SKILLS TO NOT FALL BEHIND IN SCHOOL

A PREPRINT

Felipe Maia Polo*
Institute of Mathematics and Statistics
University of São Paulo
São Paulo, Brazil
felipemaiapolo@gmail.com

January 6, 2020

ABSTRACT

Many recent studies emphasize how important the role of cognitive and social-emotional skills can be in determining people's quality of life. Although skills are of great importance in many aspects, in this paper we will focus our efforts to better understand the relationship between several types of skills with academic progress delay. Our dataset contains the same students in 2012 and 2017, and we consider that there was a academic progress delay for a specific student if he/she progressed less than expected in school grades. Our methodology primarily includes the use of a Bayesian logistic regression model and our results suggest that both cognitive and social-emotional skills may impact the conditional probability of falling behind in school, and the magnitude of the impact between the two types of skills can be comparable.[2]

Keywords Education · Bayes · Progress · Skills

1 Introduction

In a recent report by the World Economic Forum [1] there is a clear paradigm shift when it comes to skills needed by today's children, youth and adults. People are rethinking the skills considered fundamental to humans in recent decades, introducing a new skill set. The authors call this core skill set "21st Century Skills": in total there are 16 skills splitted into three categories, (i) Foundational Literacies, (ii) Competencies, and (iii) Character Qualities. The first category includes skills that encompass more technical knowledge such as math, literacy and financial education. On the other hand, the second includes skills such as creativity, communication, and the third includes skills such as curiosity, leadership, and persistence. It is important to point out that in the report itself, the authors classify the 10 skills belonging to the last two categories as part of the set of Social and Emotional skills, i.e. skills related to the way people interact to the outside world and to themselves. Figure 1 was extracted from [1] and exposes the 16 skills for the 21st century:

*https://felipemaiapolo.github.io/
[2]The code used in this paper can be found in `https://github.com/felipemaiapolo/paper_skills`.

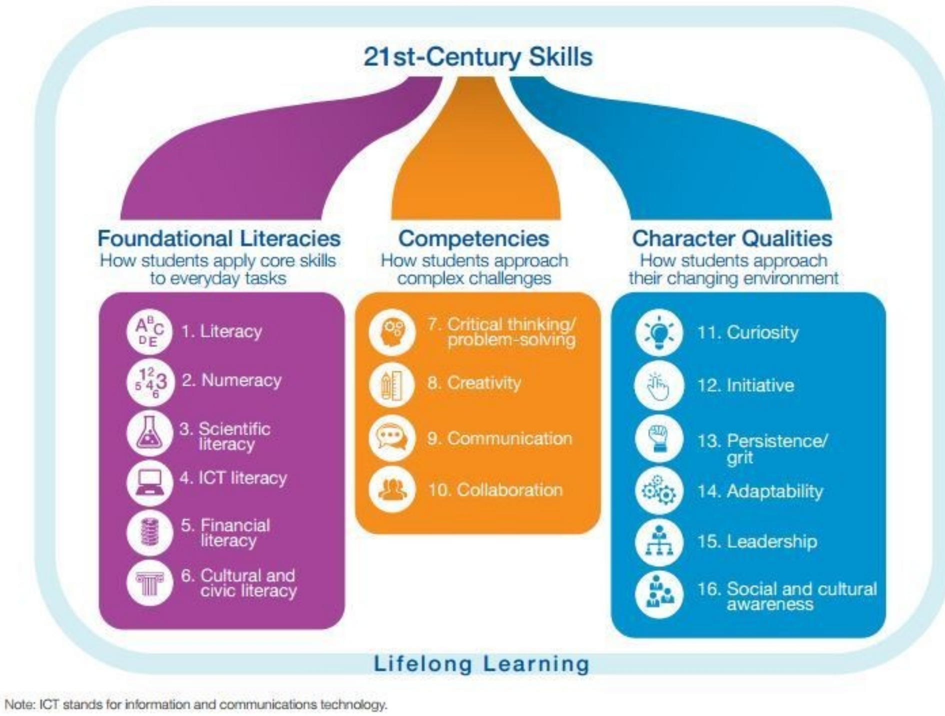

Figure 1: The 16 Skills for the 21st Century [3]

Just as we call the last 10 Social-Emotional skills we can call the first 6 as cognitive skills. Recent studies show findings by leading education researchers that these two types of skills take on great importance in defining people's future outcomes such as education, salaries and employment level [2, 3]. Cognitive skills, for example, have a major impact on income and employability while Social-Emotional skills, which account for more than 60% of fundamental 21st Century skills, have a positive impact on well-being and satisfaction and improving people's health through their lifestyle [4].

Although cognitive and social-emotional skills are of great importance in people's lives in many ways, in this paper we will focus on how these skills can affect the academic progress of students who were attending school in 2012 and 2017 in a city in the countryside state of São Paulo in Brazil. We have a rich dataset containing demographic, socioeconomic, and personal characteristics of students, and we can account for academic progress delay in a way that minimizes measurement errors - we consider that students who had academic progress delay between 2012 and 2017 were those who have progressed less in grades than expected, that is, those who fell behind in school. Our results suggest that, in fact, both cognitive and social-emotional skills impact how students progress in school.

2 Literature

In this paper we are concerned about the academic progress delays that can be caused either by school failure or temporary dropout. Failure and school failure can be explained from many angles, including personal skills, family and socioeconomic characteristics. Regarding failure, in addition to more direct variables such as cognitive skills, [5] gives evidence that variables such as income and parental education can have a large negative impact on the probability of failure. Regarding temporary dropout, some more concrete factors could be the need to enter the labor market prematurely [6], lack of motivation and expectations about future studies return [7] or low parental education, which is a great proxy for the family's permanent income. In addition, students who have failed in the past form a group which is more vulnerable and prone to drop out [8].

In a recent work [9], the author showed evidences that cognitive and social-emotional skills can have an impact on dropping out and reaching high school, which are variables of school progress. Although we use a dataset very similar to that work [9], we believe that this is an important paper to the literature because of the following factors: (i) In our work we used a larger number of social-emotional skills, which were not previously available and which are little

[3]Figure extracted from 1.

explored in literature; (ii) Our dependent variable is a more general variable, which is academic progress delay, what helps us to deliver a broader message; (iii) We added an analysis of the odds of falling behind in school, which is natural within logistic regression framework; (iv) We use a Bayesian methodology, which among many benefits, we can highlight the fact that we will obtain the entire posterior distribution of the parameters of interest and not just a point estimate.

3 Motivation

Firstly, it is important to know that a significant portion of our sample had educational progress delay between 2012 and 2017, as shown in Table 1:

	Behind	Not Behind
Proportion	0.13	0.87

Table 1: Proportion of students who fell behind in school between 2012 and 2017

Secondly, given the importance that social-emotional and cognitive skills can play in shaping people's paths, especially regarding their success during and after school, we will be inclined to better understand how these skills may relate to falling behind in school, which may be the result of failure or not. In our dataset we have twelve skills that can be considered good or bad by common sense. Two of them can be considered cognitive skills - Literacy and Numeracy[4] - and the other ten are social-emotional skills [5]: Assertiveness, Activity, Altruism, Compliance, Order, Self-Discipline, Anxiety, Depression, Aesthetics and Ideas. Two questions that will motivate us from this moment are: (i) How do these skills relate to falling behind in school? (ii) How can we compare the impacts of cognitive and social-emotional skills? In our case, what we will call academic progress delay, or to fall behind in school, is actually the age-grade distortion acquired between 2012 and 2017, i.e. if a student progressed less than five grades in five years (2012 to 2017) we say he/she fell behind or had an academic progress delay. The skills we are using in this paper were measured in 2012 according to some procedures detailed in Section 5. In Table 2, one can see the averages of various skills measured in standard deviations[6], conditioned on the variable 'Behind', which indicates students which fell behind between 2012 and 2017:

	Behind	Not Behind	Diff
Language 2012	-0.52	0.08	-0.6
Mathematics 2012	-0.44	0.07	-0.51
Activity	-0.16	0.02	-0.18
Aesthetics	0.11	-0.02	0.12
Altruism	0.19	-0.03	0.22
Anxiety	-0.11	0.02	-0.13
Assertiveness	0.27	-0.04	0.31
Compliance	-0.03	0	-0.03
Depression	0.33	-0.05	0.38
Ideas	-0.03	0	-0.04
Order	0.16	-0.02	0.19
Self-Discipline	0.05	-0.01	0.05

Table 2: Comparing the two groups regarding their personal skills

It can be seen in Table 2 that students who fell behind between 2012 and 2017 have lower average grades in language and mathematics and differ, on average, from the group that did not fall behind with respect to various social-emotional characteristics such as depression - which leads us to believe that among our students, cognitive and social-emotional skills may also be related to whether or not a student fell behind between 2012 and 2017. Although we present a characteristic of the distribution of social-emotional and cognitive skills conditional on the variable 'Behind', we will be more concerned in this paper to estimate the distribution of the 'Behind' variable conditioned to a vector of

[4]Measured by language and math scores in a standardized test.
[5]We will give a better explanation in Section 5.
[6]The variables were standardized in the sample to have zero mean and unit standard deviation.

characteristics, paying more attention to social-emotional and cognitive skills. Therefore, the descriptive analyzes done so far only serves as motivation for a more robust analysis.

4 Objective

The main objective of this paper is to better understand how cognitive and social-emotional skills relate to academic progress delay, which may be the result of failure or temporary dropout. For this purpose we use a dataset of approximately 1800 students who studied in the city of Sertãozinho in the countryside of the state of São Paulo and who were in elementary, middle school in 2012 and were re-interviewed in 2017.

5 Data

The datasets used in this study come from two field surveys conducted by LEPES/USP ("*Laboratório de Estudos e Pesquisas em Economia Social*") in the city of Sertãozinho, in the countryside of the state of São Paulo, in 2012 and 2017. In both years, information was collected about the students' family and socioeconomic context, about the situations they faced in school and about cognitive and social and emotional skills. Since cognitive and social-emotional variables are key variables in our analysis, it is important to explain in detail how they were constructed.

To assess the level of cognitive development of students in 2012 and 2017, we used the result of Language and Mathematics tests prepared by the psychometrist Dr. Ricardo Primi from items available on the platform of the National Institute for Educational Studies and Research Anísio Teixeira (INEP). The test preparation methodology includes the use of the Item Response Theory (IRT), where the final grade is not the number of correct answers, but the student's proficiency level taking into account the difficulty of the items, for example. In 2012, the students, who were in the 5th and 6th grades, answered only one test of each subject. It is important to say that the students who were in the 4th grade in 2012 answered the Big Five Inventory (BFI), discussed bellow, but did not take the exam due some bureaucratic issues.

Regarding the level of social-emotional development, the researchers who designed the research chose to use a well-established social-emotional assessment scale in the literature, which is the Big Five Inventory, or BFI, developed by [10]. Although, the BFI scale is based on the theory that a person's personality can be roughly described by big five factors, we focused in more recent measures presented in [11] which are called 'facets'. We chose to work with facets instead of using the Big Five traits because they are less broad and any result achieved could be interpreted more deeply. The ten available facets and their description, according to [12], are:

1. *Activity*: High scorers can be described as being energetic, fast-paced, and vigorous. On the other hand, low scorers can be described as unhurried, slow, and deliberate;
2. *Aesthetics*: High scorers can be described as those who value aesthetic experiences and who are moved by art and beauty. On the other hand, low scorers can be described as insensitive to art and unappreciative of beauty;
3. *Altruism*: High scorers can be described as warm, softhearted, gentle, generous, and kind. On the other hand, low scorers can be described as selfish, cynical, cold, and snobbish;
4. *Anxiety*: High scorers can be described as apprehensive, fearful, prone to worry, nervous and tense. On the other hand, low scorers can be described as calm, relaxed, stable, fearless;
5. *Assertiveness*: High scorers can be described as dominant, forceful, confident, and decisive. On the other hand, low scorers can be described as unassuming, retiring, and reticent;
6. *Compliance*: High scorers can be described as deferential, obliging, and kind. On the other hand, low scorers can be described as stubborn, demanding, headstrong, and hardhearted;
7. *Depression*: High scorers can be described as prone to feelings of guilt, sadness, hopelessness, and loneliness. On the other hand, low scorers can be described as being seldom sad, hopeful, confident, and as feeling worthwhile;
8. *Ideas*: High scorers can be described as intellectually curious, analytical, and theoretically oriented. On the other hand, low scorers can be described as pragmatic, factually oriented, and unappreciative of intellectual challenges;
9. *Order*: High scorers can be described as precise, efficient, and methodical. On the other hand, low scorers can be described as disorderly, impulsive, and careless;

10. *Self-Discipline*: High scorers can be described as organized, thorough, energetic, capable, and efficient. On the other hand, low scorers can be described as unambitious, forgetful, and absent-minded;

Firstly, we should mention that a higher score in the facets (a.k.a. social-emotional skills) can **not** always be considered a good thing. Secondly, it is important to say that we do not use the raw scores obtained after applying the scale due to the acquiescence bias. That bias is the tendency of an individual to agree or disagree with statements about his or her attitudes regardless of those [13]. According to [14], if the bias is not corrected the factor estimation may be uncorrelated with the content of the items, which would invalidate the instrument. Given that, all scores obtained for the social-emotional dimensions already discussed were corrected for acquiescence bias.

5.1 Variables description

Part of the variables selected to be used in this work was selected from the 2012 dataset, part was selected from the 2017 dataset. More information about our variables:

- Grade 2012: Variable that tells us which school grade the student was in 2012;
- Semester: Whether the student was born in the first or second semester;
- Year: born year of the student (quantitative variable);
- White: Binary variable denoting whether the student considers him/herself ethnically as white;
- Male: Binary variable denoting whether the student considers him/herself as male (opposed to female);
- Pre-K: Binary variable reported by the student whether or not he/she attended pre-kindergarten (only avaliable in 2017);
- Kinder: Binary variable reported by the student whether or not he/she attended kindergarten (only avaliable in 2017);
- Mother education: variable that tells us the level of formal education of the student's mother. For the sake of inconsistency about what the student and the student's parents were responding to, we will use a way of correcting this that was elaborated by [15]. Categories: Non-educated (omitted in regression analysis), Elementary, Middle, High, College, Unknown;
- School 2012: Categorical variable that tells us whether the student's school in 2012 was state, municipal, federal or private based on the 2012 School Census. Categories: State (omitted in regression analysis), Municipal, Federal or Private;
- Failed before 2012: Binary variable reported by the student whether or not he/she failed in school before 2012;
- Facets 2012: The social-emotional scores were standardized to have null mean and unit standard deviation in 2012 among those students who form our actual sample;
- Language and Mathematics 2012: The tests scores were standardized to have null mean and unit standard deviation in 2012 among those students who form our actual sample;
- Behind: Binary variable that tells us if the student fell behind in school between 2012 and 2017 based in his/her grades in both years;

In Appendix 10.2 one can see the main descriptive statistics of the variables used in the work.

5.2 Baseline, attrition and actual sample

Our baseline of students will consist of the 5th and 6th grade students who were present in 2012 and had no missing values in all variables except the "Pre-K", "Kinder" and "Behind" variables, which depends on the 2017 dataset. In 2012 we have 3223 students who are distributed in 2 school grades: 5th and 6th years. Among these 3223 students, 83,71% will compose our baseline according to the rule states above. Given that, we have 2698 students in our baseline, as one can check in Table 3:

	5th grade	6th grade	Total
N° Observações	1680	1543	3223
Baseline	1420	1278	2698
% Baseline	84,52%	82,82%	83,71%

Table 3: Composition of the baseline

It is important to state that our baseline is **not** our actual sample for this study for one major reason: we do not have data for some students in 2017, mainly because we could not reach them in the 2017 field research. Thus, we will use then the amount of 1780 students, which is 65,97% of the initial amount. Table 4 provides more details about all this information:

	5th grade	6th grade	Total
Baseline	1420	1278	2698
Actual sample	959	821	1780
% Actual sample	67,53%	64,24%	65,97%

Table 4: Composition of the actual sample

The problem of losing part of the sample between two time points is called attrition. When the attrition is random, that is, when the loss is independent of the characteristics of the students, we do not have many problems. However, when we lose part of the sample in a non-random way, it may be that our results are not extendable to the population of interest, that is all students from the 5th and 6th grades of the Sertãozinho school network in the countryside of the state of São Paulo. If one check in Appendix 10.1, one will see that the loss was not random: attrition was primarily related to students with less-educated mothers and those with the more problems in school.

Despite a random loss being a sufficient condition for an analysis with extendable results for the population, it is not a necessary condition in our case - I will briefly explain why. Suppose that Y is a binary variable that indicates academic progress delay between 2012 and 2017 for an individual and X a vector of characteristics of the same individual - draws of X are observable in our data. Note that our interest is to better understand the following quantity $\mathbb{P}(Y=1|X=x)$, however, we can only estimate $\mathbb{P}(Y=1|X=x, A=0)$ with our data, if A is a variable that indicates attrition. Now imagine that a Z variable vector is the only common cause of falling behind and attrition, and that U and V are other variables that impacts on progress delay or attrition as illustrated in the following causal DAG (Directed Acyclic Graph):

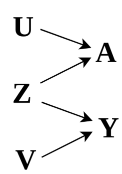

Figure 2: Attrition DAG

If the DAG is valid and Z is a vector of variables included in X, then, given $X=x$, A and Y are conditionally independent and $\mathbb{P}(Y=1|X=x) = \mathbb{P}(Y=1|X=x, A=0)$ [16]. It is reasonable to think that this condition is valid since our dataset is rich in demographic, school and personal characteristics of students. On the other hand, if the condition is not valid, we may think that our results will be extendable to a portion of the population, and it is still possible to speculate possible outcomes according to the differences shown in Appendix 10.1. Although it is not possible to estimate $\mathbb{P}(Y=1|X=x)$ directly, we will use this notation from now on taking our sample as given.

6 Methodology

6.1 The model

The model we will use in our analysis is the Bayesian Logistic Regression model. Using that model we want to directly model the logarithm of students' odds of falling behind in school given their characteristics using a linear predictor. If Y_i is a variable that indicates academic progress delay between 2012 and 2017 for student i, $x_i^\top = (1\ x_{i1}\ ...\ x_{ik})$ is a feature vector for that student, $\theta^\top = (\theta_0\ \theta_1\ ...\ \theta_k)$ a parameter vector that helps us to parameterize our model we define:

$$\pi(x_i, \theta) = \mathbb{P}(Y_i = 1 | X_i = x_i, \theta)$$

Since $\pi(x_i, \theta)$ is the probability of falling behind conditional on the characteristics x_i and on the θ parameters, we have that the log of the i student's odds to have a progress delay between 2012 and 2017, given θ, is:

$$\log\Big[O(x_i,\theta)\Big] = \log\left[\frac{\pi(x_i,\theta)}{1-\pi(x_i,\theta)}\right] = x_i^\top\theta \quad (1)$$

Although logistic regression models the log of the odds, we can obtain $\pi(x_i,\theta)$ and get the conditional probability of progress delay as follows:

$$\pi(x_i,\theta) = \sigma\big(x_i^\top\theta\big) = \frac{1}{1+\exp(-x_i^\top\theta)}$$

Where $\sigma(.)$ represents the Sigmoid function (or standard logistic distribution function).

6.2 Model Parameter Inference

In the Bayesian framework, it is common to infer about parameters as follows: (i) we adopt an prior distribution for the parameter vector, which captures our knowledge of the quantities of interest before looking at the data; (ii) after observing the data, we update our knowledge about the parameter vector by applying the Bayes Theorem.

6.2.1 Prior and posterior distributions

At first we assumed no correlation between the θ coordinates, and for each of the inputs we assumed a prior Normal distribution $N(0,1000)$, fulfilling the idea of being a weakly informative prior distribution. Our primary objective is to obtain the posterior distribution of the parameter vector, which is given by Equation 2:

$$p(\theta|Y_{1:n}=y_{1:n},X_{1:n}=x_{1:n}) = \frac{\mathbb{P}(Y_{1:n}=y_{1:n}|X_{1:n}=x_{1:n},\theta)p(\theta)}{\int\mathbb{P}(Y_{1:n}=y_{1:n}|X_{1:n}=x_{1:n},\theta)p(\theta)\mathrm{d}\theta} \quad (2)$$

Given $Y_{1:n}$ is a vector and $X_{1:n}$ is a matrix for our entire actual sample (n is the sample size). The major obstacle to calculate the posterior distribution analytically is the calculation of the integral in the denominator, which is intractable. Given that, we resort to a variation of the Hamiltonian Monte Carlo (HMC) [17] algorithm so we can sample from the posterior even without knowing its closed form. That variation of the algorithm is called "No U-Turn Samples" and it is implemented in "RStanarm" package [7] in R - that package is built upon "RStan". In our sampling process, we sampled 4 chains, each of which had a burn-in period equals 1000, a thinning of 100 and a size of 2500 (we have $L=10000$ samples in total). We conducted a convergence diagnosis of MCMC which can be viewed in more details in Appendix 10.5. After sampling a reasonable number of times from the posterior distribution, we can move on to the next step.

6.3 Analysis of Results

6.3.1 Goodness of fit

In order to understand whether the logistic regression model is a good model for our problem, we will apply a simple cross-validation procedure (one training set and one testing set) for comparing the Receiver Operating Characteristic (ROC) curves and AUC metric (Area Under Curve) between our model and a tuned Random Forest benchmark model [8], which has high predictive power. To estimate the conditional probability of falling behind in school using bayesian logistic regression, we will use the concept of posterior predictive distribution, which is calculated as follows:

$$\begin{aligned}\mathbb{P}(Y_i=1|X_i=x_i) &= \int\mathbb{P}(Y_i=1|X_i=x_i,\theta)p(\theta|Y_{1:n}=y_{1:n},X_{1:n}=x_{1:n})\mathrm{d}\theta \\ &= \mathbb{E}_{\theta|Y_{1:n},X_{1:n}}\Big[\mathbb{P}(Y_i=1|X_i=x_i,\theta)\Big|Y_{1:n}=y_{1:n},X_{1:n}=x_{1:n}\Big]\end{aligned}$$

[7] `https://cran.r-project.org/web/packages/rstanarm/index.html` - accessed in 11/11/2019

[8] The model will be tuned in a cross-validation 3-Fold procedure varying the (i) number of trees in the list c(50, 100, 150, 200, 250, 300, 400, 500, 600), (ii) the number of features used in the list c(2,3,4,5,6, 7), (iii) minimum size of nodes in c(1, 3, 5, 10, 15, 20, 30), (iv) sample fraction used in bootstrap in c(.5, .6, .8, 1) and (v) replacement in sampling in c(TRUE, FALSE).

From this moment, we can interpret i as an out-of-sample individual. As long as we expect to independently sample θ from its posterior distribution, we may resort to the Law of Large Numbers to approximate the above integral by the following mean:

$$\begin{aligned}\mathbb{P}(Y_i = 1|X_i = x_i) &\approx \frac{1}{L}\sum_{l=1}^{L}\mathbb{P}(Y_i = 1|X_i = x_i, \theta^{(l)}) \\ &= \frac{1}{L}\sum_{l=1}^{L}\sigma\big(x_i^\top \theta^{(l)}\big)\end{aligned}$$

Where $\theta \sim p(.|Y_{1:n} = y_{1:n}, X_{1:n} = x_{1:n})$, i.e., it is sampled from its posterior in a total L times. Regarding the methodology of comparison between two classifiers, we chose the ROC curves and the AUC metric because they do not depend on the cutoff threshold for classification and are easily interpretable: (i) the ROC curves graphically give us a tradoff between True Positive Rate and True Negative Rate of a binary classifier and (ii) the AUC metric is actually equivalent to the probability that a binary classifier will rank higher an instance of type '1' compared to an instance of type '0', chosen at random [18].

6.3.2 Odds analysis

The first analysis we will conduct is direct from obtaining the posterior distribution of the parameter vector. Given the property of the linearity seen in Equation 1, we have that the percentage change due to the conditional odds of falling behind due to a δ change in x_{ij}, keeping all other variables constant, is given by:

$$\Delta_\delta \% O(x_{ij}, \theta) = \frac{O(x_{ij} + \delta, \theta)}{O(x_{ij}, \theta)} - 1 = \exp\big(\delta * \theta_j\big) - 1$$

It is important to note that one important implication of linearity of the predictor is that the amount $\Delta_\delta \% O(x_{ij}, \theta)$ does not depend on i, but only on θ_j and on δ. In our analysis, we will consider $\delta = 1$, which is natural for both binary variables and quantitative variables that are measured in standard deviations. Considering $\delta = 1$, we have:

$$\Delta_1 \% O(x_{.j}, \theta) = \exp\big(\theta_j\big) - 1$$

Remember that because θ_j is a random variable, $\Delta_1 \% O(x_{.j}, \theta)$ is also a random variable and that is why we will examine its distribution directly.

6.3.3 Testing the importance of each skill variables

Our analysis regarding the importance of each of the variables in predicting progress delay in school will be analyzed in the light of a hypothesis test that we will propose inspired by [19]. In the proposed framework we have three possible hypotheses for the importance of variable j:

- H_0: the j variable is not important in determining progress delay in school;
- H_-: the j variable is negatively important in determining progress delay in school;
- H_{2+}: the j variable is positively important in determining progress delay in school;

Choosing one of the three options will be given by a decision problem under uncertainty. Recalling that θ_j assumes values in $\Theta_j = \mathbb{R}$, we want our hypothesis to be as follows:

$$\begin{cases} H_0 : \theta_j \in [\varepsilon_1,\ \varepsilon_2] \\ H_- : \theta_j \in (-\infty,\ \varepsilon_1) \\ H_+ : \theta_j \in (\varepsilon_2,\ \infty) \end{cases} \tag{3}$$

Given $\varepsilon_1 < 0$ and $\varepsilon_2 > 0$. The problem with this approach is that it is not straightforward to find ε_1 and ε_2 values that make sense. To make this task easier and the result more interpretable, first realize that $\Delta_1\%O(x_{.j}, \theta) = \exp(\theta_j) - 1$ is an increasing function in θ_j and with range equals to $\Omega_j = (-1, \infty)$. Equivalently, we can rewrite the hypothesis as follows:

$$\begin{cases} H_0 : \Delta_1\%O(x_{.j}, \theta) \in [\varepsilon_1', \ \varepsilon_2'] = \Omega_{j0} \\ H_- : \Delta_1\%O(x_{.j}, \theta) \in (-1, \ \varepsilon_1') = \Omega_{j-} \\ H_+ : \Delta_1\%O(x_{.j}, \theta) \in (\varepsilon_2', \ \infty) = \Omega_{j+} \end{cases} \tag{4}$$

In the way we lastly presented the hypotheses, one can realize it is easier to choose values of ε_1' and ε_2' that make sense. To define these values, let us remember that the percentage of students who fell behind in school between 2012 and 2017 is $p = 0.1303$, so the odds of delay is given by $\frac{p}{1-p} = 0.1499$. If p is added by d points (e.g. $d = \pm 0.01$) we get the odds to be $\frac{p+d}{1-(p+d)}$ and the percentage change in the odds will be given by:

$$\omega_d = \frac{\frac{p+d}{1-(p+d)}}{\frac{p}{1-p}} - 1 \tag{5}$$

Given constants $d_1 < 0$ and $d_2 = -d_1$, we define $\varepsilon_1' = \omega_{d_1}$ and $\varepsilon_2' = \omega_{d_2}$. The rationale can be elucidated with the following example. Suppose $d_1 = -0.01$, so $\varepsilon_1' = -8.72\%$ represents the percentage change in the odds of progress delay in school, based on our sample, due to the decreased probability of delay p by one percentage point. If $d_2 = 0.01$, then $\varepsilon_2' = 8.92\%$ represents the percentage change in odds of progress delay in school, based on our sample, due to the increased probability of delay p by one percentage point. In a way, the value of d_1 and hence the value of d_2 is still somewhat arbitrary, so in our analysis we will assume that d_1 and d_2 assume values in $\{\pm 0.01, \pm 0.02, \pm 0.03, \pm 0.04, \pm 0.05\}$ to see what our decision would be like in several scenarios.

What remains to be understood is how our decision will be made after we set d_1 and d_2. If we define a constant loss function for each of the possible decision errors, i.e. there is no reason to believe that we should value differently different types of errors, our decision to $\Delta_1\%O(x_{.j}, \theta)$ with respect to the hypotheses $H_0, \ H_-$ and H_+ is given by:

$$c_j^* = \underset{c \in \{0,-,+\}}{\text{argmax}} \ \mathbb{P}\Big(\Delta_1\%O(x_{.j}, \theta) \in \Omega_{jc}\Big) \tag{6}$$

6.3.4 Marginal effects analysis

We are interested in calculating the effect of a marginal increase in each of the skills on the predicted probability of school delay for an individual. That is, if $x_{ij} = \text{skill}_{ij}, \ j \in \{1, ..., 12\}$ denotes the value of j-th skill for i-th individual (out of sample) we want to calculate the amount:

$$\frac{\partial \mathbb{P}(Y_i = 1 | X_i = x_i)}{\partial x_{ij}}$$

By sampling $L = 10000$ from the posterior of θ, we can approximate our quantity of interest as follows:

$$
\begin{aligned}
\frac{\partial \mathbb{P}(Y_i = 1 | X_i = x_i)}{\partial x_{ij}} &= \frac{\partial}{\partial x_{ij}} \mathbb{E}_{\theta | Y_{1:n}, X_{1:n}} \Big[\mathbb{P}(Y_i = 1 | X_i = x_i, \theta) \Big| Y_{1:n} = y_{1:n}, X_{1:n} = x_{1:n} \Big] \\
&\approx \frac{\partial}{\partial x_{ij}} \frac{1}{L} \sum_{l=1}^{L} \mathbb{P}(Y_i = 1 | X_i = x_i, \theta^{(l)}) \\
&= \frac{1}{L} \sum_{l=1}^{L} \frac{\partial \mathbb{P}(Y_i = 1 | X_i = x_i, \theta^{(l)})}{\partial x_{ij}} \\
&= \frac{1}{L} \sum_{l=1}^{L} \frac{\partial \sigma\big(x_i^\top \theta^{(l)}\big)}{\partial x_{ij}} \\
&= \frac{1}{L} \sum_{l=1}^{L} \sigma\big(x_i^\top \theta^{(l)}\big) \Big[1 - \sigma\big(x_i^\top \theta^{(l)}\big)\Big] \theta_j^{(l)}
\end{aligned}
$$

Note that the marginal effects depends on the characteristics of each individual, so we will analyse the distribution of the marginal effects across our sample.

7 Results

7.1 Posterior distribution analysis

The first step in understanding our results is to analyze the posterior distribution of parameters related to skill variables [9]. In order to make the results more interpretable and easy to understand, let us look at the marginal posterior distributions, not taking in consideration the possible relationships between the parameters:

	Lower Bound - HPD 95%	Upper Bound - HPD 95%
Language 2012	-0.44	-0.08
Mathematics 2012	-0.58	-0.17
Activity	-0.22	0.07
Aesthetics	-0.11	0.19
Altruism	-0.21	0.1
Anxiety	-0.19	0.09
Assertiveness	0.1	0.41
Compliance	-0.19	0.11
Depression	0.16	0.45
Ideas	-0.06	0.23
Order	-0.06	0.23
Self-Discipline	-0.19	0.11

Table 5: HPD interval for marginal posterior distributions

In Table 5 one can see the HPD (Highest Posterior Density) intervals for the posterior distributions of our parameters - we highlight the fact that all parameters of our interest had a unimodal marginal posterior distribution. An interesting point to note is that only four of the marginal distinctions of the parameters of interest had HPD intervals that did not include zero - this gives us clues as to which may be the most important variables in predicting progress delay in school. Two of the variables we have more evidence that can predict our variable of interest have a negative impact on progress delay in school (language and math scores) and two other variables we have the most evidence that can explain our variable of interest have a positive impact on progress delay in school (depression and assertiveness scores). We will put more effort to understand the importance of variables in the next sections. In Table 6 one can see the main statistics for marginal posterior distributions:

[9]Complete results can be found in Appendix 10.3

	Mean	1st Qu.	Median	3rd Qu.
Language 2012	-0.27	-0.33	-0.27	-0.21
Mathematics 2012	-0.38	-0.45	-0.37	-0.31
Activity	-0.08	-0.13	-0.08	-0.03
Aesthetics	0.04	-0.01	0.04	0.09
Altruism	-0.05	-0.11	-0.05	0
Anxiety	-0.05	-0.09	-0.05	0
Assertiveness	0.25	0.19	0.25	0.3
Compliance	-0.04	-0.09	-0.04	0.01
Depression	0.31	0.26	0.31	0.36
Ideas	0.08	0.03	0.08	0.13
Order	0.09	0.04	0.09	0.14
Self-Discipline	-0.04	-0.09	-0.04	0.02

Table 6: Basic statistics for marginal posterior distributions

7.2 Goodness of fit

A fundamental thing to make our results interesting is the ability of our model to fit well the data we are using. To see how well our model fits the data, we proposed a cross-validation scheme comparing the predictive power of our model with a benchmark model, which in this case will be a Random Forest model. With respect to the Bayesian logistic regression model, we use the predictive distribution of Y given $X = x$. In relation to the Random Forest model, we used the combination 'number of trees'=200, 'number of variables per tree'=2, 'minimum node size'=1, 'replacement in sampling'=True, 'fraction sampled in bootstrap'=0.6 chose by means of a grid search in the training set (3-fold CV) in order to maximize the AUC metric. In Figure 3 one can see a comparison between the two models:

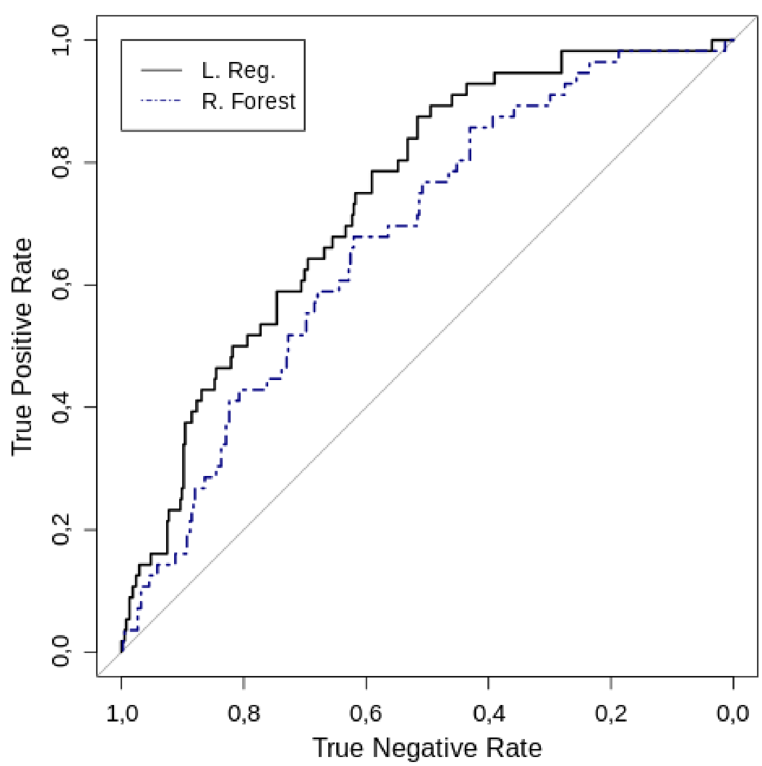

Figure 3: ROC curves - Logistic Regression Vs. Random Forest

Looking at the figure above, it is possible to see that our Bayesian logistic regression model performed better in a classification task when compared to a Random Forest tuned model, which we traditionally consider with great predictive power. We can reiterate our result by looking at Table 7:

	AUC
Logistic Regression	0.74
Random Forest	0.68

Table 7: AUC - Logistic Regression Vs. Random Forest

The great lesson we get from the results seen in this section is that our model could fit well to our data even when compared to more flexible models like a tree ensemble model, that is the random forest.

7.3 Odds Analysis

Although we get an important message by looking at the results displayed in Tables 5 and 6, interpretation may not be as straightforward as desired, especially when parameters assume larger magnitudes. Thus, using the methodology outlined in Section 6.3.2, we obtain the distribution of percentage changes in odds (conditional on the x feature vector) of falling behind in school given an increase in the magnitude of a standard deviation in a skill of interest. In Table 8 we can see the HPD intervals of the percentage change distributions in the odds of falling behind in school due to an increase in the magnitude of a standard deviation in a skill of interest:

	Lower Bound - HPD 95%	Upper Bound - HPD 95%
Language 2012	-0.37	-0.1
Mathematics 2012	-0.44	-0.16
Activity	-0.21	0.06
Aesthetics	-0.11	0.2
Altruism	-0.19	0.11
Anxiety	-0.18	0.09
Assertiveness	0.1	0.5
Compliance	-0.18	0.11
Depression	0.17	0.56
Ideas	-0.07	0.25
Order	-0.07	0.25
Self-Discipline	-0.17	0.12

Table 8: HPD interval for percentage change in the odds of study delay (given x) from a change in one standard deviation of a skill score

An important thing to note is that if the HPD intervals present in Table 5 do not contain zero, the HPD intervals in Table 8 should not contain it either, for a simple mathematical property. Thus, the four skills that bring us the most evidence of their importance are the same as before: language score, math score, depression score and assertiveness score. In this case, for example, an increase in the math score with magnitude of one standard deviation changes the predicted conditional odds of falling behind in school at a magnitude between -44% and -16% with a probability of 0.95. Regarding the depression score, an increase in the magnitude of one standard deviation of this score changes the predicted conditional odds of falling behind in school at a magnitude between 17% and 56% with probability of 0.95. The same exercise can be done for other skills. In Table 9, we can see some important statistics:

	Mean	1st Qu.	Median	3rd Qu.
Language 2012	-0.24	-0.28	-0.24	-0.19
Mathematics 2012	-0.31	-0.36	-0.31	-0.26
Activity	-0.07	-0.12	-0.08	-0.03
Aesthetics	0.04	-0.01	0.04	0.09
Altruism	-0.05	-0.1	-0.05	0
Anxiety	-0.04	-0.09	-0.05	0
Assertiveness	0.29	0.21	0.28	0.36
Compliance	-0.03	-0.09	-0.04	0.01
Depression	0.36	0.29	0.36	0.43
Ideas	0.09	0.03	0.08	0.14
Order	0.09	0.04	0.09	0.15
Self-Discipline	-0.03	-0.08	-0.04	0.02

Table 9: Main statistics for percentage change in the odds of study delay (given x) from a change in one standard deviation of a skill score

The first thing to note is that the odds percentage change distributions are more or less 'balanced' in the sense that the averages approach the medians and the quartiles are more or less symmetrical with respect to the median. One big message we can get from Table 9 is that, according to the selected statistics, some social-emotional variables can be as important as cognitive skills (math and language) in defining progress delay in school, which is very interesting, given that we traditionally give more importance to the cognitive ones.

7.4 Testing variables effects

In the author's point of view, in addition to having an overview of how factors can determine educational progress delay, it is interesting to determine a rule of choice in order to test which factors really matter. In order to make the choices, let's follow the hypothesis testing framework proposed in Section 6.3.3. As already mentioned, we performed tests for different values of d, which are exposed on the upper horizontal axis of Table 10:

|d|	0,01	0,02	0,03	0,04	0,05
Language 2012	-	-	0	0	0
Mathematics 2012	-	-	-	0	0
Activity	0	0	0	0	0
Aesthetics	0	0	0	0	0
Altruism	0	0	0	0	0
Anxiety	0	0	0	0	0
Assertiveness	+	+	+	0	0
Compliance	0	0	0	0	0
Depression	+	+	+	0	0
Ideas	0	0	0	0	0
Order	+	0	0	0	0
Self-Discipline	0	0	0	0	0

Table 10: Testing variables effects in different scenarios

In Table 10 we can see how our decision - regarding the importance of each skill in determining educational progress delay - varies as one chooses different values of d. In the table above, the symbol "0" means that, by adopting variations in percentage points for the prior probability of educational progress delay with magnitude $|d|$, the variable of interest has no importance in predicting educational progress delay. The symbols "-" and "+" states that the variable of interest is important (negatively and positively) in predicting educational progress delay when adopting variations in percentage points for the prior probability of educational progress delay with magnitude $|d|$. It is interesting to note that in addition to the four variables we have already mentioned, the social-emotional score of "Order" seems to be related to school delay, which is not very intuitive. However, this importance seems to be less robust when compared to the importance of the other four already mentioned variables.

7.5 Marginal effects

Finally, we will present the results regarding the methodology presented in Section 6.3.4. The marginal effects we calculate give us the rate of change in conditional probability of school delay for an infinitesimal variation in the j skill score, which is a quantitative variable. As we have seen, this rate of change depends on the characteristics of the individuals in our sample - in the end of the day, we will have a distribution of marginal effects in our sample. In Table 11 one can check the HPD intervals for the marginal effects of each variable:

	Lower Bound - HPD 95%	Upper Bound - HPD 95%
Portuguese 2012	-0.06	0
Mathematics 2012	-0.08	0
Activity	-0.02	0
Aesthetics	0	0.01
Altruism	-0.01	0
Anxiety	-0.01	0
Assertiveness	0	0.06
Compliance	-0.01	0
Depression	0	0.07
Ideas	0	0.02
Order	0	0.02
Self-Discipline	-0.01	0

Table 11: HPD intervals for marginal effects across samples

It can be seen in Table 11 that the language and math scores variables and the depression and assertiveness scores continue to stand out from the other variables. Regarding the language score, for example, the result tells us that a change in ϵ standard deviations in the score makes us expect a change that can range from 0 to $-0.06 * \epsilon$ percentage points in conditional probability of falling behind in school, with a probability of 0.95. For the result to make sense, ϵ must be a small number, but even for $\epsilon = 1$ we get a reasonable approximation. Regarding the assertiveness score, a variation in ϵ standard deviations in the score makes us expect a change that can range from 0 to $0.06 * \epsilon$ percentage points in the conditional probability of falling behind in school, with 0.95 of probability. In Table 12, we have some important statistics:

	Mean	1st Qu.	Median	3rd Qu.
Language 2012	-0.03	-0.04	-0.02	-0.01
Mathematics 2012	-0.04	-0.05	-0.03	-0.02
Activity	-0.01	-0.01	-0.01	0
Aesthetics	0	0	0	0.01
Altruism	-0.01	-0.01	0	0
Anxiety	0	-0.01	0	0
Assertiveness	0.02	0.01	0.02	0.03
Compliance	0	-0.01	0	0
Depression	0.03	0.01	0.03	0.04
Ideas	0.01	0	0.01	0.01
Order	0.01	0	0.01	0.01
Self-Discipline	0	-0.01	0	0

Table 12: Statistics for marginal effects across samples

The results presented in Table 12 summarize everything presented so far. We emphasize again that an interesting result obtained is that, by comparing the magnitude of the effects, social-emotional skills are as important as cognitive skills in determining educational progress delay.

8 Conclusion and Discussion

The main objective of this study was to analyse the relationship between student skills and the fact of falling behind in school, since traditionally in the literature, only the relationship between the contextual variables (family and school) with school delay is evaluated - It is also important to mention the fact that It is hard to have access of a dataset containing a social-emotional assessment of students. To tackle our objective, we used a longitudinal dataset built in two field researches (2012 and 2017) in the city of Sertãozinho, in the countryside of the state of São Paulo. According to the results achieved, we can highlight the following findings, which summarize the contributions of this paper: (i) not all skills are important in determining academic progress delay, in fact, we had only four (out of a dozen) that seem to be more important, namely language and math skills (negative impact on delay) and depression and assertiveness (positive impact on delay); (ii) in terms of magnitude of impact, we can say that social-emotional skills are as important as cognitive skills in determining academic progress delay.

An additional point that was not covered during the development of this paper is that the results found potentially have a causal interpretation, although I believe that a more detailed study has to be done later. I say this because the context variables are exogenous, we got a great diversity of control variables and, in my view, the only obstacle would be the fact that social-emotional skills have an impact on the probability of falling behind in school through academic performance and did not model that relationship directly. In fact, the hypothesis that social-emotional skills cause cognitive skills is a relevant and supported hypothesis in the literature [20, 21] - the opposite way is still more questionable. During the experimental phase of this work, I estimated a model without math and language grades, so we would no longer have undesirable control blocking the a path between social-emotional skills and school progress delay. The results were almost identical (posterior distributions) for the social-emotional variables. Because of this, I once again highlight the potential for causal interpretation brought by the results, even if this subject should be the main topic of another paper.

To conclude, I think this work may have an influence on the debate of educational public policies in Brazil and in the world since it deals with current and important issues besides being able to change the way we look at education and what kind of skills we would like to develop in young people.

9 Acknowledgments

I would like to thank CAPES (Higher Education Personnel Improvement Coordination), which is financing my master's degree in statistics at IME/USP and the LEPES/USP (Laboratório de Estudos e Pesquisas em Economia Social) for the dataset used.

References

[1] World Economic Forum. *New vision for education: Unlocking the potential of technology*. British Columbia Teachers' Federation, 2015.

[2] James J Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482, 2006.

[3] Tim Kautz, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, National Bureau of Economic Research, 2014.

[4] Koji Miyamoto, M Huerta, Katarzyna Kubacka, Hiroko Ikesako, and E Oliveira. Skills for social progress: the power of social and emotional skills, 2015.

[5] Fernanda Leite Lopez de Leon and Naércio Aquino Menezes-Filho. Reprovação, avanço e evasão escolar no brasil. 2002.

[6] Miguel G ARROYO. Educação e exclusão da cidadania: In buffa. *E.: ARROYO, M*, 1995.

[7] Zvi Eckstein and Kenneth I Wolpin. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339, 1999.

[8] Tufi Machado Soares, Neimar da Silva Fernandes, Mariana Calife Nóbrega, and Alexandre Chibebe Nicolella. Fatores associados ao abandono escolar no ensino médio público de minas gerais. *Educação e Pesquisa*, 41(3):757–772, 2015.

[9] Vitor Augusto Carlos. O papel da repetência escolar sobre variáveis de fluxo: uma análise sobre o abandono escolar e chegada ao ensino médio. Master's thesis, Universidade de São Paulo, 2019.

[10] Oliver P John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

[11] Christopher J Soto and Oliver P John. Ten facet scales for the big five inventory: Convergence with neo pi-r facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, 43(1):84–90, 2009.

[12] Ralph L Piedmont. *The revised NEO Personality Inventory: Clinical and research applications*. Springer Science & Business Media, 2013.

[13] John D Winkler, David E Kanouse, and John E Ware. Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67(5):555, 1982.

[14] Felipe Valentini. Influência e controle da aquiescência na análise fatorial. *Avaliação Psicológica*, 16(2), 2017.

[15] João Pedro S. Lavinas, Marcos P.C. Costa, and Daniel D. Santos. Ambiente familiar e proficiencia escolar: o papel das habilidades socioemocionais. *Anais do SJPED, Volume 5, Número 5, Santa Maria*, 2018.

[16] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

[17] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

[18] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[19] Luis Gustavo Esteves, Rafael Izbicki, Julio Michael Stern, and Rafael Bassi Stern. Pragmatic hypotheses in the evolution of science. *Entropy*, 21(9):883, 2019.

[20] Flavio Cunha and James J Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008.

[21] Flavio Cunha, James J Heckman, and Susanne M Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

[22] Michael Betancourt. Diagnosing suboptimal cotangent disintegrations in hamiltonian monte carlo. *arXiv preprint arXiv:1604.00695*, 2016.

[23] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019.

[24] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc. *arXiv preprint arXiv:1903.08008*, 2019.

10 Appendix

10.1 Attrition

	Attrition	Non-Attrition	Diff.
Unknown	0.06	0.01	0.05
Non-educated	0.14	0.15	-0.01
Elementary	0.34	0.29	0.05
Middle	0.19	0.22	-0.03
High	0.19	0.26	-0.07
College	0.08	0.07	0.01

Table 13: Comparing the two groups

	Attrition	Non-Attrition	Diff.
State	0.31	0.27	0.04
Municipal	0.62	0.63	-0.01
Private	0.08	0.1	-0.02

Table 14: Comparing the two groups

	Attrition	Non-Attrition	Diff.
White	0.36	0.41	-0.04
Man	0.55	0.48	0.07
Failed before 2012	0.47	0.24	0.23

Table 15: Comparing the two groups

In Table 16, we standardize the scores (mean = 0, std = 1) before selecting only those we could find in 2017:

	Attrition	Non-Attrition	Diff.
Language 2012	-0.27	0.14	-0.41
Mathematics 2012	-0.23	0.12	-0.35
Activity	-0.08	0.04	-0.12
Aesthetics	0.1	-0.05	0.15
Altruism	0.04	-0.02	0.06
Anxiety	0	0	0
Assertiveness	0.05	-0.03	0.08
Compliance	-0.04	0.02	-0.07
Depression	0.07	-0.04	0.11
Ideas	-0.04	0.02	-0.05
Order	0.07	-0.04	0.11
Self-Discipline	0.03	-0.02	0.05

Table 16: Comparing the two groups

10.2 Descriptive statistics

	Proportion
Unknown	0.01
Non-educated	0.15
Elementary	0.29
Middle	0.22
High	0.26
College	0.07

Table 17: Descriptive statistics in actual sample

	Proportion
State	0.27
Municipal	0.63
Private	0.1

Table 18: Descriptive statistics in actual sample

	Proportion
White	0.41
Man	0.48
Failed before 2012	0.24
Pre-K	0.48
Kinder	0.86

Table 19: Descriptive statistics in actual sample

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Portuguese 2012	-4.04	-0.48	0.11	0	0.72	1.82
Mathematics 2012	-3.34	-0.64	0.03	0	0.71	2.08
Activity	-3.5	-0.7	0.04	0	0.76	2.62
Aesthetics	-3.45	-0.64	-0.05	0	0.66	3.33
Altruism	-3.02	-0.7	-0.05	0	0.6	3.97
Anxiety	-3.87	-0.42	0.01	0	0.73	3.27
Assertiveness	-3.21	-0.62	-0.05	0	0.65	3.39
Compliance	-2.68	-0.7	-0.02	0	0.71	2.8
Depression	-2.45	-0.74	-0.16	0	0.57	5.5
Ideas	-4.09	-0.62	0	0	0.64	3.31
Order	-2.51	-0.75	-0.15	0	0.64	4.37
Self-Discipline	-3.13	-0.63	0	0	0.65	3.14

Table 20: Descriptive statistics in actual sample

10.3 Posterior distribution analysis

	Lower Bound - HPD 95%	Upper Bound - HPD 95%
Intercept	-11.02	-4.99
Grade 2012	0.62	1.4
Birth Year	0	0
Birth Semester	-0.43	0.17
School 2012: Municipal	0.49	1.38
School 2012: Private	0.74	1.94
Ethinicity: White	-0.5	0.13
Gender: Male	0.16	0.79
Failed Before 2012	0.13	0.79
Mother Educ.: Unknown	-4.67	0.33
Mother Educ.: Elementary	0.01	0.93
Mother Educ.: Middle	-0.12	0.87
Mother Educ.: High	-0.53	0.52
Mother Educ.: College	-1.75	0.03
Childhood Educ: Pre-k	-0.24	0.36
Childhood Educ: Kinder	-0.25	0.61
Language 2012	-0.44	-0.08
Mathematics 2012	-0.58	-0.17
Activity	-0.22	0.07
Aesthetics	-0.11	0.19
Altruism	-0.21	0.1
Anxiety	-0.19	0.09
Assertiveness	0.1	0.41
Compliance	-0.19	0.11
Depression	0.16	0.45
Ideas	-0.06	0.23
Order	-0.06	0.23
Self-Discipline	-0.19	0.11

Table 21: HPD interval for marginal posterior distributions

	Mean	1st Qu.	Median	3rd Qu.
Intercept	-8.05	-9.09	-8.03	-7.02
Grade 2012	1.01	0.88	1.01	1.15
Birth Year	0	0	0	0
Birth Semester	-0.14	-0.24	-0.14	-0.03
School 2012: Municipal	0.94	0.78	0.94	1.09
School 2012: Private	1.33	1.13	1.34	1.54
Ethinicity: White	-0.19	-0.3	-0.19	-0.08
Gender: Male	0.47	0.35	0.47	0.58
Failed Before 2012	0.46	0.34	0.46	0.57
Mother Educ.: Unknown	-1.97	-2.7	-1.82	-1.04
Mother Educ.: Elementary	0.46	0.29	0.46	0.62
Mother Educ.: Middle	0.4	0.23	0.4	0.57
Mother Educ.: High	-0.01	-0.19	-0.01	0.18
Mother Educ.: College	-0.86	-1.16	-0.84	-0.54
Childhood Educ: Pre-k	0.07	-0.04	0.06	0.17
Childhood Educ: Kinder	0.17	0.02	0.16	0.31
Language 2012	-0.27	-0.33	-0.27	-0.21
Mathematics 2012	-0.38	-0.45	-0.37	-0.31
Activity	-0.08	-0.13	-0.08	-0.03
Aesthetics	0.04	-0.01	0.04	0.09
Altruism	-0.05	-0.11	-0.05	0
Anxiety	-0.05	-0.09	-0.05	0
Assertiveness	0.25	0.19	0.25	0.3
Compliance	-0.04	-0.09	-0.04	0.01
Depression	0.31	0.26	0.31	0.36
Ideas	0.08	0.03	0.08	0.13
Order	0.09	0.04	0.09	0.14
Self-Discipline	-0.04	-0.09	-0.04	0.02

Table 22: Basic statistics for marginal posterior distributions

10.4 Testing variables impacts

	0,01	0,02	0,03	0,04	0,05
Intercept	-	-	-	-	-
Grade 2012	+	+	+	+	+
Birth Year	0	0	0	0	0
Birth Semester	-	0	0	0	0
School 2012: Municipal	+	+	+	+	+
School 2012: Private	+	+	+	+	+
Ethinicity: White	-	-	0	0	0
Gender: Male	+	+	+	+	+
Failed Before 2012	+	+	+	+	+
Mother Educ.: Unknown	-	-	-	-	-
Mother Educ.: Elementary	+	+	+	+	+
Mother Educ.: Middle	+	+	+	+	+
Mother Educ.: High	-	0	0	0	0
Mother Educ.: College	-	-	-	-	-
Childhood Educ: Pre-k	+	0	0	0	0
Childhood Educ: Kinder	+	+	0	0	0
Language 2012	-	-	0	0	0
Mathematics 2012	-	-	-	0	0
Activity	0	0	0	0	0
Aesthetics	0	0	0	0	0
Altruism	0	0	0	0	0
Anxiety	0	0	0	0	0
Assertiveness	+	+	+	0	0
Compliance	0	0	0	0	0
Depression	+	+	+	0	0
Ideas	0	0	0	0	0
Order	+	0	0	0	0
Self-Discipline	0	0	0	0	0

Table 23: Testing variables impacts

10.5 Convergence diagnosis of MCMC

10.5.1 Divergent transitions No U-Turn Sampler

In addition to the general methods for diagnosing convergence of MCMC algorithms, we have some specific methods for the "No U-Turn Sampler" algorithm. The Stan package automatically offers ways to infer about method convergence. The most straightforward way is to check graphically if the package reports "Divergent transitions" [17, 22, 23]. According to the package documentation itself [10]: "A divergence arises when the simulated Hamiltonian trajectory departs from the true trajectory as measured by departure of the Hamiltonian value from its initial value. When this divergence is too high, the simulation has gone off the rails and cannot be trusted". In the figure below we have that every tick represents a parameter on the horizontal axis and each line represents an isolated sample. One can see that the package did not identify divergent transitions compared to the examples in [23]:

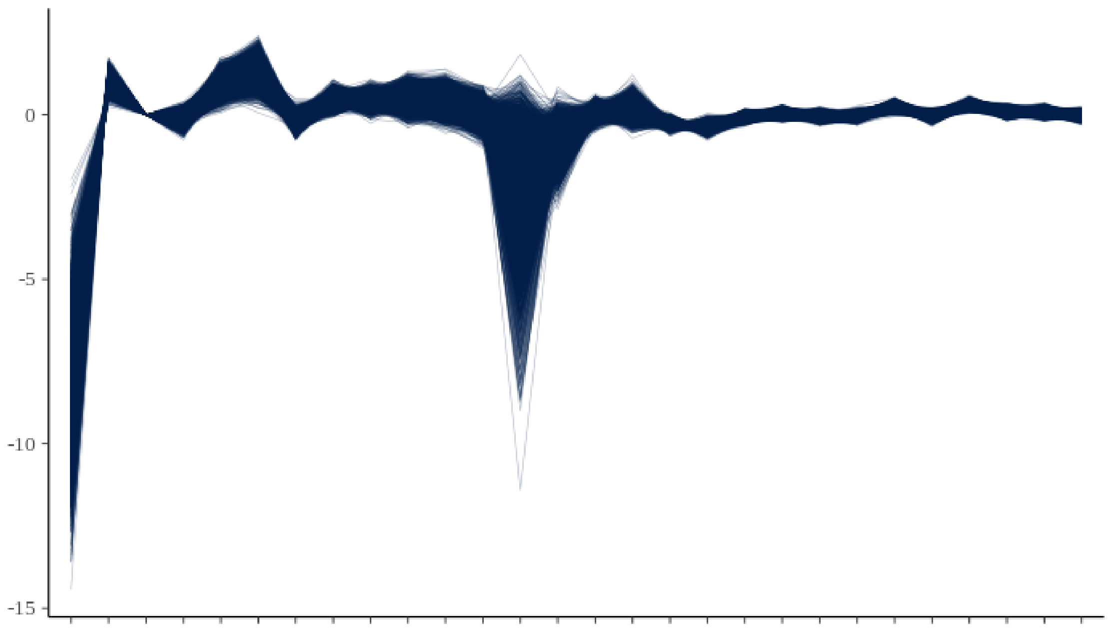

Figure 4: Divergent transitions No U-Turn Sampler

[10] https://mc-stan.org/docs/2_19/reference-manual/divergent-transitions.html#fn24

10.5.2 $\hat{R}$ and effective sample size

$\hat{R}$ is a classic diagnosis statistic for multiple chains of MCMC algorithms. A new version is presented in [24] and corrects some points made by the authors. If $\hat{R}$ is too close to 1, we have evidence that the chains have converged to the same distribution. In addition to $\hat{R}$, in the table below we can check the effective sample size, which is very close to 100% for all variables, indicating that autocorrelations within chains are low. Then, we have two more evidences that we sampled from the distributions of interest:

	Rhat	Bulk_ESS	Tail_ESS
Intercept	1	4994	4928
Grade 2012	1	4852	4839
Birth Year	1	5111	5008
Birth Semester	1	4639	4872
School 2012: Municipal	1	5190	5031
School 2012: Private	1	4823	4648
Ethinicity: White	1	5084	4911
Gender: Male	1	4856	4780
Failed Before 2012	1	5142	4724
Mother Educ.: Unknown	1	5094	4391
Mother Educ.: Elementary	1	4849	4910
Mother Educ.: Middle	1	4836	4907
Mother Educ.: High	1	4796	4400
Mother Educ.: College	1	5050	4870
Childhood Educ: Pre-k	1	5016	5119
Childhood Educ: Kinder	1	5125	4758
Language 2012	1	4604	4791
Mathematics 2012	1	4971	4755
Activity	1	4975	5016
Aesthetics	1	4786	4763
Altruism	1	5082	4951
Anxiety	1	5023	4863
Assertiveness	1	5248	4952
Compliance	1	4707	4820
Depression	1	4910	5030
Ideas	1	4704	4646
Order	1	4603	4952
Self-Discipline	1	5080	4792

Table 24: $\hat{R}$ and effective sample size

10.5.3 Comparing chains samplings graphically: rank plots

The idea here is to compare the samples from the 4 chains graphically. The work [24] recommends that we abandon traditional Trace Plots and move on to the following Rank Plots. The idea here is: (i) we pool the samples of all chains into a single vector creating a rank (by the sampled values), (ii) compare the chains individually with the initial rank, creating a frequency histogram to evaluate whether the chains were sampled from the same distribution. If we observe uniform rank distributions across all chains, we can conclude that they were able to sample from the same distribution. In the graphs below, at least visually, we have evidence that the chains sample from the same distribution and we are more confident about the convergence of the method:

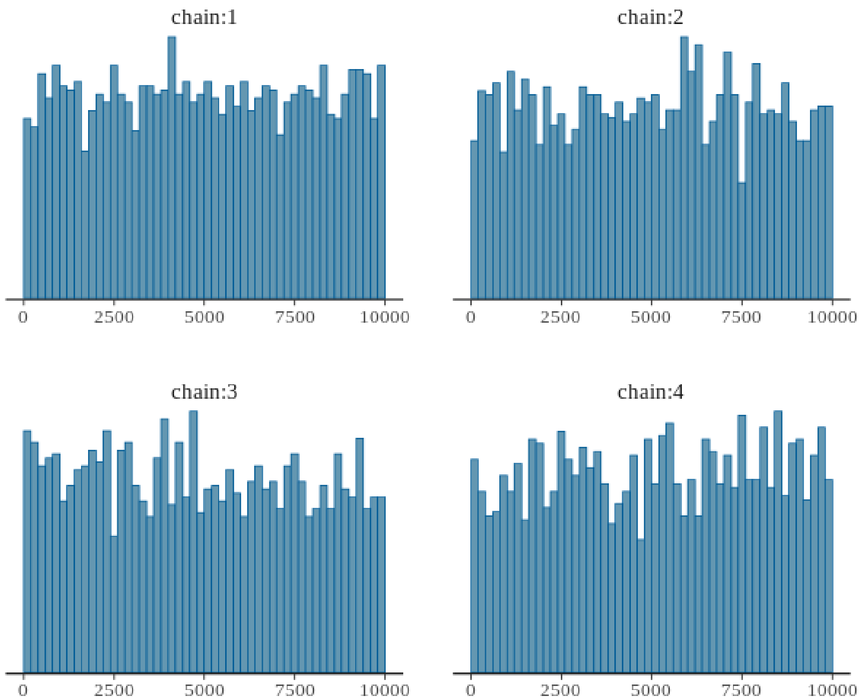

Figure 5: Intercept

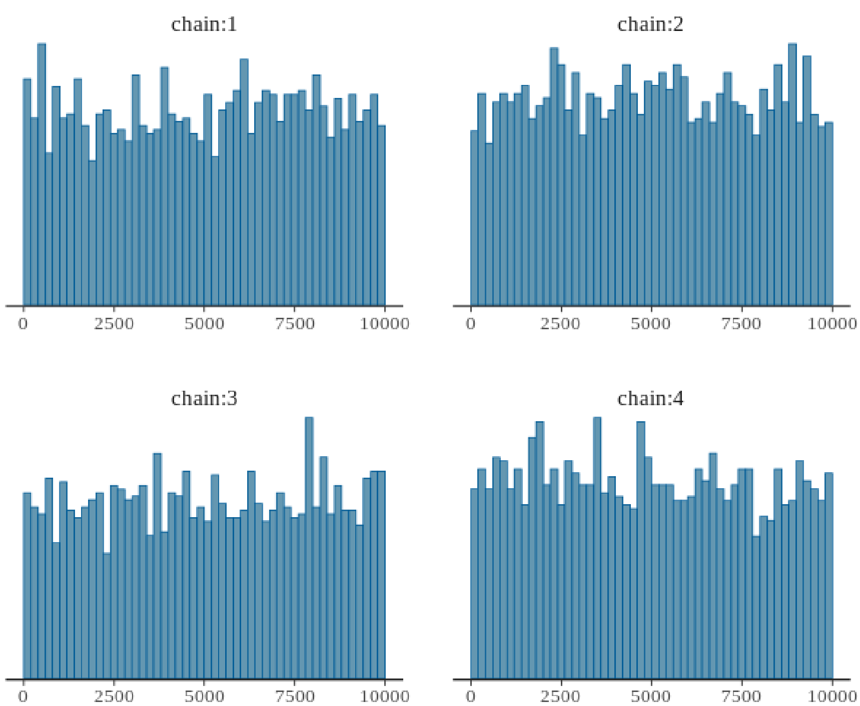

Figure 6: grade 2012

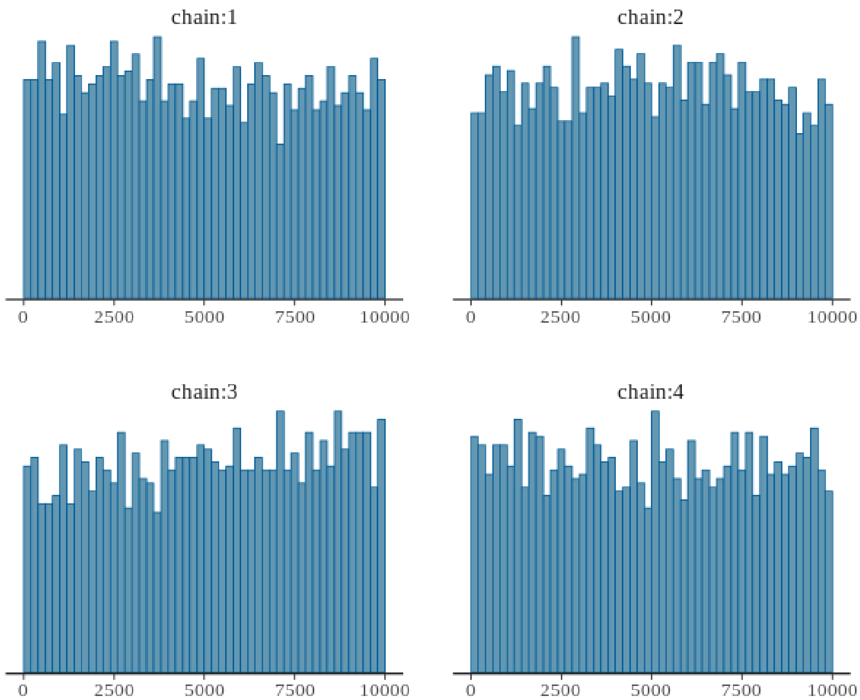

Figure 7: year

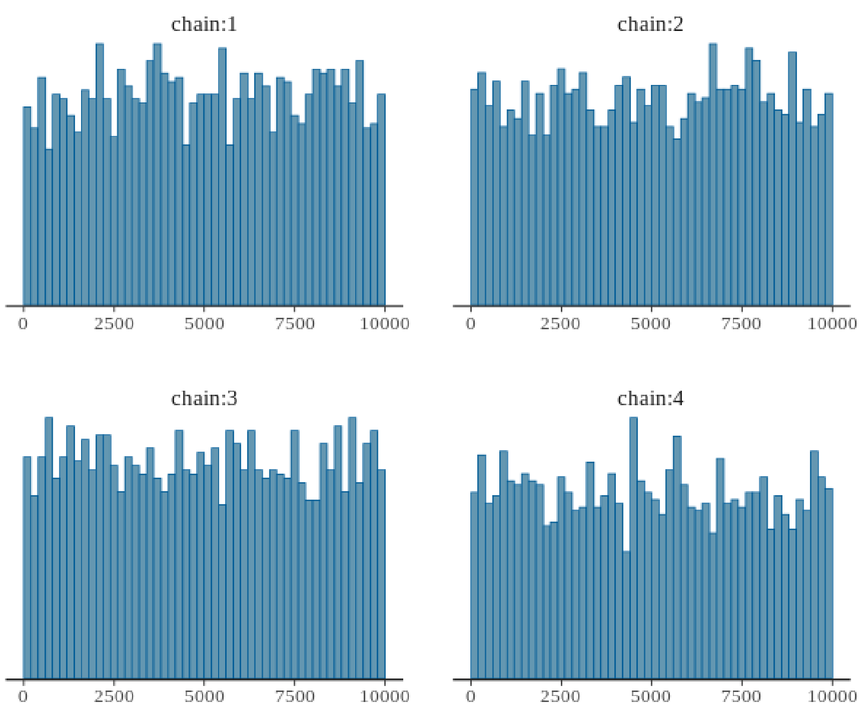

Figure 8: semester

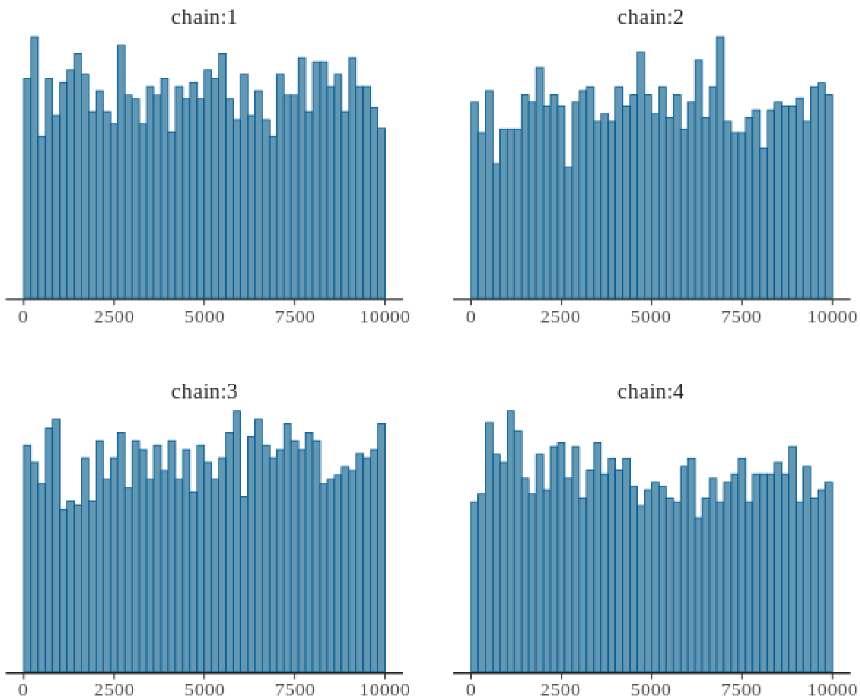

Figure 9: school 2012 - municipal

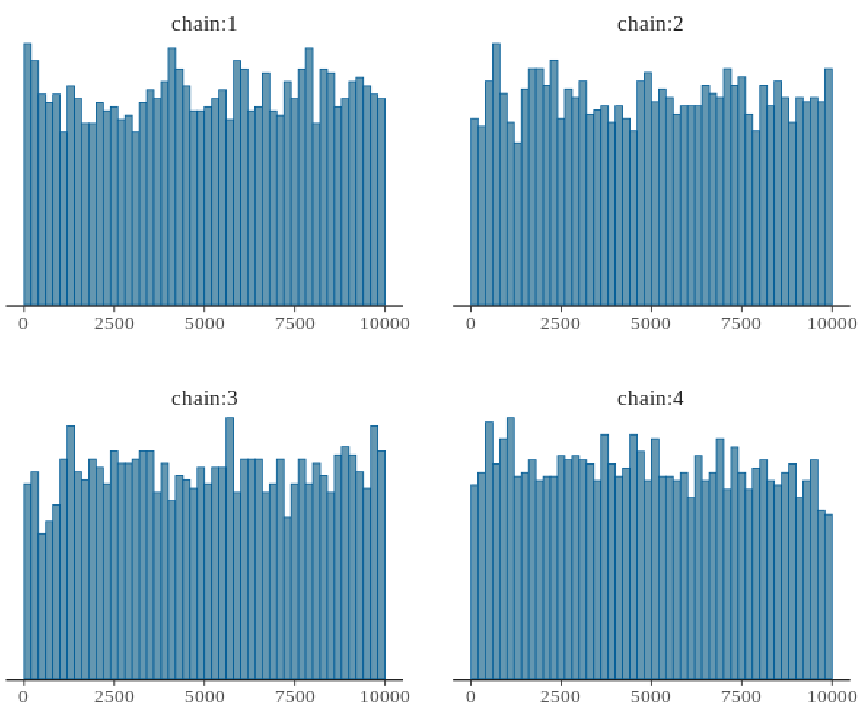

Figure 10: school 2012 - privada

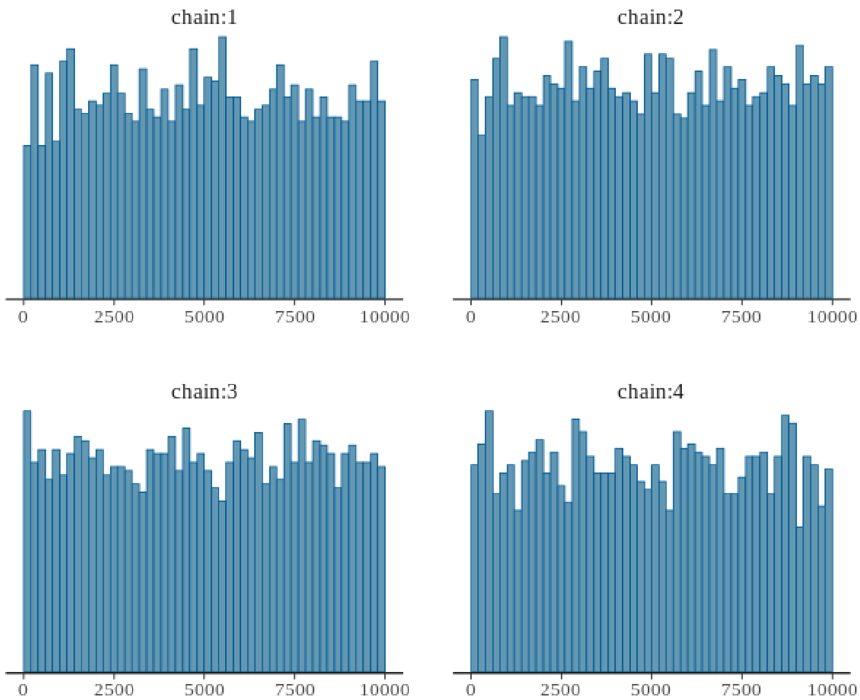

Figure 11: white

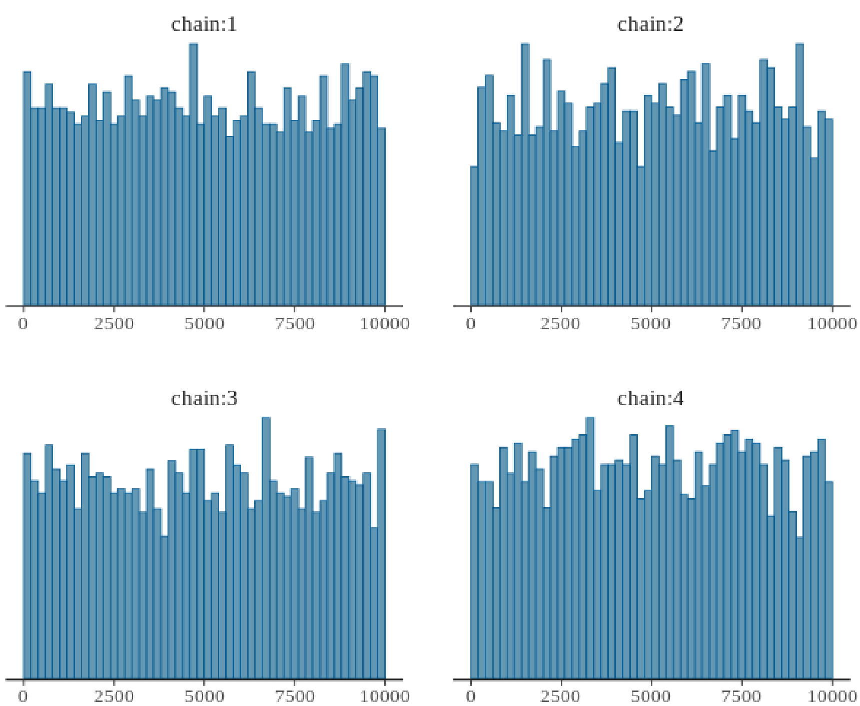

Figure 12: male

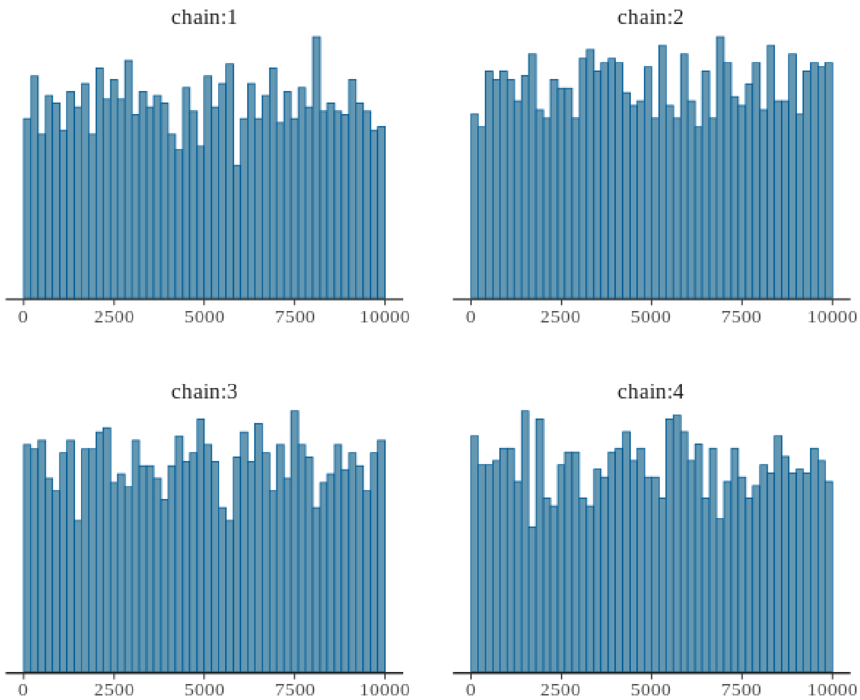

Figure 13: failed before 2012

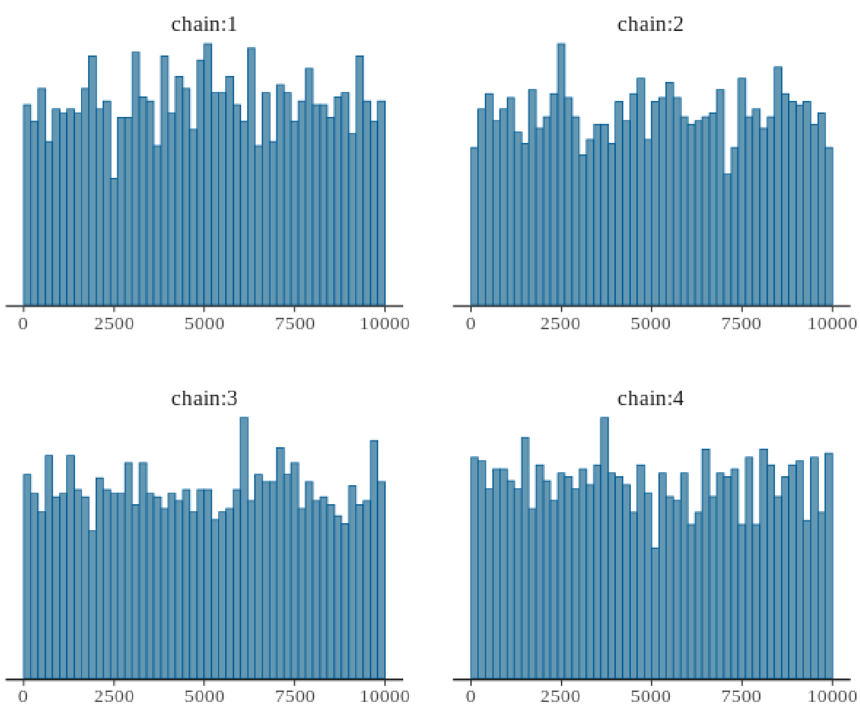

Figure 14: mother educ - ef1

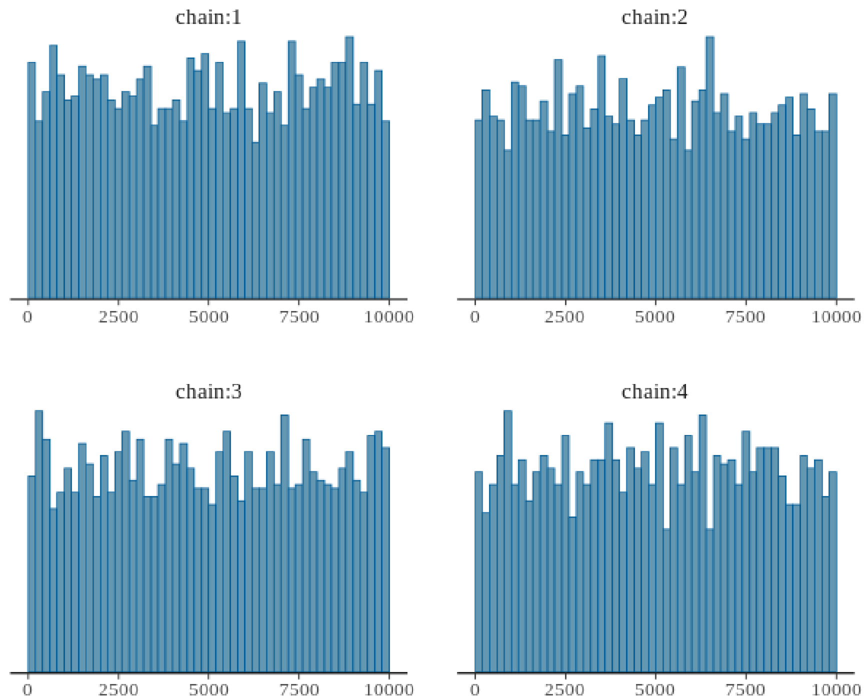

Figure 15: mother educ - ef2

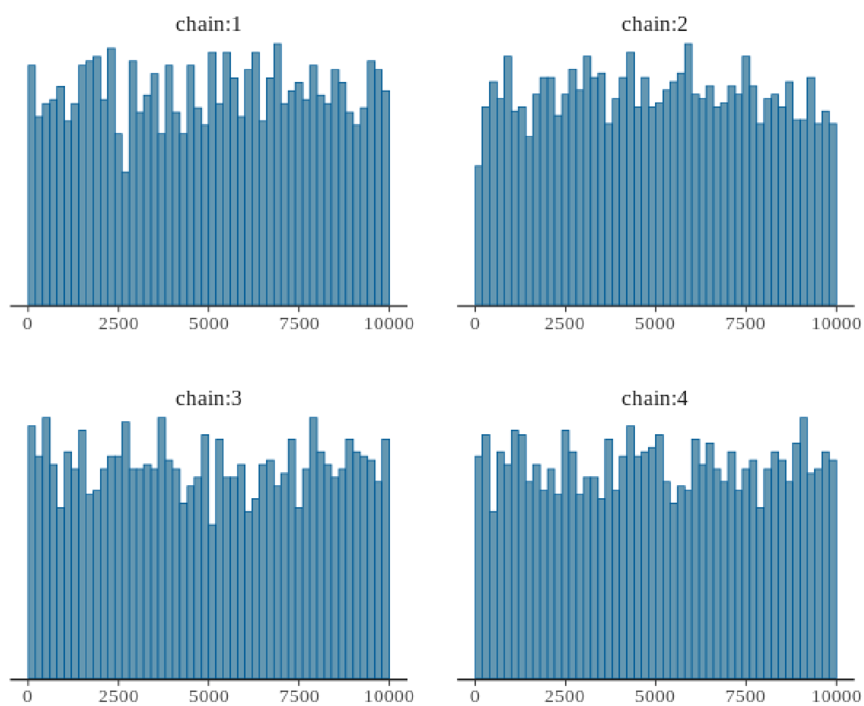

Figure 16: mother educ - em

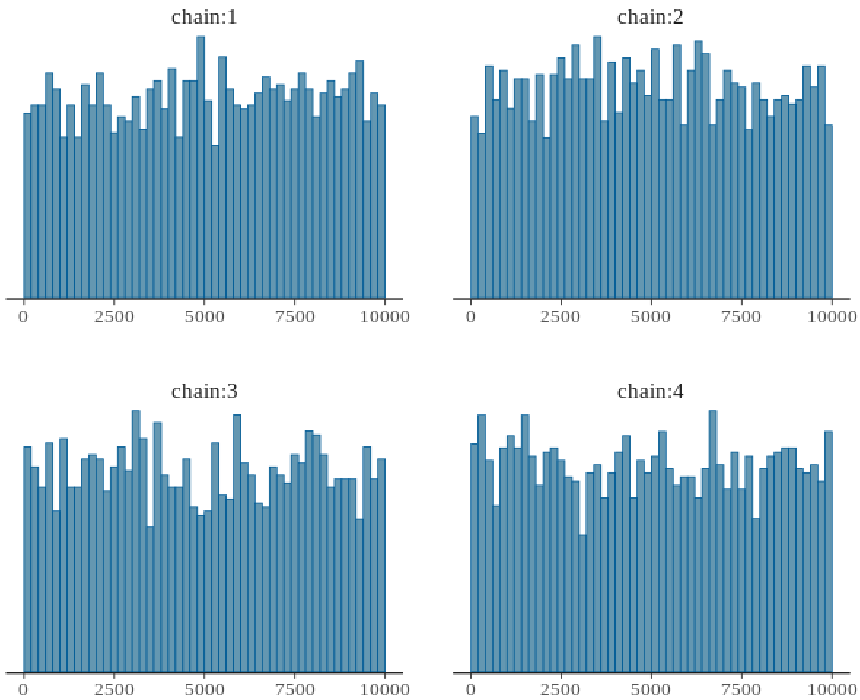

Figure 17: mother educ - nao sabe

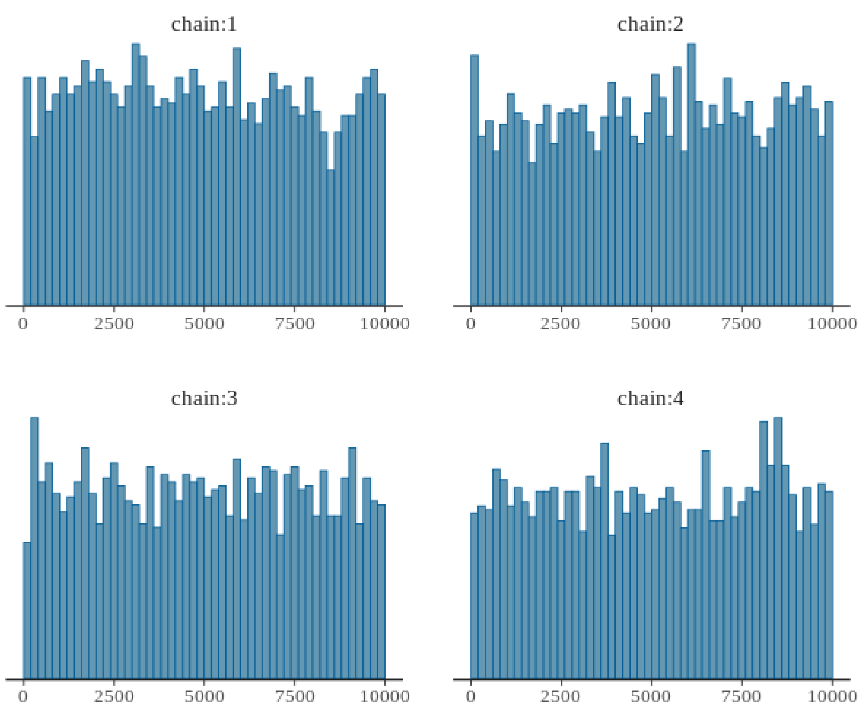

Figure 18: mother educ - superior

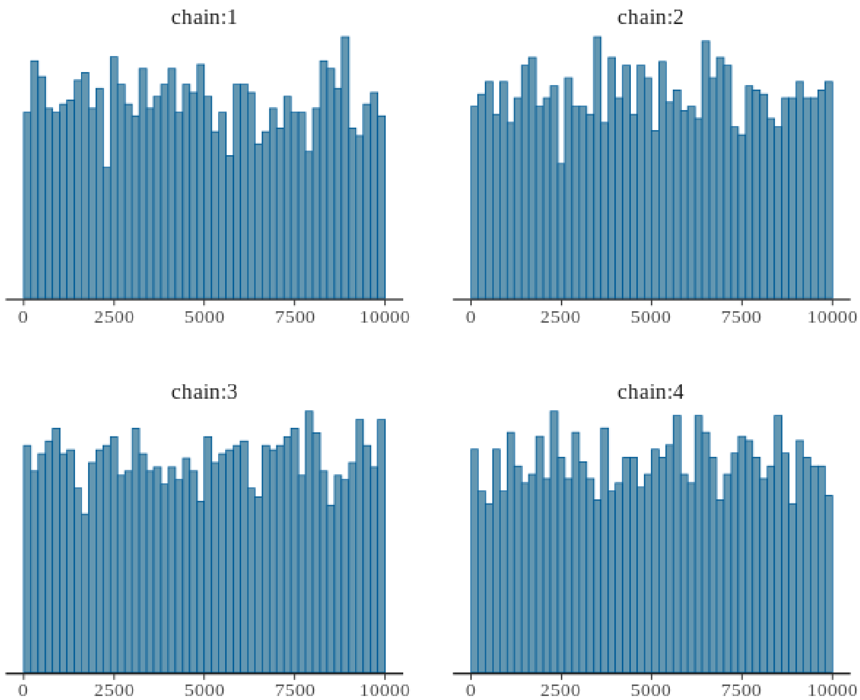

Figure 19: pre k

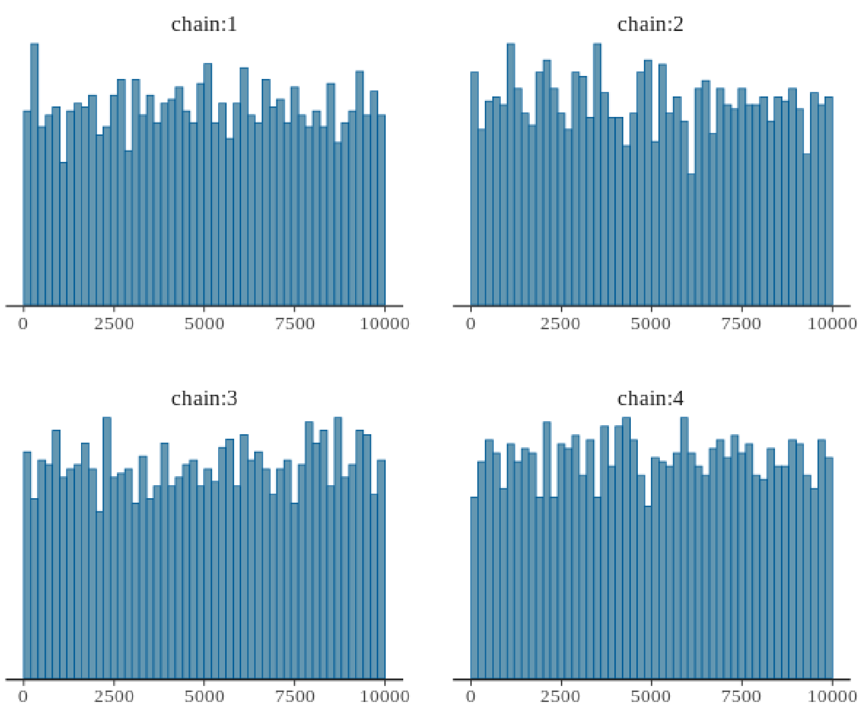

Figure 20: kinder

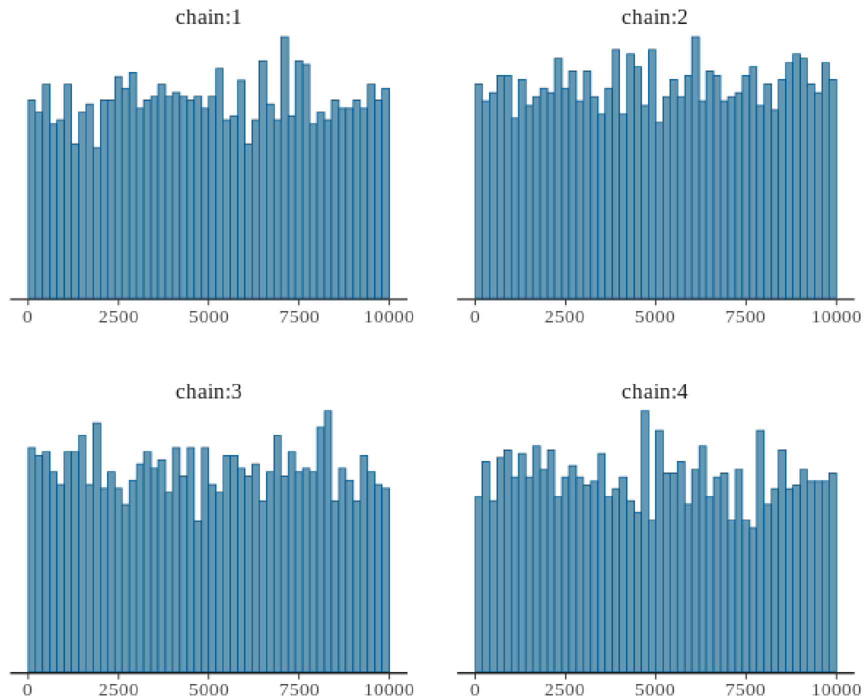

Figure 21: lang 2012

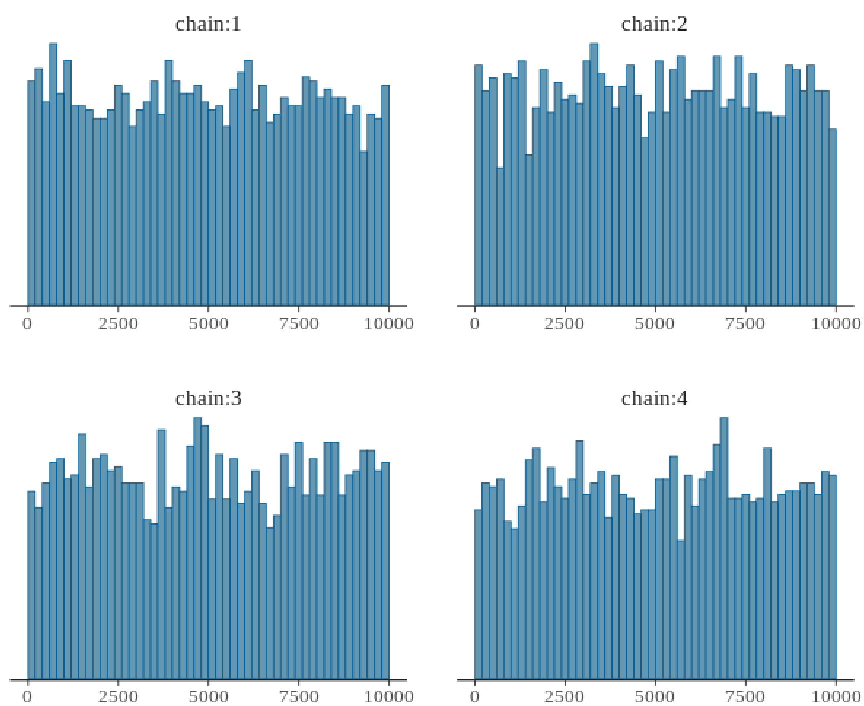

Figure 22: math 2012

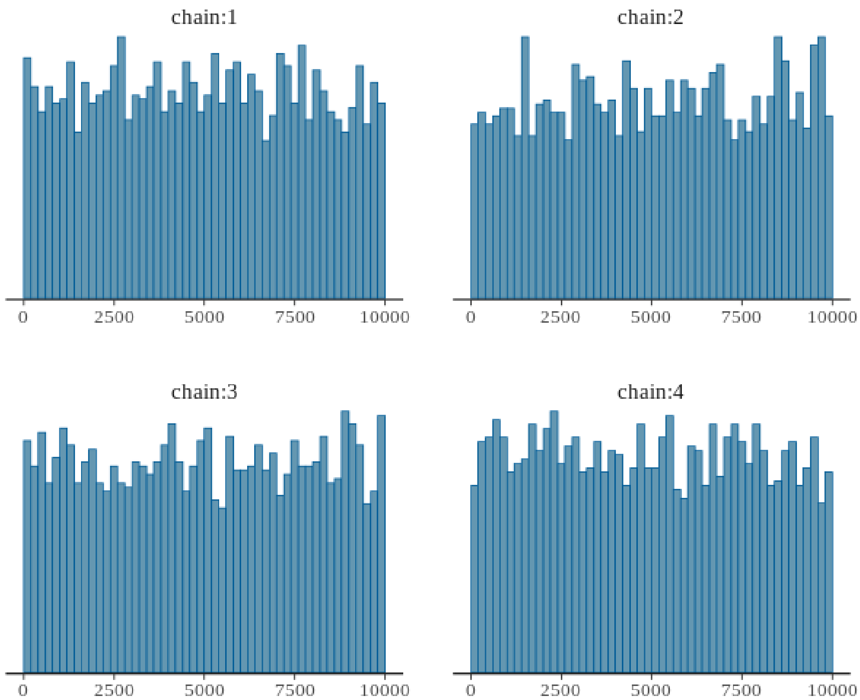

Figure 23: act 2012

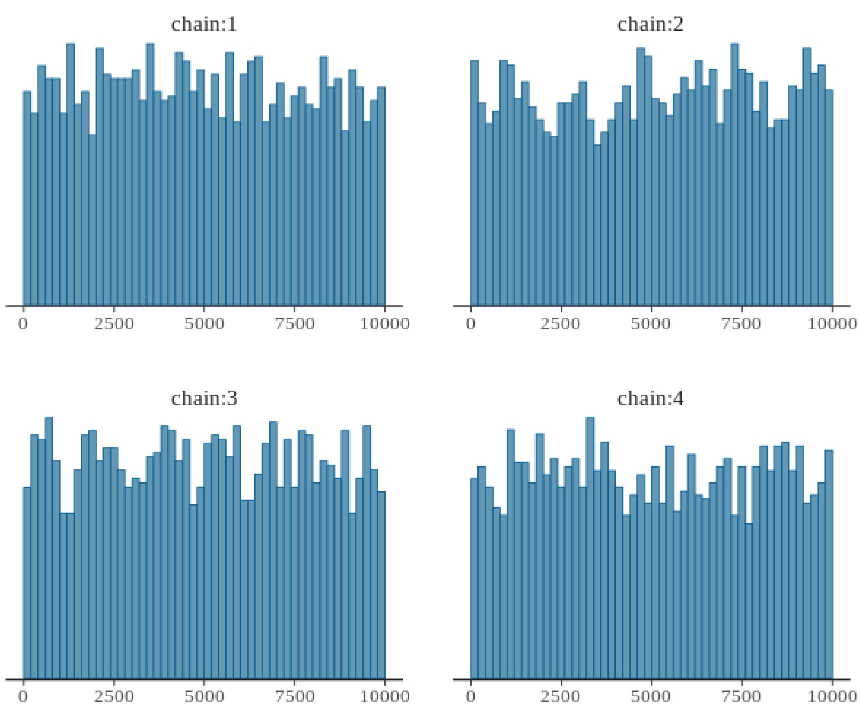

Figure 24: aes 2012

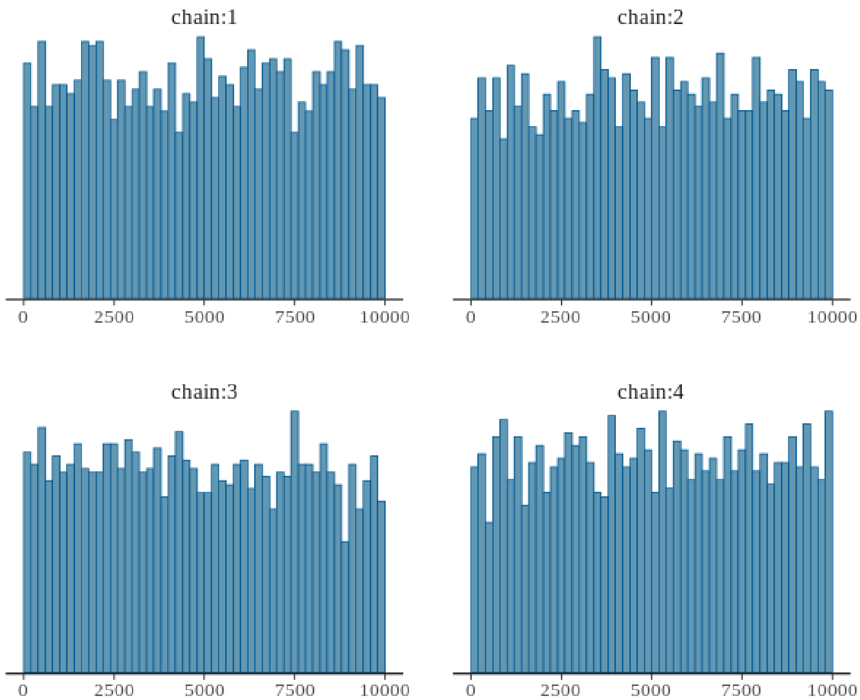

Figure 25: alt 2012

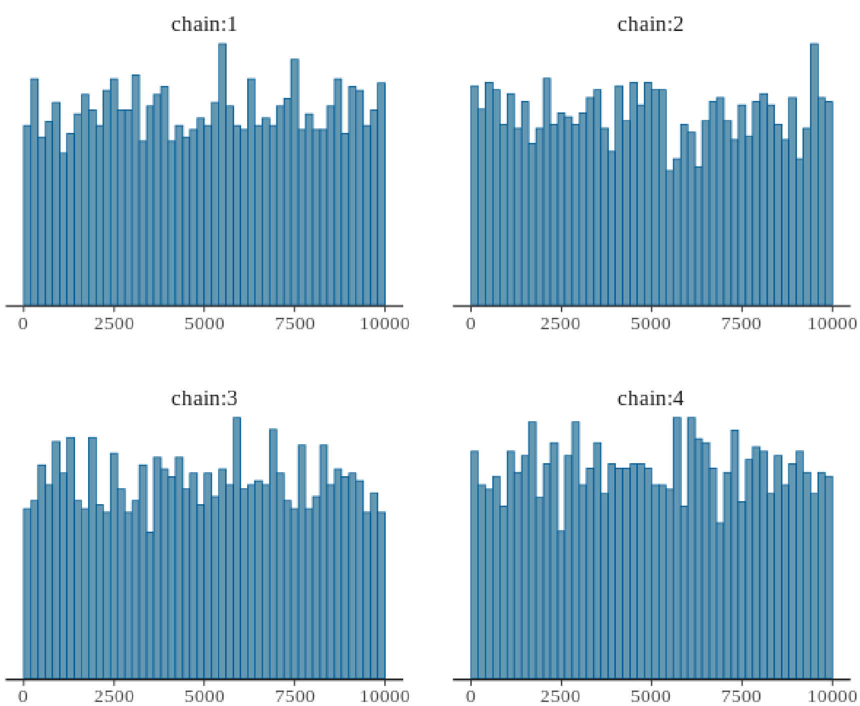

Figure 26: anx 2012

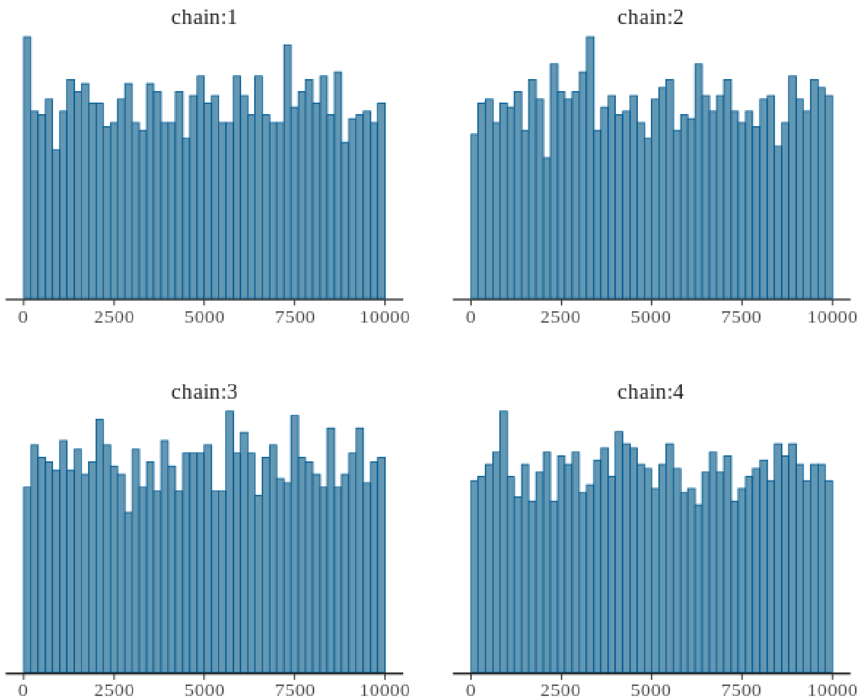

Figure 27: ass 2012

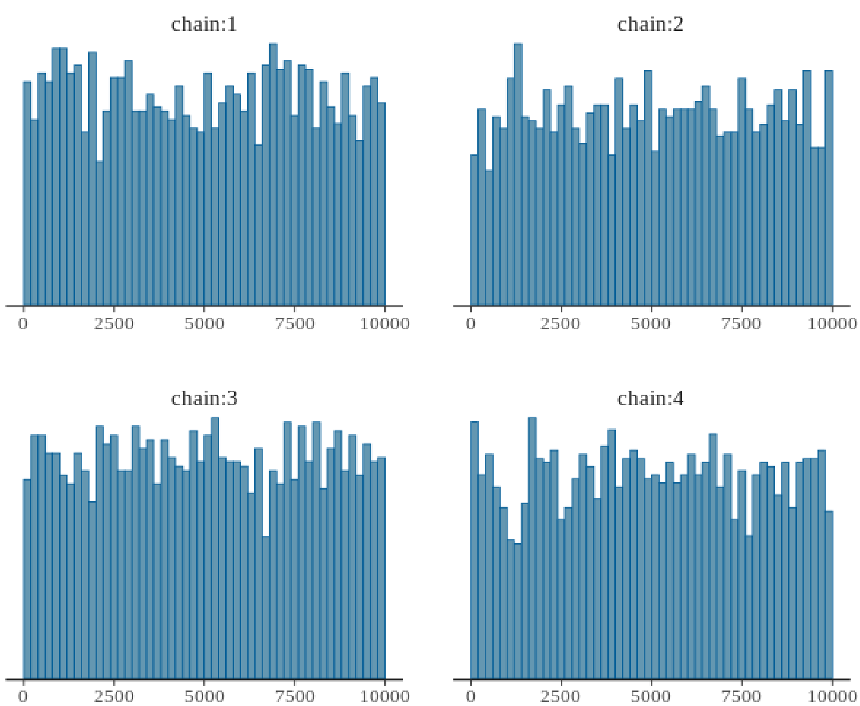

Figure 28: cmp 2012

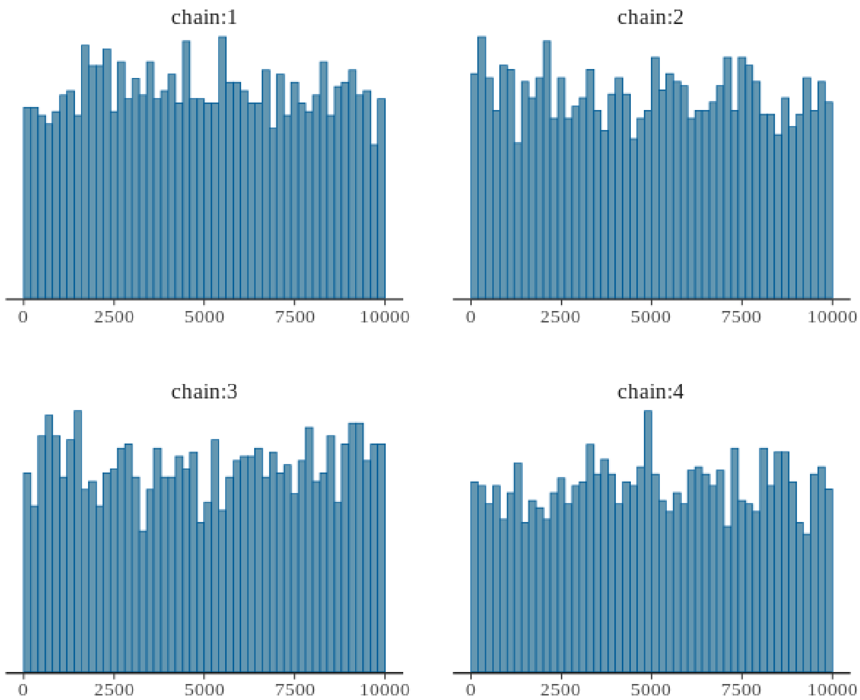

Figure 29: dep 2012

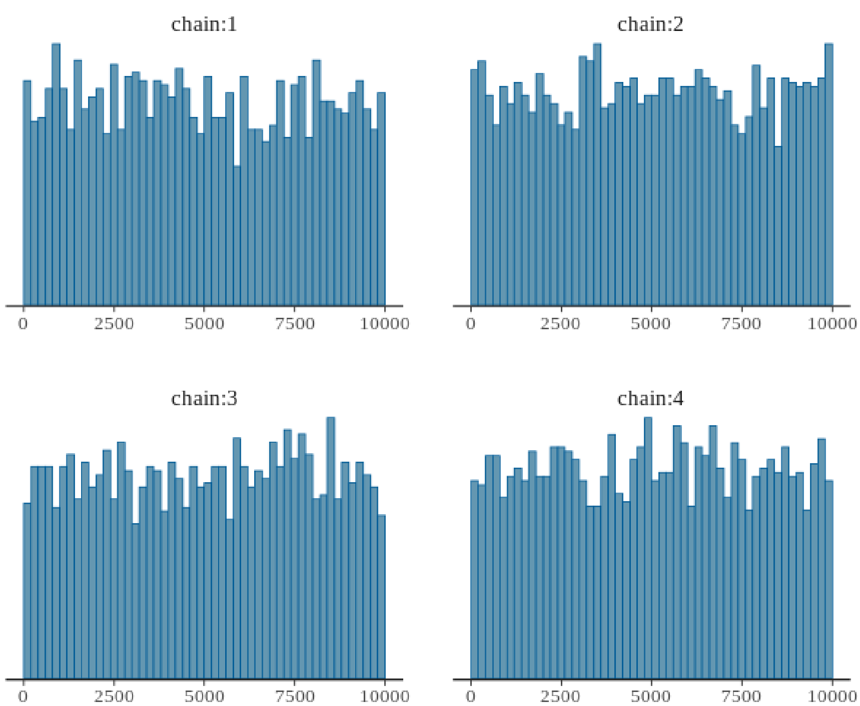

Figure 30: ids 2012

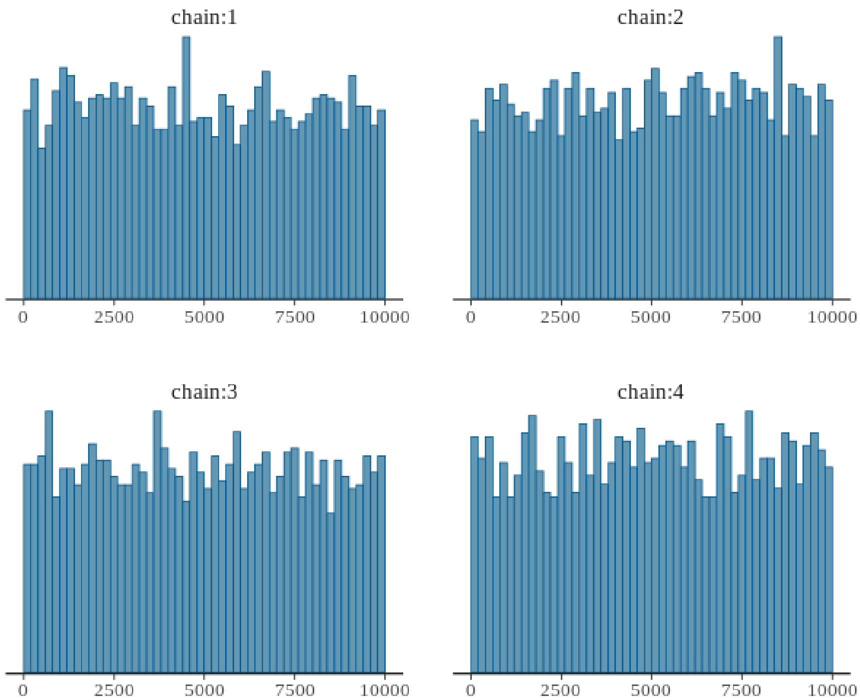

Figure 31: ord 2012

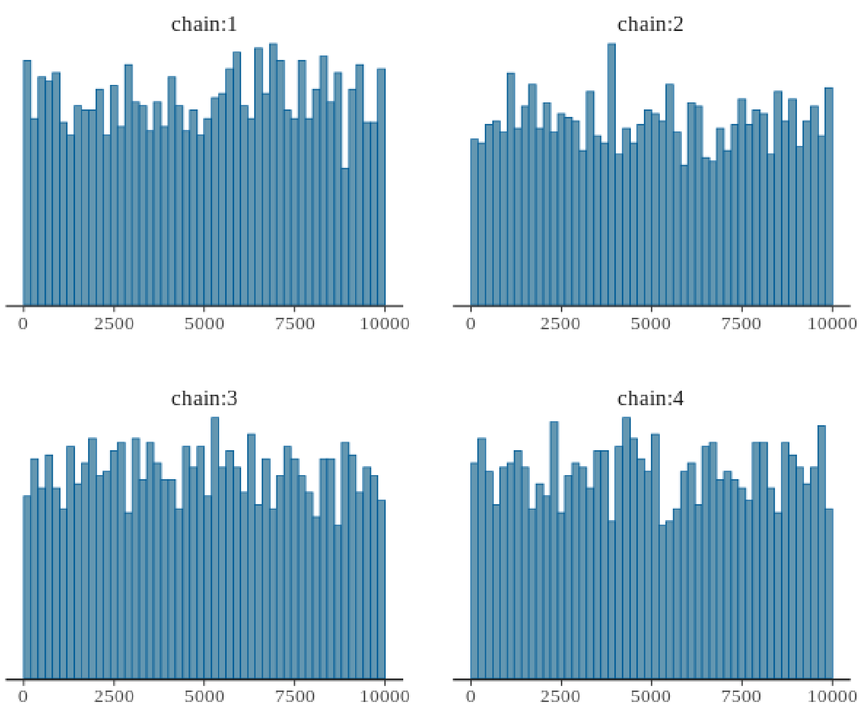

Figure 32: sfd 2012

10.5.4 Autocorrelation

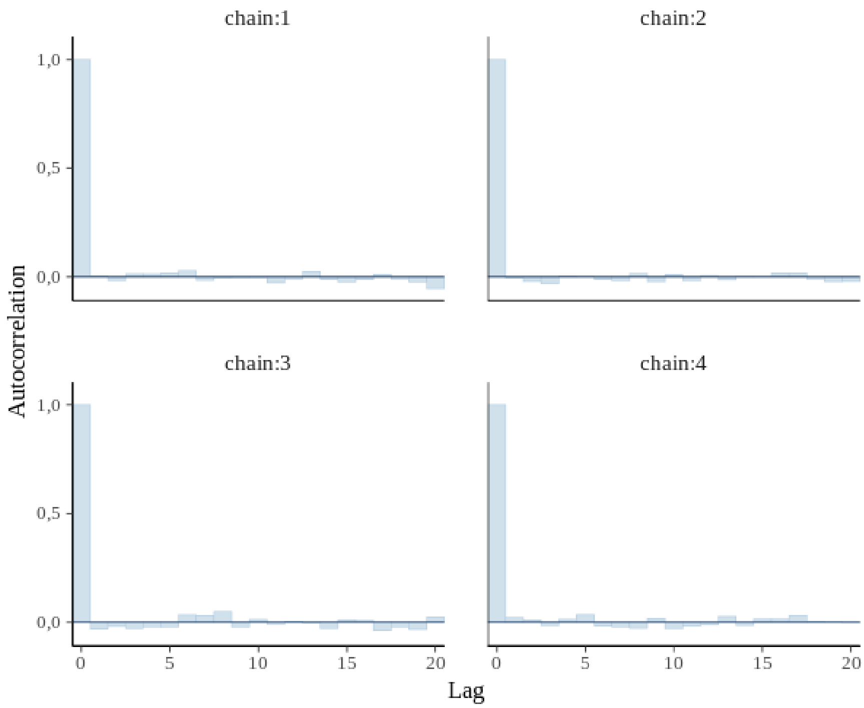

Figure 33: Intercept

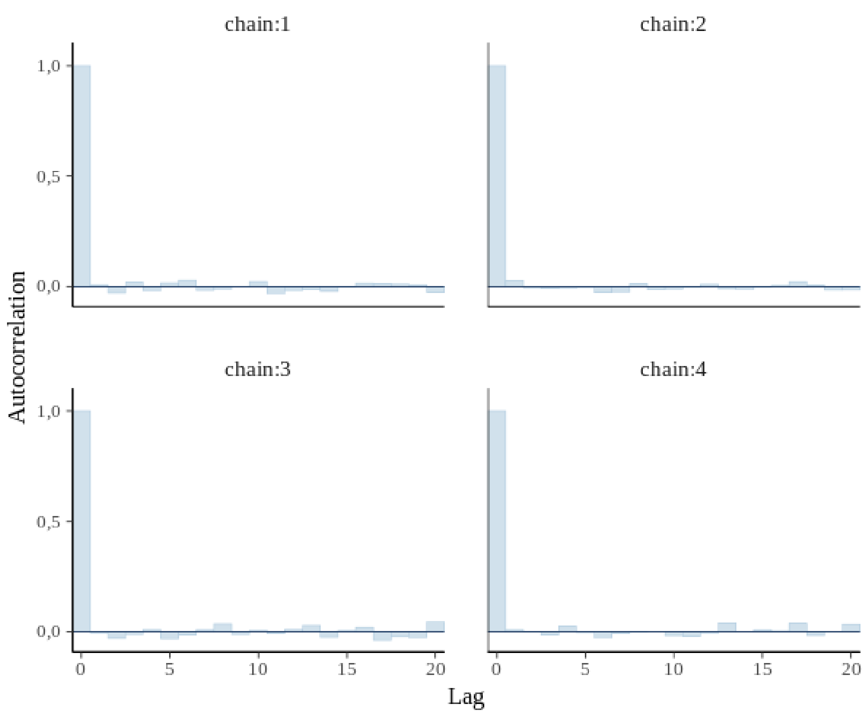

Figure 34: grade 2012

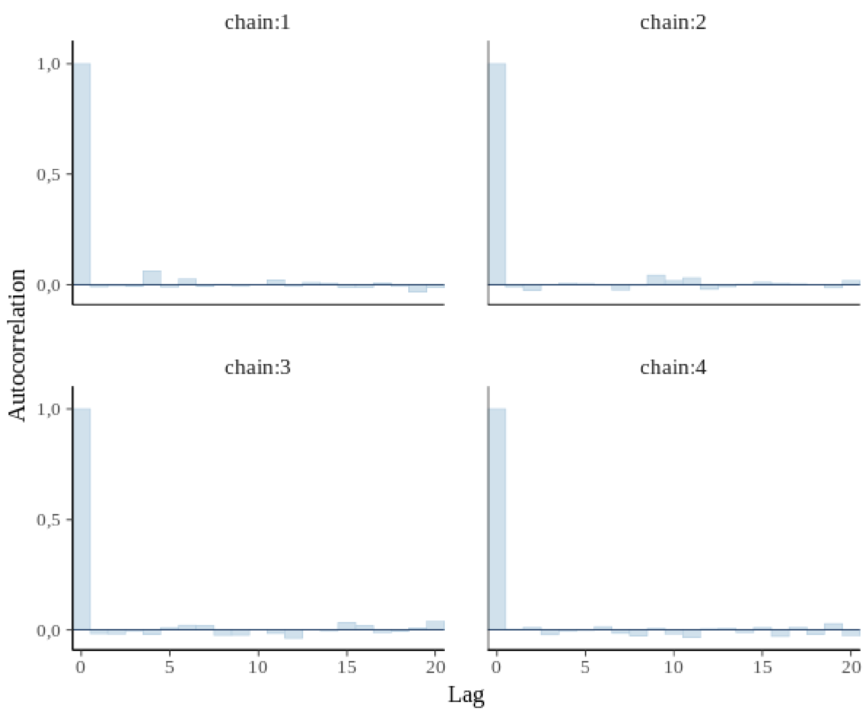

Figure 35: year

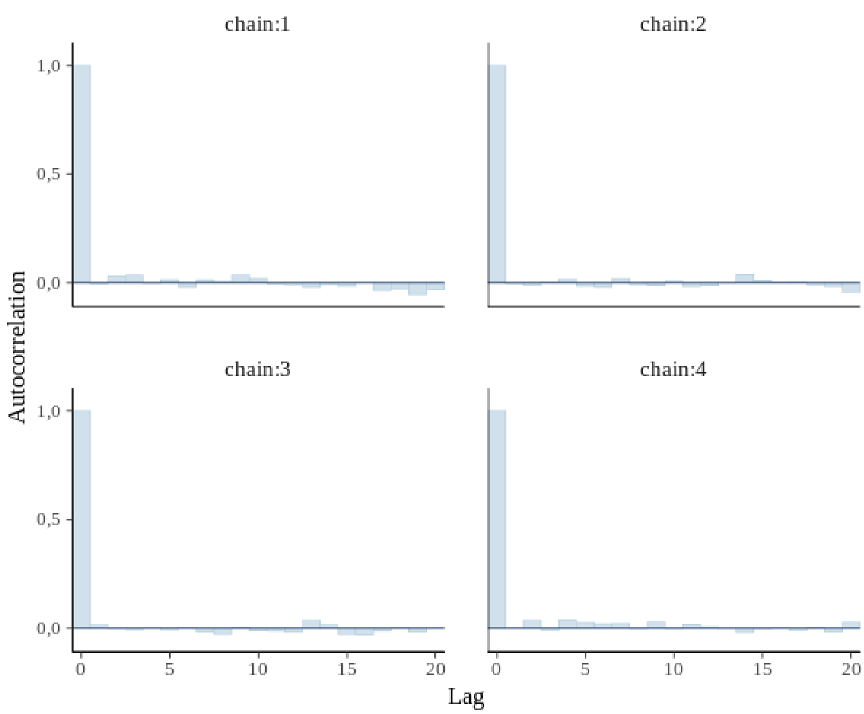

Figure 36: semester

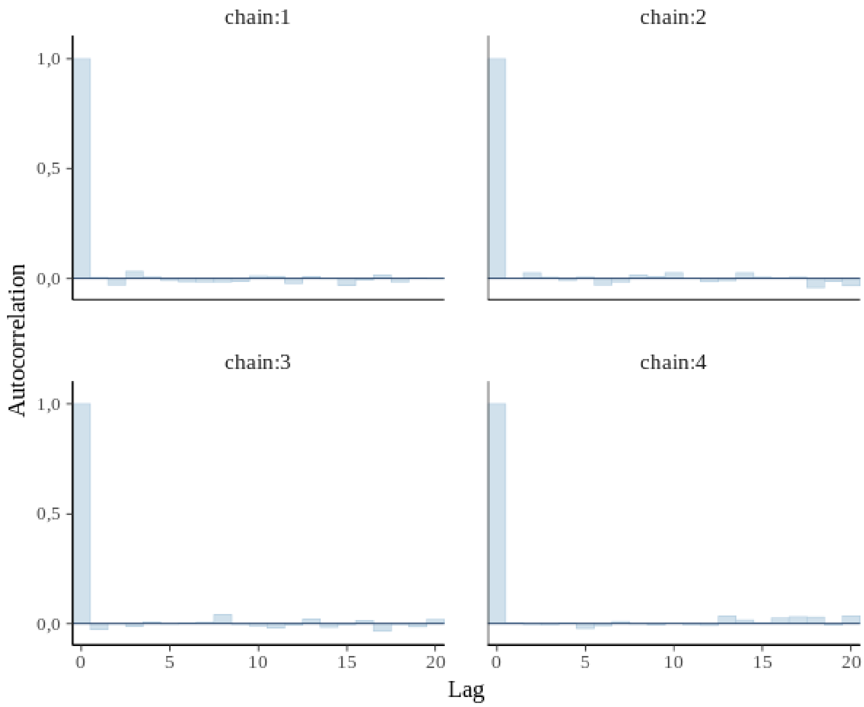

Figure 37: school 2012 - municipal

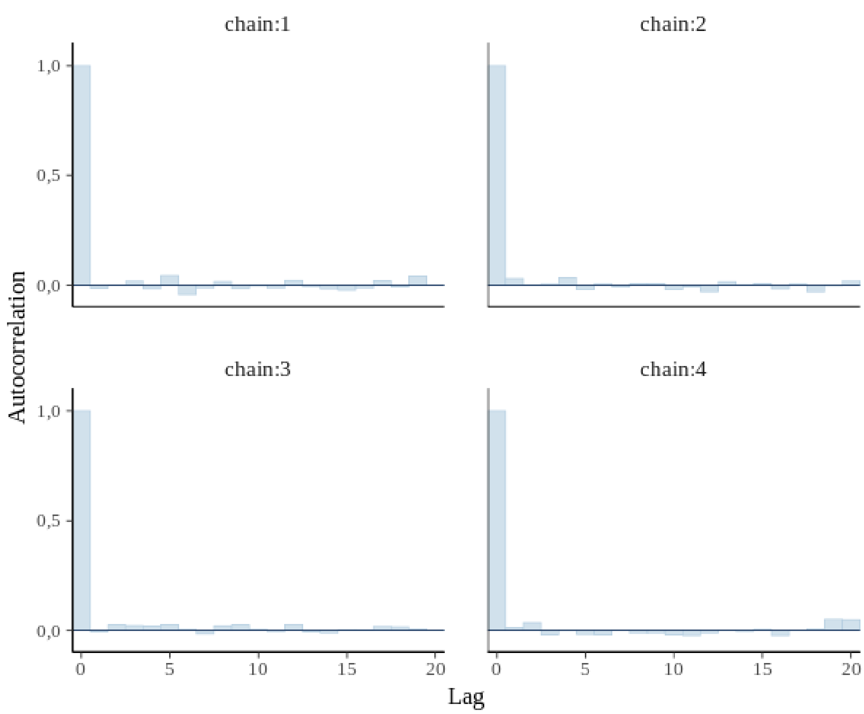

Figure 38: school 2012 - privada

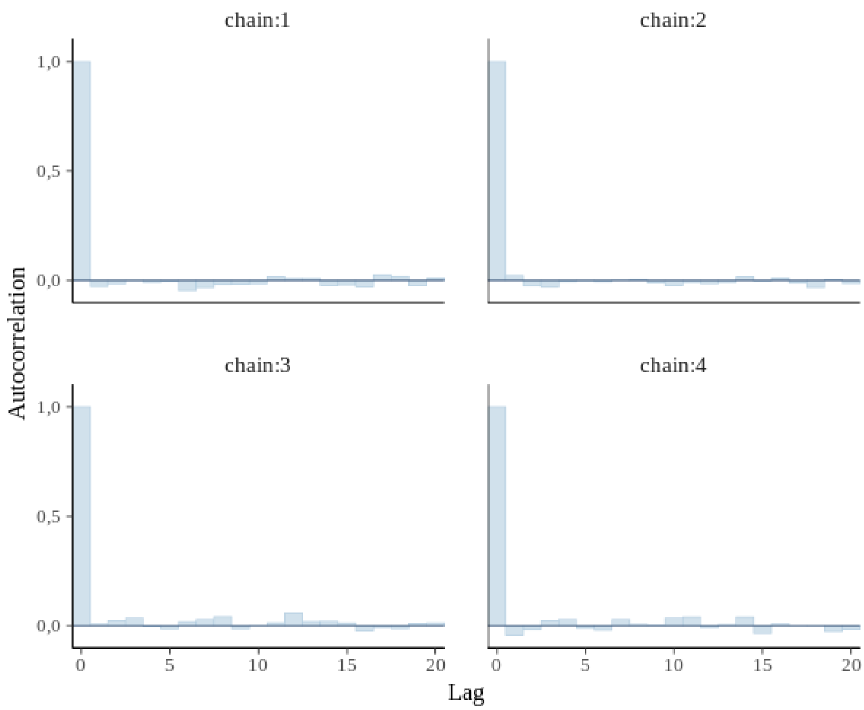

Figure 39: white

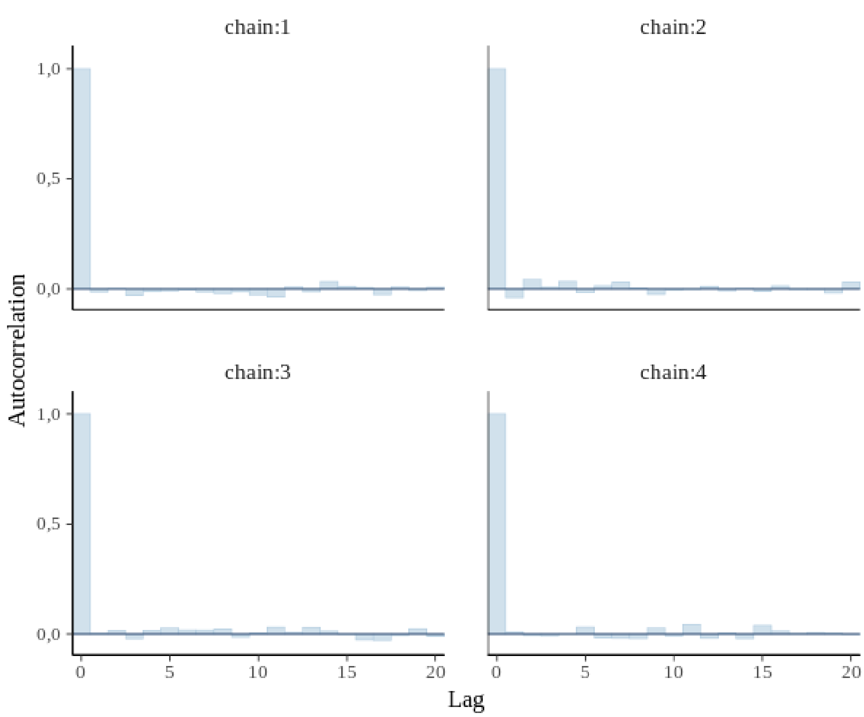

Figure 40: male

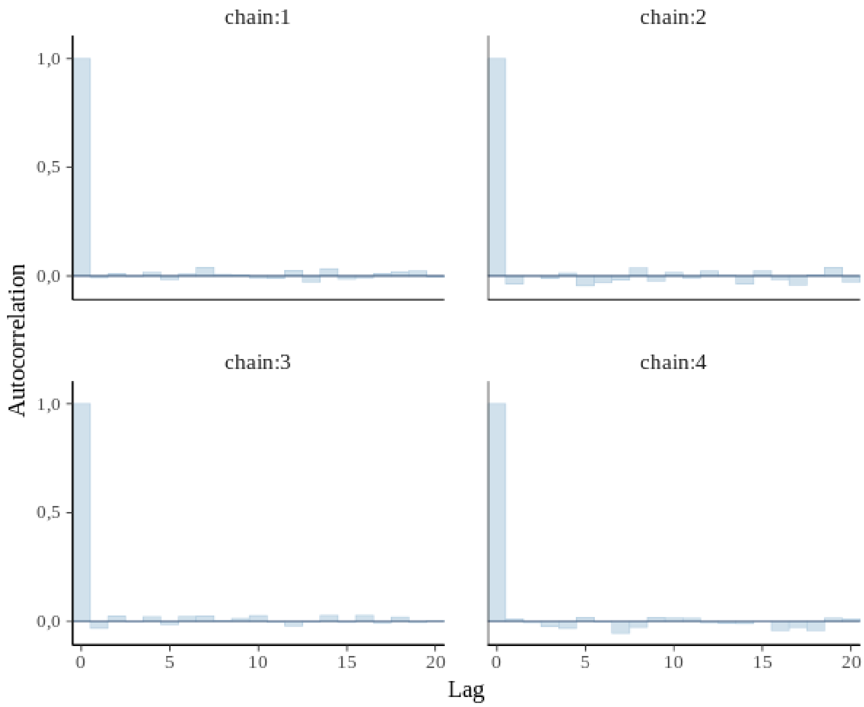

Figure 41: failed before 2012

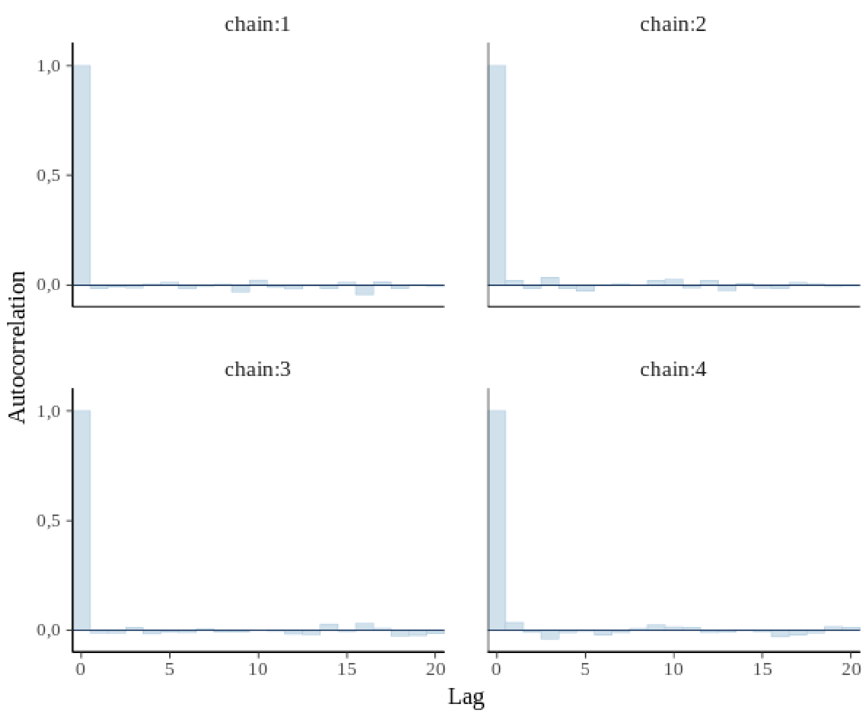

Figure 42: mother educ - ef1

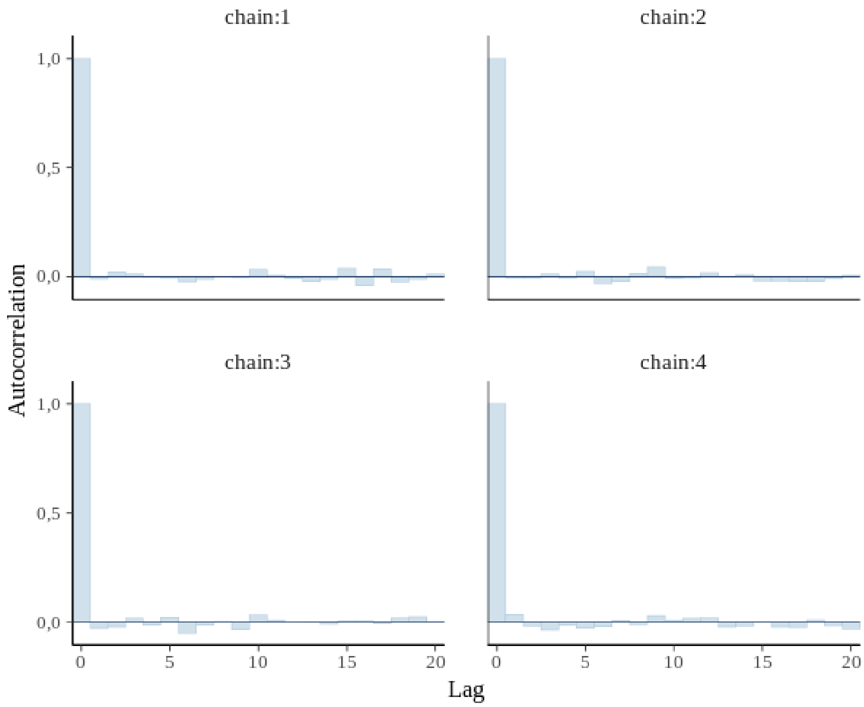

Figure 43: mother educ - ef2

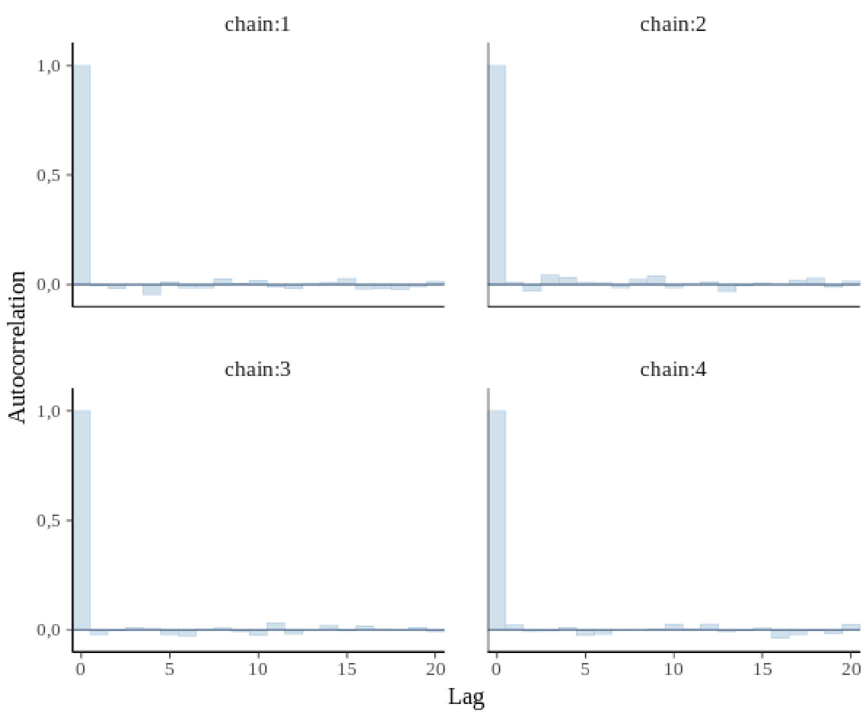

Figure 44: mother educ - em

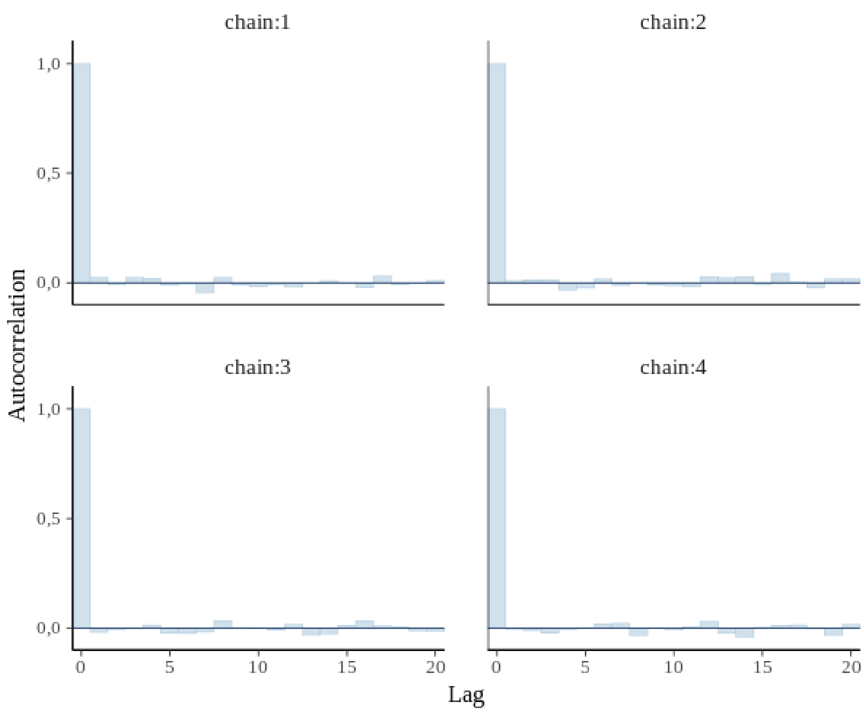

Figure 45: mother educ - nao sabe

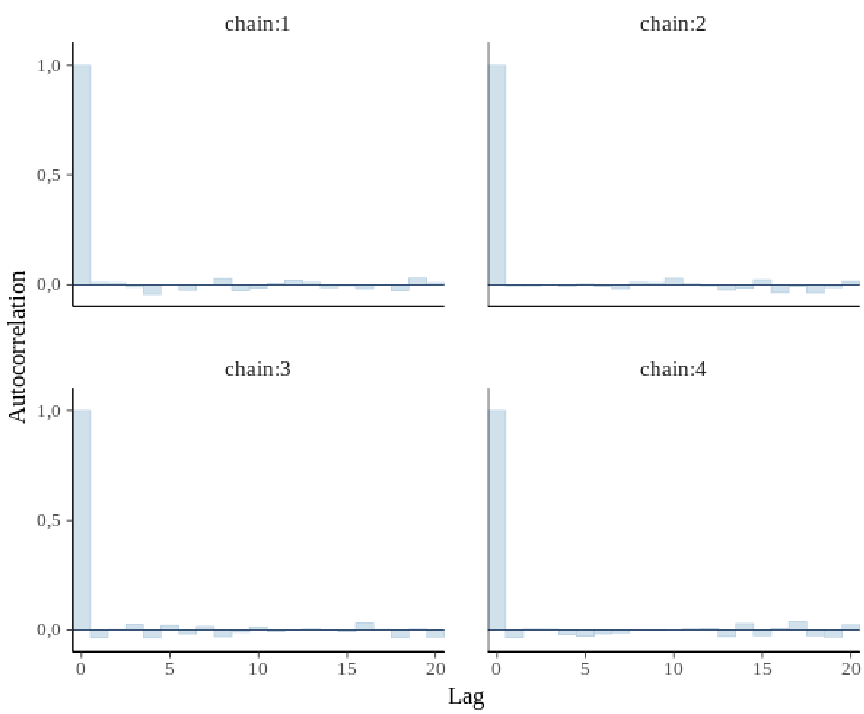

Figure 46: mother educ - superior

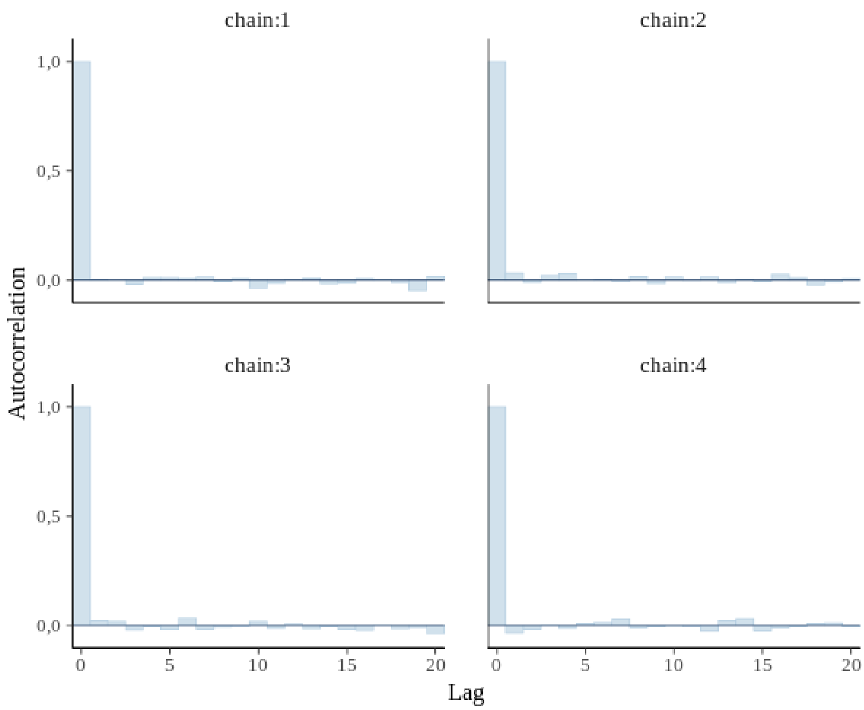

Figure 47: pre k

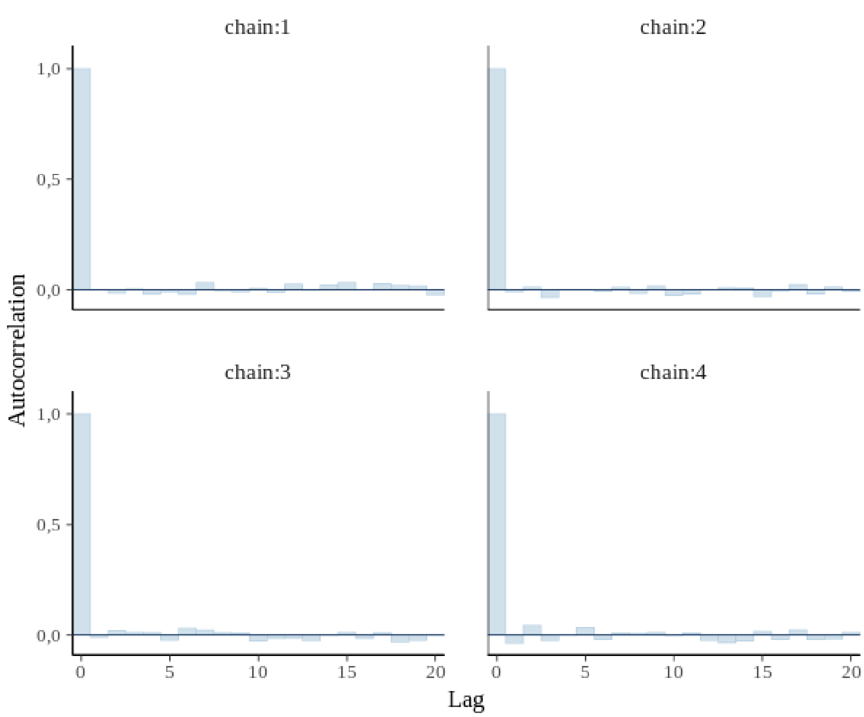

Figure 48: kinder

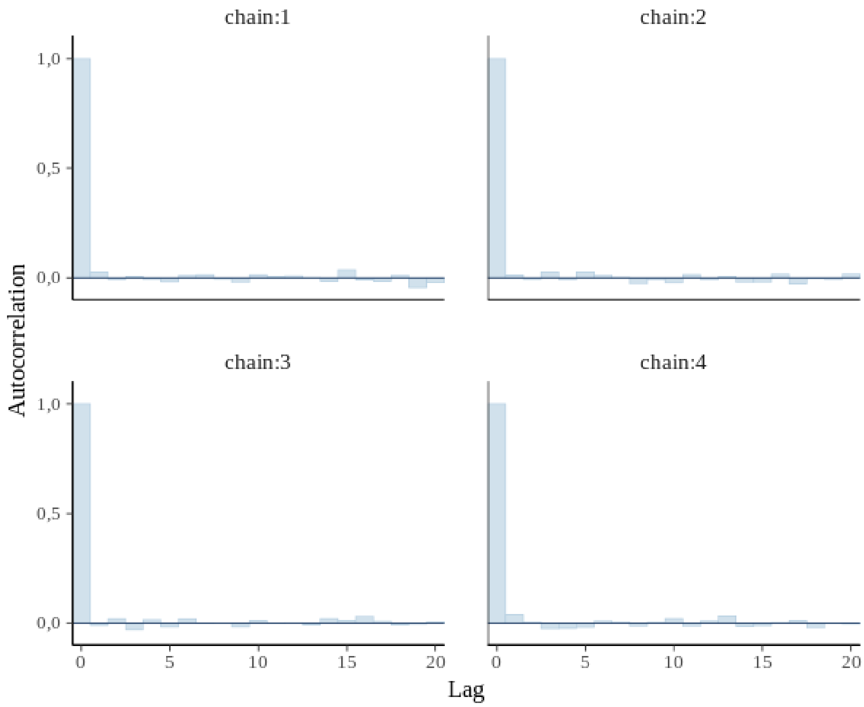

Figure 49: lang 2012

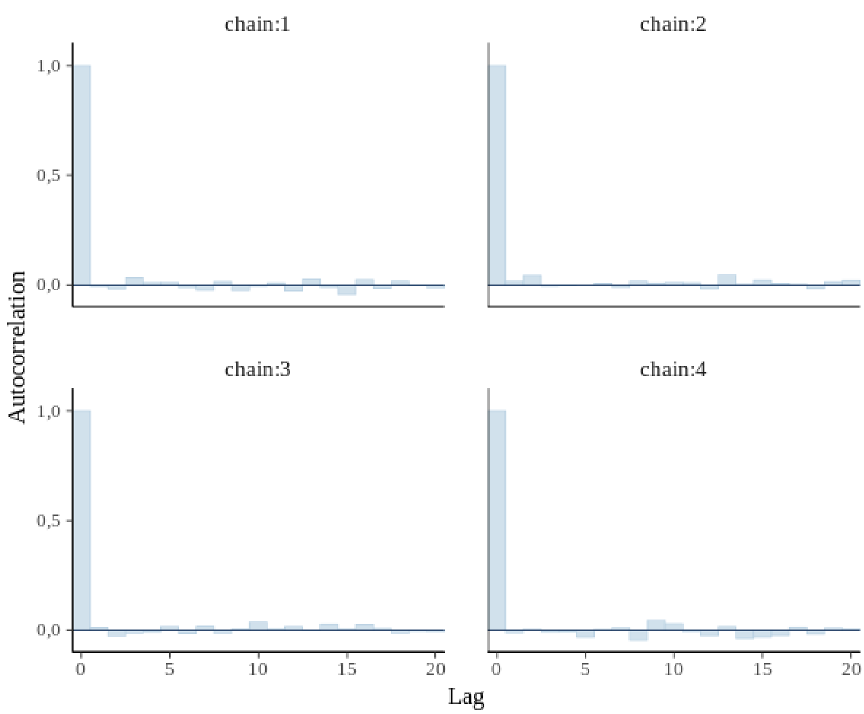

Figure 50: math 2012

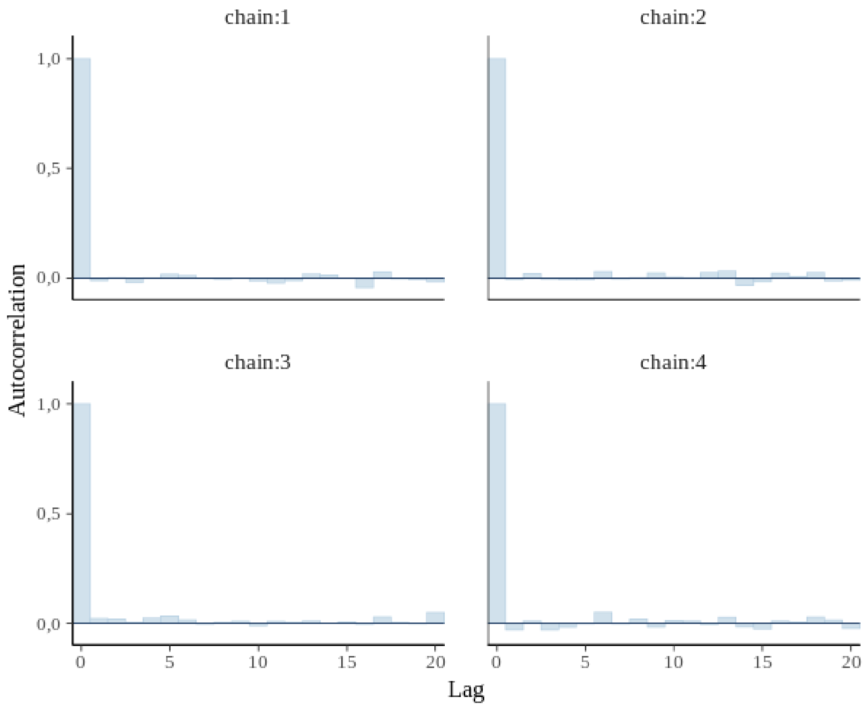

Figure 51: act 2012

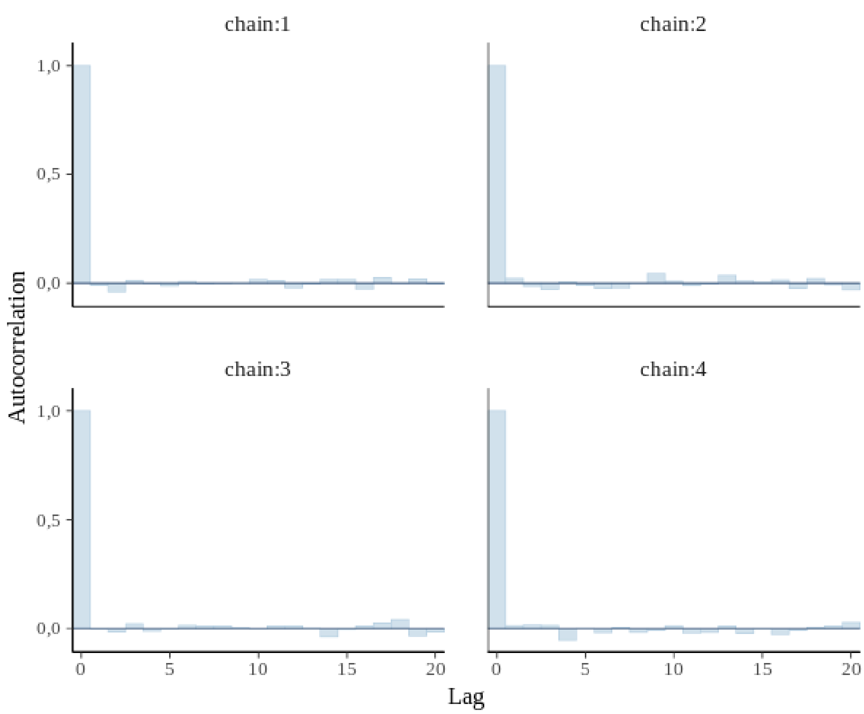

Figure 52: aes 2012

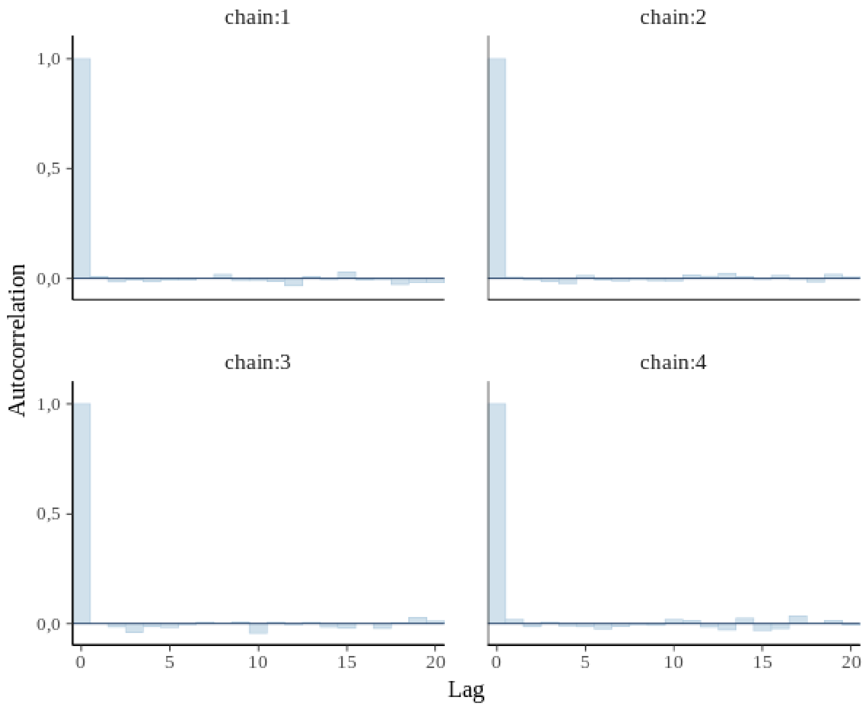

Figure 53: alt 2012

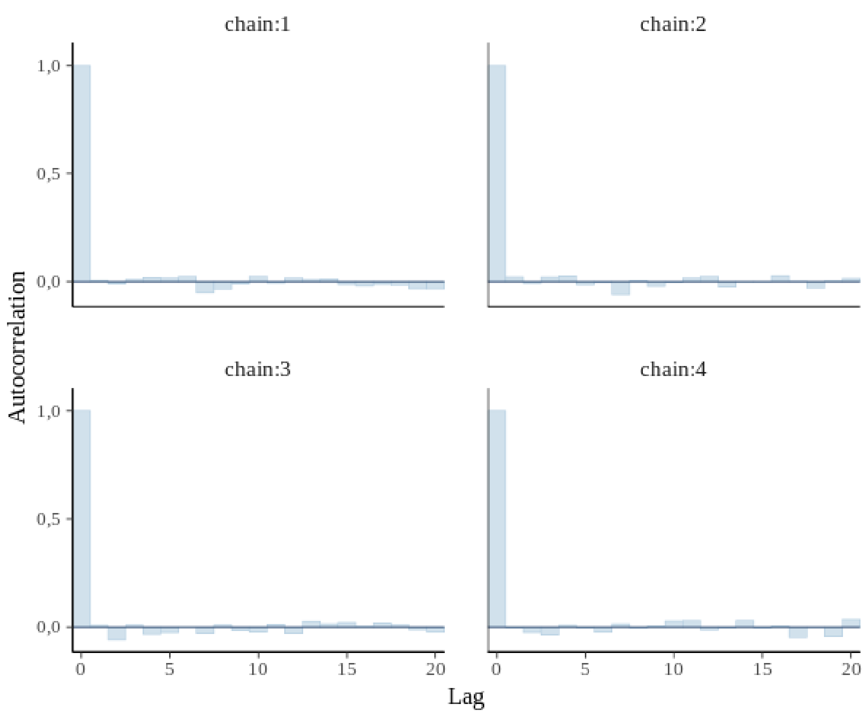

Figure 54: anx 2012

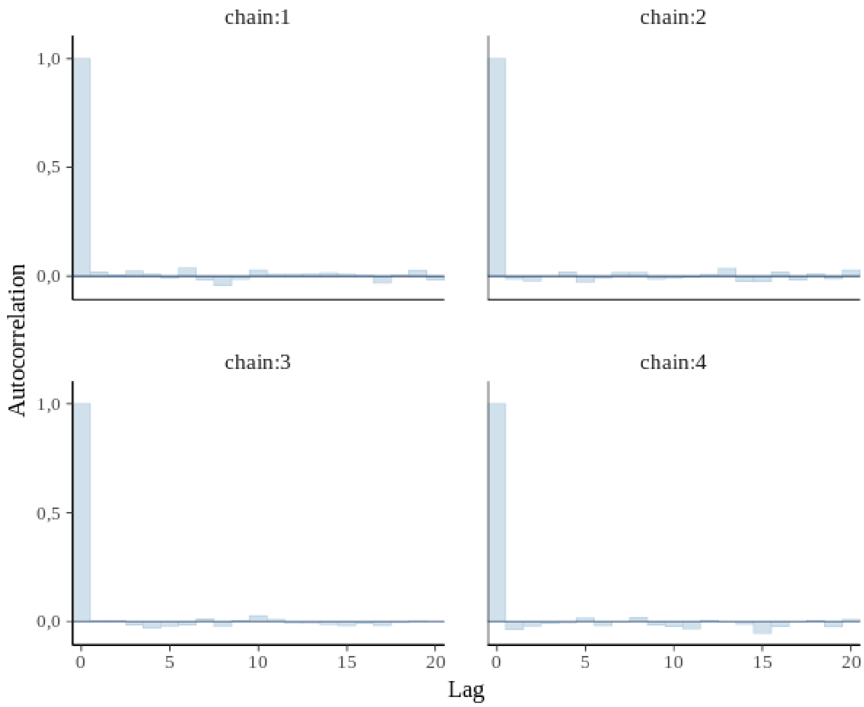

Figure 55: ass 2012

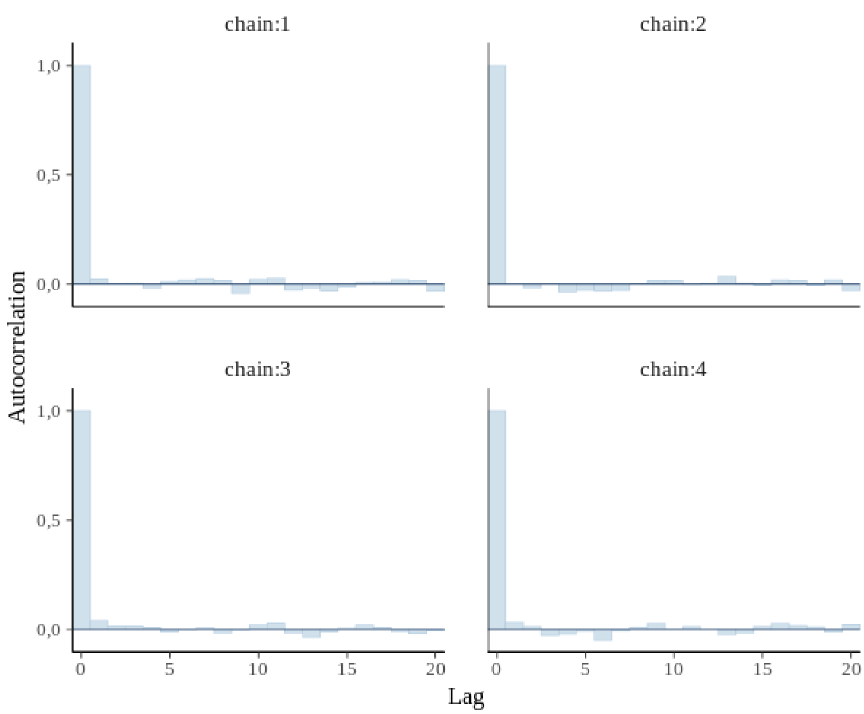

Figure 56: cmp 2012

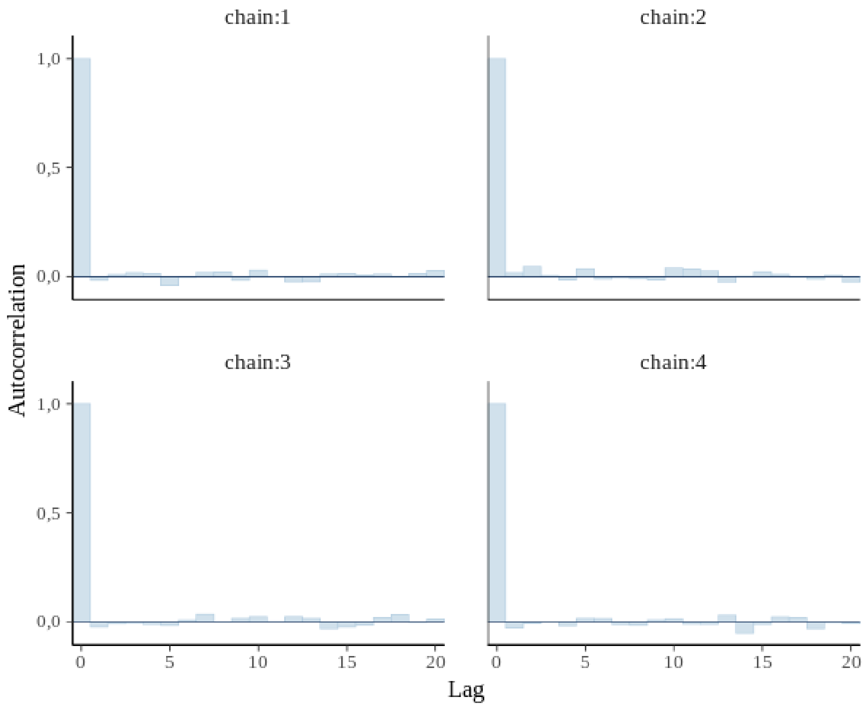

Figure 57: dep 2012

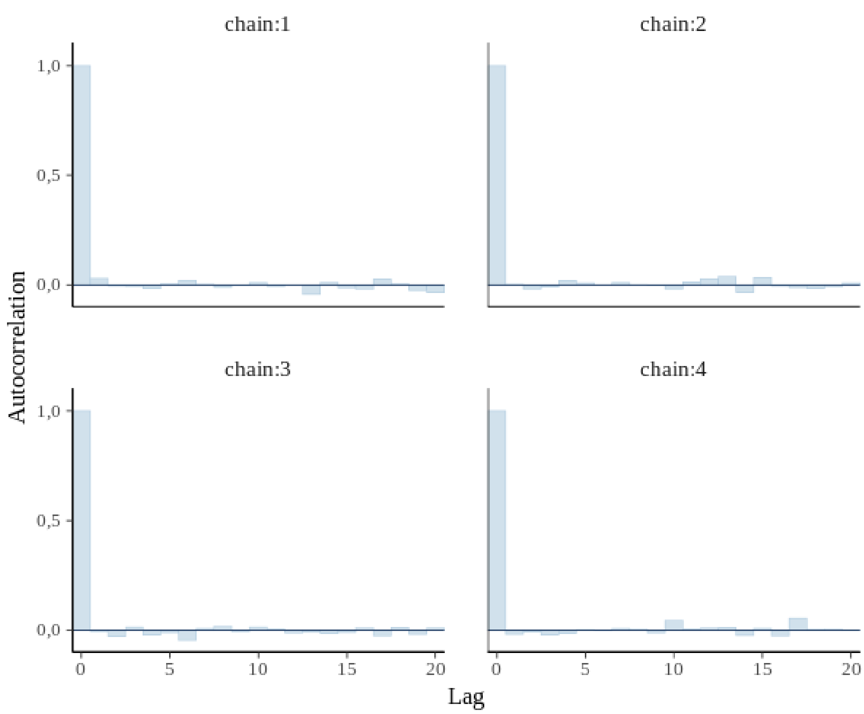

Figure 58: ids 2012

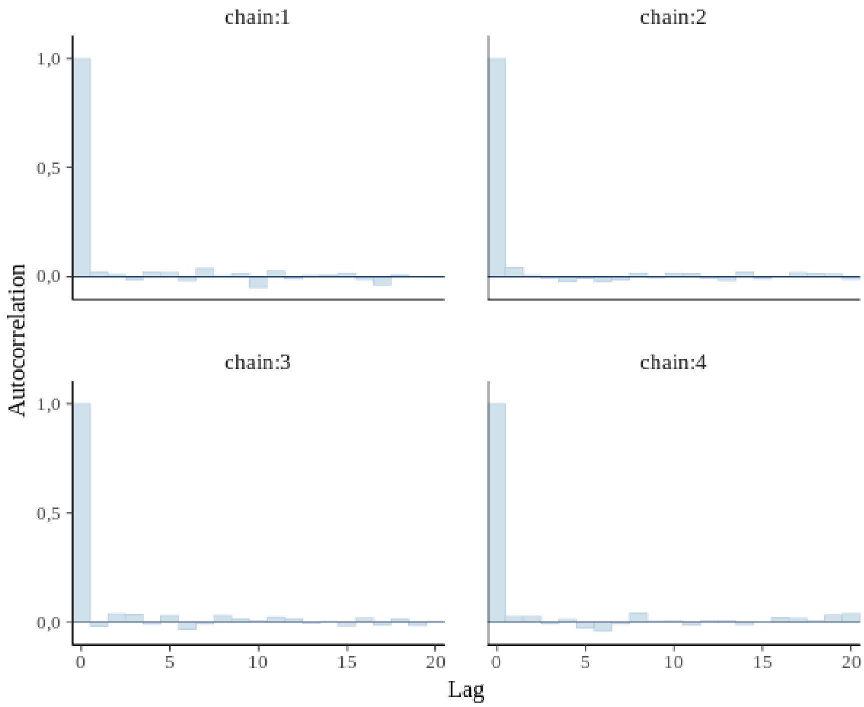

Figure 59: ord 2012

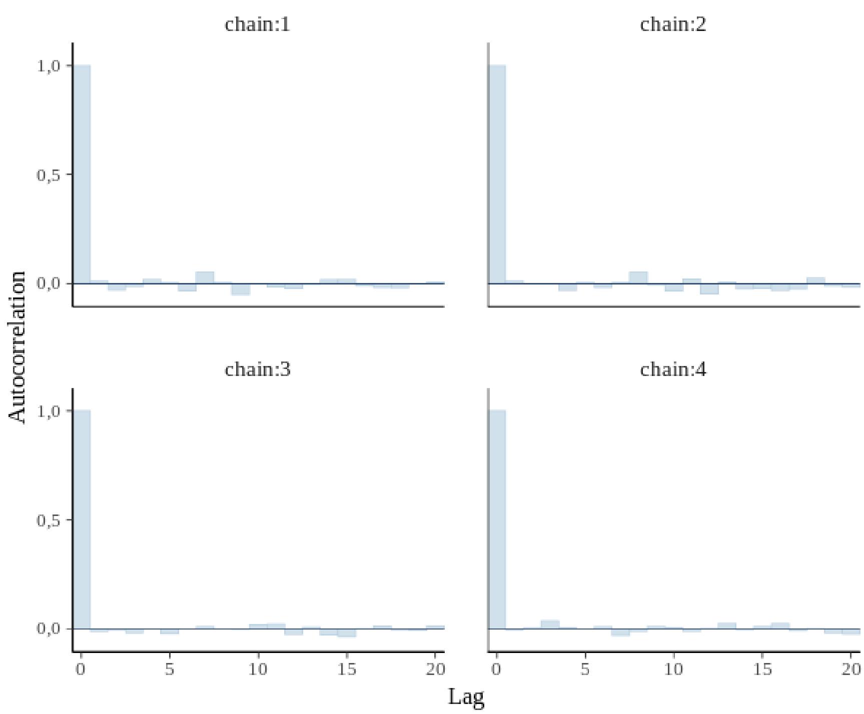

Figure 60: sfd 2012