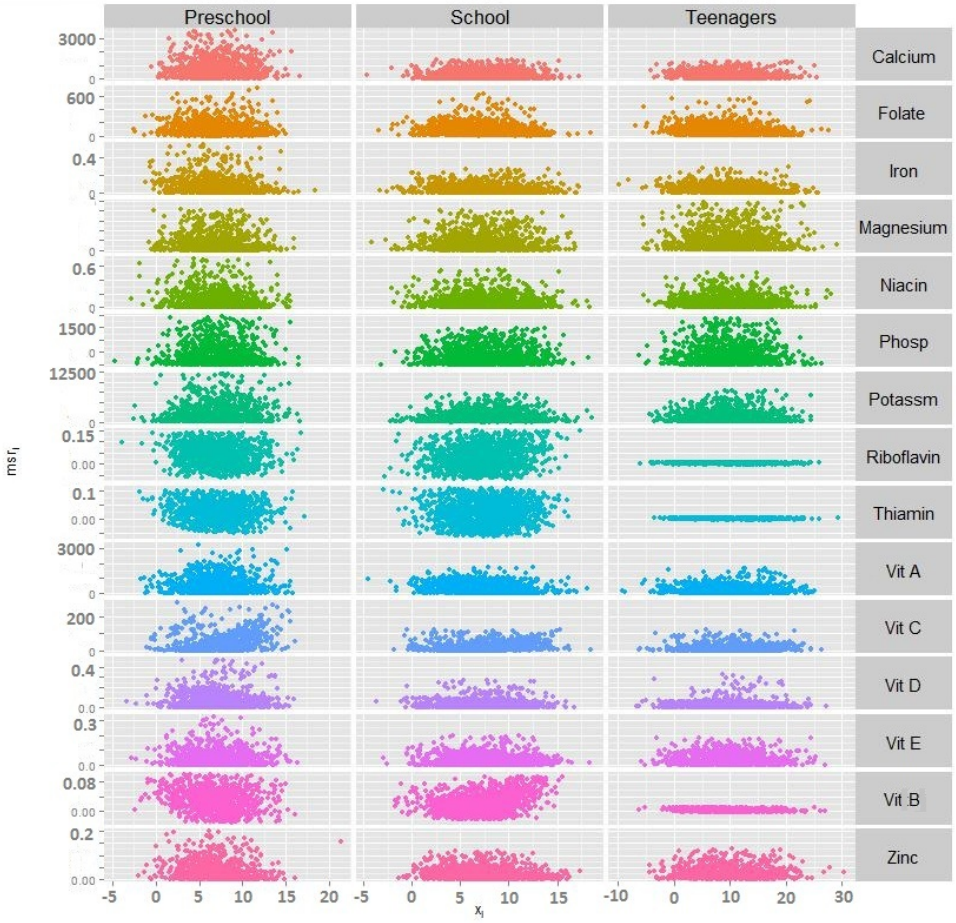


Figure 5.21: Model 2: msr_i vs $\hat{x}_i$.

While both models appear to fit the data reasonably well, we wish to decide on one of the two in order address the subtantive problem that motivated this thesis. To do so, we carry out model comparison in Section 5.4.

5.4 Model comparison

Previously, model (5.2) and (5.3) were fitted, the convergence of the chains for each parameter and the residuals were analyzed. Now we select one model in order to determine the association between regular intake of sugary drinks and nutrient density of the diet. The model will be assessed by using Deviance Information Criteria (DIC) [7]. In Section 3.1.1 the correspondent parameter vector for the model is $\theta = (\beta_0, \beta_x, \beta_z, \sigma^2_{w_i}, \sigma^2_{u_i}, \alpha_0, \alpha_1)$ for $i = 1, \ldots, n$. For the model is Section 3.1.2 $\theta = (\beta_0, \beta_x, \sigma^2_w, \sigma^2_x, \mu_x)$. It will be necessary to sample from the posterior distribution of θ in order to obtain θ^l and compute:

$$DIC = \bar{D}(y) + p_D. \tag{5.9}$$

Where

$$\bar{D}(y) = \frac{1}{L}\sum_{l=1}^{L} D(y, \theta^l)$$

$$D(y, \theta) = -2log(y|\theta)$$

$$p_D = \bar{D}(y) - D_{\hat{\theta}}(y) \tag{5.10}$$

$$D(y, \theta) = -2log(y|\theta)$$

$$D_{\hat{\theta}}(y) = D(y, \hat{\theta}). \tag{5.11}$$

In order to compute the DIC of the models a number sample $L = 20000$ was considered. The results are in table 5.7, the shaded numbers are the lowest DIC between the two models.

Nutrient	Adolescents		School		Preschool	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Calcium	17122	19920	19677	22440	15577	19227
Iron	9083	8593	9590	9667	8070	8154
Magnesium	15436	16480	17263	18059	13135	14225
Phosphorus	17601	19929	19590	22099	15352	17626
Potassium	19454	21746	21764	24208	19192	19829
Zinc	8343	9063	9014	9774	7455	7383
Vit C	15146	16839	17035	18709	13718	15456
Thiamin	5737	8421	7285	14388	7332	11201
Riboflavin	6209	9149	7443	14703	7665	11505
Niacin	9657	9341	10583	10483	8630	8427
Vit B	6842	8861	7556	13258	7296	10822
Folate	16579	18953	18718	21013	14665	16555
Vit A	17580	21309	20524	24301	15465	18438
Vit E	8895	8525	9597	8849	7898	7512
Vit D	8958	8742	9797	9912	8185	8665

Table 5.7: Deviance Information Criteria (DIC).

Given the DIC of both models in question and taking into account that the performance of the models is almost the same but considering a more realistic model it can be concluded that:
Model 1 will be applied in all groups of age for nutrients calcium, magnesium, phosphorus, potassium, zinc, vitamin C, thiamin, riboflavin, vitamin B, folate and vitamin A. Model 2 will be fit in all the age groups for niacin and vitamin E.

For iron, we fit Model 1 in the group of Adolescents and Model 2 in the two other groups.

5.5 Results

Once we have selected a model to fit in each age group and nutrient, we estimated the posterior distributions of model parameters as described earlier. The interpretation and conclusions will be given next.

Most of the sugary drinks consumed in Mexico contribute negligible amounts of nutrition and high amounts of calories to the diet of children and adolescents. For this reason, our working hypothesis is that higher consumption of sugary drinks is associated with lower nutrient density in the diet. That is, we expect that consumption of sugary drinks displaces nutrients. In fact, the results we obtained in the analysis of consumption information obtained in ENSANUT provide evidence in support of our working hypothesis for almost all nutrients. Somewhat surprisingly, we found that BMI did not help to explain usual nutrient density. The estimated 95% credible sets for the coefficients of the linear model fitted are shown in Table 5.8.

The sugary drinks we included in this analysis are: sodas, industrialized flavored water and artificial fruit/vegetables drinks. These beverages have different nutritional content (in all cases, nutritional content is low with the exception of vitamin C) and are consumed in different quantities depending on the age of the individuals. Therefore it is reasonable to find differences in the association between the usual intake of sugary drinks and the usual nutrient density in the diet for the different age groups. In this light, we interpret results separately for each age group.

- Adolescents: The association between usual intake of sugary drinks and the calcium, iron, magnesium, phosphorus, potassium, zinc, niacin and folate density in the usual diet is negative and statistically significant. There is no significant association between consumption of sugary drinks and the diet density of thiamin, riboflavin, vitamin A, B, D and E . The density of vitamin C in the diet increases significantly with consumption of sugary drinks; this can be explained by the fact that some artificial fruit and vegetable drinks are fortified with vitamin C.

- School-aged children: In this group, the nutrient density of calcium, iron, magnesium, phosphorus, folate, niacin and vitamin A are negatively and significantly associated with the usual consumption of sugary drinks. The association between density of potassium, zinc, thiamin, riboflavin, vitamin B, E and D and intake of sugary drinks is not statistically significant. As in the case of adolescents, sugary drinks are associated with higher vitamin C density in the diet.

- Preschool children: In this group, we find a statistically significant and negative association between the densities of magnesium, zinc, niacin and folate with consumption of sugary drinks. In this age group, the density of both potassium and vitamin C increase with increased consumption of sugary drinks. The rest of the nutrient densities are not significantly associated with consumption of sugary drinks.

To summarize, usual consumption of sugary drinks is associated with lower nutrient densities in the diet of adolescents and school-age children for several important nutrients. With the exception of vitamin C, increased consumption of sugary drinks does not improve the

diet density of any of the micro nutrients considered in ENSANUT. If we take into account that sugary drinks contribute a significant proportion of the calories to the diet of children and adolescents in Mexico and that over-weight and obesity have reached epidemic proportions, we have to conclude that there is no benefit whatsoever associated with the consumption of these drinks.

Nutrient	Group	Intercept	β_z	β_z
Calcium	Adolescents	(41.9 , 52.8)	(-0.9 , -0.4)	(-0.3 , 0.1)
	School	(41.1 , 54.3)	(-1.1 , -0.019)	(-0.3 , 0.3)
	Preschool	(12.8 , 56.2)	(-2 , 0.3)	(0.6 , 3)
Iron	Adolescents	(0.7 , 0.79)	(-0.02 , -0.01)	- -
	School	(0.6 , 0.8)	(-0.015 , -0.003)	(-0.006 , 0.004)
	Preschool	(0.5 , 0.9)	(-0.02 , 0)	(-0.011 , 0.016)
Magnesium	Adolescents	(17.2 , 20.5)	(-0.2 , -0.1)	(-0.1 , 0.004)
	School	(16.6 , 20)	(-0.3 , -0.013)	(-0.2 , -0.006)
	Preschool	(16.2 , 22.6)	(-0.4 , -0.004)	(-0.3 , 0.041)
Phosphorus	Adolescents	(67.4 , 80.5)	(-0.7 , -0.1)	(-0.4 , 0.2)
	School	(67.7 , 80.8)	(-1.1 , -0.017)	(-0.5 , 0.2)
	Preschool	(58.6 , 88.9)	(-1.5 , 0.1)	(-0.8 , 0.8)
Potassium	Adolescents	(116.5 , 143)	(-1.3 , -0.1)	(-1.1 , 0.1)
	School	(116.7 , 145)	(-1.3 , 1)	(-1.3 , 0.03)
	Preschool	(66.8 , 144)	(1.2 , 5.5)	(-1.6 , 2.7)
Zinc	Adolescents	(0.4 , 0.6)	(-0.008 , -0.001)	(-0.003 , 0.004)
	School	(0.4 , 0.6)	(-0.009 , 0.001)	(-0.003 , 0.006)
	Preschool	(0.4 , 0.8)	(-0.013 , 0.001)	(-0.01 , 0.013)
Vitamin C	Adolescents	(1.5 , 4.3)	(0.043 , 0.2)	(-0.013 , 0.1)
	School	(1.2 , 4.6)	(0.3 , 0.7)	(-0.1 , 0.029)
	Preschool	(-4.5 , 4.2)	(0.6 , 1.1)	(-0.1 , 0.3)
Thiamin	Adolescents	(0.033 , 0.1)	(-0.002 , 0.001)	(-0.002 , 0.001)
	School	(0.028 , 0.1)	(-0.003 , 0.002)	(-0.002 , 0.002)
	Preschool	(-0.1 , 0.3)	(-0.007 , 0.006)	(-0.013 , 0.01)
Riboflavin	Adolescents	(0.037 , 0.1)	(-0.002 , 0.001)	(-0.002 , 0.002)
	School	(0.033 , 0.1)	(-0.004 , 0.002)	(-0.002 , 0.003)
	Preschool	(-0.1 , 0.3)	(-0.008 , 0.006)	(-0.014 , 0.011)
Niacin	Adolescents	(0.82 , 0.93)	(-0.01 , -0.004)	- -
	School	(0.82 , 1.01)	(-0.03 , -0.001)	- -
	Preschool	(0.87 , 1.08)	(-0.04 , -0.01)	- -
Vitamin B6	Adolescents	(0.045 , 0.1)	(-0.003 , 0.001)	(-0.002 , 0.002)
	School	(0.1 , 0.2)	(-0.004 , 0.002)	(-0.003 , 0.003)
	Preschool	(-0.033 , 0.3)	(-0.007 , 0.005)	(-0.01 , 0.008)
Folate	Adolescents	(13.7 , 19.8)	(-0.4 , -0.1)	(-0.1 , 0.1)
	School	(17.2 , 24.4)	(-0.7 , -0.1)	(-0.3 , 0.1)
	Preschool	(16.1 , 31.9)	(-0.9 , -0.1)	(-0.6 , 0.3)
Vitamin A	Adolescents	(21.9 , 33.8)	(-0.5 , 0.046)	(-0.3 , 0.2)
	School	(26.1 , 39.8)	(-1.1 , -0.012)	(-0.3 , 0.4)
	Preschool	(10.5 , 49)	(-1.2 , 0.8)	(-0.3 , 1.8)
Vit E	Adolescents	(0.37 , 0.45)	(-0.01 , 0.002)	- -
	School	(0.35 , 0.45)	(-0.01 , 0.01)	- -
	Preschool	(0.39 , 0.54)	(-0.02 , 0.004)	- -
Vitamin D	Adolescents	(0.22 , 0.31)	(-0.01 , 0.001)	- -
	School	(0.1 , 0.3)	(-0.011 , 0.002)	(-0.003 , 0.009)
	Preschool	(-0.1 , 0.4)	(-0.013 , 0.007)	(-0.004 , 0.027)

Table 5.8: For a fixed intercept and β_z, every increase in the $\%$ of kcal ingested through sugary drinks, nutrient density chance β_x units.

Chapter 6

Conclusions

In this work, we propose two heteroscedastic measurement error models to describe the association between consumption of sugary drinks and nutritional content of the diet of children and adolescents in Mexico. The data for analysis consist of one or two measurements of daily intake for a sample of ENSANUT 2013 participants. To estimate model parameters, we adopt a Bayesian approach with both informative and non-informative priors. The two models we propose are reasonable given the known sources of variability in this problem:

- Measurement error in both the response variable and the predictor (known as day to day variability in intakes or within-person variability).
- The variability in intakes between individuals.

When the day-to-day variability in the replicates is high, the credible intervals for the within-person variances may not include the true value of the variances. This is because for some individuals, the variance in sugary drink or nutrient intakes across days is very high, and this significantly increases the value of the mean (across persons) within-person variances. However, the credible intervals for coefficients of the linear regression do cover the means.

If the prior distributions of the variances of the errors proposed in Chapter 3 are changed for a inverse Chi-square, as was proposed by Castro (2012) [1], the results of simulations will have insignificant differences when the value of the hyperparameters are accurate. However the estimation of hyperparameter using method of moments does not work very well in simulations. This fact make difficult to propose values of hyperparameters for each nutrient and group age. For this reason it was more practical to propose an inverse gamma distribution as a prior distribution of the variances of the errors. The difference between the two types of prior is one of re-scaling of parameters in any case.

We briefly considered fitting a model under the assumption of homogeneous measurement error variances across individuals. We abandoned this approach as unrealistic, since it is clear from the ENSANUT data that the within-person variances in both nutrient density and consumption of sugary drinks are heterogeneous. From a technical point of view, fitting the model with a homogeneity assumption for the measurement errors would require a different formulation to avoid identifiability issues.

For a majority of the nutrients considered in ENSANUT and for children and adolescents in Mexico, we found that higher consumption of sugary drinks are negatively and significantly associated with lower nutrient density in the diet. For most of the other nutrients, there was no association between nutritional quality of the diet and consumption of sugary drinks. That is, the calories provided by sugary drinks are “empty” calories, in that they do not add any nutritional value to the diet of children and adolescents. The only exception is vitamin C; in all age groups, higher consumption of sugary drinks is significantly associated with higher vitamin C density in the diet. This is explained by the fact that artificial fruit and vegetable drinks in Mexico tend to have added vitamin C.

Appendix A

Results

A.1 Model 1

Nutrient	Group	α_0	α_1	β_0	β_x	β_z	$\sigma_{\tilde{u}_i}$	$\sigma_{\tilde{w}_i}$	σ_q	σ_x
	Adolescents	(4.7 , 8.3)	(0.1 , 0.2)	(41.9 , 52.8)	(-0.9 , -0.4)	(-0.3 , 0.1)	(3.409 , 3.586)	(3.875 , 4.214)	(16.54 , 18.138)	(4.767 , 5.518)
Calcium	School	(4.9 , 7.7)	(-0.005 , 0.1)	(41.1 , 54.3)	(-1.1 , -0.019)	(-0.3 , 0.3)	(3.62 , 3.796)	(4.356 , 4.683)	(18.389 , 20.045)	(2.794 , 3.497)
	Preschool	(4.3 , 10.4)	(-0.2 , 0.2)	(12.8 , 56.2)	(-2 , 0.3)	(0.6 , 3)	(4.084 , 4.325)	(4.635 , 5.438)	(29.401 , 32.475)	(2.537 , 3.37)
	Adolescents	(4.7 , 8.3)	(0.1 , 0.2)	(0.6 , 0.8)	(-0.011 , -0.003)	(-0.005 , 0.004)	(3.403 , 3.578)	(0.364 , 0.381)	(0.12 , 0.145)	(4.795 , 5.544)
Iron	School	(5 , 7.8)	(-0.01 , 0.1)	(0.6 , 0.8)	(-0.015 , -0.003)	(-0.006 , 0.004)	(3.617 , 3.793)	(0.34 , 0.353)	(0.116 , 0.14)	(2.8 , 3.51)
	Preschool	(4.5 , 10.4)	(-0.2 , 0.2)	(0.5 , 0.9)	(-0.02 , 0)	(-0.011 , 0.016)	(4.082 , 4.32)	(0.357 , 0.374)	(0.143 , 0.179)	(2.564 , 3.384)
	Adolescents	(4.7 , 8.4)	(0.1 , 0.2)	(17.2 , 20.5)	(-0.2 , -0.1)	(-0.1 , 0.004)	(3.397 , 3.572)	(1.935 , 2.04)	(4.7 , 5.211)	(4.831 , 5.593)
Magnesium	School	(5.1 , 7.9)	(-0.015 , 0.1)	(16.6 , 20)	(-0.3 , -0.013)	(-0.2 , -0.006)	(3.622 , 3.798)	(1.912 , 2.007)	(4.251 , 4.7)	(2.784 , 3.474)
	Preschool	(4.1 , 10.1)	(-0.2 , 0.2)	(16.2 , 22.6)	(-0.4 , -0.004)	(-0.3 , 0.041)	(4.086 , 4.325)	(1.678 , 1.771)	(4.37 , 4.871)	(2.54 , 3.363)
	Adolescents	(4.7 , 8.4)	(0.1 , 0.2)	(67.4 , 80.5)	(-0.7 , -0.1)	(-0.4 , 0.2)	(3.399 , 3.575)	(4.33 , 4.613)	(20.513 , 22.375)	(4.818 , 5.577)
Phosphorus	School	(5 , 7.9)	(-0.018 , 0.1)	(67.7 , 80.8)	(-1.1 , -0.017)	(-0.5 , 0.2)	(3.62 , 3.796)	(4.269 , 4.506)	(18.37 , 19.912)	(2.783 , 3.485)
	Preschool	(4.3 , 10.3)	(-0.2 , 0.2)	(58.6 , 88.9)	(-1.5 , 0.1)	(-0.8 , 0.8)	(4.083 , 4.321)	(3.735 , 3.942)	(21.439 , 23.536)	(2.548 , 3.373)
	Adolescents	(4.7 , 8.4)	(0.1 , 0.2)	(116.5 , 143)	(-1.3 , -0.1)	(-1.1 , 0.1)	(3.406 , 3.583)	(6.625 , 7.216)	(40.61 , 44.889)	(4.773 , 5.531)
Potassium	School	(5 , 7.9)	(-0.017 , 0.1)	(116.7 , 145)	(-1.3 , 1)	(-1.3 , 0.03)	(3.623 , 3.8)	(5.997 , 6.468)	(39.821 , 43.3)	(2.766 , 3.467)
	Preschool	(4.4 , 10.4)	(-0.2 , 0.2)	(66.8 , 144)	(1.2 , 5.5)	(-1.6 , 2.7)	(4.079 , 4.317)	(6.377 , 7.004)	(51.637 , 57.577)	(2.58 , 3.396)
	Adolescents	(4.7 , 8.4)	(0.1 , 0.2)	(0.4 , 0.6)	(-0.008 , -0.001)	(-0.003 , 0.004)	(3.402 , 3.577)	(0.303 , 0.317)	(0.109 , 0.129)	(4.793 , 5.554)
Zinc	School	(5 , 7.8)	(-0.009 , 0.1)	(0.4 , 0.6)	(-0.009 , 0.001)	(-0.003 , 0.006)	(3.62 , 3.795)	(0.286 , 0.299)	(0.104 , 0.122)	(2.791 , 3.486)
	Preschool	(4.4 , 10.4)	(-0.2 , 0.2)	(0.4 , 0.8)	(-0.013 , 0.001)	(-0.01 , 0.013)	(4.08 , 4.32)	(0.295 , 0.308)	(0.117 , 0.141)	(2.57 , 3.393)
	Adolescents	(4.6 , 8.3)	(0.1 , 0.2)	(1.5 , 4.3)	(0.043 , 0.2)	(-0.013 , 0.1)	(3.403 , 3.578)	(1.836 , 1.979)	(4.033 , 4.541)	(4.804 , 5.561)
Vitamin C	School	(5 , 7.8)	(-0.01 , 0.1)	(1.2 , 4.6)	(0.3 , 0.7)	(-0.1 , 0.029)	(3.631 , 3.807)	(1.946 , 2.076)	(3.78 , 4.311)	(2.748 , 3.422)
	Preschool	(4.5 , 10.6)	(-0.2 , 0.2)	(-4.5 , 4.2)	(0.6 , 1.1)	(-0.1 , 0.3)	(4.077 , 4.317)	(2.455 , 2.66)	(4.449 , 5.32)	(2.589 , 3.436)
	Adolescents	(4.6 , 8.3)	(0.1 , 0.2)	(0.033 , 0.1)	(-0.002 , 0.001)	(-0.002 , 0.001)	(3.405 , 3.581)	(0.116 , 0.123)	(0.068 , 0.076)	(4.783 , 5.54)
Thiamin	School	(5 , 7.9)	(-0.015 , 0.1)	(0.028 , 0.1)	(-0.003 , 0.002)	(-0.002 , 0.002)	(3.622 , 3.798)	(0.193 , 0.231)	(0.073 , 0.082)	(2.766 , 3.473)
	Preschool	(4.2 , 10.1)	(-0.2 , 0.2)	(-0.1 , 0.3)	(-0.007 , 0.006)	(-0.013 , 0.01)	(4.084 , 4.324)	(0.398 , 0.44)	(0.107 , 0.127)	(2.539 , 3.371)
	Adolescents	(4.7 , 8.3)	(0.1 , 0.2)	(0.037 , 0.1)	(-0.002 , 0.001)	(-0.002 , 0.002)	(3.405 , 3.581)	(0.135 , 0.145)	(0.072 , 0.081)	(4.778 , 5.537)
Riboflavin	School	(5.1 , 7.8)	(-0.013 , 0.1)	(0.033 , 0.1)	(-0.004 , 0.002)	(-0.002 , 0.003)	(3.621 , 3.798)	(0.21 , 0.252)	(0.074 , 0.084)	(2.783 , 3.485)
	Preschool	(4.3 , 10.6)	(-0.2 , 0.2)	(-0.1 , 0.3)	(-0.008 , 0.006)	(-0.014 , 0.011)	(4.083 , 4.324)	(0.432 , 0.481)	(0.11 , 0.131)	(2.537 , 3.374)
	Adolescents	(4.7 , 8.3)	(0.1 , 0.2)	(0.6 , 0.9)	(-0.013 , -0.002)	(-0.004 , 0.008)	(3.403 , 3.578)	(0.45 , 0.47)	(0.14 , 0.175)	(4.795 , 5.55)
Niacin	School	(5 , 7.9)	(-0.014 , 0.1)	(0.7 , 1)	(-0.017 , 0.001)	(-0.006 , 0.006)	(3.619 , 3.795)	(0.447 , 0.466)	(0.137 , 0.169)	(2.795 , 3.492)
	Preschool	(4.4 , 10.4)	(-0.2 , 0.2)	(0.8 , 1.4)	(-0.024 , 0)	(-0.032 , 0.001)	(4.081 , 4.319)	(0.464 , 0.486)	(0.152 , 0.193)	(2.574 , 3.392)
	Adolescents	(4.6 , 8.2)	(0.1 , 0.2)	(0.045 , 0.1)	(-0.003 , 0.001)	(-0.002 , 0.002)	(3.404 , 3.581)	(0.167 , 0.177)	(0.079 , 0.09)	(4.783 , 5.541)
Vitamin B	School	(5 , 7.8)	(-0.012 , 0.1)	(0.1 , 0.2)	(-0.004 , 0.002)	(-0.003 , 0.003)	(3.622 , 3.798)	(0.208 , 0.23)	(0.079 , 0.09)	(2.776 , 3.475)
	Preschool	(4.7 , 10.6)	(-0.2 , 0.1)	(-0.033 , 0.3)	(-0.007 , 0.005)	(-0.01 , 0.008)	(4.083 , 4.324)	(0.354 , 0.392)	(0.103 , 0.121)	(2.544 , 3.378)
	Adolescents	(4.6 , 8.3)	(0.1 , 0.2)	(13.7 , 19.8)	(-0.4 , -0.1)	(-0.1 , 0.1)	(3.404 , 3.581)	(3.043 , 3.364)	(8.849 , 10.044)	(4.783 , 5.539)
Folate	School	(5 , 7.9)	(-0.014 , 0.1)	(17.2 , 24.4)	(-0.7 , -0.1)	(-0.3 , 0.1)	(3.622 , 3.796)	(3.238 , 3.508)	(9.116 , 10.328)	(2.785 , 3.478)
	Preschool	(4.8 , 10.6)	(-0.2 , 0.1)	(16.1 , 31.9)	(-0.9 , -0.1)	(-0.6 , 0.3)	(4.084 , 4.324)	(3.149 , 3.403)	(10.887 , 12.183)	(2.529 , 3.363)
	Adolescents	(4.7 , 8.3)	(0.1 , 0.2)	(21.9 , 33.8)	(-0.5 , 0.046)	(-0.3 , 0.2)	(3.407 , 3.584)	(4.881 , 5.627)	(18.154 , 19.935)	(4.764 , 5.524)
Vitamin A	School	(5 , 7.8)	(-0.012 , 0.1)	(26.1 , 39.8)	(-1.1 , -0.012)	(-0.3 , 0.4)	(3.62 , 3.796)	(5.561 , 6.42)	(18.565 , 20.367)	(2.797 , 3.488)
	Preschool	(4.4 , 10.4)	(-0.2 , 0.2)	(10.5 , 49)	(-1.2 , 0.8)	(-0.3 , 1.8)	(4.086 , 4.325)	(4.645 , 5.2)	(26.037 , 29)	(2.541 , 3.368)
	Adolescents	(4.7 , 8.4)	(0.1 , 0.2)	(0.2 , 0.5)	(-0.006 , 0.003)	(-0.003 , 0.006)	(3.404 , 3.581)	(0.354 , 0.369)	(0.121 , 0.147)	(4.788 , 5.543)
Vitamin E	School	(5 , 7.9)	(-0.016 , 0.1)	(0.2 , 0.4)	(-0.007 , 0.005)	(-0.001 , 0.011)	(3.623 , 3.799)	(0.342 , 0.355)	(0.116 , 0.139)	(2.768 , 3.467)
	Preschool	(4.3 , 10.1)	(-0.2 , 0.2)	(0.3 , 0.7)	(-0.011 , 0.006)	(-0.019 , 0.009)	(4.082 , 4.322)	(0.347 , 0.362)	(0.13 , 0.16)	(2.552 , 3.383)
	Adolescents	(4.6 , 8.4)	(0.1 , 0.2)	(0.1 , 0.3)	(-0.005 , 0.003)	(-0.005 , 0.003)	(3.405 , 3.581)	(0.358 , 0.374)	(0.118 , 0.143)	(4.783 , 5.542)
Vitamin D	School	(5.1 , 7.8)	(-0.011 , 0.1)	(0.1 , 0.3)	(-0.011 , 0.002)	(-0.003 , 0.009)	(3.621 , 3.797)	(0.37 , 0.386)	(0.117 , 0.14)	(2.78 , 3.483)
	Preschool	(4.2 , 10.5)	(-0.2 , 0.2)	(-0.1 , 0.4)	(-0.013 , 0.007)	(-0.004 , 0.027)	(4.083 , 4.324)	(0.38 , 0.399)	(0.142 , 0.178)	(2.548 , 3.369)

Table A.1: Credible Intervals.

A.2 Model 2

Nutrient	Group	β_0	β_x	μ_x	σ_{u_i}	σ_x	σ_w
Calcium	Adolescents	(42.73 ,49.98)	(-0.99 ,-0.31)	(9.88 ,10.32)	(3.41 ,3.59)	(4.8 ,5.56)	(22.42 ,24.25)
	School	(41.94 ,51.47)	(-0.88 ,0.35)	(7.35 ,7.78)	(3.62 ,3.8)	(2.77 ,3.47)	(23.05 ,24.79)
	Preschool	(36.03 ,75.57)	(-2.24 ,3.28)	(6.84 ,7.36)	(4.09 ,4.33)	(2.53 ,3.36)	(46.88 ,50.99)
Iron	Adolescents	(0.7 ,0.79)	(-0.02 ,-0.01)	(9.88 ,10.33)	(3.41 ,3.59)	(4.79 ,5.55)	(0.28 ,0.3)
	School	(0.74 ,0.88)	(-0.03 ,-0.01)	(7.35 ,7.78)	(3.62 ,3.8)	(2.78 ,3.48)	(0.3 ,0.32)
	Preschool	(0.78 ,0.97)	(-0.04 ,-0.01)	(6.83 ,7.36)	(4.08 ,4.32)	(2.53 ,3.35)	(0.36 ,0.4)
Magnesium	Adolescents	(17.07 ,18.77)	(-0.28 ,-0.12)	(9.91 ,10.35)	(3.4 ,3.58)	(4.85 ,5.61)	(5.82 ,6.3)
	School	(15.7 ,18.03)	(-0.28 ,0.02)	(7.34 ,7.78)	(3.62 ,3.8)	(2.76 ,3.47)	(5.38 ,5.79)
	Preschool	(16.12 ,18.81)	(-0.4 ,-0.03)	(6.83 ,7.36)	(4.09 ,4.33)	(2.51 ,3.34)	(5.04 ,5.48)
Phosphorus	Adolescents	(67.86 ,74.76)	(-0.66 ,-0.03)	(9.9 ,10.35)	(3.4 ,3.58)	(4.84 ,5.61)	(23.56 ,25.46)
	School	(65.9 ,75.19)	(-0.98 ,0.21)	(7.34 ,7.78)	(3.62 ,3.8)	(2.76 ,3.47)	(21.14 ,22.73)
	Preschool	(69.33 ,81.1)	(-1.67 ,-0.05)	(6.84 ,7.37)	(4.08 ,4.32)	(2.52 ,3.36)	(22.12 ,24.04)
Potassium	Adolescents	(113.5 ,127.5)	(-1.46 ,-0.17)	(9.88 ,10.33)	(3.41 ,3.59)	(4.8 ,5.57)	(46.62 ,50.38)
	School	(107.3 ,126.25)	(-0.91 ,1.54)	(7.34 ,7.77)	(3.63 ,3.8)	(2.76 ,3.45)	(44.36 ,47.68)
	Preschool	(98.38 ,128.43)	(1.43 ,5.58)	(6.85 ,7.37)	(4.08 ,4.32)	(2.55 ,3.38)	(54.57 ,59.54)
Zinc	Adolescents	(0.54 ,0.66)	(-0.01 ,-0.002)	(9.89 ,10.34)	(3.41 ,3.58)	(4.82 ,5.58)	(0.34 ,0.37)
	School	(0.57 ,0.71)	(-0.02 ,-0.01)	(7.35 ,7.78)	(3.62 ,3.8)	(2.78 ,3.48)	(0.32 ,0.35)
	Preschool	(0.6 ,0.77)	(-0.03 ,-0.01)	(6.84 ,7.36)	(4.08 ,4.32)	(2.55 ,3.37)	(0.27 ,0.29)
Vit C	Adolescents	(3.02 ,5.19)	(0.06 ,0.26)	(9.91 ,10.35)	(3.4 ,3.58)	(4.85 ,5.6)	(6.92 ,7.48)
	School	(-0.67 ,2.9)	(0.47 ,0.94)	(7.34 ,7.77)	(3.63 ,3.8)	(2.75 ,3.45)	(6.55 ,7.14)
	Preschool	(-3.65 ,2.01)	(0.94 ,1.74)	(6.88 ,7.4)	(4.07 ,4.31)	(2.62 ,3.44)	(8.16 ,9.22)
Thiamin	Adolescents	(0.08 ,0.15)	(-0.01 ,0)	(9.88 ,10.33)	(3.41 ,3.59)	(4.8 ,5.57)	(0.27 ,0.29)
	School	(-0.23 ,0.42)	(-0.03 ,0.05)	(7.34 ,7.77)	(3.63 ,3.8)	(2.76 ,3.45)	(1.5 ,1.61)
	Preschool	(-0.05 ,0.6)	(-0.06 ,0.03)	(6.83 ,7.36)	(4.09 ,4.33)	(2.51 ,3.35)	(1.39 ,1.51)
Riboflavin	Adolescents	(0.08 ,0.18)	(-0.01 ,0)	(9.89 ,10.33)	(3.41 ,3.59)	(4.8 ,5.57)	(0.33 ,0.36)
	School	(-0.3 ,0.44)	(-0.03 ,0.07)	(7.34 ,7.77)	(3.63 ,3.8)	(2.76 ,3.46)	(1.66 ,1.78)
	Preschool	(-0.03 ,0.73)	(-0.07 ,0.04)	(6.84 ,7.36)	(4.09 ,4.33)	(2.53 ,3.36)	(1.61 ,1.75)
Niacin	Adolescents	(0.82 ,0.93)	(-0.01 ,-0.004)	(9.89 ,10.33)	(3.41 ,3.58)	(4.8 ,5.57)	(0.37 ,0.4)
	School	(0.82 ,1.01)	(-0.03 ,-0.001)	(7.35 ,7.78)	(3.62 ,3.8)	(2.79 ,3.49)	(0.42 ,0.45)
	Preschool	(0.87 ,1.08)	(-0.04 ,-0.01)	(6.85 ,7.37)	(4.08 ,4.32)	(2.54 ,3.38)	(0.41 ,0.45)
Vitamin B6	Adolescents	(0.11 ,0.2)	(-0.01 ,0)	(9.88 ,10.33)	(3.41 ,3.59)	(4.81 ,5.57)	(0.31 ,0.34)
	School	(-0.2 ,0.24)	(-0.004 ,0.05)	(7.34 ,7.77)	(3.63 ,3.8)	(2.74 ,3.45)	(1.03 ,1.11)
	Preschool	(0.04 ,0.62)	(-0.05 ,0.03)	(6.84 ,7.36)	(4.09 ,4.33)	(2.53 ,3.36)	(1.24 ,1.34)
Folate	Adolescents	(17.15 ,21.82)	(-0.58 ,-0.15)	(9.89 ,10.33)	(3.41 ,3.58)	(4.8 ,5.57)	(15.7 ,16.96)
	School	(18.69 ,24.7)	(-0.93 ,-0.16)	(7.34 ,7.78)	(3.62 ,3.8)	(2.77 ,3.47)	(14.26 ,15.35)
	Preschool	(20.3 ,27.86)	(-1.19 ,-0.15)	(6.84 ,7.36)	(4.09 ,4.33)	(2.52 ,3.36)	(14.43 ,15.69)
Vit A	Adolescents	(25.94 ,37.53)	(-0.84 ,0.23)	(9.89 ,10.33)	(3.41 ,3.59)	(4.8 ,5.57)	(38.54 ,41.65)
	School	(32.58 ,51.13)	(-2.23 ,0.14)	(7.34 ,7.78)	(3.62 ,3.8)	(2.76 ,3.47)	(44.34 ,47.69)
	Preschool	(37.19 ,54.23)	(-1.72 ,0.62)	(6.83 ,7.36)	(4.09 ,4.33)	(2.52 ,3.36)	(33.39 ,36.25)
Vit E	Adolescents	(0.37 ,0.45)	(-0.01 ,0.002)	(9.89 ,10.33)	(3.41 ,3.58)	(4.81 ,5.57)	(0.27 ,0.29)
	School	(0.35 ,0.45)	(-0.01 ,0.01)	(7.34 ,7.78)	(3.63 ,3.8)	(2.76 ,3.46)	(0.23 ,0.25)
	Preschool	(0.39 ,0.54)	(-0.02 ,0.004)	(6.84 ,7.36)	(4.08 ,4.32)	(2.54 ,3.37)	(0.28 ,0.3)
Vit D	Adolescents	(0.22 ,0.31)	(-0.01 ,0.001)	(9.88 ,10.33)	(3.41 ,3.59)	(4.8 ,5.57)	(0.31 ,0.34)
	School	(0.25 ,0.39)	(-0.01 ,0.003)	(7.35 ,7.78)	(3.62 ,3.8)	(2.77 ,3.48)	(0.34 ,0.36)
	Preschool	(0.39 ,0.63)	(-0.03 ,0.001)	(6.83 ,7.36)	(4.09 ,4.33)	(2.52 ,3.34)	(0.47 ,0.52)

Table A.2: Credible Intervals.

Appendix B

Compute Residuals

B.1 Model 1

Following the idea in [2] where given the measurement error model

$$\begin{aligned} y_i &= \beta_0 + x_i\beta_x + q_i, \quad q_i|\sigma_q^2 \overset{iid}{\sim} (0, \sigma_q^2), \\ Y_{ij} &= y_i + w_{ij}, \quad w_{ij}|\sigma_{w_i}^2 \overset{iid}{\sim} N(0, \sigma_{w_i}^2) \\ X_{ij} &= x_i + u_{ij}, \quad u_{ij}|\sigma_{u_i}^2 \overset{iid}{\sim} N(0, \sigma_{u_i}^2). \end{aligned} \tag{B.1}$$

Let

$$r_i = \bar{Y}_{ij} - \hat{y}_i$$

$$e_i = q_i + w_{ij} - \beta_x u_{ij}$$

then

$$\begin{aligned} Var(e_i) &= \sigma_q^2 + \sigma_w^2 + \beta_x^2\sigma_u^2 - 2\beta_x cov(X_{ij}, Y_{ij}) \\ &= \sigma_q^2 + \sigma_w^2 + \beta_x^2\sigma_u^2 - 2\beta_x^2\sigma_x^2 \end{aligned} \tag{B.2}$$

approximately (B.1)estimate r_i^2 but changes in r_i non necessarily reflect changes in $Var(e_i)$ when the measurement error variances or covariance change over i. In that case we consider a modified squared residual

$$msr_i = r_i^2 - (\sigma_q^2 + \sigma_w^2 + \beta_x^2\sigma_{u_i}^2 - 2\beta_x^2\sigma_x^2)$$

To compute msr_i in a bayesian way one have simulate from the posterior of $\theta = (\beta_x, \sigma_x, \sigma_w, \sigma_q)$,$\eta = (\alpha_*, \sigma_x)$, and simulate from the posterior of x_i in next way:

1. Select the size K for the sample.
2. Simulate $\theta^{(k)}$ from the posterior distribution of θ,$k = 1, \ldots, K$.
3. Simulate $\hat{x}_i$ using $\eta^{(i)}$, $i = 1, \ldots N_g$. N_g is the number of observations X_{ij} for age group g.

4. Compute $msr_i^{(k)}$ using $\theta^{(k)}$ for every $\hat{x}_i$, then $msr_i = \frac{\sum_K^{k=1} msr_i^{(k)}}{K}$. At the end $msr = (msr_1, \ldots, mrs_{N_g})$

B.2 Model 2

Considering the model

$$
\begin{aligned}
Y_{ij} = \beta_0 + x_i\beta_x + w_{ij}, \quad w_{ij}|\sigma_w^2 &\overset{iid}{\sim} (0, \sigma_w^2) \\
X_{ij} = x_i + u_{ij}, \quad u_{ij}|\sigma_{u_i}^2 &\overset{iid}{\sim} (0, \sigma_u^2) \\
x_i|\mu_x, \sigma_x^2 &\overset{iid}{\sim} N\big(\mu_x, \sigma_x^2\big) \\
Y_{ij}|\beta_x, x_i, \sigma_w^2 &\overset{iid}{\sim} N\big(\beta_0 + x_i\beta_x, \sigma_w^2\big) \\
X_{ij}|x_i, \sigma_{u_i}^2 &\overset{iid}{\sim} N\big(x_i, \sigma_u^2\big)
\end{aligned} \tag{B.3}
$$

and following the idea to compute the residual of model 1, it will be obtained

$$r_i = \bar{Y_{ij}} - \hat{Y}_{ij}$$

$$e_i = w_{ij} - \beta_x u_i$$

because w_{ij}, u_{ij} are independent

$$Var(e_i) = \sigma_w^2 + \beta_x^2\sigma_u^2.$$

Therefore

$$msr_i = r_i^2 - (\sigma_w^2 + \beta_x^2\sigma_{u_i}^2).$$

To compute the msr_i the algorithm in the last section have to be used.

Appendix C

R code

C.1 Data Simulated

C.1.1 Data Simulated code Model 1

The library used to simulate from a inverse gamma distribution.

```
library(MCMCpack)
```

It is simulated 1000 observations.

```
set.seed(1985)
N <- 1000
```

The simulation of data is made according to model (3.13) using priors in (3.2.1). The value of the parameters are next:

```
B0 <- 70
Bz <- -0.1
Bz <- -.5

a0 <- 5
a1 <- 1.5

sigmax   <-5
sigmau   <- rinvgamma(N,sqrt(6),sqrt(36))
sigmaeps <- 21
sigmay   <- rinvgamma(N,sqrt(6),sqrt(49))
```

Then the observable varaible Z and the not observable variables x_i and y_i are simulated:

```
Z <- matrix(rnorm(N,1,1),ncol=1)
X <- matrix(rnorm(N,a0+a1*Z,sigmax) ,ncol=1)
y <- rnorm(N,B0+x2*X + Bz*Z[,1],sigmaeps)
```

Here will be simulated X_{ij} and Y_{ij}, each one with at most two replicates per observation i.

```
reps <- 2
W <- numeric(0)
  for (i in 1:reps)
    W <- cbind(W,rnorm(N,X,sigmau))

#Adding some missing observations
W[sample(1:N,N/10*9),2] <- NA
```

```
Y <- numeric(0)
 for (i in 1:reps)
   Y <- cbind(Y,rnorm(N,y,sigmay))

#Adding some missing observations
Y[sample(1:N,N/10*9),2] <- NA
```

In order to compute the hyperparameters of the prior distribution of the prior distribution of σ_{u_i} and σ_{w_i}, it will be computed the estimator of moments respectively.

```
WW <-W[!is.na(W[,2]),]
YY <-Y[!is.na(Y[,2]),]
sdw<-apply(WW,1,sd)
sdY<-apply(YY,1,sd)

media.sdw <-mean(sdw)
sd.sdw    <-sd(sdw)
mw        <-media.sdw

# Estimator of the shape for sigma_w_i
alfaw <-(mw^2*length(sdw))/sum(sdw^2)+2

# Estimator of the scale for sigma_w_i
betaw <- mw*(alfaw-1)

my <-mean(sdY)
sy <-sd(sdY)

# Estimator of the shape for sigma_u_i
alfay <-(my^2)*length(sdY)/sum(sdY^2)+2

# Estimator of the scale for sigma_u_i
betay <- my*(alfay-1)
```

C.1.2 Data Simulated code Model 2

It is simulated 1000 observationsas according to (3.13) using priors in (3.2.1). The value of the parameters are below,

```
N    <- 1000
reps <- 2

B0 <- 70
B1 <- -.5

sigmax <-5
```

```
mux    <-5
sigmau <- rinvgamma(N,sqrt(6.5),sqrt(60))
sigmay <- 0.69
```

The variables x_i and y_i are simulated:

```
x <-rnorm(mux,sigmax)
X <- matrix(rnorm(N,x,sigmax) ,ncol=1)
y <- B0+B1*X
```

Now variables X_{ij} and Y_{ij} are simulated:

```
W <- numeric(0)
for (i in 1:reps)
  W <- cbind(W,rnorm(N,X,sigmau))

# Adding some missing observations
W[sample(1:N,N/10*9),2] <- NA


Y <- numeric(0)
for (i in 1:reps)
  Y <- cbind(Y,rnorm(N,y,sigmay))

##Adding some missing observations
Y[sample(1:N,N/10*9),2] <- NA
```

It will be necessary to give of estimation of μ_x in order to give a prior value of this parameter. For this reason it is compute $\bar{X}_{ij}$ as estimation of μ_x.

```
mW<-apply(W,1,function(x) mean(x,na.rm=T))
mpob=mean(mW)
```

C.2 R code for Gibbs Sampler

Before to execute a Gibbs Sampler code it will be necessary to call libraries that will be required

```
library(rjags)
library(coda)
```

Then the data are charged:

```
data<-read.csv("data_fin_dep.csv",head=T)
hyp<-read.csv("hyperparameter.csv",head=T)

z<-data[,c(18,72,26)]
Y1<-data[,c(32:46)]
Y2<-data[,c(53:67)]
W1<-data[,68]
W2<-data[,69]
```

C.2.1 Gibbs Sampler code Model 1

Remember that it will be analyzed fifteen nutrients per three age group. For this reason will be used a loop in order to execute the Gibbs Sampler for each data set automatically.

```
for (i in 1:n){
  for(j in 1:3){
  datos2<-cbind(z[,2],Y1[,i],Y2[,i],W1,W2)
  nombre<-c(ndem[2],nz[2],nY1[i],nY2[i],nW1,nW2)

  analiz<-datos2[datos$\$$GrupoPOB==j,]

  Y<-analiz[,3:4]
  Z<-analiz[,2]
  W<-analiz[,5:6]
  N<-dim(analiz)[1]

  # Values of hyperparameters of the inverse gamma distribution.
  # These values are save in  hyperparameter.csv file.

  u1<-hyp[1,2*j-1]
  u2<-hyp[2,j*2]
  p1<-hyp[i+1,2*j-1]
  p2<-hyp[i+1,2*j]


  dataList = list(
    Y = Y,
    Z = Z,
    W=W,
    nalphas=2,
    nbetas=3,
    Nobservations=N,
    Nreplications=2,
    u1=u1,
    u2=u2,
    py1= p1,
    py2=p2  )

 # Here begin the Gibb Samper rutine

model= "model
{#BEGIN MODEL

for (i in 1:Nobsevations)
 {#BEGIN FOR i in 1:Nobservations
#Outcome model
Y[i,1]~ dnorm(y[i],tauy[i])
Y[i,2]~ dnorm(y[i],tauy[i])
y[i]~dnorm(meany[i],taueps)
```

```
meany[i]<-beta[1]+beta[2]*X[i]+beta[3]*Z[i]

#Replication model
for (j in 1:Nreplications)
{W[i,j]~dnorm(X[i],tauu[i])}

#Exposure model
X[i]~dnorm(meanX[i],taux)
meanX[i]<-alpha[1]+alpha[2]*Z[i]

} #END FOR i in 1:Nobservations


#Noninformative priors on the model parameters

 for (j in 1:Nobservations)
{
tauy[j]~dgamma(py1,py2)
tauu[j]~dgamma(u1,u2)
 }


taueps~dgamma(3,1)
taux~dgamma(3,1)


#Priors for alpha and beta
for (i in 1:nalphas)
{alpha[i]~dnorm(0,1.0E-6)}

for (i in 1:nbetas)
{beta[i]~dnorm(0,1.0E-6)}

#Deterministic transformations: standard deviations

sigmaeps<-1/sqrt(taueps)
sigmax<-1/sqrt(taux)

m.sigmau<-mean(1/sqrt(tauu))
m.sigmay<-mean(1/sqrt(tauy))


}#END MODEL"

   m=jags.model(textConnection(model),dataList,n.adapt=10000,n.chains=2)
   update(m, 10000) # burn in

  # DIC especifications
   dic.pd <- dic.samples(m, 5000, "pD")
```

```
   penalty<-sum(dic.pd[[1]])
   m.deviance<-sum(dic.pd[[2]])
   DIC<-penalty+m.deviance


  # Parameters to print
   parameters= c("beta","alpha","sigmax","sigmaeps","m.sigmau","m.sigmay")

  # Chains especifications
   res1 = coda.samples(m,parameters, n.iter=100000, nthin=200,n.burnin=10000)

  # Compute Gelman test
gelman.diag(mcmc.list(res1))

  # Compute Geweke test
    geweke.diag(mcmc.list(res1))
  }}
```

C.2.2 Gibbs Sampler code Model 2

The estimation of parameters in model(3.13) using priors in (3.2.1) is given next. Fist it is necessary to call data:

```
library(rjags)
library(coda)

datos<-read.csv("data_fin_dep.csv",head=T)
hyp<-read.csv("hyperparameter.csv",head=T)
para<-des[,8:13]
datos<-datos[,-1]

bmi<-datos$\$$peso/(datos$\$$talla/100)^2

data<-cbind(datos,bmi)
variable<-names(data)

demo<-data[,c(1,20)]
z<-data[,c(18,72)]
Y1<-data[,c(32:46)]
Y2<-data[,c(53:67)]
W1<-data[,68]
W2<-data[,69]

n<-length(Y1)
```

Now the loop that contain the Gibbs Sampler algorithm is given:

```
for (i in 1:n){
  for(j in 1:3){
datos2<-cbind(demo[,2],z[,2],Y1[,i],Y2[,i],W1,W2)
nombre<-c(ndem[2],nz[j],nY1[i],nY2[i],nW1,nW2)

analiz<-datos2[datos$\$$GrupoPOB==j,]

Y<-analiz[,3:4]
Z<-analiz[,2]
W<-analiz[,5:6]
N<-dim(analiz)[1]
u1<-para[1,2*j-1]
u2<-para[2,j*2]
e1<-para[i+1,2*j-1]
e2<-para[i+1,2*j]

mW<-apply(W,1,function(x) mean(x,na.rm=T))
mpob<-mean(mW)
sW<-sd(mW)


dataList = list(
Y = Y,
W=W,
nbetas=2,
Nobservations=N,
Nreplications=2,
u1=u1,
u2=u2,
e1= e1,
e2=e2,
mpob=mpob
)

model= "model
{#BEGIN MODEL
for (i in 1:Nobservations)
{#BEGIN FOR i in 1:Nobservations
#Outcome model
Y[i,2]~ dnorm(meany[i],tauy)
Y[i,1]~ dnorm(meany[i],tauy)
meany[i]<-beta[1]+beta[2]*X[i]

#Replication model
for (j in 1:Nreplications)
{W[i,j]~dnorm(X[i],tauu[i])}

#Exposure model
X[i]~dnorm(meanX,taux)
```

```
}#END FOR i in 1:Nobservations

meanX~dnorm(mx,taux)

#Noninformative priors on the model parameters
for (i in 1:Nobservations){
tauu[i]~dgamma(u1,u2)}
tauy~dgamma(e1,e2)
taux~dgamma(3,1)
mx~dnorm(mpob,.1)


#Priors for alpha and beta
for (i in 1:nbetas)
{beta[i]~dnorm(0,1.0E-6)}

#Deterministic transformations: standard deviations

sigmau<-mean(1/sqrt(tauu))
mediaX<-mean(X)
sigmay<-mean(1/sqrt(tauy))
sigmax<-1/sqrt(taux)


}#END MODEL"

m=jags.model(textConnection(model),dataList,n.adapt=1,n.chains=2)
update(m, 20000) # burn in

parameters= c("beta","sigmau","sigmax","sigmay","mediaX")
res1 = coda.samples(m,parameters, n.iter=100000, nthin=200,n.burnin=50000)

# Compute DIC
  dic.pd <- dic.samples(m, 20000, "pD") # Deviance Information Criterion
  penalty<-sum(dic.pd[[1]])
  m.deviance<-sum(dic.pd[[2]])
  DIC<-penalty+m.deviance

#Gelman ans Geweke test
 gelman.diag(mcmc.list(res1))
 geweke.diag(mcmc.list(res1))

  }}
```

Bibliography

[1] M. de Castro, H. Bolfarine and M. Galea. *Bayesian Inference in Measurement Error Models for replicated Data*. Envirometrics Vol. 24, 22-30, 2012.

[2] J. Buonaccorsi. *Measurement Error Models, Methods and Applications*. Ed. Chapman and Hall Books, 2010.

[3] R. Carroll, K. Roeder and L. Wasserman. *Flexible Parametric Measurement Error Models*. Biometrics 55, 44-54, 1999.

[4] C. Robert and G. Cassella. *Monte Carlo Statistical Methods, Second Edition.* ed. Springer, 2004.

[5] L. Freedman, A. Schastzkint, D. Midthune and V. Kipnis. *Dealing with Dietary Measurement Error in Nutritional Cohort Studies*. Journal of the National Cancer Institute Vol.103, Issue 14, 2011.

[6] W. Fuller. *Measurement Error Models.* ed. John Wiley and Sons, 1987.

[7] A. Gelman, J. Carlin and H. Stern. *Bayesian Data Analysis, Second Edition.* ed. Chapman and Hall CRC, 2004.

[8] S. Nusser, A. Carriquiry, K. Dodd and W. Fuller. *A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distribution*. Journal of the American Statistical Association Vol. 91, No.436, 1996.

[9] D. Ruppert, L. Stefansky, C. Crainiceanu. *Measurement Error in Nonlinear Models A Modern Perspective*, Second Edition. ed. Chapman and Hall CRC, 2006.

[10] Instituto Nacional de Salud Pública. *Manual para la Aplicación del Cuestionario de 24 Horas.* ENSANUT, 2012.

Websites

[11] Package rjags. rjags - `http://cran.r-project.org/web/packages/rjags/rjags.pdf`.

[12] ENSANUT 2012 - `http://ensanut.insp.mx`.

[13] CODA - `http://cran.r-project.org/web/packages/coda/coda.pdf`.