

A GUIded tour of Bayesian regression

Andrés Ramírez-Hassan* Mateo Graciano-Londoño†

December 23, 2019

Abstract

This paper presents a graphical user interface (GUI) to carry out a Bayesian regression analysis in a very friendly environment without any programming skills (drag and drop). This paper is designed for teaching and applied purposes at an introductory level; we present the basic theory underlying all regression models that we developed in our GUI, which in turn is based on an interactive web application using shiny, and libraries from R. We carry out some applications to highlight the potential of our GUI for applied researchers and practitioners.

Keywords: Bayesian regression, graphical user interface.

*Department of Economics, School of Economics and Finance, Universidad EAFIT, Cra. 49 7 sur 50 Medellín, Antioquia, Colombia. e-mail: aramir21@eafit.edu.co URL: www.besmarter-team.org.

†Department of Mathematical Engineering, School of Sciences, Universidad EAFIT, Cra. 49 7 sur 50 Medellín, Colombia; email: magralo@gmail.com

1 Introduction

The main objective of this paper is to present an open source teaching graphical user interface (GUI) to implement Bayesian regression analysis using cross sectional and longitudinal data. We present a tutorial for implementing these models in our GUI, some applications, and the basic theoretical framework. Therefore, practitioners and applied researchers can apply Bayesian regression analysis understanding its theoretical foundation without requiring programming skills. The latter seems to be a significant impediment to increasing the use of the Bayesian framework (Woodward, 2005; Karabatsos, 2016).

Table 1 shows the available graphical user interfaces for carrying out Bayesian regression analysis. ShinyStan (Stan Development Team, 2017) is a very flexible open source program, but users are required to have some programming skills. BugsXLA (Woodward, 2005) is open source, but less flexible. However, users do not need to have programming skills. Bayesian regression: Nonparametric and parametric models (Karabatsos, 2016) is a very flexible and friendly GUI that is based on MATLAB Compiler for a 64-bit Windows computer. Its focus is on Bayesian nonparametric regressions, and it can be thought of for users who have mastered basic parametric models, such as the ones that we show in our GUI. On the other hand, MATLAB toolkit, Stata and BayES are not open source.

We developed our GUI based on an interactive web application using shiny (Chang, 2018), and some libraries in R (R Core Team, 2018). The specific libraries and commands that are used in our GUI can be seen in Table 2. It has nine univariate models, four multivariate, three hierarchical longitudinal, Bayesian bootstrap, and six Bayesian model averaging frameworks. In addition, it gives basic summaries and diagnostics of the posterior chains, as well as the posterior chains themselves, and different plots, such as trace, autocorrelation and densities. In terms of its flexibility and possibilities, our GUI lies between ShinyStan and BugsXLA: users are not required to have any programming skills, but it is not as advanced as Karabatsos (2016)'s software. However, our GUI can be run in any operating system. Our GUI, which we call BEsmarter,¹ is freely available at <https://github.com/besmarter/BSTApp>; so users have access to all our code and datasets.

After this brief introduction, we present in Section 2 our GUI, and how to use it. Section 3 presents some empirical examples to illustrate the potential use of our GUI. Sections 4 and 5 are more theory oriented. Section 4 introduces some general theoretical and computational aspects of the Bayesian framework, and Section 5 presents the specific technical details of the models implemented in our GUI. Lastly, Section 6 presents some concluding remarks and future developments.

2 Using BEsmarter

Simulated and applied datasets are in the folders *DataSim* (see Table 3 for details), and *DataApp* (see Table 4 for details), respectively. The former folder also includes the files that

¹Bayesian econometrics: Simulations, models and applications to research, teaching and encoding with responsibility.

were used to simulate different processes, so, the population parameters are available, and as a consequence these files can be used as a pedagogical tool to show some statistical properties of the inferential frameworks available in our GUI. The latter folder contains the datasets used in our applications in Section 3. Users should use these datasets as templates as a guide to the structure of their own datasets. Simply type `shiny::runGitHub("besmarter/BSTApp", launch.browser=T)` in the R package console or any R code editor to run our GUI.

After this, users can see a new window where a presentation of our research team is displayed. In addition, the top panel in Figure 1 shows the class of models that can be estimated in our GUI.

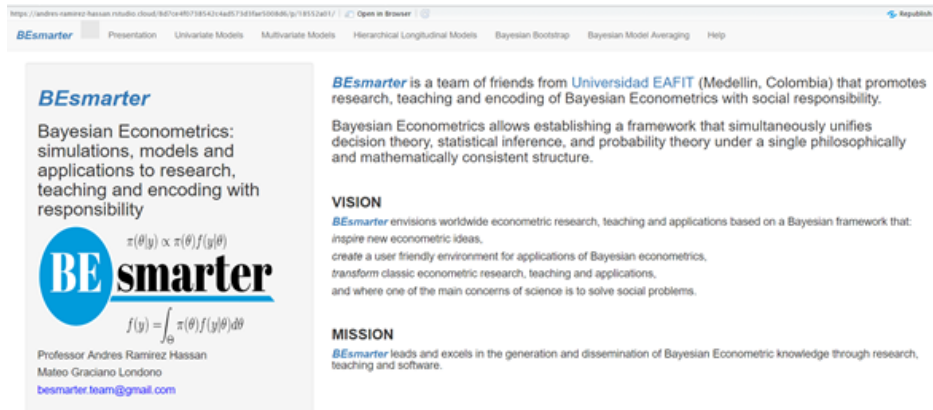


Figure 1: BESmarter GUI

The selection indicates univariate models in that the radio button on the left hand side shows the specific models inside this generic class. In particular, users can see that the normal model is selected from inside the class of univariate models. See Figure 2.

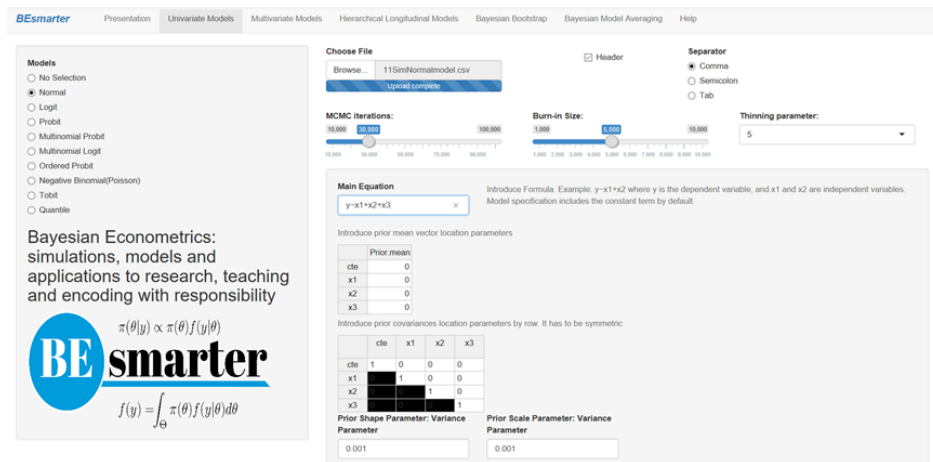


Figure 2: Univariate models: Specification

Then, the right hand side panel displays a widget to upload the input dataset, which should be a csv file with headers in the first row. Users also should select the kind of separator used in the input file: comma, semicolon, or tab (use the folders *DataSim* and *DataApp* for the input file templates). Range sliders help to set the number of iterations of the MCMC and the amount of burn-in, and the thinning parameter can be selected as well (see Section 4 for technical details). After this, users should specify the equation. This is expressed in the format used by R (see Main equation box in Figure 2, $y \sim x1 + x2 + x3$). Note that the class of univariate models includes the intercept by default, except ordered probit, where the specification has to do this explicitly, that is, ordered probit models do not admit an intercept, for identification issues (see Section 5.1). Hence, users should write down specifically this fact ($y \sim x1 + x2 + x3 - 1$). Finally, users should define the hyperparameters of the prior; for instance, in the normal-inverse gamma model, these are the mean, covariance, shape, and scale (see Figure 2). However, users should take into account that our GUI has “non-informative” hyperparameters by default in all our modelling frameworks, so the last part is not a requirement.

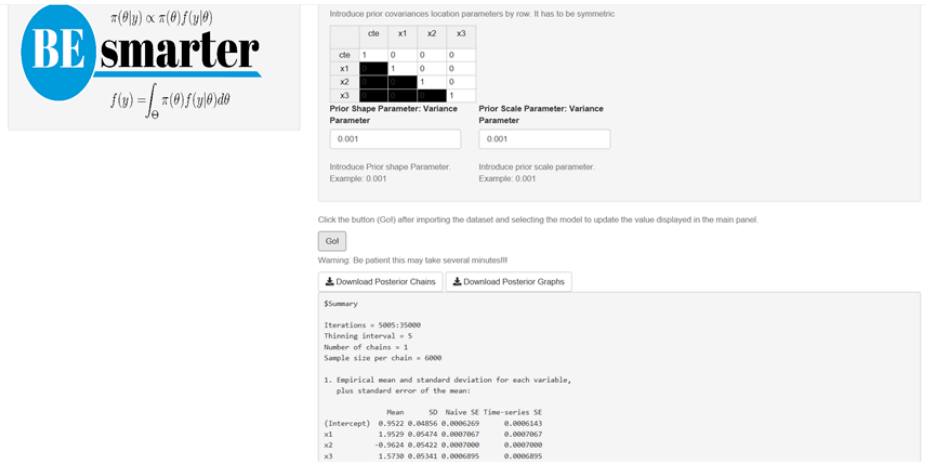


Figure 3: Univariate models: Results

After this specification process, users should click the Go! button to initiate the estimation. Our GUI displays the summary statistics and convergence diagnostics after this process is finished (see Figure 3). There are also widgets to download posterior chains (csv file) and graphs (pdf and eps files). Note that the order of the coefficients in the results (summary, posterior chains, and graphs) is first for the location parameters, and then for the scale parameters.

Multinomial models (probit and logit) require a dataset file to have in the first column the dependent variable, then alternative specific regressors (for instance alternatives’ prices), and finally, non-alternative regressors (for instance, income). Specification also requires defining the base category, number of alternatives (this is also required in ordered probit), number of alternative specific regressors, and number of non-alternative regressors (see Figure 4). Multinomial logit also allows defining a tuning parameter, the number of degrees of freedom

in this case, for the Metropolis–Hastings algorithm (see Sections 4 and 5 for details). This is a feature in our GUI when the estimation of the models is based on the Metropolis–Hastings algorithm. The order of the coefficients in the results of these models is first the intercepts, and then the non-alternative specific regressors, which are in order of the category, and lastly, the coefficients for the alternative specific regressors. Note that the non-alternative specific regressors associated with the base category are equal to zero (they do not appear in the results). In addition, some coefficients of the main diagonal of the covariance matrix are constant due to identification issues in multinomial and multivariate probit models.

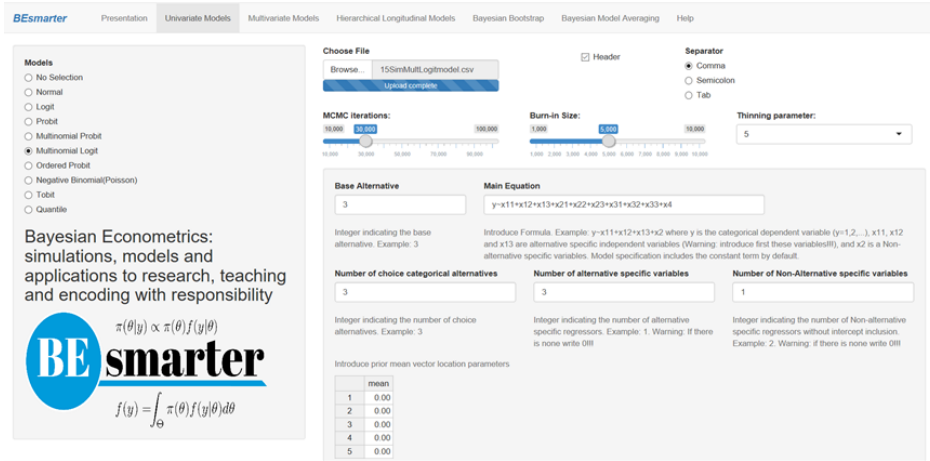


Figure 4: Univariate models: Multinomial

In the case of the negative binomial model, users should set a dispersion parameter (α , see the negative binomial model). User should also set the censorship points and quantiles in the Tobit and quantile models, respectively.

Figure 5 displays the multivariate regression setting. In this case, the input file should have first the dependent variables, and then the regressors. If there are intercepts in each equation, there should be a column of 1's after the dependent variables in the input file. The user also has to set the number of dependent variables, the number of regressors, if necessary include the intercept, and the values of the hyperparameters (see Figure 5).

The input file in seemingly unrelated regressions should have first the dependent variables, and then the regressors by equation, including the intercept in each equation if necessary (column of 1's). Users should define the number of dependent variables (equations), the number of total regressors, that is, the sum of all regressors associated with the equation (if necessary include intercepts, each intercept is an additional regressor), and the number of regressors by equation (if necessary include the intercept). Users can also set the values of the hyperparameters if there is prior information.

The results of the simple multivariate and seemingly unrelated regressions show first the posterior location parameters by equation, and then the posterior covariance matrix.

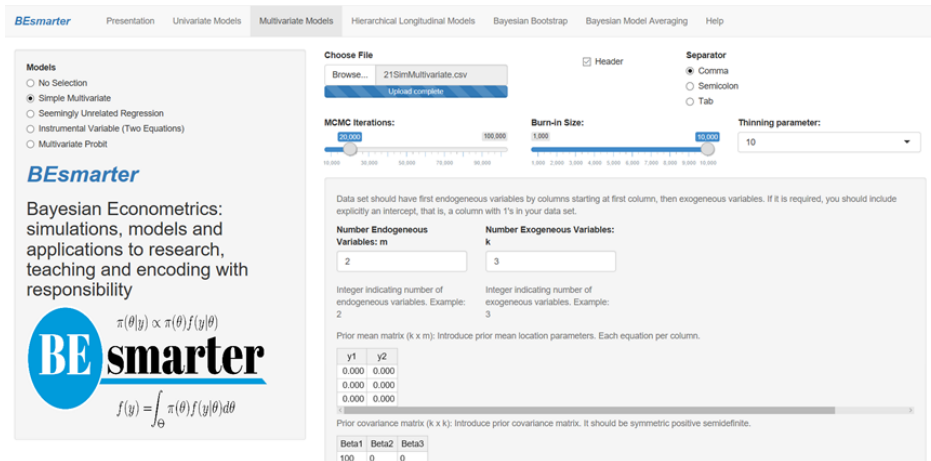


Figure 5: Multivariate models: Simple multivariate

In the instrumental variable setting, users should specify the main equation and the instrumental equation. This setting includes intercepts by default. The first variable on the right hand side in the main equation has to be the variable with endogeneity issues. In the instrumental equation box, the dependent variable is the variable with endogeneity issues as a function of the instruments. Users can also specify the values of the hyperparameters if they have prior information. The input file should have the dependent variable, the endogenous regressor, the instruments, and the exogenous regressors. The results first list the posterior estimates of the endogenous regressor, then the location parameters of the auxiliary regression (instrumental equation), and the location parameters of the exogenous regressors. Last is the posterior covariance matrix.

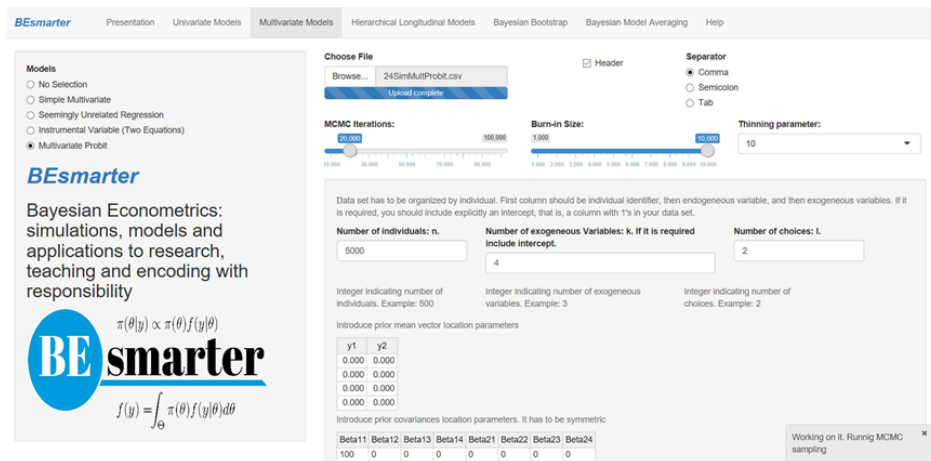


Figure 6: Multivariate models: Multivariate probit

The multivariate probit model requires an input dataset ordered by unit, for instance three choices implies repeat each unit three times. The first column has to be the identifi-

cation of each unit; users should use ordered integers, then the dependent variable, just one vector, composed of 0's and 1's, then the regressors, which should include a column of 1's for the intercepts. Users should set the number of units, number of regressors, and number of choices (see Figure 6). The results first display the posterior location parameters by equation, and then the posterior covariance matrix.

The input files for hierarchical longitudinal models should have first the dependent variable, then the regressors and a cross sectional identifier ($i = 1, 2, \dots, m$). It is not a requirement to have a balanced dataset: n_i can be different for each i (see subsection 5.3 for technical details). Users should specify the fixed part equation and the random part equation, both in R format. In case of only requiring random intercepts, do not introduce anything in the latter part (see Figure 7). Users should also type the name of the cross sectional identifier variable. The results displayed and the posterior graphs are associated with the fixed effects and covariance matrix. However, users can download the posterior chains of all posterior estimates: fixed and random effects, and covariance matrix.

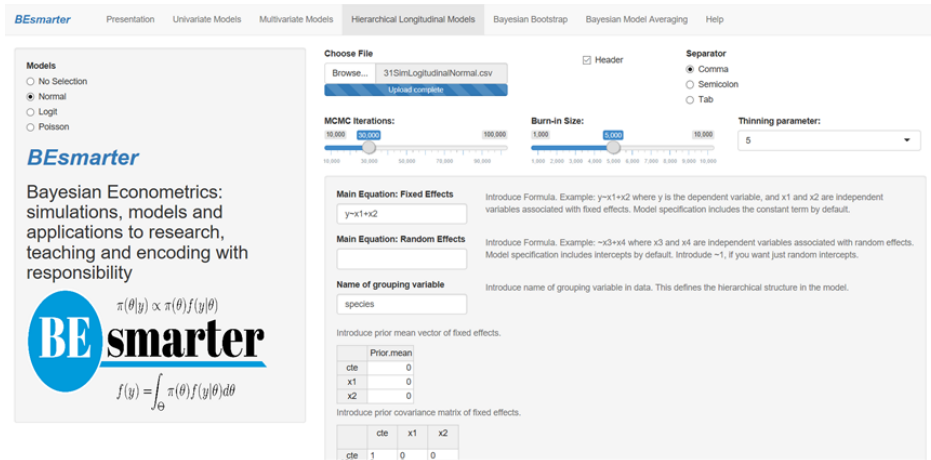


Figure 7: Hierarchical longitudinal models: Specification

Bayesian bootstrap only requires uploading a dataset, specifying the number of iterations of the MCMC, the resampling size, and the equation (see Figure 8). The input file has the same structure as the file used in the univariate normal model.

Bayesian model averaging based on a Gaussian distribution can be carried out using the Bayesian information criterion (BIC) approximation, Markov chain Monte Carlo model composition (MC3), or instrumental variables (see Figure 9). The former two approaches require an input dataset where the first column is the dependent variable, and then, the potentially important regressors. Users should set the band width model selection parameter (O_R) and number of iterations for BIC and MC3, respectively. The results include the posterior inclusion probability ($p^i = 0$), expected value (EV), and standard deviation (SD) of the coefficients associated with each regressor. The BIC framework also displays the most relevant models, including the number of regressors, the coefficient of determination (R^2), the BIC, and the

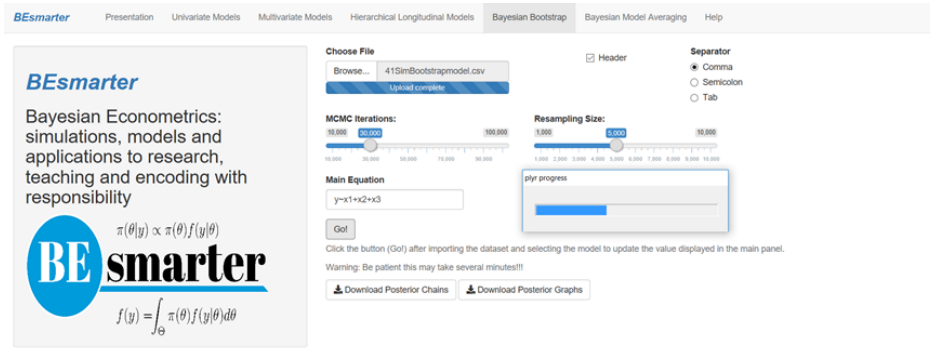


Figure 8: Bayesian bootstrap: Specification

posterior model probability. Users can download two csv files: *Best models* and *Descriptive statistics coefficients*. The former is a 0-1 matrix such that the columns are the regressors and the rows are the models; a 1 indicates the presence of a specific regressor in a specific model, 0 otherwise. Note that the last column of this file is the posterior model probability for each model (row). The latter file shows the posterior inclusion probabilities, expected values, and standard deviations associated with each regressor, taking into account the BMA procedure based on the best models.

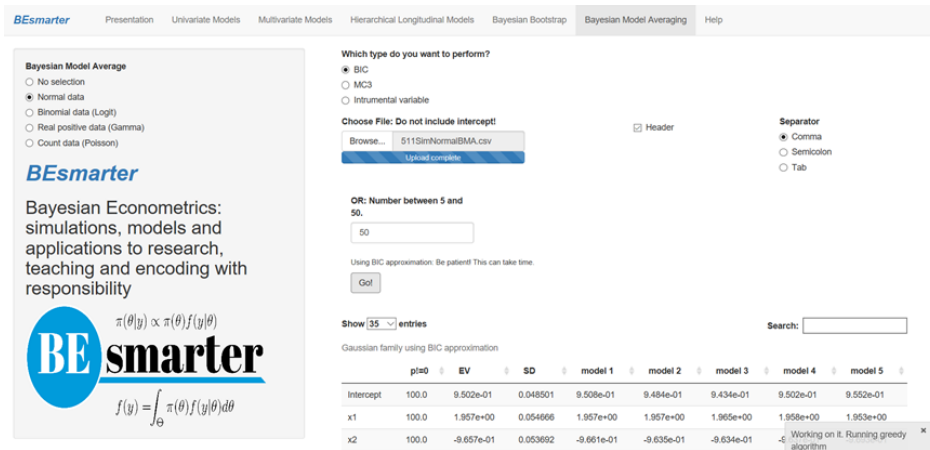


Figure 9: Bayesian model averaging: Specification and results

Bayesian model averaging with endogeneity issues requires two input files. The first one has the dependent variable in the first column, the next columns are the regressors with endogeneity issues, and then the exogenous regressors. The user should include a column of 1's if an intercept is required. The second input file has all the instruments. Users should also introduce the number of regressors with endogeneity issues (see Figure 10).

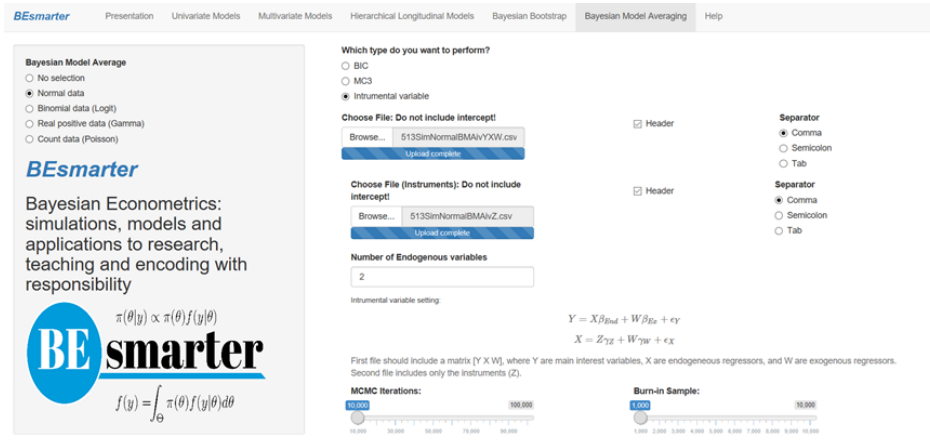


Figure 10: Bayesian model averaging: Instrumental variable specification

The results include the posterior inclusion probabilities and expected values for each regressor. The user can find the results of the main equation, and then of the auxiliary equations. Users can download csv files of BMA results for both the second stage (main equation) and the first stage (auxiliary equations). In addition, users can download the posterior chains of the location parameters of the main equation, β_l , $l = 1, 2, \dots, \dim\{\beta\}$, the location parameters of the auxiliary equations, $\gamma_{j,i}$, $j = 1, 2, \dots, \dim\{\beta_s\}$ where $\dim\{\beta_s\}$ is the number of regressors with endogeneity issues, $i = 1, 2, \dots, \dim\{\gamma\}$, where $\dim\{\gamma\}$ is the number of regressors in the auxiliary regressors (exogeneous regressors + instruments), and the elements of the covariance matrix $\sigma_{j,k}$ (see subsection 5.5 for technical details).

Bayesian model averaging based on BIC approximation for non-linear models, logit, gamma, and Poisson, requires an input dataset where the first column is the dependent variable, and the other columns are the potentially relevant regressors. Users should specify the band width model selection parameters, which are also referred to as Occam's window parameters (O_R and O_L). Our GUI displays the PIP ($p^l = 0$), the expected value of the posterior coefficients (EV), and the standard deviation (SD). In addition, users can see the results associated with the models with the highest posterior model probabilities, and download csv files with the results of specifications of the best models, and descriptive statistics of the posterior coefficients from the BMA procedure. These files are similar to the results of the BIC approximation of the Gaussian model.

User should also note that sometimes our GUI shuts down. In our experience, this is due to computational issues using the implicit commands that we call when estimating some models, for instance, computationally singular systems, missing values where TRUE/FALSE needed, L-BFGS-B needs finite values of "fn", NA/NaN/Inf values, or Error in backsolve. Sometimes these issues can be solved by adjusting the dataset, for instance, avoiding high levels of multicollinearity. In addition, users can identify these problems by checking the console of their rstudio cloud sections, where the specific folder/file where the issue happened is specified. In any case, we would appreciate your feedback to improve and enhance our

GUI.

3 Applications using BEsmarter

The main purpose of this section is to illustrate the potential of our GUI to carry out some applications. We encourage users to replicate these applications, as we do not display in figures most of the results, due to space limitations.² In addition, there are technical aspects that are covered in the next two sections.

3.1 Univariate models

Continuous response: The market value of soccer players in Europe

We use the dataset *1ValueFootballPlayers.csv*, which was provided by [Serna Rodríguez et al. \(2018\)](#), to find the determinants of high performance soccer players in the five most important national leagues in Europe.

The specification to enter in the main equation box is

$$\log(\text{Value}) \sim \text{Perf} + \text{Perf}^2 + \text{Age} + \text{Age}^2 + \text{NatTeam} + \text{Goals} + \text{Goals}^2 + \text{Exp} + \text{Exp}^2 + \text{Assists},$$

where Value is the market value in Euros (2017), Perf is a measure of performance, Age is the players' age in years, NatTem is an indicator variable that takes the value of 1 if the player has been on the national team, Goals is the number of goals scored by the player during his career, Exp is his experience in years, and Assists is the number of assist made by the player in the 2015–2016 season. All variables followed by a 2 are squared variables.

We initially assume that there are no censorship problems, the effect of the regressors are the same through the support of the dependent variable, and the dependent variable obeys a normal distribution. So, we ran a normal-inverse gamma model using 30,000 MCMC iterations plus a burn-in equal to 5,000, and a thinning parameter equal to 1 using the default hyperparameters.

The results suggest that age, squared age, national team, goals, experience, and squared experience are relevant regressors. For instance, we found that the 2.5% and 97.5% percentiles of the posterior estimate associated with the variable Goals are 4.57e-03 and 1.82e-02. These values can be used to find the 95% symmetric credible interval. This means that there is a 0.95 probability that the population parameter lies in (4.57e-03, 1.82e-02), which would suggest that this variable is relevant to explain the market value of a soccer player.³ We also found that the effect of having been on the national team has a 95% credible interval equal to (0.58, 1.04) with a median equal to 0.81, that is, an increase of the market value of the player of 124.8% ($\exp(0.81) - 1$) compared with a player that has not ever been on a national team. The posterior distribution of this variable can be seen in Figure 11. This

²Take into account that as inference in Bayesian models is based on simulation methods, results do not coincide 100%.

³Users should take into account that formal inference (hypothesis tests) in a Bayesian framework are based on Bayes factors. See subsection 5.5 for details.

graph is automatically generated by our GUI, and can be downloaded in the zip file named *Posterior Graphs.csv*; but we should take into account that the national team is the sixth variable, remember that by default the intercept is the first variable.

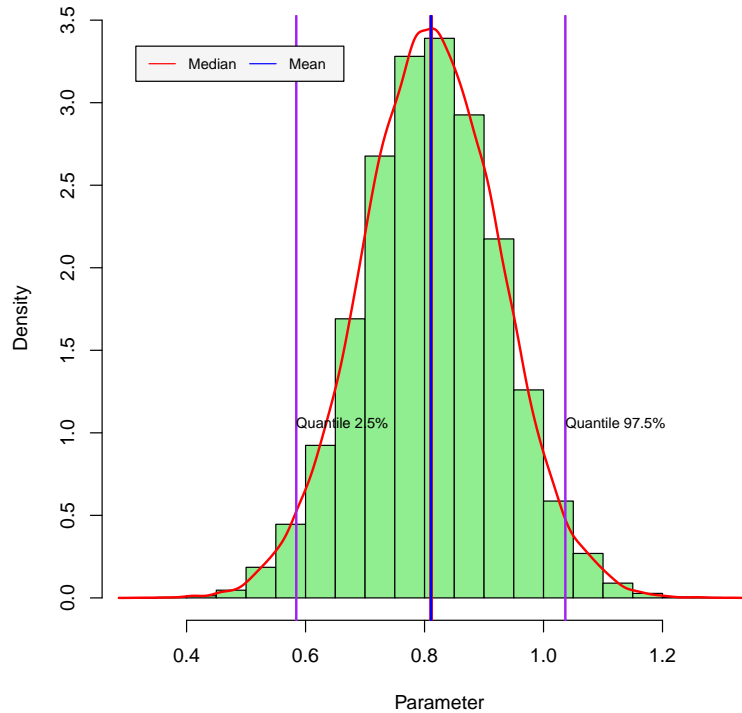


Figure 11: Posterior distribution: National team

A good advantage of the Bayesian framework is that we can easily calculate the posterior distribution of functions of the parameter estimates, for instance, the age that maximizes the market value of a soccer player, $\text{OptAge} = -\frac{\beta_{Age}}{2\beta_{Age2}}$. We can estimate this using the posterior chains that can be downloaded from our GUI. This is in the file named *Posterior chains.csv*. Figure 12 shows this posterior distribution, where we observe that the mean value is equal to 24.31 years, and the 95% symmetric credible interval is (23.28, 25.36).

We can also see some convergence diagnostics from this application. In particular, the Geweke (1992) test indicates that there is no statistically significant difference at 5% between the first 10% of the posterior chains and the last 50% of the posterior chains. This is due to the fact that the absolute value of all the statistical tests are less than 1.96 (the value that defines the critical region in a normal distribution for a bilateral test at the 5% significance level). The Raftery and Lewis (1992) tests indicate dependence factors very close to 1 in all cases, and as a consequence lower than 5, which means a low level of autocorrelation of the posterior draws. Lastly, all the posterior chains passed the Heidelberger and Welch (1983)

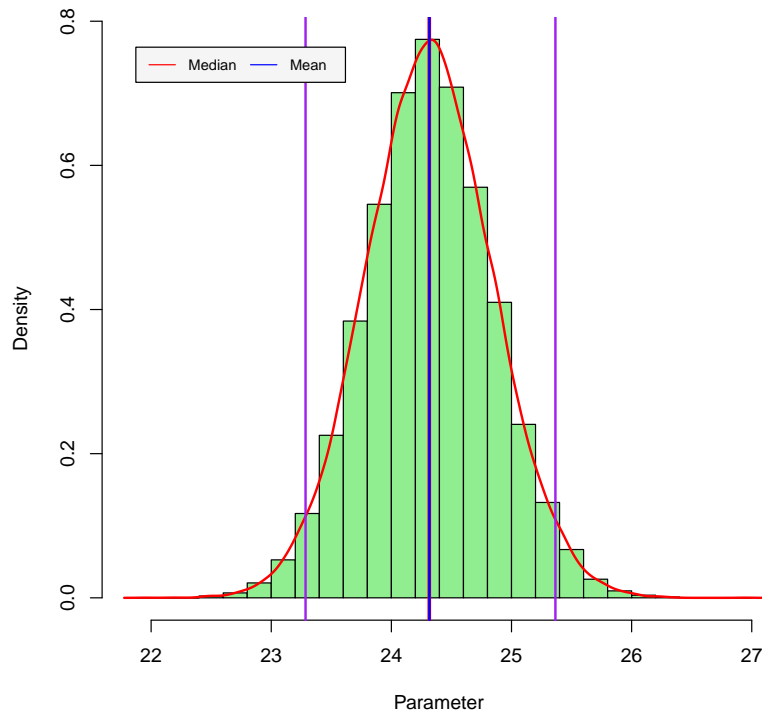


Figure 12: Posterior distribution: Age to maximize market value of soccer players

test, indicating that it seems that the posterior draws come from stationary distributions. See Section 4 for details regarding these tests.

Let’s assume that we only have the market value of soccer players whose value is greater than €1,000,000, which means that approximately 21.5% of our sample is censored. Estimating a normal-inverse gamma model without taking into account the censoring issue would mean inconsistent parameter estimates. For instance, we estimated a normal-inverse gamma model having as dependent variable $\log(\text{ValueCens})$, which is the censored dependent variable, using the same setting as the baseline framework. We found that age, squared age, national team, goals and experience are potentially relevant variables for predicting the market value, but this exercise suggests that squared experience is not relevant, a variable that was relevant in our previous estimation without censoring issues. Therefore, we estimated a Tobit model where $\log(\text{ValueCens})$ is the dependent variable, which is left censored at $\log(1,000,000) \approx 13.82$, with the same MCMC setting and hyperparameters as the baseline estimation. All convergence diagnostics seem good, and we got the same potentially relevant variables as in the baseline estimation, except for squared experience.

Now let’s check if the marginal effects of the regressors are not constant over the support of the dependent variable. For instance, we want to check if the marginal effect of goals varies with the market value of the soccer player. So, we can estimate a Bayesian quantile

regression. In particular, we estimated models at the 0.1, 0.5 (median) and 0.9 quantiles. We found that age, squared age, and national team are potentially relevant regressors to explain these quantiles. For instance, the age that maximizes the market value is approximately 24.5 years in all these three quantiles. However, goals is only relevant when we estimated the median model, which in general has better convergence diagnostics and narrower credible intervals. Observe that experience is not relevant in quantile regressions, whereas this variable is relevant in mean regressions.

Lastly, we carried out a Bayesian bootstrap, which means that we did not assume any particular distribution for the dependent variable. In particular, we set 20,000 iterations with a resample size equal to 1,000 (see subsection 5.4 for technical details). We used the same specification as in the normal-inverse gamma model.

The results show the posterior mean estimates, the highest posterior density credible intervals at 95%, and some percentiles that can be used to obtain the 95% symmetric credible interval. It seems that age, squared age, national team, goals, experience, and squared experience are statistically significant variables to explain the market value of a soccer player. Observe that these variables were also relevant in the normal-inverse gamma model. For instance, the highest density and symmetric credible intervals for national team are the same, (0.61, 1.08). This is also similar to the 95% credible interval using the normal-inverse gamma model. All convergence statistics seem good, which suggests that the posterior draws come from stationary distributions.

Binary response: Determinants of hospitalization in Medellín

We use the dataset named *2HealthMed.csv*, which was provided by [Ramírez Hassan et al. \(2013\)](#). Our dependent variable is a binary indicator with a value equal to 1 if an individual was hospitalized in 2007, and 0 otherwise.

The equation to enter in the main equation box is

$$\text{Hosp} \sim \text{SHI} + \text{Female} + \text{Age} + \text{Age}^2 + \text{Est}2 + \text{Est}3 + \text{Fair} + \text{Good} + \text{Excellent},$$

where SHI is a binary variable equal to 1 if the individual is in a subsidized health care program and 0 otherwise, Female is an indicator of gender, Age in years, Age² is squared age, Est2 and Est3 are indicators of socio-economic status, the reference is Est1, which is the lowest, and self perception of health status where bad is the reference.

We ran this application using a logit model with 30,000 MCMC iterations plus a burn-in equal to 10,000, a thinning parameter equal to 5, and a tuning parameter for the Metropolis–Hastings algorithm equal to 1.01. This implies an effective sample size equal to 6,000. It seems from our results that female and health status are relevant variables for hospitalization, as their 95% credible intervals do not cross 0. Women have a higher probability of being hospitalized than do men, and people with bad self perception of health condition also have a higher probability of being hospitalized. Observe that we can use the posterior chains, which can be downloaded from our GUI, to obtain the posterior distributions of the marginal effects without extra computational burden.

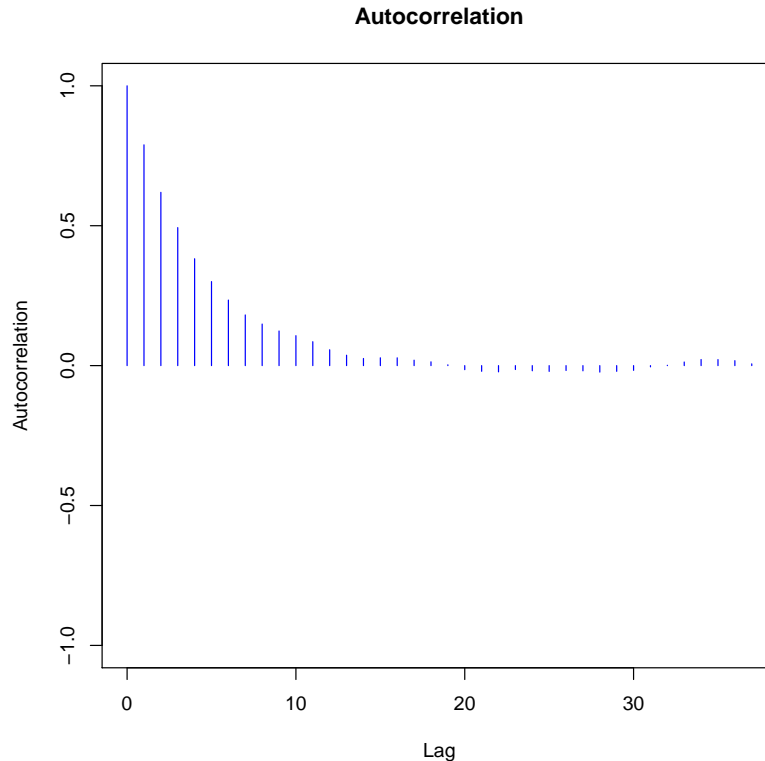


Figure 13: Autocorrelation plot: Female, logit with MCMC iterations 30,000

Regarding the convergence diagnostics, we observe that the posterior chains passed the Geweke (1992) test, most of them passed the Heidelberger and Welch (1983) test, but there is a high level of autocorrelation as the dependence factors are around 30. This also can be seen in Figure 13, which is automatically generated by our GUI, where we have the autocorrelation plot of Female.

We can increase the number of MCMC iterations and the thinning parameter, or play around with the tuning parameter, to decrease the level of autocorrelation. For instance, we ran again this application with the same settings, except that we ran 100,000 iterations with a thinning parameter equal to 50. This setting decreases enormously the level of autocorrelation.

We also carried out this application using the probit model with the baseline setting of the logit model. We got the same results regarding potentially relevant predictors.⁴ However, the probit model does not require a tuning parameter in its MCMC algorithm, which in turn generates less autocorrelated chains, as can be seen in Figure 14.

Multinomial response: Choice of fishing mode

⁴Remember that in this model our GUI displays the posterior results according to the order in the equation.

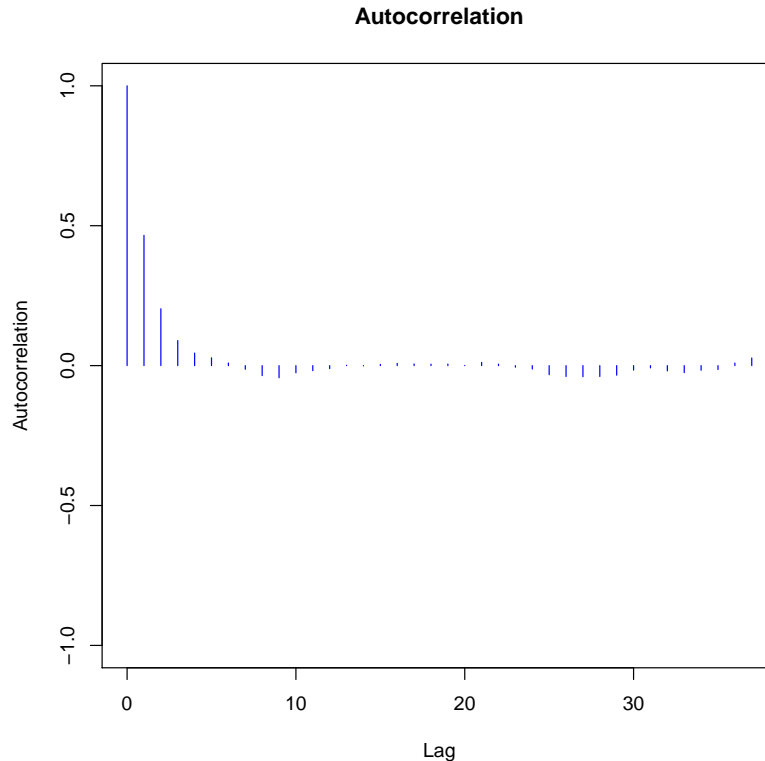


Figure 14: Autocorrelation plot: Female, probit with MCMC iterations 30,000

We used in these applications the dataset *3Fishing.csv* from [Cameron and Trivedi \(2005, p. 491\)](#). The dependent variable is mutually exclusive alternatives regarding fishing modes (mode), where beach is equal to 1, pier is equal to 2, private boat is equal to 3, and chartered boat (baseline alternative) is equal to 4. The specification to enter in the main equation box is

$$\begin{aligned} \text{mode} \sim & \text{price.beach} + \text{price.pier} + \text{price.boat} + \text{price.charter} \\ & + \text{catch.beach} + \text{catch.pier} + \text{catch.boat} + \text{catch.charter} + \text{income}, \end{aligned}$$

where `price.beach`, `price.pier`, `price.boat` and `price.charter` are the prices for each of the alternative fishing modes; `catch.beach`, `catch.pier`, `catch.boat` and `catch.charter` are the catch rates for each mode, and `income` is the individual's income.

We estimated a multinomial probit model where chartered boat is the base category, the number of choice categories is four, there are two alternative-specific regressors (price and catch rate), and one non alternative-specific regressor (income). This setting involves the estimation of eight location parameters: three intercepts, three for income, one for price, and one for catch rate. This is the order of the posterior chains. Note that the location coefficients are set equal to 0 for the baseline category. For multinomial models, we strongly recommend using the last category as the baseline (see subsection 5.1 for details). The multinomial probit

model also gives posterior estimates for a 3×3 covariance matrix (four alternatives minus one), where the element (1,1) is equal to 1 due to identification issues, and elements 2 and 4 are the same, as well as 3 and 7, and 6 and 8, due to symmetry.⁵ Observe that this identification restriction implies NaN values in Geweke (1992) and Heidelberg and Welch (1983) tests for element (1,1) of the covariance matrix, and just eight dependence factors associated with the remaining elements of the covariance matrix.

We ran 100,000 MCMC iterations plus 10,000 as burn-in with a thinning parameter equal to 5, where all priors use default values for the hyperparameters. We found that the 95% credible intervals of the coefficient associated with income for beach and private boat alternatives are equal to (1.31e-05, 9.284e-05) and (4.05e-05, 1.50e-04). This suggests that the probability of choosing these alternatives increases compared to a chartered boat when income increases. In addition, an increase in the price and/or a decrease in the catch rate for specific fishing alternatives imply lower probabilities of choosing them as the 95% credible intervals are (-1.00e-02, -4.90e-01) and (1.50e-01, 4.89e-01), respectively.

We also estimated this specification using the multinomial logit model. We got the warning message “Error: missing value where TRUE/FALSE needed” in many tries. However, we can download the posterior chains and graphs, and observed that the mixing properties of the posterior draws are very bad, that is, we rejected the null hypothesis that the posterior draws come from stationary distributions. In general, we found that the mixing properties of these models (multinomial probit and logit) were bad in this application. It should also be taken into account that when warning messages are displayed in our GUI, there is a high chance that there are convergence issues of the posterior chains. So, the results are not trustworthy.

Ordered and count response: Determinants of preventive health care in Medellín

We used the file named *2HealthMed.csv* in these applications. First, we used an ordered probit model where the dependent variable is *MedVisPrevOr*, which is an ordered variable equal to 1 if the individual did not visit a physician in 2007 for preventive reasons, 2 if the individual visited once in that year, and so on, until it is equal to 6 for visiting five or more times. The latter category is 1.6% of the sample. Observe that the dependent variable has six categories. The equation to enter in the main equation box is

$$\text{MedVisPrevOr} \sim \text{SHI} + \text{Female} + \text{FemaleAge} + \text{Age} + \text{Age2} + \text{Est2} + \text{Est3} + \text{Fair} + \text{Good} + \text{Excellent} \\ + \text{PriEd} + \text{HighEd} + \text{VocEd} + \text{UnivEd} - 1,$$

where *FemaleAge* is the interaction between female and age, and the last four variables are indicators of educational achievement: primary, high school, vocational, and university, the reference is not obtaining any degree. All other variables were described previously in the binary response applications. Observe the -1 at the end of the specification. This is because ordered probit models do not identify the intercept, which is implicit in the cutoffs (see subsection 5.1).

⁵Remember that this is the order in the pdf, eps and csv files that can be downloaded from our GUI.

We ran this application with 50,000 MCMC iterations plus 10,000 as burn-in, and thinning parameter equal to 5. This setting means 10,000 effective posterior draws. We used the default settings for the hyperparameters of the prior and the tuning parameter for the Metropolis–Hastings algorithm.

The results suggest that young women and older individuals (at decreasing rate) in the subsidized health program, characterized in the second socioeconomic status with increasing self perception of health condition, and not having high school as their highest education degree, have a higher probability of visiting a physician for preventive health aims.

However, there are some convergence issues with the posterior chains of the health self perception category. This can be seen in the [Geweke \(1992\)](#) test due to these estimates having statistical tests greater than 1.96 in magnitude. These values implies rejection of the null hypothesis of equal means between the first 10% posterior draws and the last 50% posterior draws. Dependence factors are close to 5, except for excellent health self perception. The [Heidelberger and Welch \(1983\)](#) test does not give evidence to reject the null hypothesis of draws from stationary chains.

We also got the posterior estimates of the cutoffs in the ordered probit model. These estimates are necessary to calculate the probability that an individual is in a specific category of visiting physicians. Due to identification restrictions, the first cutoff is set equal to 0. That is why we have NaN values in [Geweke \(1992\)](#) and [Heidelberger and Welch \(1983\)](#) tests, and we observe only four values in the [Raftery \(1993\)](#) test, which correspond to the remaining free cutoffs. It seems that these cutoff estimates have some convergence issues when taking as diagnostic tool the [Geweke \(1992\)](#) test. Their dependence factors are also very high.

We ran this application again using 100,000 MCMC iterations plus a burn-in of 10,000 and a thinning parameter equal to 10. With this setting, the convergence properties of all the posterior draws were improved. We got similar results regarding the posterior descriptive statistics. In addition, we can use the posterior chains to calculate the posterior distribution of any function of the parameter estimates, such as the marginal effects.

Now we estimate a negative binomial model where the dependent variable (MedVisPrev) is the number of preventive health care visits to physicians in 2007. The equation to enter into the main equation box is

$$\text{MedVisPrev} \sim \text{SHI} + \text{Female} + \text{Age} + \text{Est2} + \text{Est3} + \text{Fair} + \text{Good} + \text{Excellent} + \text{PriEd} + \text{HighEd} + \text{VocEd} + \text{UnivEd},$$

We initially included in our specification the squared age; however, the algorithm had issues with infinite values. This can be an issue with highly collinear regressors.

We ran this application using 20,000 MCMC iterations, a burn-in equal to 5,000, and a thinning parameter equal to 5. We set the default priors, that is, mean equal to 0, an identity matrix as covariance matrix for the location parameters, and shape and scale hyperparameters equal to 0.001 for the the parameter associated with the number of failures until the experiment stopped in the negative binomial distribution. The tuning parameters of the

Metropolis–Hastings algorithm were also the defaults.

The results suggest that individuals characterized in stratum 3 visit physicians less frequently for preventive purposes than individuals in other strata. The 95% credible interval for Est3 is (-0.15, -0.06). On the other hand, individuals whose maximum level of education is primary and vocational visit physicians more frequently than individuals without any educational achievement. The median estimates for these coefficients are 0.08 and 0.24, respectively. Observe that these coefficients are the semi-elasticities of the odds ratios, success vs. failure, which in turn is associated with the mean value of the negative binomial variable.⁶ So, individuals whose maximum level of education is primary visit physicians on average 8% more than individuals without any education achievement.

In general, the posterior chains do not reject the null hypothesis of stationarity, as the values of the Geweke (1992) tests are lower than 1.96 in magnitude, and the p -values of Heidelberger and Welch (1983) are greater than 0.05. However, the dependence factor is greater than 30, which means a high level of autocorrelation of the posterior chains, as can be seen from the autocorrelation plots (not displayed here, but generated automatically using our GUI).

We carried out the same exercise except that the prior mean vector was equal to 0.5, and the covariance matrix to $diag\{1000\}$, which means a priori a positive effect of the regressors, but more uncertainty regarding the location of the prior mean. We got similar results as with the baseline exercise. This is because with an increasing sample size, the posterior estimates are consistent, and the prior information has less weight on the posterior results. The sample size is 12,975 in this dataset.

3.2 Multivariate models

Continuous responses: The effect of institutions on per capita GDP

To illustrate the potential of our GUI to estimate multivariate models, we used the dataset provided by Acemoglu et al. (2001), who analyzed the effect of property rights on economic growth.

First of all, we used the dataset *5Institutions.csv* to estimate the following set of equations:

$$\log(\text{pcGDP95}_i) = \pi_0 + \pi_1 \log(\text{Mort}_i) + \pi_2 \text{Africa} + \pi_3 \text{Asia} + \pi_4 \text{Other} + e_{1i}, \quad (1)$$

$$\text{PAER}_i = \gamma_0 + \gamma_1 \log(\text{Mort}_i) + e_{2i}, \quad (2)$$

where pcGDP95, PAER and Mort are the per capita GDP in 1995, the average index of protection against expropriation between 1985 and 1995, and the settler mortality rate during

⁶The mean of a negative binomial random variable is $\frac{\theta_i \gamma}{1 - \theta_i}$, where θ_i is the success probability and γ is the number of failures until the experiment stopped. Following the definitions of subsection 5.1 regarding negative binomial models, and after some simple manipulations, we have that $\log \mathbb{E}[Y_i | \mathbf{x}_i] = \mathbf{x}'_i \beta$.

the time of colonization. Africa, Asia and Other are dummies for continents, with America as the baseline group.

As there are different sets of regressors in each equation, and we suspect there is correlation between the stochastic errors of these two equations, we should estimate a seemingly unrelated regressions (SUR) model.

We should take into account that there are two equations: the first one has five regressors, including the intercept, and the second equation has two regressors (intercept plus the mortality rate). We used default values for the hyperparameters, this implies “vague” prior information, and hence an “objective” Bayesian approach.

We set 10,000 MCMC iterations plus 1,000 burn-in iterations, and a thinning parameter equal to 1. It seems that this setting gives posterior chains that converge to stationary distributions. All stationary tests do not reject the null hypothesis of “stationarity,” and the mixing properties look good (dependence factors close to 1, autocorrelation and trace plots seem to indicate no autocorrelation).

The most important parameters are the effect of the mortality rate on gross domestic product and property rights. Their 95% credible intervals are (-0.67, -0.29) and (-0.85, -0.35), respectively (second and seventh parameters). This suggests that the settler mortality rate during the time of colonization is negatively associated with economic growth and property rights. In addition, the 95% credible interval of the covariance between the stochastic errors of these two equations is (0.33, 0.88), which suggests that there is statistically significant evidence of correlation between the equations.

The previous set of equations can be considered as a restricted reduced form system, where the coefficients of the continents are set equal to 0 in the property rights equation. We can think in the following system of structural equations as producing the previous, but unrestricted, reduced form system,

$$\log(\text{pcGDP95}_i) = \beta_0 + \beta_1 \text{PAER}_i + \beta_2 \text{Africa} + \beta_3 \text{Asia} + \beta_4 \text{Other} + u_{1i}, \quad (3)$$

$$\text{PAER}_i = \alpha_0 + \alpha_1 \log(\text{pcGDP95}_i) + \alpha_2 \log(\text{Mort}_i) + u_{2i}. \quad (4)$$

We used the file *4Institutions.csv*, which has the structure to estimate multivariate Bayesian regressions using our GUI, to identify the causal effect of property rights on per capita GDP. In particular, we use the same MCMC and hyperparameters setting as in the previous exercise to obtain the posterior estimates of the reduced system without imposing zero restrictions of the effect of continents on property rights. The structural parameter β_1 is equal to π_1/γ_1 .⁷ We used the posterior draws automatically generated by our GUI to obtain the posterior

⁷Substituting Equation 4 into Equation 3 and comparing with Equation 1 yields $\pi_1 = \frac{\beta_1 \alpha_2}{1 - \beta_1 \alpha_1}$. Solving for the PAER as a function of the exogenous regressors in the structural system, and comparing with Equation 2, yields $\gamma_1 = \frac{\alpha_2}{1 - \beta_1 \alpha_1}$. Observe one needs independent equations ($\beta_1 \alpha_1 \neq 1$), and the exclusion restriction ($\alpha_2 \neq 0$).

chain of this structural parameter, which are the causal effects that [Acemoglu et al. \(2001\)](#) wanted to identify. The 95% credible interval is (0.56, 2.93), the posterior mean value is 1.12, and the median value is 0.98. If we estimate a multivariate system without taking into account the dummy variables associated with the continents, the causal effect has a 95% credible interval (0.68, 1.43) with posterior mean and median values equal to 0.94 and 0.97, respectively. Observe that the length of the second interval is shorter than the first. This is because the dummy variables of the continents are not statistically relevant for the property rights equation. As a consequence, the former estimation is less efficient.

Observe that we also obtain the posterior draws of the covariance matrix of these two reduced form equations from our GUI. All the convergence diagnostics indicate that the posterior draws (location and scale parameters) seem to come from stationary distributions.

Another way to identify the causal effect of property rights on per capita GDP is using instrumental variables. Therefore, we used the file *6Institutions.csv* to estimate Equation 3 using the mortality rate as an instrument for property rights. The equation to enter in the main equation box is

$$\text{logpcGDP95} \sim \text{PAER} + \text{Africa} + \text{Asia} + \text{Other},$$

and the equation to enter in the instrumental equation box is

$$\text{PAER} \sim \text{logMort}.$$

We used 20,000 MCMC iterations plus a burn-in equal to 5,000, and a thinning parameter equal to 5. So, the effective length of the posterior draws is 4,000. Using the default hyperparameters, the 95% credible interval of the coefficient associated with the endogenous variable, which are the first to be displayed in our descriptive and diagnostic statistics, is (0.55, 1.21), and the mean value is equal to 0.82. So, this is the effect of property rights on per capita GDP. Our GUI display next the posterior results associated with the instrumental equation, there we obtained a 95% credible interval equal to (-0.83, -0.35) for the effect of the mortality rate on the property rights. This suggests that the instrument is not weak. Then, we obtained the posterior results for the exogenous regressors in the main equation, which suggest that Africa and Asia dummies variables have negative effects on per capita GDP. Finally, we got the posterior estimates for the covariance matrix, which suggest that there is a negative covariance between the GDP equation and PAER equation, the 95% credible interval is (-1.50, -0.26).

All posterior draws seem to come from stationary distributions. However, there are high levels of autocorrelation in some posterior chains, as suggested by the dependence factors and posterior plots.

Binary responses: Self selection in hospitalization due to a subsidized health care program in Medellín

We finish this subsection using the dataset *7HealthMed.csv* where the dependent variable is equal to $y = [\text{Hosp SHI}]'$ where Hosp is equal to 1 if an individual was hospitalized in 2007, 0 otherwise, and SHI is equal to 1 if the individual had subsidized health insurance that year,

and 0 otherwise. Recall that our application in binary response models was to uncover the determinants of hospitalization in Medellín (Colombia), where one of the regressors was a binary indicator of being in a subsidized health care program. We can use a bivariate probit model if we suspect there is a dependence regarding the decisions involving these two variables. We would expect a priori that being in a subsidized health care program would imply a higher probability of being hospitalized *ceteris paribus*. However, if an individual expects to be hospitalized in the future, and the factors that drive this decision are unobserved to the econometrician, we would have a feedback effect from being hospitalized on being in a subsidized health care program.

We took into account 9 regressors: a constant, female, age, squared age, self perception of health status taking as reference bad (four categories), and the proportion of the individual's age spent living in her/his neighborhood (PTL). The last variable tries to take into account the social capital that can affect being in the subsidized health insurance program, as the target population is identified by the local government (?). We have 12975 individuals chosen two options (subsidized regime and hospitalization).

We set 20,000 MCMC iterations plus 1,000 iterations as burn-in, and a thinning parameter equal to 5. This implies an effective length of the posterior chains equal to 4,000 draws. We also used default values for the hyperparameters of the prior distributions. In general, the convergence diagnostics seem good, except that there is a high level of autocorrelation for the posterior chain of the correlation between the two equations, as indicated by the dependence factors, and the trace and autocorrelation plots. Observe that the tests of Geweke (1992) and Heidelberger and Welch (1983) have NaN values for the elements (1, 1) and (2, 2) of the covariance matrix, as these parameters were set equal to 1 due to identification restrictions. This also means just two values for the dependence factors, which are actually the same due to symmetry.

The results suggest that only female is relevant to explain hospitalization. The 95% credible interval is (3.13e-02, 0.22). Observe that only 3.11% of the sample has been hospitalized. Probit models are not well designed for this kind of dataset, but our main purpose is to illustrate the use of our GUI. On the other hand, the results suggest that age, squared age, and the proportion of age spent living in the neighborhood are statistically relevant to explain enrollment in the subsidized program. Their 95% credible intervals are (2.63e-01, 3.27e-01), (-8.05e-03, -2.05e-03) and (0.15, 0.27), respectively. The latter result seems to support the social capital hypothesis. Lastly, the 95% credible interval for the correlation between the two binary equations is (-0.07, 0.06), suggesting that there is no self selection regarding these two decisions (hospitalization and subsidized insurance). So, it seems that it is better to estimate univariate binary models for each of these dependent variables, for the sake of parsimony.

3.3 Hierarchical longitudinal models

Normal model: The relation between productivity and public investment

We used the dataset named *8PublicCap.csv* used by Ramírez Hassan (2017) to analyze the relation between public investment and gross state product in the setting of a spatial panel dataset consisting of 48 US states from 1970 to 1986. In particular, the specification to type

into the main equation box of fixed effects is

$$\log(\text{gsp}) \sim \log(\text{pcap}) + \log(\text{pc}) + \log(\text{emp}) + \text{unemp},$$

where `gsp` is the gross state product, `pcap` is public capital, and `pc` is private capital all in US\$, `emp` is employment (people), and `unemp` is the unemployment rate in percentage.

We left empty the main equation box of random effects as we assumed that the unobserved heterogeneity is not associated with any particular regressors. This means that we control for the unobserved heterogeneity using just the constant terms. The variable which identifies the units is `id`.

We ran this application using 10,000 MCMC iterations plus a burn-in equal to 5,000 iterations, and a thinning parameter equal to 1. We also used the default values for the hyperparameters of the prior distributions. It seems that all posterior draws come from stationary distributions, as suggested by the diagnostics and posterior plots.

The 95% symmetric credible intervals for public capital, private capital, employment, and unemployment, are (-2.54e-02, -2.06e-02), (2.92e-01, 2.96e-01), (7.62e-01, 7.67e-01) and (-5.47e-03, -5.31e-03), respectively. The posterior mean elasticity estimate of public capital to `gsp` is -0.023, that is, an increase by 1% in public capital means a 0.023% decrease in gross state product. The posterior mean estimates of private capital and employment elasticities are 0.294 and 0.765, respectively. In addition, a 1% increase in the unemployment rate means a decrease of 0.54% in `gsp`. It seems that all these variables are statistically relevant. In addition, the posterior mean estimates of the variance associated with the unobserved heterogeneity and stochastic errors are 1.06e-01 and 1.45e-03. We obtained the posterior chain of the proportion of the variance associated with the unobserved heterogeneity (see Figure 15). The 95% symmetric credible interval is (0.98, 0.99) for this proportion, that is, unobserved heterogeneity is very important to explain the total variability.

Binary and count models: The effects of health care reforms and doctor visits in Germany

We used the dataset *9VisitDoc.csv* provided by Winkelmann (2004) (see <http://qed.econ.queensu.ca/jae/2004-v19.4/winkelmann/> for details). First of all, we analyzed the determinants of a binary variable (`DocVis`) which is equal to 1 if an individual visited a physician in the last three months, and 0 otherwise. The dataset contains 32,837 observations of 9,197 individuals in an unbalanced panel over the years 1995–1999 from the German Socioeconomic Panel Data.

The specification to type into the main equation box of the fixed effects is

$$\text{DocVis} \sim \text{Age} + \text{Male} + \text{Educ} + \text{Married} + \text{Sozh} + \text{LogInc},$$

where this specification controls for age, a gender indicator, years of schooling, an indicator of marital status, whether the individual receives welfare payments, and the logarithm of monthly gross income.

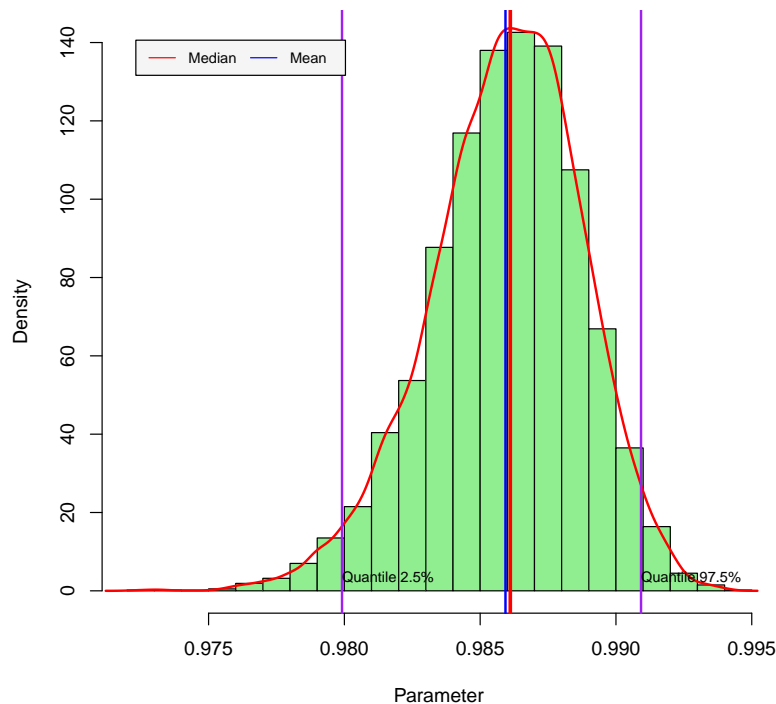


Figure 15: Posterior distribution: Proportion of variance associated with unobserved heterogeneity

We assumed that the unobserved heterogeneity is associated with self perception of health status. So, we type \sim GoodHealth + BadHealth into the main equation box of random effects in the hierarchical logit model. In addition, we typed id in the name of the grouping variable, as this identifies the individuals in our dataset. We set 10,000 MCMC iterations plus 1,000 burn-in, and a thinning parameter equal to 1. We also set all hyperparameters to their default values. Bayesian inference in this model takes several minutes.

The results suggest that age and income increase the probability of visiting the physician, the posterior estimates have 95% symmetric credible intervals equal to $(2.25e-02, 2.93e-02)$ and $(8.05e-02, 2.48e-01)$, whereas men have a lower probability of visiting a physician, the 95% credible interval is $(-1.02, -0.93)$. The posterior mean of the variances of the unobserved heterogeneity associated with the bad and good health status are 0.51 and 0.88, and the variance of the unobserved heterogeneity not associated with these variables is 1.81. The posterior mean of the variance associated with the stochastic error is 0.14.

We ran the previous setting taking as dependent variable the number of doctor visits in the last three months (DocNum), so we used the hierarchical Poisson model. The results suggest that on average older individuals visit physicians more frequently; the 95% credible interval is $(9.87e-03, 1.31e-02)$. In particular, one more year of age increases by 1.15% the number of

visits, on average. On the other hand, men have, on average, 39.7% ($\exp(-0.506) - 1$) fewer visits to physicians than do women. The 95% credible interval is (-0.54, -0.47). It seems that income is not a relevant regressor in this setting, the 95% credible interval is (-0.01, 0.06). We found in this application that the covariance between the unobserved heterogeneity between self perceptions of health status is negative; the 95% credible interval is (-0.34, -0.18) with a posterior median equal to -0.26.

3.4 Bayesian model averaging

Continuous response: Determinants of export diversification

We used the dataset provided by [Jetter and Ramírez Hassan \(2015\)](#) to analyze the determinants of export diversification. The dataset named *10ExportDiversificationHHI.csv* contains information about 36 potential determinants of export diversification measured using the Herfindahl–Hirschman Index (avghhi) for 104 countries (see [Jetter and Ramírez Hassan \(2015\)](#) for details). This setting implies 68.7 billion models (2^{36}).

We implemented three Bayesian model average (BMA) strategies: Bayesian information criterion approximation (BIC), Markov chain Monte Carlo model composition (MC³), and instrumental variable (IVBMA). The former takes into account possible endogeneity between export diversification and gross domestic product.

Regarding BMA using the BIC approximation, we set 50 (default value) for OR. This parameter defines the number of best models to take into account in our BMA strategy (see subsection 5.5 for details). We obtained a table where we can see the posterior inclusion probability (PIP), expected value, standard deviation, and posterior mean estimates associated with the best models for each variable. The best models are defined using posterior model probabilities, which appear at the bottom of the table, where we also see the number of variables associated with each model as well as the coefficients of determination and BIC values. Our GUI also produces two csv files. The first one is *Best Models.csv*, where we have by row the best models, and the variables by columns, a 1 indicates the presence of the specific variable in the model’s specification, and a 0 its absence. The last column in this file is the posterior model probability. The second one is *Descriptive Statistics.csv*, where we see the posterior inclusion probability, expected value, and standard deviation of each variable.

Following [Kass and Raftery \(1995\)](#)’s suggestions, we found that there is very strong evidence that being a former colony of Portugal, the total net primary enrollment, and the total natural resources rents as percentage of GDP are determinants of export diversification. Their expected values are 0.15, -0.006 and 0.008, respectively, which means that there are negative effects of having been a colony of Portugal and of having natural resources on export diversification. Recall that higher values of HHI indicate less diversification.

We also ran this application using the MC3 strategy with 10,000 MCMC iterations. We got results similar to those with the BIC approximation.

We estimated an instrumental variable BMA to take into account possible endogeneity between export diversification and GDP using 20,000 MCMC iterations plus a burn-in equal to 5,000, where there is one endogenous variable (GDP). In particular, we used the files *11ExportDiversificationHHI.csv* and *12ExportDiversificationHHIInstr.csv*. The first file has the dependent variable in the first column (avhhi) followed by the endogenous variable (avlgdpcap), the constant term (a column of 1's) and exogenous regressors. The second file has the instrumental variables, which are geographical, cultural, and colonial factors.

Our GUI displays first the outcomes of the second stage equation (main equation), and then, the first stage equation (instrumental equation). We can download three csv files: *BMA Results First Stage.csv*, *BMA Results Second Stage.csv* and *Posterior chains.csv*. The first two files have the same structure: posterior inclusion probabilities and expected values. We can see from these files that educational levels and governance performance are the most important variables to foster gross domestic product (PIP=100), primary enrollment fosters export diversification (PIP=79.5), whereas natural resources discourages it (PIP=97.1). The latter file has the posterior draws where the name *beta* is associated with the variables in the main equation (second stage), and *gamma* is associated with the instrumental variable equation (first stage). Lastly, we have the posterior draws of the covariance matrix of the stochastic errors in the first and second stage equations. We can see in Figure 16 this posterior density, which has a 95% symmetric credible interval equal to (-0.014, 0.024), suggesting that there are no endogeneity issues.

Binary response: Determinants of Internet adoption in Medellín

We used the file *13InternetMed.csv* provided by Ramírez-Hassan (2019). The dependent variable is an indicator of Internet adoption (internet) for 5,000 households in Medellín (Colombia) during the period 2006–2014. This dataset contains information about 18 potential determinants. This means 262,144 (2^{18}) potential models just taking into account variable uncertainty.

The results displayed in our GUI are the posterior inclusion probabilities, posterior means and standard deviations, as well as the posterior means of the parameters associated with the best models, which are found using posterior model probabilities. We also generated the files *Best Models.csv* and *Descriptive Statistics.csv*. Using the default values for the Occam's window parameters, we observed that the best model has a posterior model probability (PMP) equal to 31%, and the second best model has a PMP equal to 30%. It seems that age, squared age, years of education of the head of the household, total expenses, having pay TV, any household member studying, and number of children in the household are statistically relevant determinants of Internet adoption. For instance, taking into account that we have the logarithm of household expenses as regressor, we obtained an odds ratio to income elasticity equal to 1.48, that is, a 1% income increase implies a 1.48% odds ratio increase. This parameter estimate takes into account model (variable) uncertainty.

Continuous positive response: The market for soccer players in Europe

We used the dataset *14ValueFootballPlayers.csv* where there are 26 potential determinants of the market value of a stratified sample of 335 soccer players in the five most impor-

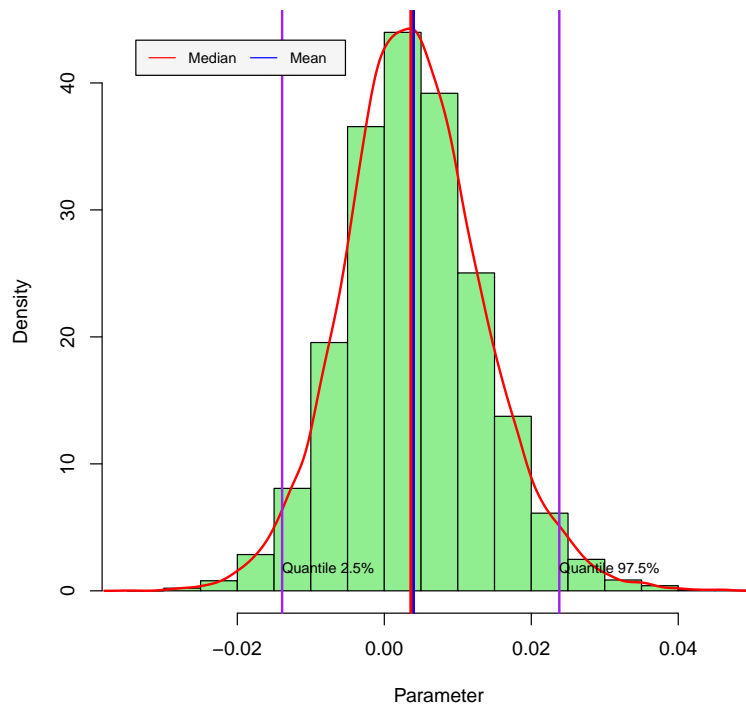


Figure 16: Posterior distribution: Covariance between HHI and GDP in instrumental variable Bayesian model average

tant leagues in Europe (see [Serna Rodríguez et al. \(2018\)](#) for details). This setting implies 67,108,864 (2^{26}) potential models, which are analyzed using the BIC approximation of BMA assuming a gamma distribution, and setting default values for Occam’s window.

The results suggest that performance, age at decreasing rate, having participated in senior and under-21 national team, scored goals, participating in the UEFA champions league in the 2015–2016 season, and team experience, increase market value. The best ten models add 33% of the posterior model probabilities. In addition, we used the same dataset to carry out BMA using the BIC approximation assuming a normal distribution, and a BMA using MC3, obtaining the same relevant variables. The results of estimating a generalized linear model with the gamma family using this dataset suggest that there are other relevant variables, such as experience, squared experience, conmebol and premier league indicators, public attendance, and change of team in the last season, which are also statistically significant variables at 5%. However, the latter procedure does not take into account model (variable) uncertainty. As a consequence, there is an underestimation of the standard errors, which implies an overestimation of the statistical tests to reject the null hypothesis of no relevant regressor.

Count response: Determinants of number of children

We used the dataset *15Fertile2.csv* from ([Wooldridge, 2012](#), p. 547) which has informa-

tion about 1,781 women from Botswana in 1988 (for details, see <https://rdr.io/cran/wooldridge/man/fertil2.html>, and take into account that we deleted some variables and omitted observations with NA values). In this case our dependent variable is the number of children ever born (ceb), which is a count variable, as a function of 19 potential determinants. We found that the best model, for which the PMP is equal to 27%, has as regressors age, squared age, age at first birth, use birth control, husband’s years of education, woman’s years of education, and living in an urban area. The first five variables have PIPs equal to 100. For instance, we found that women using birth control have approximately 15% fewer children on average than women who are not using birth control. The results of estimating a generalized linear model with the Poisson family using this dataset suggest that having a bicycle is a relevant determinant of the number of children, at the 5% significance level. However, this variable has a PIP equal to 19.3%.

4 Bayesian framework: A brief summary of theory and computation

The theoretical point of departure is the Bayes rule, which formally establishes how prior beliefs are updated with new information (data),

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where $\pi(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution, $\pi(\boldsymbol{\theta})$ is the prior, $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood, and $p(\mathbf{y}) = \int_{\Theta} p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal likelihood.

The Bayesian framework allows reporting the full posterior distribution. In case only one value of the posterior distribution should be reported, decision theory offers an elegant framework, for instance, a quadratic loss function implies reporting the posterior mean, whereas an absolute value loss function implies reporting the posterior median. Another way to summarize a posterior distribution is through intervals (univariate) or regions (multivariate). For instance, we can report the $(1-\alpha)\%$ symmetric credible interval, where each tail of the distribution contains $(\alpha/2)\%$ of the probability, or the highest posterior density credible interval, $\Theta_{\alpha} = \{\boldsymbol{\theta} \in \Theta : \pi(\boldsymbol{\theta}|\mathbf{y}) \geq k(\alpha)\}$, $k(\alpha)$ is the largest constant such that $p(\Theta_{\alpha}|\mathbf{y}) \geq 1 - \alpha$.

There are other appealing characteristics of this statistical framework, namely: there is a solid probability theory framework for hypothesis testing based on posterior odds (H_1 vs H_2), $PO_{12} = \frac{p(H_1|\mathbf{y})}{p(H_2|\mathbf{y})} = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_2)} \times \frac{\pi(H_1)}{\pi(H_2)}$, the first term is called the Bayes factor, and the second term is the prior odds, $\pi(H_j)$ is the prior probability of hypothesis H_j , and $p(\mathbf{y}|H_j) = \int_{\Theta_j} p(\mathbf{y}|\boldsymbol{\theta}_j, H_j)\pi(\boldsymbol{\theta}_j|H_j)d\boldsymbol{\theta}_j$ (Kass and Raftery, 1995). With this framework it is also easy to take into account model uncertainty based on posterior model probabilities, $p(M_j|\mathbf{y}) = \frac{\pi(M_j)p(\mathbf{y}|M_j)}{\sum_{m=1}^M \pi(M_m)p(\mathbf{y}|M_m)}$, where $\pi(M_j)$ is the prior model probability, and $p(\mathbf{y}|M_j) = \int_{\Theta_j} p(\mathbf{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$ is the marginal likelihood under model j , $j = 1, 2, \dots, M$. In addition, it is also easy to carry out an inference of nonlinear functions of parameter estimates without requiring extra computational effort (re-sampling techniques) or asymptotic

results (plug-in approach based on delta method). In a Bayesian setting, we can have parameters that are not identified, but carry out an inference of functions of these parameters which are well identified. Finally, predictive distributions acknowledge parameter uncertainty, $\pi(\mathbf{y}^{New}|\mathbf{y}) = \int_{\Theta} f(\mathbf{y}^{New}|\mathbf{y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \approx \frac{1}{S} \sum_{s=1}^S f(\mathbf{y}^{New}|\mathbf{y}, \boldsymbol{\theta}^{(s)})$, $\boldsymbol{\theta}^{(s)} \sim \pi(\boldsymbol{\theta}|\mathbf{y})$.

The main issue regarding estimation in the Bayesian approach is how to obtain draws from the posterior distribution when this distribution is not a standard one. The Metropolis–Hastings (M–H) algorithm is one of the most popular techniques to carry out this task (Metropolis et al., 1953; Hastings, 1970). Given a target distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$, such that $\pi(d\boldsymbol{\theta}^c|\mathbf{y}) = \int_{\Theta} p(\boldsymbol{\theta}, d\boldsymbol{\theta}^c)\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ where $p(\boldsymbol{\theta}, d\boldsymbol{\theta}^c)$ is a conditional distribution function (transition kernel) that represents the probability of moving from $\boldsymbol{\theta}$ to a point in $d\boldsymbol{\theta}^c$, the M–H algorithm establishes

$$p(\boldsymbol{\theta}, d\boldsymbol{\theta}^c) = q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)d\boldsymbol{\theta}^c + \left[1 - \int_{\Theta} q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)d\boldsymbol{\theta}^c\right] \delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}^c),$$

where $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ is a proposed density function to draw candidates ($\boldsymbol{\theta}^c$), $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = \min \left\{ \frac{\pi(\boldsymbol{\theta}^c|\mathbf{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)}{\pi(\boldsymbol{\theta}|\mathbf{y})q(\boldsymbol{\theta}^c, \boldsymbol{\theta})}, 1 \right\} = \min \left\{ \frac{p(\mathbf{y}|\boldsymbol{\theta}^c)\pi(\boldsymbol{\theta}^c)q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)}{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}^c, \boldsymbol{\theta})}, 1 \right\}$ is the probability of moving from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^c$,⁸ $\delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}^c)$ is equal to 1 if $\boldsymbol{\theta} \in d\boldsymbol{\theta}^c$ and 0 otherwise, and $\left[1 - \int_{\Theta} q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)d\boldsymbol{\theta}^c\right]$ is the probability that the chain remains at $\boldsymbol{\theta}$ (Chib and Greenberg, 1995).⁹ The intuition behind the construction of this transition kernel is to transform a transition kernel that does not satisfy the reversibility condition into one that satisfies it.¹⁰ The algorithm is the following:

Algorithm A1 The Metropolis–Hastings algorithm

- 1: Initialized at an arbitrary value $\boldsymbol{\theta}^{(0)}$.
 - 2: Draw $\boldsymbol{\theta}^c$ from $q(\boldsymbol{\theta}^{(s-1)}, \cdot)$, $s = 1, 2, \dots, S$, and u from $\mathcal{U}(0, 1)$.
 - 3: If $u \leq \alpha(\boldsymbol{\theta}^{(s-1)}, \boldsymbol{\theta}^c)$, set $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^c$, else $\boldsymbol{\theta}^{(s)} = \boldsymbol{\theta}^{(s-1)}$.
 - 4: Repeat this process S times, $s = 1, 2, \dots, S$ iterations.
 - 5: Return the values $\{\boldsymbol{\theta}^{(b)}, \boldsymbol{\theta}^{(b+d)}, \dots, \boldsymbol{\theta}^{(S)}\}$ where b and d are burn-in and thinning parameters, respectively.
-

Observe that the M–H algorithm does not depend on the marginal likelihood, therefore it is not necessary to calculate the integral, which can be very computationally demanding. This algorithm requires setting the proposed density, and generates autocorrelated chains by construction. The chain remains at $\boldsymbol{\theta}^{(s-1)}$ if $\boldsymbol{\theta}^c$ is rejected. It is necessary to discard some initial draws in order to avoid their dependence on the initial value. There must be enough iterations to achieve convergence to the target distribution.

Regarding the density of the proposal, we implement in this GUI the random-walk proposal and the tailored proposal. The former generates $\boldsymbol{\theta}^c = \boldsymbol{\theta}^{(s-1)} + \boldsymbol{\epsilon}^{(s)}$, where the distribution of $\boldsymbol{\epsilon}^{(s)}$ is specified. If this distribution is symmetric around zero, then $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = q(\boldsymbol{\theta}^c, \boldsymbol{\theta})$.

⁸ $\pi(\boldsymbol{\theta}|\mathbf{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) \neq 0$ is usually satisfied in practice.

⁹Observe that $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ depends on \mathbf{y} , and the proposed $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ can depend or not depend on \mathbf{y} .

¹⁰Intuitively, the reversibility condition, $\pi(\boldsymbol{\theta}|\mathbf{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = \pi(\boldsymbol{\theta}^c|\mathbf{y})q(\boldsymbol{\theta}^c, \boldsymbol{\theta})$, implies that the probability of moving from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^c$ is equal that the probability of moving from $\boldsymbol{\theta}^c$ to $\boldsymbol{\theta}$.

The tailored proposal is based on a fat-tailed distribution, whose mean is the maximum of the logarithm of the posterior distribution, $\hat{\boldsymbol{\theta}}$, and scale matrix equal to the negative of the inverse Hessian at the maximum, $\left(-\frac{\partial^2 \ln \pi(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \Big|_{\hat{\boldsymbol{\theta}}}\right)^{-1}$.

A crucial point associated with the proposed densities is the acceptance rate. Low or high acceptance rates are not ideal. A low rate implies poor mixing, that is, the chain does not move through the support of the posterior distribution. A high acceptance rate implies that the chain will converge too slowly. A sensible value depends on the dimension of the parameter space. A rule of thumb in case the dimension is less than or equal to 2 is 0.50. If the dimension is greater than 2, the acceptance rate should be around 0.25 (Roberts et al., 1997).

Regarding convergence issues, we implement some diagnostics in order to check the adequacy of the posterior chains (Plummer et al., 2016). In particular, trace plots should look stable, and autocorrelation plots should decrease very quickly. In addition, we implement Geweke’s test (Geweke, 1992), which is a simple two-sample test of means. If the mean of the fraction in the first window (10%) is not significantly different from the mean of the fraction in the second window (50%), then we conclude that the target distribution has converged. The test of Raftery and Lewis (Raftery and Lewis, 1992) is designed to calculate the approximate number of iterations (S), burn-in (b), and thinning parameter (d) required to estimate $p[H(\boldsymbol{\theta}) \leq h]$, $H(\boldsymbol{\theta}) : \mathcal{R}^k \rightarrow \mathcal{R}$, given a specific quantile of interest (q), precision (r), and probability (p). Their diagnostic is based on the dependence factor, $I = \frac{S+b}{S_{Min}}$, $S_{Min} = \Phi^{-1}\left(\frac{1}{2}(p+1)\right)^2 q(1-q)/r^2$, $\Phi(\cdot)$ is the normal distribution function. Values of I much greater than 5 indicate a high level of dependence. Heidelberger and Welch’s test (Heidelberger and Welch, 1983) is based on a Cramer-von-Mises statistic to test the null hypothesis that the sampled values, $\boldsymbol{\theta}^{(s)}$, come from a stationary distribution, $CVM(B_S) = \int_0^1 B_S(t)^2 dt$, $B_S(t) = (S_{[St]} - [St] \bar{\boldsymbol{\theta}}^S) / (Sp(0))^{1/2}$, $S_S = \sum_{s=1}^S \boldsymbol{\theta}^{(s)}$, $\bar{\boldsymbol{\theta}}^S = S_S/S$, $p(0)$ is the spectral density at 0, and $0 \leq t \leq 1$. $B_S(t)$ converges in distribution to the Brownian bridge under the null hypothesis. This test is recursively applied, until either the null hypothesis is not rejected, or $t = 50\%$ of the chain has been discarded. Then, the half-width test calculates a 95% confidence interval for the mean, using the portion of the chain which passed the stationarity test. If the ratio between the half-width of this interval and the mean is lower than 0.1, this test is passed.

Another popular MCMC algorithm that is extensively used in our models is the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). This can be seen as a particular case of the M–H algorithm where the acceptance rate is equal to 1 (Gelman and Rubin, 1992; Robert and Casella, 2004). This is based on the fact that the full conditional distributions perfectly summarize the joint density (Hammersley–Clifford theorem),¹¹ as a consequence, it is necessary to have the full set of conditional posterior distributions to implement it. The algorithm is the following:

We should highlight that the Gibbs sampler can be applied in settings where the condi-

¹¹This statement requires the positivity condition, if $\pi(\boldsymbol{\theta}_l|\mathbf{y}) > 0$, then $\pi(\boldsymbol{\theta}|\mathbf{y}) > 0$, $l = 1, 2, \dots, k$.

¹²Note that $\boldsymbol{\theta}_l$, $l = 1, 2, \dots, k$ can be either scalars or vectors in Algorithm A2.

Algorithm A2 The Gibbs sampler algorithm

- 1: Given $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k]$, initialized at an arbitrary value $\boldsymbol{\theta}_{-1}^{(0)}$, $\boldsymbol{\theta}_{-l}^{(0)} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{l-1}, \boldsymbol{\theta}_{l+1}, \dots, \boldsymbol{\theta}_k]$.¹²
 - 2: Draw $\boldsymbol{\theta}_1^{(s)}$ from $\pi(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_{-1}^{(s-1)}; \mathbf{y})$.
 - 3: Draw $\boldsymbol{\theta}_2^{(s)}$ from $\pi(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(s)}, \boldsymbol{\theta}_3^{(s-1)}, \dots, \boldsymbol{\theta}_k^{(s-1)}; \mathbf{y})$.
 - 4: \vdots
 - 5: Draw $\boldsymbol{\theta}_k^{(s)}$ from $\pi(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{-k}^{(s)}; \mathbf{y})$.
 - 6: Repeat this process S times, $s = 1, 2, \dots, S$ iterations.
 - 7: Return the values $\{\boldsymbol{\theta}^{(b)}, \boldsymbol{\theta}^{(b+d)}, \dots, \boldsymbol{\theta}^{(S)}\}$ where b and d are burn-in and thinning parameters, respectively.
-

tional posterior distributions are not standard. In this case, Metropolis-within-Gibbs algorithm can be adopted. Rather than simulating $\boldsymbol{\theta}_l^{(s)} \sim \pi(\boldsymbol{\theta}_l | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{l-1}, \boldsymbol{\theta}_{l+1}, \dots, \boldsymbol{\theta}_k)$, we can implement a M–H step. The theoretical validity of this strategy is exactly the same as with the original Gibbs sampler (Robert and Casella, 2010).

After this brief introduction to the general Bayesian framework, we present specific details of models in our GUI in the next section.

5 Models

5.1 Univariate models

Normal–Inverse Gamma: The Gaussian linear model specifies $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ such that $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ is an stochastic error, \mathbf{X} is an $n \times k$ matrix of regressors, $\boldsymbol{\beta}$ is a k dimensional vector of coefficients, \mathbf{y} is an n dimensional vector of a dependent variable, and n is the number of units.

The conjugate independent priors for the parameters are $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$. Given the likelihood function, $p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}$, the conditional posterior distributions are

$$\boldsymbol{\beta} | \sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \sigma^2 \mathbf{B}),$$
¹³

$$\sigma^2 | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \mathcal{IG}(\alpha^*/2, \delta^*/2),$$
¹⁴

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \sigma^{-2} \mathbf{X}'\mathbf{X})^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2} \mathbf{X}'\mathbf{X}\mathbf{y})$, $\alpha^* = \alpha_0 + n$ and $\delta^* = \delta_0 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

We can employ the Gibbs sampler in this model due to having standard conditional posterior distributions.

¹³ \mathcal{N} denotes a normal density.

¹⁴ \mathcal{IG} denotes an inverse gamma density.

Logit: In the logit model the dependent variable is binary, then it follows a Bernoulli distribution, $y_i \stackrel{i.i.d.}{\sim} \mathcal{B}(\pi_i)$,¹⁵ that is $p(y_i = 1) = \pi_i$, such that $\pi_i = \frac{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}$. We specify a Gaussian distribution as prior $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$.

The logit model does not have a standard posterior distribution. Then, a random walk Metropolis–Hastings algorithm is used to obtain draws from the posterior distribution (Martin et al., 2011, 2018). In particular, the proposal is multivariate Gaussian centered at the current value, with covariance matrix $\mathbf{S}(\mathbf{B}_0^{-1} + \widehat{\boldsymbol{\Sigma}}^{-1})^{-1} \mathbf{S}$, where $\mathbf{S} = s \mathbf{I}_{\dim \boldsymbol{\beta}}$, $s > 0$ is a tuning parameter,¹⁶ and $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix from the maximum likelihood estimation.

Probit: The probit model also has as dependent variable a binary outcome. In this case, there is a latent variable (y_i^* , unobserved) that defines the structure of the estimation problem. In particular, $y_i = \begin{cases} 0, & y_i^* \leq 0 \\ 1, & y_i^* > 0 \end{cases}$ such that $y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + \mu_i$, $\mu_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$.¹⁷ This implies $p(y_i = 1) = \pi_i = \Phi(\mathbf{x}'_i \boldsymbol{\beta})$.

Albert and Chib (1993) implemented data augmentation (Tanner and Wong, 1987) to apply a Gibbs sampling algorithm in this model. Augmenting this model with y_i^* , we can have the likelihood contribution from observation i , $p(y_i | y_i^*) = 1_{y_i=0} 1_{y_i^* \leq 0} + 1_{y_i=1} 1_{y_i^* > 0}$, where 1_A is an indicator function that takes the value of 1 when condition A is satisfied.

The posterior distribution is $\pi(\boldsymbol{\beta}, \mathbf{y}^* | \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^n [1_{y_i=0} 1_{y_i^* \leq 0} + 1_{y_i=1} 1_{y_i^* > 0}] \times \mathcal{N}_n(\mathbf{y}^* | \mathbf{X} \boldsymbol{\beta}, \mathbf{I}_n) \times \mathcal{N}_n(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \mathbf{B}_0)$ when taking a Gaussian distribution as prior $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$. This implies

$$y_i^* | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{TN}_{(-\infty, 0]}(\mathbf{x}'_i \boldsymbol{\beta}, 1), & y_i = 0 \\ \mathcal{TN}_{(0, \infty)}(\mathbf{x}'_i \boldsymbol{\beta}, 1), & y_i = 1 \end{cases}, \quad \text{18}$$

$$\boldsymbol{\beta} | \mathbf{y}^*, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}),$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \mathbf{X}' \mathbf{X})^{-1}$, and $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}' \mathbf{y}^*)$.

Multinomial probit: The multinomial probit model is used to model mutually exclusive discrete outcomes or qualitative response variables. We observe $y_{il} = \begin{cases} 1, & y_{il}^* \geq \max(y_{i\ell}^*) \\ 0, & \text{otherwise} \end{cases}$ such that $\mathbf{y}_i^* = \mathbf{X}_i \boldsymbol{\delta} + \boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, \mathbf{y}_i^* is an unobserved latent L dimensional vector, $\mathbf{X}_i = [(1 \ \mathbf{c}'_i) \otimes \mathbf{I}_L \ \mathbf{A}_i]$ is an $L \times j$ matrix of regressors for each alternative, $l = 1, 2, \dots, L$, $j = L \times (1 + \dim \{\mathbf{c}_i\}) + a$, \mathbf{c}_i is a vector of the individuals' specific characteristics, \mathbf{A}_i is an $L \times a$ matrix of alternative-varying regressors, a is the number of alternative-varying regressors, and $\boldsymbol{\delta}$ is a j dimensional vector of parameters. We take into account simultaneously the alternative-varying regressors (alternative attributes) and alternative-invariant regressors

¹⁵B denotes a Bernoulli density.

¹⁶Tuning parameters should be set in such a way that one obtains reasonable diagnostic criteria.

¹⁷The variance in this model is set to 1 due to identification restrictions. Observe that multiplying y_i^* by a positive constant does not affect y_i .

¹⁸ \mathcal{TN} denotes a truncated normal density.

(individual characteristics).¹⁹ \mathbf{y}_i^* can be stacked up into a multiple regression with correlated stochastic errors, $\mathbf{y}^* = \mathbf{X}\boldsymbol{\delta} + \boldsymbol{\mu}$, where $\mathbf{y}^{*'} = [\mathbf{y}_1^{*'}, \mathbf{y}_2^{*'}, \dots, \mathbf{y}_n^{*'}]$, $\mathbf{X}' = [\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n]$, and $\boldsymbol{\mu}' = [\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2, \dots, \boldsymbol{\mu}'_n]$.

Following the practice of expressing y_{il}^* relative to y_{iL}^* by letting $\mathbf{w}'_i = [w_{i1}, w_{i2}, \dots, w_{iL-1}]$, $w_{il} = y_{il}^* - y_{iL}^*$, we can write $\mathbf{w}_i = \mathbf{R}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, where $\mathbf{R}_i = [(1 \ \mathbf{c}'_i) \otimes \mathbf{I}_{L-1} \ \boldsymbol{\Delta} \mathbf{A}_i]$ is an $(L-1) \times k$ matrix where $\boldsymbol{\Delta} \mathbf{A}_i = \mathbf{A}_{li} - \mathbf{A}_{Li}$, $l = 1, 2, \dots, L-1$, that is, the last row of \mathbf{A}_i is subtracted from each row of \mathbf{A}_i , and $\boldsymbol{\beta}$ is a k dimensional vector, $k = (L-1) \times (1 + \dim\{\mathbf{c}_i\}) + a$. Observe that $\boldsymbol{\beta}$ contains the same last a elements as $\boldsymbol{\delta}$, that is, alternative specific attributes coefficients, but the first $(L-1) \times (1 + \dim\{\mathbf{c}_i\})$ -th elements are $\delta_{jl} - \delta_{jL}$, $j = 1 + \dim\{\mathbf{c}_i\}$, $l = 1, 2, \dots, L-1$, that is, the difference between the coefficients of each qualitative response and the L -th alternative for the individuals' characteristics. This makes it difficult to interpret the multinomial probit coefficients. Note that in multinomial models, for each alternative specific attribute, it is only required to estimate one coefficient for all alternatives, whereas for individuals' characteristics (non-alternative specific regressors), it is necessary to estimate $L-1$ coefficients (the coefficient of the base alternative is set equal to 0).

The likelihood function in this model is $p(\boldsymbol{\beta}, \boldsymbol{\Omega} | \mathbf{y}, \mathbf{R}) = \prod_{i=1}^n \prod_{l=1}^L p_{il}^{y_{il}^*}$ where $p_{il} = p(y_{il}^* \geq \max(\mathbf{y}_i^*))$. We assume independent priors, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\boldsymbol{\Omega}^{-1} \sim \mathcal{W}(\alpha_0, \boldsymbol{\Sigma}_0)$.²⁰ We can employ Gibbs sampling in this model because this is a standard Bayesian linear regression model when data augmentation in \mathbf{w} is used. The posterior conditional distributions are

$$\boldsymbol{\beta} | \boldsymbol{\Omega}, \mathbf{w} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}),$$

$$\boldsymbol{\Omega}^{-1} | \boldsymbol{\beta}, \mathbf{w} \sim \mathcal{W}(\alpha^*, \boldsymbol{\Sigma}^*),$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \mathbf{X}^{*'} \mathbf{X}^*)^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^{*'} \mathbf{w}^*)$, $\boldsymbol{\Omega}^{-1} = \mathbf{C}' \mathbf{C}$, $\mathbf{X}_i^{*'} = \mathbf{C}' \mathbf{R}_i$, $\mathbf{w}_i^* = \mathbf{C}' \mathbf{w}_i$, $\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \mathbf{X}_2^* \\ \vdots \\ \mathbf{X}_n^* \end{bmatrix}$, $\alpha^* = \alpha_0 + n$, $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_0 + \sum_{i=1}^n (\mathbf{w}_i - \mathbf{R}_i \boldsymbol{\beta})' (\mathbf{w}_i - \mathbf{R}_i \boldsymbol{\beta}))^{-1}$.

We can collapse the multinomial vector \mathbf{y}_i into the indicator variable $d_i = \sum_{l=1}^{L-1} l \times I_{\max(\mathbf{w}_i) = w_{il}}$.²¹ Then the distribution of $\mathbf{w}_i | \boldsymbol{\beta}, \boldsymbol{\Omega}^{-1}, d_i$ is an $L-1$ dimensional Gaussian distribution truncated over the appropriate cone in \mathcal{R}^{L-1} . McCulloch and Rossi (1994) propose drawing from the univariate conditional distributions $w_{il} | \mathbf{w}_{i,-l}, \boldsymbol{\beta}, \boldsymbol{\Omega}^{-1}, d_i \sim \mathcal{TN}_{I_{il}}(m_{il}, \tau_{il}^2)$, where $I_{il} = \begin{cases} w_{il} > \max(\mathbf{w}_{i,-l}, 0), & d_i = l \\ w_{il} < \max(\mathbf{w}_{i,-l}, 0), & d_i \neq l \end{cases}$, and permuting the columns and rows of $\boldsymbol{\Omega}^{-1}$ so that the l -th column and row is the last,

$$\boldsymbol{\Omega}^{-1} = \begin{bmatrix} \boldsymbol{\Omega}_{-l,-l} & \boldsymbol{\omega}_{-l,l} \\ \boldsymbol{\omega}_{l,-1} & \boldsymbol{\omega}_{l,l} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Omega}_{-l,-l}^{-1} + \mathbf{f}' \mathbf{E} \mathbf{f} & -\mathbf{f} \tau_{ll}^{-2} \\ -\tau_{ll}^{-2} \mathbf{f}' & \tau_{ll}^{-2} \end{bmatrix}$$

¹⁹Note that this model is not identified if $\boldsymbol{\Sigma}$ is unrestricted. The likelihood function is the same if a scalar random variable is added to each of the L latent regressions.

²⁰ \mathcal{W} denotes the Wishart density.

²¹Observe that the identification issue in this model is due to scaling w_{il} by a positive constant and does not change the value of d_i .

where $\mathbf{f} = \boldsymbol{\Omega}_{-l,-l}\boldsymbol{\omega}_{-l,l}$, $\tau_{ll}^2 = \omega_{ll} - \boldsymbol{\omega}_{l,-l}\boldsymbol{\Omega}_{-l,-l}^{-1}\boldsymbol{\omega}_{-l,l}$, $m_{il} = \mathbf{r}'_{il}\boldsymbol{\beta} + \mathbf{f}'(\mathbf{w}_{i,-l} - \mathbf{R}_{i,-l}\boldsymbol{\beta})$, $\mathbf{w}_{i,-l}$ is an $L - 2$ dimensional vector of all components of \mathbf{w}_i excluding w_{il} , \mathbf{r}_{il} is the l -th row of \mathbf{R}_i , $l = 1, 2, \dots, L - 1$.

The identified parameters are obtained by normalizing with respect to one of the diagonal elements $\frac{1}{\omega_{1,1}^{0.5}}\boldsymbol{\beta}$ and $\frac{1}{\omega_{1,1}}\boldsymbol{\Omega}$.²²

Multinomial logit: The multinomial logit model is also used to model mutually exclusive discrete outcomes or qualitative response variables. We consider the multinomial mixed logit model (not to be confused with the random parameters logit model), that is, we take into account simultaneously alternative-varying regressors (conditional) and alternative-invariant regressors (multinomial).²³

In this setting there are L mutually exclusive alternatives, and the dependent variable y_i is equal to l if the l th alternative is chosen by individual i , $l = \{1, 2, \dots, L\}$. The likelihood function is $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^n p(y_i = l|\boldsymbol{\beta}, \mathbf{X})$, where $p(y_i = l|\boldsymbol{\beta}, \mathbf{X}) = \frac{\exp\{\mathbf{x}'_{il}\boldsymbol{\beta}\}}{\sum_{j=1}^L \exp\{\mathbf{x}'_{ij}\boldsymbol{\beta}\}}$, \mathbf{X} is an $nL \times k$ matrix, $k = k_1 + k_2$ is the total number of regressors, k_1 and k_2 are the number of alternative-varying and alternative-invariant regressors, respectively. In addition, we assume $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ as prior distribution.

As the multinomial logit model does not have a standard posterior distribution, we propose a “tailored” Metropolis–Hastings algorithm as the proposed distribution, in particular, a multivariate Student’s t distribution with v degrees of freedom (tuning parameter), mean equal to the maximum likelihood estimator, and scale equal to the inverse of the Hessian matrix.

Ordered probit: The ordered probit model is used when there is a natural order in the categorical response variable. In this case, there is a latent variable $y_i^* = \mathbf{x}'_i\boldsymbol{\beta} + \mu_i$, $\mu_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ such that $y_i = l$ if and only if $\alpha_{l-1} < y_i^* \leq \alpha_l$, $l = \{1, 2, \dots, L\}$, where $\alpha_0 = -\infty$, $\alpha_1 = 0$ and $\alpha_L = \infty$.²⁴ Then, $p(y_i = l) = \Phi(\alpha_l - \mathbf{x}'_i\boldsymbol{\beta}) - \Phi(\alpha_{l-1} - \mathbf{x}'_i\boldsymbol{\beta})$, and the likelihood function is $p(\boldsymbol{\beta}, \boldsymbol{\alpha}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^n p(y_i = l|\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{X})$. The independent priors of this model are $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\gamma}_0, \boldsymbol{\Gamma}_0)$, $\boldsymbol{\gamma} = [\gamma_2, \gamma_3, \dots, \gamma_{L-1}]'$, such that $\boldsymbol{\alpha} = \left[\exp\{\gamma_2\}, \sum_{l=2}^3 \exp\{\gamma_l\}, \dots, \sum_{l=2}^{L-1} \exp\{\gamma_l\} \right]'$.

This model does not have a standard conditional posterior distribution for $\boldsymbol{\gamma}(\boldsymbol{\alpha})$, but it does have a standard conditional distribution for $\boldsymbol{\beta}$ once data augmentation is used. Then, we use a Metropolis-within-Gibbs sampling algorithm. In particular, we use Gibbs sampling algorithms to draw $\boldsymbol{\beta}$ and \mathbf{y}^* ,

$$\boldsymbol{\beta}|\mathbf{y}^*, \boldsymbol{\alpha}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}),$$

²²Our GUI takes into account this identification restriction to display the outcomes of the posterior chains.

²³The multinomial mixed logit model can be implemented as a conditional logit model.

²⁴Identification issues necessitate setting the variance in this model equal to 1 and $\alpha_1 = 0$. Observe that multiplying y_i^* by a positive constant or adding a constant to all of the cut-offs and subtracting the same constant from the intercept does not affect y_i .

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{y}^*)$, and $y_i^* | \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{y}, \mathbf{X} \sim \mathcal{TN}_{(\alpha_{y_i-1}, \alpha_{y_i})}(\mathbf{x}'_i \boldsymbol{\beta}, 1)$.

We use a random-walk Metropolis–Hastings algorithm for γ that has as proposal a Gaussian distribution with mean equal to the current value, and covariance matrix $s^2(\boldsymbol{\Gamma}_0^{-1} + \hat{\boldsymbol{\Sigma}}_\gamma^{-1})^{-1}$, where $s > 0$ is a tuning parameter, and $\hat{\boldsymbol{\Sigma}}_\gamma$ is the sample covariance matrix associated with γ from the maximum likelihood estimation.

Negative binomial: The dependent variable in the negative binomial model is a non-negative integer or count. In contrast to the Poisson model, the negative binomial model takes into account over-dispersion. The Poisson model has equal mean and variance.

We assume that $y_i \stackrel{i.n.d.}{\sim} \mathcal{NB}(\gamma, \theta_i)$, that is, the density function for individual i is $\frac{\Gamma(y_i + \gamma)}{\Gamma(\gamma) y_i!} \theta_i^{y_i} (1 - \theta_i)^\gamma$, the success probability is $\theta_i = \frac{\lambda_i}{\lambda_i + \gamma}$, $\lambda_i = \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}$ and $\gamma = \exp\{\alpha\}$. The independent priors for this model are $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\alpha \sim \mathcal{G}(\alpha_0, \delta_0)$.²⁵

This model does not have standard conditional posterior distributions, so we use a random-walk Metropolis–Hastings algorithm where the proposed distribution for $\boldsymbol{\beta}$ is Gaussian centered at the current stage with covariance matrix $s_\beta^2 \hat{\boldsymbol{\Sigma}}_\beta$ where s_β is a tuning parameter and $\hat{\boldsymbol{\Sigma}}_\beta$ is the maximum likelihood covariance estimator. In addition, the proposal for α is normal centered at the current value, with variance $s_\alpha^2 \hat{\sigma}_\alpha^2$ where s_α is a tuning parameter and $\hat{\sigma}_\alpha^2$ is the maximum likelihood variance estimator.

Tobit: The dependent variable is partially observed in Tobit models due to sampling schemes, whereas the regressors are completely observed. In particular,

$$y_i = \begin{cases} L, & y_i^* < L \\ y_i^*, & L \leq y_i^* < U \\ U, & y_i^* \geq U \end{cases},$$

where $y_i^* \stackrel{i.n.d.}{\sim} \mathcal{N}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$.²⁶

We use conjugate independent priors $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$, and data augmentation using \mathbf{y}_C^* such that $y_{C_i}^* \stackrel{i.n.d.}{\sim} \mathcal{N}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2)$, $y_{C_i} = \{y_{C_i^L}^* \cup y_{C_i^U}^*\}$ are lower and upper censored data. This allows implementing the Gibbs sampling algorithm (Chib, 1992). Then, $\pi(\boldsymbol{\beta}, \sigma^2, \mathbf{y}^* | \mathbf{y}, \mathbf{X}) \propto \prod_{i=1}^n \left[1_{y_i=L} 1_{y_{C_i^L}^* < L} + 1_{L \leq y_i < U} + 1_{y_i=U} 1_{y_{C_i^U}^* \geq U} \right] \mathcal{N}(y_i^* | \mathbf{x}'_i \boldsymbol{\beta}, \sigma^2) \times \mathcal{N}_n(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \mathbf{B}_0) \times \mathcal{IG}(\sigma^2 | \alpha_0/2, \delta_0/2)$

The posterior distributions are

$$y_{C_i}^* | \boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{TN}_{(-\infty, L)}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2), & y_i = L \\ \mathcal{TN}_{[U, \infty)}(\mathbf{x}'_i \boldsymbol{\beta}, \sigma^2), & y_i = U \end{cases},$$

$$\boldsymbol{\beta} | \sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \sigma^2 \mathbf{B}),$$

²⁵ \mathcal{G} denotes a gamma density.

²⁶We can set L or U equal to $-\infty$ or ∞ to model data censored in just one side.

$$\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \mathcal{IG}(\alpha^*/2, \delta^*/2),$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}'\mathbf{y})$, $\alpha^* = \alpha_0 + n$ and $\delta^* = \delta_0 + (\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})$.

Quantile: In quantile regression the location parameters vary according to the quantile of the dependent variable. Let $q_\tau(\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}_\tau$ denote the τ -th ($0 < \tau < 1$) quantile regression function of y_i given \mathbf{x}_i such that $y_i = \mathbf{x}_i'\boldsymbol{\beta}_\tau + \mu_i$ where $\int_{-\infty}^0 f_\tau(\mu_i)d\mu_i = \tau$. In particular, $f_\tau(\mu_i) = \tau(1-\tau)\exp\{\mu_i(\tau - I_{\mu_i < 0})\}$ (asymmetric Laplace distribution). [Kozumi and Kobayashi \(2011\)](#) propose the location-scale mixture of normals with a representation given by $\mu_i = \theta e_i + \psi\sqrt{e_i}z_i$ where $\theta = \frac{1-2\tau}{\tau(1-\tau)}$, $\psi^2 = \frac{2}{\tau(1-\tau)}$, $e_i \sim \mathcal{E}(1)$ and $z_i \sim \mathcal{N}(0, 1)$, $e_i \perp z_i$.²⁷ As a consequence of this representation and the fact that the sample is i.i.d., $p(\mathbf{y}|\boldsymbol{\beta}_\tau, \mathbf{e}, \mathbf{X}) \propto \left(\prod_{i=1}^n e_i^{-1/2}\right) \exp\left\{-\sum_{i=1}^n \frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau - \theta e_i)^2}{2\psi^2 e_i}\right\}$.

Taking as prior a normal distribution for $\boldsymbol{\beta}_\tau$, that is, $\boldsymbol{\beta}_\tau \sim \mathcal{N}(\boldsymbol{\beta}_{\tau 0}, \mathbf{B}_{\tau 0})$, and using data augmentation for \mathbf{e} , we can implement a Gibbs sampling algorithm in this model. The posterior distributions are

$$\boldsymbol{\beta}_\tau|\mathbf{e}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}_\tau^*, \mathbf{B}_\tau),$$

$$e_i|\boldsymbol{\beta}_\tau, \mathbf{y}, \mathbf{X} \sim \mathcal{GIG}(1/2, \alpha_i^*, \delta_i^*),$$
²⁸

where $\mathbf{B}_\tau = \left(\mathbf{B}_{\tau 0}^{-1} + \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i'}{\psi^2 e_i}\right)^{-1}$, $\boldsymbol{\beta}_\tau^* = \mathbf{B}_\tau \left(\mathbf{B}_{\tau 0}^{-1}\boldsymbol{\beta}_{\tau 0} + \sum_{i=1}^n \frac{\mathbf{x}_i(y_i - \theta e_i)}{\psi^2 e_i}\right)$, $\alpha_i^* = ((y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau)^2 / \psi^2)^{0.5}$ and $\delta_i^* = (2 + \theta^2 / \psi^2)^{0.5}$.

5.2 Multivariate models

Multivariate regression: This model is used when there are m multiply dependent variables which share the same set of regressors, and their stochastic errors are contemporaneously correlated. In particular, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m]$ is an $n \times m$ matrix that is generated by $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}$ where \mathbf{X} is an $n \times k$ matrix, $\mathbf{B} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_m]$ is a $k \times m$ matrix of parameters, and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$ is a matrix of stochastic random errors such that $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $i = 1, 2, \dots, n$ is each row of \mathbf{U} . Then, $p(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{Y}, \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}^{-1}\right]$, where tr denotes the trace operator.

[Rossi et al. \(2005\)](#) propose the natural conjugate priors $\pi(\text{vec}(\mathbf{B})|\boldsymbol{\Sigma}) \sim \mathcal{N}(\text{vec}(\mathbf{B}_0), \boldsymbol{\Sigma} \otimes \boldsymbol{\Delta}_0)$ and $\pi(\boldsymbol{\Sigma}) \sim \mathcal{IW}(\alpha_0, \boldsymbol{\Sigma}_0)$ where vec is the vectorization operator and \otimes is the Kronecker product. Therefore, the conditional posterior distributions are

$$\boldsymbol{\beta}|\boldsymbol{\Sigma}, \mathbf{Y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \boldsymbol{\Sigma} \otimes \boldsymbol{\Delta}^*),$$

$$\boldsymbol{\Sigma}|\mathbf{Y}, \mathbf{X} \sim \mathcal{IW}(\alpha^*, \boldsymbol{\Sigma}^*),$$
²⁹

²⁷ \mathcal{E} denotes an exponential density.

²⁸ \mathcal{GIG} denotes a generalized inverse Gaussian density.

²⁹ \mathcal{IW} denotes an inverse Wishart density.

where $\mathbf{\Delta}^* = (\mathbf{X}'\mathbf{X} + \mathbf{\Delta}_0^{-1})^{-1}$, $\boldsymbol{\beta}^* = \text{vec}(\mathbf{B}^*)$, $\mathbf{B}^* = \mathbf{\Delta}^*(\mathbf{\Delta}_0^{-1}\mathbf{B}_0 + \mathbf{X}'\mathbf{X}\hat{\mathbf{B}})$, $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\alpha^* = \alpha_0 + n$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_0 + (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$.

We can use a Gibbs sampling algorithm in this model since the conditional posterior distributions are standard.

Seemingly unrelated regression: In this model there are m dependent variables with potentially different regressors, and the stochastic errors are contemporaneously correlated. In particular, $\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\mu}_j$, where $\boldsymbol{\beta}_j$ is a k_j vector, $j = 1, 2, \dots, m$. Setting $\boldsymbol{\mu}_i = [\mu_{1i}, \mu_{2i}, \dots, \mu_{mi}]'$ such that $\boldsymbol{\mu}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and stacking the m equations, we can write $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ where $\mathbf{y} = [\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_m]'$, $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_m]'$ is a K dimensional vector, $K = \sum_{j=1}^m k_j$, \mathbf{X} is an $mn \times K$ block diagonal matrix composed of \mathbf{X}_j and $\boldsymbol{\mu} = [\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2, \dots, \boldsymbol{\mu}'_m]'$ such that $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$. Then, $p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$.

Using independent priors $\pi(\boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\pi(\boldsymbol{\Sigma}^{-1}) \sim \mathcal{W}(\alpha_0, \boldsymbol{\Sigma}_0)$, the posterior distributions are

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}^*),$$

$$\boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \mathcal{W}(\alpha^*, \boldsymbol{\Sigma}^*),$$

where $\mathbf{B}^* = (\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{X} + \mathbf{B}_0^{-1})^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}^*(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathbf{y})$, $\alpha^* = \alpha_0 + n$ and $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_0^{-1} + \mathbf{U}'\mathbf{U})^{-1}$, where \mathbf{U} is an $n \times m$ matrix whose columns are $\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}_j$.

We can employ a Gibbs sampling algorithm with this model since the conditional posterior distributions are standard.

Instrumental variables: This model is used when there are endogeneity problems caused by feedback, omitted relevant variables, or measurement error in the regressors. So, we specify the dependent variable as a linear function of one endogenous regressor and some exogenous regressors. That is, $y_i = \mathbf{x}'_{ei}\boldsymbol{\beta}_1 + \beta_s x_{si} + \mu_i$ where $x_{si} = \mathbf{x}'_{ei}\boldsymbol{\gamma}_1 + \mathbf{z}'_i\boldsymbol{\gamma}_2 + v_i$, \mathbf{x}_s is the variable which generates the endogeneity issues, such that \mathbf{x}_e are k_1 exogenous regressors and \mathbf{z} are k_2 instruments. Assuming $(u_i, v_i)' \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = [\sigma_{lm}]$, $l, m = 1, 2$, the likelihood function is $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (y_i - \mathbf{x}'_i\boldsymbol{\beta}, x_{si} - \mathbf{w}'_i\boldsymbol{\gamma}) \boldsymbol{\Sigma}^{-1} \begin{pmatrix} y_i - \mathbf{x}'_i\boldsymbol{\beta} \\ x_{si} - \mathbf{w}'_i\boldsymbol{\gamma} \end{pmatrix}\right\}$ where $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \beta_s]'$, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2]'$, $\mathbf{x}_i = [\mathbf{x}'_{ei}, x_{si}]'$ and $\mathbf{w}_i = [\mathbf{x}'_{ei}, \mathbf{z}'_i]'$.

We get standard conditional posterior densities using the following independent priors $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\gamma}_0, \mathbf{G}_0)$, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\alpha_0, \boldsymbol{\Sigma}_0)$. In particular,

$$\boldsymbol{\beta} | \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}^*)$$

$$\boldsymbol{\gamma} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} \sim \mathcal{N}(\boldsymbol{\gamma}^*, \mathbf{G}^*)$$

$$\boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} \sim \mathcal{W}(\alpha^*, \boldsymbol{\Sigma}^*)$$

where $\mathbf{B}^* = (\omega_1^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i + \mathbf{B}_0^{-1})^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}^* \left(\mathbf{B}_0^{-1}\boldsymbol{\beta}_0 + \omega_1^{-1} \sum_{i=1}^n \left[\mathbf{x}_i \left(y_i - \frac{\sigma_{12}(x_{si} - \mathbf{w}'_i\boldsymbol{\gamma})}{\sigma_{22}} \right) \right] \right)$, $\omega_1 = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, $\mathbf{G}^* = (\omega_2^{-1} \sum_{i=1}^n \mathbf{w}_i \mathbf{w}'_i + \mathbf{G}_0^{-1})^{-1}$, $\boldsymbol{\gamma}^* = \mathbf{G}^* \left(\mathbf{G}_0^{-1}\boldsymbol{\gamma}_0 + \omega_2^{-1} \sum_{i=1}^n \left[\mathbf{w}_i \left(x_{si} - \frac{\sigma_{12}(y_i - \mathbf{x}'_i\boldsymbol{\beta})}{\sigma_{11}} \right) \right] \right)$,

$$\omega_2 = \sigma_{22} - \sigma_{12}^2 / \sigma_{11}, \alpha^* = \alpha_0 + n \text{ and } \Sigma^* = \left[\Sigma_0^{-1} + \sum_{i=1}^n \begin{pmatrix} y_i - \mathbf{x}'_i \boldsymbol{\beta} \\ x_{si} - \mathbf{w}'_i \boldsymbol{\gamma} \end{pmatrix} (y_i - \mathbf{x}'_i \boldsymbol{\beta}, x_{si} - \mathbf{w}'_i \boldsymbol{\gamma}) \right]^{-1}.$$

We also use a Gibbs sampling algorithm in this model since we have standard conditional posterior distributions.

Multivariate probit: In the multivariate probit model (Edwards and Allenby, 2003), the response variable $y_{il} = \{0, 1\}$ indicates that individual (unit) i makes binary choices regarding alternatives l , $i = 1, 2, \dots, n$, $l = 1, 2, \dots, L$.³⁰ In particular, $y_{il} = \begin{cases} 0, & y_{il}^* \leq 0 \\ 1, & y_{il}^* > 0 \end{cases}$

such that $\mathbf{y}_i^* = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\mu}_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, \mathbf{y}_i^* is an unobserved latent L dimensional vector, \mathbf{X}_i is an $L \times K$ design matrix of regressors, $K = L \times k$, k is the number of regressors, and $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_k]'$, where the $\boldsymbol{\beta}_j$ make up an L dimensional vector of coefficients, $j = 1, 2, \dots, k$. We simultaneously take into account the alternative-varying regressors (alternative attributes) and alternative-invariant regressors (individual characteristics).

The likelihood function in this model is $p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \prod_{l=1}^L p_{il}^{y_{il}}$ where $p_{il} = p(y_{il}^* \geq 0)$. Observe that $p(\mathbf{y}_i^* \geq 0) = p(\boldsymbol{\Lambda} \mathbf{y}_i^* \geq 0)$, where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_{il}\}$, $\lambda_{il} > 0$. This generates identification issues. We follow the post processing strategy proposed by Edwards and Allenby (2003) to get identified parameters, that is, $\boldsymbol{\beta} = \text{vec}\{\mathbf{D}\boldsymbol{\Gamma}\}$ and the correlation matrix $\mathbf{R} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}$, where $\mathbf{D} = \text{diag}\{\sigma_{ll}\}^{-1/2}$ and $\boldsymbol{\Gamma} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_k]$.³¹

We assume independent priors, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\alpha_0, \boldsymbol{\Sigma}_0)$. We can employ Gibbs sampling in this model because this is a standard Bayesian linear regression model when data augmentation in \mathbf{y}^* is used. The posterior conditional distributions are

$$\begin{aligned} \boldsymbol{\beta} | \boldsymbol{\Sigma}, \mathbf{w} &\sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}), \\ \boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \mathbf{w} &\sim \mathcal{W}(\alpha^*, \boldsymbol{\Sigma}^*), \\ y_{il}^* | \mathbf{y}_{i,-l}^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}, \mathbf{y}_i &\sim \mathcal{TN}_{I_{il}}(m_{il}, \tau_{il}^2) \end{aligned}$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \mathbf{X}^{*'} \mathbf{X}^*)^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^{*'} \mathbf{y}^{**})$, $\boldsymbol{\Sigma}^{-1} = \mathbf{C}' \mathbf{C}$, $\mathbf{X}_i^{*'} = \mathbf{C}' \mathbf{X}_i$,

$$\mathbf{y}_i^{**} = \mathbf{C}' \mathbf{y}_i^*, \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \mathbf{X}_2^* \\ \vdots \\ \mathbf{X}_n^* \end{bmatrix}, \alpha^* = \alpha_0 + n, \boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_0 + \sum_{i=1}^n (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta})' (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}))^{-1},$$

$I_{il} = \begin{cases} y_{il}^* > 0, & y_{il} = 1 \\ y_{il}^* \leq 0, & y_{il} = 0 \end{cases}$, $m_{il} = \mathbf{x}'_{il} \boldsymbol{\beta} + \mathbf{f}'(\mathbf{y}_{i,-l}^* - \mathbf{X}_{i,-l} \boldsymbol{\beta})$, $\mathbf{y}_{i,-l}^*$ is an $L-1$ dimensional vector of all components of \mathbf{y}_i^* excluding y_{il}^* , \mathbf{x}_{il} is the l -th row of \mathbf{X}_i , $\mathbf{X}_{i,-l}$ is \mathbf{X}_i after deleting the

l -th row, $\tau_{il}^2 = 1/\sigma^{ll}$, and $\mathbf{f} = -\sigma^{ll} \boldsymbol{\omega}_{l,-l}$, σ^{jl} is the jl -th element of $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \omega'_1 \\ \omega'_2 \\ \vdots \\ \omega'_L \end{bmatrix}$, $\boldsymbol{\omega}'_{l,-l}$

³⁰These alternatives are not mutually exclusive.

³¹In a Bayesian setting, we can have a non identified model; however, the posterior of the model parameters exists given a proper prior distribution (Edwards and Allenby, 2003).

is the l -th row of $\boldsymbol{\Sigma}^{-1}$ extracting the l -th element.

5.3 Hierarchical longitudinal models

Normal: The hierarchical longitudinal normal model establishes $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\mathbf{b}_i + \boldsymbol{\mu}_i$ where \mathbf{y}_i are n_i vectors corresponding to units $i = 1, 2, \dots, m$, \mathbf{X}_i are $n_i \times k$ matrices, $\boldsymbol{\beta}$ is a k dimensional vector of “fixed” effects, \mathbf{W}_i are $n_i \times q$ matrices associated with random effects (q typically less than k), \mathbf{b}_i is a q dimensional vector of unit-specific random effects such that $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, and $\boldsymbol{\mu}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ are stochastic errors.

We use standard conjugate prior distributions following [Chib and Carlin \(1999\)](#). In particular, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$, $\sigma^2 \sim \mathcal{IG}(\alpha_0, 1/\delta_0)$ and $\mathbf{D} \sim \mathcal{IW}(d_0, d_0 \mathbf{D}_0)$, and taking into account that $\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{D}, \sigma^2 \sim \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i)$ where $\mathbf{V}_i = \sigma^2 \mathbf{I} + \mathbf{W}_i \mathbf{D} \mathbf{W}_i'$, we have

$$\begin{aligned} \boldsymbol{\beta} | \sigma^2, \mathbf{D}, \mathbf{y}, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}), \\ \mathbf{b}_i | \boldsymbol{\beta}, \sigma^2, \mathbf{D}, \mathbf{y}, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\mathbf{b}_i^*, \mathbf{B}_i), \\ \mathbf{D} | \mathbf{b} &\sim \mathcal{IW}(d^*, \mathbf{D}^*), \\ \sigma^2 | \boldsymbol{\beta}, \mathbf{D}, \mathbf{b}, \mathbf{y}, \mathbf{X}, \mathbf{W} &\sim \mathcal{IG}(\alpha^*, \delta^*), \end{aligned}$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i)$, $\mathbf{B}_i = (\mathbf{D}^{-1} + \sigma^{-2} \mathbf{W}_i' \mathbf{W}_i)^{-1}$, $\mathbf{b}_i^* = \mathbf{B}_i(\sigma^{-2} \mathbf{W}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}))$, $d^* = d_0 + m$ and $\mathbf{D}^* = d_0 \mathbf{D}_0 + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i'$, $\alpha^* = \alpha_0 + \frac{1}{2} \sum_{i=1}^m n_i$ and $\delta^* = 1/\delta_0 + \frac{1}{2} \sum_{i=1}^m (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)$.

Logit: The hierarchical longitudinal logit model establishes $y_{ij} \sim \mathcal{B}(\pi_{ij})$ where $\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \equiv \mathbf{y}_{ij}^* = \mathbf{x}_{ij}' \boldsymbol{\beta} + \mathbf{w}_{ij}' \mathbf{b}_i + \mu_{ij}$, $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n_i$ where \mathbf{x}_{ij} are k dimensional vectors of regressors, $\boldsymbol{\beta}$ is a k dimensional vector of “fixed” effects, \mathbf{w}_{ij} are q dimensional vectors of regressors associated with random effects (q typically less than k), \mathbf{b}_i is a q dimensional vector of unit-specific random effects such that $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, and $\mu_{ij} \sim \mathcal{N}(0, \sigma^2)$ are stochastic errors.

We use standard conjugate prior distributions following [Chib and Carlin \(1999\)](#). In particular, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$, $\sigma^2 \sim \mathcal{IG}(\alpha_0, 1/\delta_0)$ and $\mathbf{D} \sim \mathcal{IW}(d_0, d_0 \mathbf{D}_0)$, and taking into account that $\mathbf{y}_i^* | \boldsymbol{\beta}, \mathbf{D}, \sigma^2 \sim \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i)$ where $\mathbf{V}_i = \sigma^2 \mathbf{I} + \mathbf{W}_i \mathbf{D} \mathbf{W}_i'$, we have

$$\begin{aligned} \mathbf{y}_{ij}^* | \boldsymbol{\beta}, \sigma^2, \mathbf{b}, \mathbf{y}, \mathbf{X}, \mathbf{W} &\propto \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \times \mathcal{N}(\mathbf{x}_{ij}' \boldsymbol{\beta} + \mathbf{w}_{ij}' \mathbf{b}_i, \sigma^2), \\ \boldsymbol{\beta} | \sigma^2, \mathbf{D}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}), \\ \mathbf{b}_i | \boldsymbol{\beta}, \sigma^2, \mathbf{D}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\mathbf{b}_i^*, \mathbf{B}_i), \\ \mathbf{D} | \mathbf{b} &\sim \mathcal{IW}(d^*, \mathbf{D}^*), \\ \sigma^2 | \boldsymbol{\beta}, \mathbf{D}, \mathbf{b}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{IG}(\alpha^*, \delta^*), \end{aligned}$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i^*)$, $\mathbf{B}_i = (\mathbf{D}^{-1} + \sigma^{-2} \mathbf{W}_i' \mathbf{W}_i)^{-1}$, $\mathbf{b}_i^* = \mathbf{B}_i(\sigma^{-2} \mathbf{W}_i' (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}))$, $d^* = d_0 + m$ and $\mathbf{D}^* = d_0 \mathbf{D}_0 + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i'$,

$$\alpha^* = \alpha_0 + \frac{1}{2} \sum_{i=1}^m n_i \text{ and } \delta^* = 1/\delta_0 + \frac{1}{2} \sum_{i=1}^m (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)' (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i).$$

We can implement a Gibbs sampling algorithm for the latter four standard conditional posterior distributions. However, this model has been augmented with the latent variable $\mathbf{y}^* = [y_{ij}^*]$ which should be drawn using a random-walk Metropolis–Hastings algorithm such that the proposed distribution is Gaussian with mean y_{ij}^* and variance equal to 1, that is, $y_{ij}^{*c} = y_{ij}^* + \epsilon_{ij}$ where $\epsilon_{ij} \sim \mathcal{N}(0, 1)$ which implies $\pi_{ij} = \frac{1}{1+e^{-y_{ij}^*}}$ and $\pi_{ij}^c = \frac{1}{1+e^{-y_{ij}^{*c}}}$.

Poisson: The hierarchical longitudinal Poisson model establishes $y_{ij} \sim \mathcal{P}(\lambda_{ij})$,³² where $\log(\lambda_{ij}) \equiv y_{ij}^* = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{w}'_{ij} \mathbf{b}_i + \mu_{ij}$, $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n_i$ where \mathbf{x}_{ij} are k dimensional vectors of regressors, $\boldsymbol{\beta}$ is a k dimensional vector of “fixed” effects, \mathbf{w}_{ij} are q dimensional vectors of regressors associated with random effects (q typically less than k), \mathbf{b}_i is a q dimensional vector of unit-specific random effects such that $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, and $\mu_{ij} \sim \mathcal{N}(0, \sigma^2)$ are stochastic errors.

We use standard conjugate prior distributions following Chib and Carlin (1999). In particular, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{B}_0)$, $\sigma^2 \sim \mathcal{IG}(\alpha_0, 1/\delta_0)$ and $\mathbf{D} \sim \mathcal{IW}(d_0, d_0 \mathbf{D}_0)$, and taking into account that $\mathbf{y}_i^* | \boldsymbol{\beta}, \mathbf{D}, \sigma^2 \sim \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}_i)$ where $\mathbf{V}_i = \sigma^2 \mathbf{I} + \mathbf{W}_i \mathbf{D} \mathbf{W}_i'$, we have

$$\begin{aligned} y_{ij}^* | \boldsymbol{\beta}, \sigma^2, \mathbf{b}, \mathbf{y}, \mathbf{X}, \mathbf{W} &\propto \lambda_{ij}^{y_{ij}^*} e^{-\lambda_{ij}^*} \times \mathcal{N}(\mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{w}'_{ij} \mathbf{b}_i, \sigma^2), \\ \boldsymbol{\beta} | \sigma^2, \mathbf{D}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}), \\ \mathbf{b}_i | \boldsymbol{\beta}, \sigma^2, \mathbf{D}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{N}(\mathbf{b}_i^*, \mathbf{B}_i), \\ \mathbf{D} | \mathbf{b} &\sim \mathcal{IW}(d^*, \mathbf{D}^*), \\ \sigma^2 | \boldsymbol{\beta}, \mathbf{D}, \mathbf{b}, \mathbf{y}^*, \mathbf{X}, \mathbf{W} &\sim \mathcal{IG}(\alpha^*, \delta^*), \end{aligned}$$

where $\mathbf{B} = (\mathbf{B}_0^{-1} + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}$, $\boldsymbol{\beta}^* = \mathbf{B}(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i^*)$, $\mathbf{B}_i = (\mathbf{D}^{-1} + \sigma^{-2} \mathbf{W}_i' \mathbf{W}_i)^{-1}$, $\mathbf{b}_i^* = \mathbf{B}_i(\sigma^{-2} \mathbf{W}_i' (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}))$, $d^* = d_0 + m$ and $\mathbf{D}^* = d_0 \mathbf{D}_0 + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i'$, $\alpha^* = \alpha_0 + \frac{1}{2} \sum_{i=1}^m n_i$ and $\delta^* = 1/\delta_0 + \frac{1}{2} \sum_{i=1}^m (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)' (\mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)$.

We can implement a Gibbs sampling algorithm for the latter four standard conditional posterior distributions. However, this model has been augmented with the latent variable $\mathbf{y}^* = [y_{ij}^*]$ which should be drawn using a random-walk Metropolis–Hastings algorithm such that the proposal distribution is Gaussian with mean y_{ij}^* and variance equal to 1, that is, $y_{ij}^{*c} = y_{ij}^* + \epsilon_{ij}$ where $\epsilon_{ij} \sim \mathcal{N}(0, 1)$ which implies $\lambda_{ij} = e^{y_{ij}^*}$ and $\lambda_{ij}^c = e^{y_{ij}^{*c}}$.

5.4 Bayesian bootstrap

We implement the Bayesian bootstrap (Rubin, 1981) for linear regression models. In particular, the Bayesian bootstrap simulates the posterior distributions assuming that the sample cumulative distribution function (cdf) is the population cdf (this assumption is also implicit

³² \mathcal{P} denotes a Poisson density.

in the frequentist bootstrap (Efron, 1979)).

Given $y_i \stackrel{i.n.d.}{\sim} \mathcal{F}$ where \mathcal{F} does not define a particular parametric family of distributions, $i = 1, 2, \dots, n$, but sets $E(Y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$, such that \mathbf{x}_i is a k dimensional vector of regressors and $\boldsymbol{\beta}$ is a k dimensional vector of parameters, the Bayesian bootstrap generates posterior probabilities for each y_i where the values of Y that are not observed have zero posterior probability.

The algorithm to implement the Bayesian bootstrap is the following:

Algorithm A3 Bayesian bootstrap in linear regression

- 1: Draw $\mathbf{g} \sim Dir(\alpha_1, \alpha_2, \dots, \alpha_n)$ such that $\alpha_i = 1 \forall i$.
- 2: $\mathbf{g} = (g_1, g_2, \dots, g_n)$ is the vector of probabilities to attach to $(y_1, \mathbf{x}'_1), (y_2, \mathbf{x}'_2), \dots, (y_n, \mathbf{x}'_n)$ for each Bayesian bootstrap replication.
- 3: Sample (y_i, \mathbf{x}'_i) , $i = 1, 2, \dots, n$ S_1 times with replacement and probabilities g_i .
- 4: Estimate $\boldsymbol{\beta}$ using ordinary least squares in the model $E(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, \mathbf{y} being an S_1 dimensional vector of realizations of \mathbf{Y} , and \mathbf{X} an $S_1 \times k$ matrix from the previous stage.*
- 5: Repeat this process S_2 times.
- 6: The distribution of $\boldsymbol{\beta}^{(s_2)}$ is the Bayesian distribution of $\boldsymbol{\beta}$.

*Ordinary least squares is the posterior mean of $\boldsymbol{\beta}$ using Jeffrey's prior in a linear regression.

5.5 Bayesian model averaging

Bayesian model averaging (BMA) is an approach which takes into account model uncertainty. In particular, we consider uncertainty in the regressors (variable selection) in a regression framework where there are k possible explanatory variables. This implies $\mathcal{M} = \{M_1, M_2, \dots, M_{2^k}\}$ potential models indexed by parameters $\boldsymbol{\theta}_m$, $m = 1, 2, \dots, 2^k$. Following Simmons et al. (2010), the posterior model probability is

$$\pi(M_j|\mathbf{y}) = \frac{p(\mathbf{y}|M_j)\pi(M_j)}{\sum_{m=1}^{2^k} p(\mathbf{y}|M_m)\pi(M_m)},$$

where $\pi(M_j)$ is the prior model probability,³³ $p(\mathbf{y}|M_j) = \int_{\Theta_j} p(\mathbf{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$ is the marginal likelihood, and $\pi(\boldsymbol{\theta}_j|M_j)$ is the prior distribution of $\boldsymbol{\theta}_j$.

Following Raftery (1993), the posterior distribution of $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \sum_{m=1}^{2^k} \pi(\boldsymbol{\theta}_m|\mathbf{y}, M_m)\pi(M_m|\mathbf{y})$$

where $\pi(\boldsymbol{\theta}_m|\mathbf{y}, M_m)$ is the posterior distribution of $\boldsymbol{\theta}$ under model m , $E(\boldsymbol{\theta}|\mathbf{y}) = \sum_{m=1}^{2^k} \hat{\boldsymbol{\theta}}_m\pi(M_m|\mathbf{y})$, $Var(\boldsymbol{\theta}|\mathbf{y}) = \sum_{m=1}^{2^k} \pi(M_m|\mathbf{y})\widehat{Var}(\boldsymbol{\theta}|\mathbf{y}, M_m) + \sum_{m=1}^{2^k} \pi(M_m|\mathbf{y})(\hat{\boldsymbol{\theta}}_m - E(\boldsymbol{\theta}|\mathbf{y}))^2$, $\hat{\boldsymbol{\theta}}_m$ and $\widehat{Var}(\boldsymbol{\theta}|\mathbf{y}, M_m)$

³³We attach equal prior probabilities to each model. However, this choice gives more prior probability to the set of models of medium size (think about the k -th row of Pascal's triangle). An interesting alternative is to use the Beta-Binomial prior proposed by Ley and Steel (2009).

are the posterior mean and variance under model m , respectively.

The posterior variance highlights how the BMA method takes into account model uncertainty. The first term is the weighted variance of each model, averaged over all potential models, and the second term indicates how stable the estimates are across models. The more the estimates differ between models, the greater is the posterior variance.

The posterior predictive distribution is

$$\pi(\mathbf{y}^{New}|\mathbf{y}) = \sum_{m=1}^{2^k} p_m(\mathbf{y}^{New}|\mathbf{y}, M_m)\pi(M_m|\mathbf{y})$$

where $p_m(\mathbf{y}^{New}|\mathbf{y}, M_m) = \int_{\Theta_m} p(\mathbf{y}^{New}|\mathbf{y}, \boldsymbol{\theta}_m, M_m)\pi(\boldsymbol{\theta}_m|\mathbf{y}, M_m)d\boldsymbol{\theta}_m$ is the posterior predictive distribution under model m .

Another important statistic in BMA is the posterior inclusion probability associated with variable \mathbf{x}_l , $l = 1, 2, \dots, k$, which is

$$PIP(\mathbf{x}_l) = \sum_{m=1}^{2^k} \pi(M_m|\mathbf{y}) \times I_{l,m},$$

where $I_{l,m} = \left\{ \begin{array}{ll} 1 & \text{if } \mathbf{x}_l \in M_m \\ 0 & \text{if } \mathbf{x}_l \notin M_m \end{array} \right\}$.

Kass and Raftery (1995) suggest that posterior inclusion probabilities (PIP) less than 0.5 are evidence against the regressor, $0.5 \leq PIP < 0.75$ is weak evidence, $0.75 \leq PIP < 0.95$ is positive evidence, $0.95 \leq PIP < 0.99$ is strong evidence, and $PIP \geq 0.99$ is very strong evidence.

BMA allows incorporating model uncertainty in a regression framework, but sometimes it is desirable to select just one model. Two compelling alternatives are the model with the highest posterior model probability, and the median probability model. The latter is the model which includes every predictor that has posterior inclusion probability higher than 0.5. The first model is the best alternative for prediction in the case of a 0–1 loss function (Clyde and George, 2004), whereas the second is the best alternative when there is a quadratic loss function in prediction (Barbieri and Berger, 2004).

There are two main computational issues in implementing BMA. First, the marginal likelihood $p(\mathbf{y}|M_j) = \int_{\Theta_j} p(\mathbf{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$ most of the time does not have an analytic solution, and second, the number of models in the model space is 2^k , which sometimes can be enormous.

The Bayesian information criterion is a possible solution for the first issue. Defining $h(\boldsymbol{\theta}|M_j) = -\frac{\log(p(\mathbf{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j))}{n}$, then $p(\mathbf{y}|M_j) = \int_{\Theta_j} \exp\{-nh(\boldsymbol{\theta}|M_j)\} d\boldsymbol{\theta}_j$. If n is sufficiently large ($n \rightarrow \infty$), we can make the following assumptions (Hoeting et al., 1999):

- We can use the Laplace method for approximating integrals (Tierney and Kadane, 1986).
- The posterior mode is reached at the same point as the maximum likelihood estimator (MLE), denoted by $\hat{\boldsymbol{\theta}}_{MLE}$.

We get the following results under these assumptions:

$$p(\mathbf{y}|M_j) \approx \left(\frac{2\pi}{n}\right)^{k_j/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-nh(\hat{\boldsymbol{\theta}}_j^{MLE}|M_j)\right\}, \quad n \rightarrow \infty,$$

where $\boldsymbol{\Sigma}$ is the Hessian matrix of $h(\hat{\boldsymbol{\theta}}_j^{MLE}|M_j)$, and $k_j = \dim\{\boldsymbol{\theta}_j\}$.

This implies

$$\log(p(\mathbf{y}|M_j)) \approx \frac{k_j}{2} \log(2\pi) - \frac{k_j}{2} \log(n) - \frac{1}{2} \log(|\boldsymbol{\Sigma}|) + \log(p(\mathbf{y}|\hat{\boldsymbol{\theta}}_j^{MLE}, M_j)) + \log(\pi(\hat{\boldsymbol{\theta}}_j^{MLE}|M_j)), \quad n \rightarrow \infty.$$

Since $\frac{k_j}{2} \log(2\pi)$ and $\log(\pi(\hat{\boldsymbol{\theta}}_j^{MLE}|M_j))$ are constants as functions of \mathbf{y} , and $|\boldsymbol{\Sigma}|$ is bounded by a finite constant, we have

$$\log(p(\mathbf{y}|M_j)) \approx \frac{k_j}{2} \log(n) + \log(p(\mathbf{y}|\hat{\boldsymbol{\theta}}_j^{MLE}, M_j)) = -\frac{BIC}{2}, \quad n \rightarrow \infty.$$

The second computational issue, which is related to the size of the model space (2^k), is basically a problem of ranking models. This can be tackled using different approaches, such as Occam's window criterion (Madigan and Raftery, 1994; Raftery et al., 1997), reversible jump Markov chain Monte Carlo computation (Green, 1995), Markov chain Monte Carlo model composition (Madigan et al., 1995b), and multiple testing using intrinsic priors (Casella and Moreno, 2006) or nonlocal prior densities (Jhonson and Rossell, 2012). In this GUI we focus on Occam's window and Markov chain Monte Carlo model composition.³⁴

In Occam's window, a model is discarded if its predictive performance is much worse than that of the best model (Madigan and Raftery, 1994; Raftery et al., 1997). Thus, models not belonging to $\mathcal{M}' = \left\{M_j : \frac{Max_r \pi(M_r|\mathbf{y})}{\pi(M_j|\mathbf{y})} \leq c\right\}$ should be discarded, where c is chosen by the user (Madigan and Raftery (1994) propose $c = 20$). In addition, complicated models than are less supported by the data than simpler models are also discarded, that is, $\mathcal{M}'' = \left\{M_j : \exists M_m \in \mathcal{M}', M_m \subset M_j, \frac{\pi(M_m|\mathbf{y})}{\pi(M_j|\mathbf{y})} > 1\right\}$. Then, the set of models used in BMA is $\mathcal{M}^* = \mathcal{M}' \cap \mathcal{M}''^c \in \mathcal{M}$. Raftery et al. (1997) find that the number of models in \mathcal{M}^* is normally less than 25.

However, the previous theoretical framework requires calculating $Max_r \pi(M_r|\mathbf{y})$, which implies calculating all possible models in \mathcal{M} . This is computationally burdensome. Hence,

³⁴Variable selection (model selection) is a topic related to model uncertainty. Approaches such as stochastic search variable selection (spike and slab) (George and McCulloch, 1993, 1997; Ishwaran and Rao, 2005) and Bayesian Lasso (Park and Casella, 2008) are good examples of how to tackle this issue. Future developments will include these approaches.

a heuristic approach is proposed by [Raftery et al. \(2012\)](#) based on ideas of [Madigan and Raftery \(1994\)](#). The search strategy is based on a series of nested comparisons of ratios of posterior model probabilities. Let M_0 be a model with one regressor less than model M_1 . If $\log(\pi(M_0|\mathbf{y})/\pi(M_1|\mathbf{y})) > \log(O_R)$, then M_1 is rejected and M_0 is considered; if $\log(\pi(M_0|\mathbf{y})/\pi(M_1|\mathbf{y})) \leq -\log(O_L)$, then M_0 is rejected, and M_1 is considered; and if $\log(O_L) < \log(\pi(M_0|\mathbf{y})/\pi(M_1|\mathbf{y})) \leq \log(O_R)$, M_0 and M_1 are considered. Here O_R is a number specifying the maximum ratio for excluding models in Occam’s window, and $O_L = 1/O_R^2$ is defined by default in [Raftery et al. \(2012\)](#). The search strategy can be “up,” adding one regressor, or “down,” dropping one regressor (see [Madigan and Raftery \(1994\)](#), down and up algorithms for details). The leaps and bounds algorithm ([Furnival and Wilson, 1974](#)) is implemented to improve the computational efficiency of this search strategy ([Raftery et al., 2012](#)). Once the set of potentially acceptable models is defined, we discard all the models that are not in \mathcal{M}' , and the models that are in \mathcal{M}'' where 1 is replaced by $\exp\{O_R\}$ due to the leaps and bounds algorithm giving an approximation to BIC, so as to ensure that no good models are discarded.

The second approach that we consider in this GUI to tackle the model space size issue is Markov chain Monte Carlo model composition ([Madigan et al., 1995a](#)). In particular, given the space of models \mathcal{M} , we simulate a chain of M_s models, $s = 1, 2, \dots, S \ll 2^k$, where the algorithm randomly extracts a candidate model M_c from a neighborhood of models ($nb d(M)$) that consists of the actual model itself and the set of models with either one variable more or one variable less ([Raftery et al., 1997](#)). Therefore, there is a transition kernel in the space of models $q(M \rightarrow M_c)$, such that $q(M \rightarrow M_c) = 0 \forall M_c \notin nb d(M)$ and $q(M \rightarrow M_c) = \frac{1}{|nb d(M)|} \forall M \in nb d(M)$, $|nb d(M)|$ being the number of neighbors of M . This candidate model is accepted with probability

$$\alpha(M_{s-1}, M_c) = \text{Min} \left\{ \frac{|nb d(M)|p(\mathbf{y}|M_c)\pi(M_c)}{|nb d(M_c)|p(\mathbf{y}|M_{(s-1)})\pi(M_{(s-1)})}, 1 \right\}.$$

Observe that by construction $|nb d(M)| = |nb d(M_c)| = k$, except in extreme cases where a model has only one regressor or has all regressors.

Normal: The Gaussian linear model specifies $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ such that $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and the conjugate priors for the parameters are $\boldsymbol{\beta}|\sigma^2 \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 \mathbf{B}_0)$ and $\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$. Given the likelihood function, $p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}$, the marginal distribution $p(\mathbf{y})$ is $\mathcal{T}(\mathbf{X}\boldsymbol{\beta}_0, \frac{\alpha_0(\mathbf{I} + \mathbf{X}\mathbf{B}_0\mathbf{X}')}{\delta_0}, \delta_0)$.

- Bayesian information criterion: We implement the BIC approximation in the Gaussian linear model using the Occam’s window approach.
- Markov chain Monte Carlo model composition (MC3): We implement the Gaussian linear model using MC3 using the marginal multivariate Student’s t distribution.
- Instrumental variables: Consider the previous instrumental variable framework (see subsection 5.2) assuming $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(3, \mathbf{I})$ ([Karl and Lenkoski, 2012](#)). Then, the posterior distributions are

$$\begin{aligned}\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} &\sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{B}^*) \\ \boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} &\sim \mathcal{N}(\boldsymbol{\gamma}^*, \mathbf{G}^*) \\ \boldsymbol{\Sigma}^{-1}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}, \mathbf{X}, \mathbf{Z} &\sim \mathcal{W}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}^*)\end{aligned}$$

$$\begin{aligned}\text{where } \mathbf{B}^* &= (\omega_1^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' + \mathbf{I})^{-1}, \boldsymbol{\beta}^* = \mathbf{B}^* \left(\omega_1^{-1} \sum_{i=1}^n \left[\mathbf{x}_i \left(y_i - \frac{\sigma_{12}(x_{si} - \mathbf{w}_i' \boldsymbol{\gamma})}{\sigma_{22}} \right) \right] \right), \\ \omega_1 &= \sigma_{11} - \sigma_{12}^2 / \sigma_{22}, \mathbf{G}^* = (\omega_2^{-1} \sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i' + \mathbf{I})^{-1}, \boldsymbol{\gamma}^* = \mathbf{G}^* \left(\omega_2^{-1} \sum_{i=1}^n \left[\mathbf{w}_i \left(x_{si} - \frac{\sigma_{12}(y_i - \mathbf{x}_i' \boldsymbol{\beta})}{\sigma_{11}} \right) \right] \right), \\ \omega_2 &= \sigma_{22} - \sigma_{12}^2 / \sigma_{11}, \boldsymbol{\alpha}^* = 3+n \text{ and } \boldsymbol{\Sigma}^* = \left[\mathbf{I} + \sum_{i=1}^n \begin{pmatrix} y_i - \mathbf{x}_i' \boldsymbol{\beta} \\ x_{si} - \mathbf{w}_i' \boldsymbol{\gamma} \end{pmatrix} (y_i - \mathbf{x}_i' \boldsymbol{\beta}, x_{si} - \mathbf{w}_i' \boldsymbol{\gamma}) \right]^{-1}.\end{aligned}$$

Lenkoski et al. (2013) propose an algorithm based on conditional Bayes factors (Dickey and Gunel, 1978) that allows embedding MC³ within a Gibbs sampling algorithm. Given the candidate (M_c^{2nd}) and actual (M_{s-1}^{2nd}) models for the iteration s in the second stage, the conditional Bayes factor is $CBF^{2nd} = \frac{p(\mathbf{y}|M_c^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}{p(\mathbf{y}|M_{s-1}^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}$, where $p(\mathbf{y}|M_c^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = \int_{\mathcal{M}^{2nd}} p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\beta}|M_c^{2nd}) d\boldsymbol{\beta} \propto |\mathbf{B}^*|^{1/2} \text{Exp} \left\{ \frac{1}{2} \boldsymbol{\beta}^{*'} \mathbf{B}^{*-1} \boldsymbol{\beta}^* \right\}$. For the first stage, $CBF^{1st} = \frac{p(\mathbf{y}|M_c^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma})}{p(\mathbf{y}|M_{s-1}^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma})}$, where $p(\mathbf{y}|M_c^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int_{\mathcal{M}^{1st}} p(\mathbf{y}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\gamma}|M_c^{1st}) d\boldsymbol{\gamma} \propto |\mathbf{G}^*|^{1/2} \text{Exp} \left\{ \frac{1}{2} \boldsymbol{\gamma}^{*'} \mathbf{G}^{*-1} \boldsymbol{\gamma}^* \right\}$.³⁵ These conditional Bayes factors assume $\pi(M^{1st}, M^{2sd}) \propto 1$. See Lenkoski et al. (2013) for more details of the instrumental variable BMA algorithm.³⁶

The Gaussian linear model is an example of a generalized linear model. A GLM is characterized by a distribution function that is in the exponential family, that is, $p_i(y_i|\theta_i, \phi) = h(y_i, \phi) \text{Exp} \left\{ (\theta_i y_i - b(\theta_i)) / a(\phi) \right\}$ (canonical representation), $y_i \stackrel{i.n.d.}{\sim} p_i$, $i = 1, 2, \dots, n$. It also has a linear predictor $\theta_i = \mathbf{x}_i' \boldsymbol{\beta}$, and a link function g such that $E(Y_i|x_i) \equiv \mu_i = b'(\theta_i) = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta})$ (g is monotonic and differentiable), and $V(Y_i) = b''(\theta_i) a(\phi)$ (McCullagh and Nelder, 1989). The identity function $\mu_i = \mathbf{x}_i' \boldsymbol{\beta}$ is the canonical link function in the case of the Gaussian model.³⁷ This statistical framework can help us to characterize:

Logit: The logit model is also a GLM, where the link function is $\mathbf{x}_i' \boldsymbol{\beta} = \log \left(\frac{\mu_i}{1 - \mu_i} \right)$. We carry out BMA using the BIC approximation and the Occam's window approach in the logit model.

Gamma: The gamma model is also a GLM, where the link function is $\mathbf{x}_i' \boldsymbol{\beta} = \mu_i^{-1}$. We carry out BMA using the BIC approximation and the Occam's window approach in the gamma model.

Poisson: The Poisson model is also a GLM, where the link function is $\mathbf{x}_i' \boldsymbol{\beta} = \log(\mu_i)$. We carry out BMA using the BIC approximation and the Occam's window approach in the Poisson model.

³⁵In the case that $\beta_s = 0$, the update is based on the seemingly unrelated regressions framework.

³⁶Koop et al. (2012) and Lenkoski et al. (2014) propose other frameworks for BMA taking into account endogeneity.

³⁷A canonical link functions is characterized by the existence of a sufficient statistic ($\mathbf{X}'\mathbf{y}$) equal in dimension to $\boldsymbol{\beta}$.

6 Concluding remarks

The Bayesian statistical framework has become very popular among scientists since the computational revolution in the 1990's. In particular, computationally burdensome procedures such as Markov chain Monte Carlo algorithms can be easily implemented nowadays. However, most of the open source software to apply these procedures requires programming skills. This may be one reason why the Bayesian framework is not very popular among applied researchers and practitioners. In this paper, we introduced a graphical user interface to implement Bayesian regression analysis under different frameworks, explaining the basic theory, so that users can understand the basic principles of Bayesian statistics and apply them easily. Our objective has been to increase the popularity of the Bayesian statistical framework among applied researchers and practitioners.

Table 1: Graphical user interfaces to carry out Bayesian regression analysis.

Name	Language	Models	Open source
ShinyStan	R+Stan	MCMC Implementation*	Yes
Bayesian regression: NP&P	MATLAB Compiler	Bayesian infinite-mixture regression Bayesian normal regression Hierarchical linear regression Binary regression Ordered regression Censoring regression Quantile regression Survival regression Density estimation Variable selection (spike-and-slab)	Yes
BugsXLA	OpenBUGS + Excel	Normal linear models GLM: Binomial GLM: Poisson GLM: Survival GLM: Multivariate categorical data Normal linear mixed Generalized linear mixed Bayesian variable selection Robust models	Yes
MATLAB toolkit: E&E ⁺	MATLAB	Linear Regression Regression with non-spherical errors Regime switch regression Regression with restricted parameters Seemingly unrelated regression (SUR) Vector AutoRegression (VAR) Instrumental variable Probit and logit Tobit Model Panel Data Analysis Stochastic search variable selection Highest posterior density (HPD) region Marginal likelihood of linear regression	No
Stata	Stata	MCMC implementation*	No
BayES	C++	Simple linear model Random-effects Random-coefficients Stochastic frontiers Inefficiency-effects Random-effects stochastic frontiers Dynamic stochastic frontier Probit and logit Random-effects probit and logit Multinomial probit and logit Ordered probit and logit Poisson and negative-binomial Type I Tobit Type II Tobit Seemingly unrelated regressions (SUR) Vector Autoregressive (VAR)	No

*User should define prior and likelihood.

⁺Toolkit on econometrics and economics teaching.

Table 2: Libraries and commands in BESmarter GUI.

Univariate models			
Model	Library	Command	Reference
Normal	MCMCpack	MCMCregress	Martin et al. (2018)
Logit	MCMCpack	MCMClogit	Martin et al. (2018)
Probit	bayesm	rbprobitGibbs	Rossi (2017)
Multinomial(Mixed) Probit	bayesm	rmnpGibbs	Rossi (2017)
Multinomial(Mixed) Logit	bayesm	rmnlIndepMetrop	Rossi (2017)
Ordered Probit	bayesm	rordprobitGibbs	Rossi (2017)
Negative Binomial(Poisson)	bayesm	rnebinRw	Rossi (2017)
Tobit	MCMCpack	MCMCtobit	Martin et al. (2018)
Quantile	MCMCpack	MCMCquantreg	Martin et al. (2018)
Multivariate models			
Model	Library	Command	Reference
Multivariate	bayesm	rmultireg	Rossi (2017)
Seemingly Unrelated Regression	bayesm	rsurGibbs	Rossi (2017)
Instrumental Variable	bayesm	rivGibbs	Rossi (2017)
Bivariate Probit	bayesm	rmvpGibbs	Rossi (2017)
Hierarchical longitudinal models			
Model	Library	Command	Reference
Normal	MCMCpack	MCMChregress	Martin et al. (2018)
Logit	MCMCpack	MCMChlogit	Martin et al. (2018)
Poisson	MCMCpack	MCMChpoisson	Martin et al. (2018)
Bayesian Bootstrap			
Model	Library	Command	Reference
Bayesian bootstrap	bayesboot	bayesboot	Baath (2018)
Bayesian model averaging			
Model	Library	Command	Reference
Normal (BIC)	BMA	bic.glm	Raftery et al. (2012)
Normal (MC ³)	BMA	MC3.REG	Raftery et al. (2012)
Normal (instrumental variables)	ivbma	ivbma	Lenkoski et al. (2013)
Logit (BIC)	BMA	bic.glm	Raftery et al. (2012)
Gamma (BIC)	BMA	bic.glm	Raftery et al. (2012)
Poisson (BIC)	BMA	bic.glm	Raftery et al. (2012)
Diagnostics			
Diagnostic	Library	Command	Reference
Trace plot	coda	traceplot	Plummer et al. (2016)
Autocorrelation plot	coda	autocorr.plot	Plummer et al. (2016)
Geweke test	coda	geweke.diag	Plummer et al. (2016)
Raftery & Lewis test	coda	raftery.diag	Plummer et al. (2016)
Heidelberger & Welch test	coda	heidel.diag	Plummer et al. (2016)

Table 3: Datasets templates in folder *DataSim*.

Univariate models		
Model	Data set file	Data set simulation
Normal	11SimNormalmodel.csv	11SimNormal.R
Logit	12SimLogitmodel.csv	12SimLogit
Probit	13SimProbitmodel.csv	13SimProbit.R
Multinomial(Mixed) Probit	14SimMultProbmodel.csv	14SimMultinomialProbit.R
Multinomial(Mixed) Logit	15SimMultLogitmodel.csv	15SimMultinomialLogit.R
Ordered Probit	16SimOrderedProbitmodel.csv	16SimOrderedProbit.R
Negative Binomial(Poisson)	17SimNegBinmodel.csv	17SimNegBin.R
Tobit	18SimTobitmodel.csv	18SimTobit.R
Quantile	19SimQuantilemodel.csv	19SimQuantile.R
Multivariate models		
Model	Data set file	Data set simulation
Multivariate	21SimMultivariate.csv	21SimMultReg.R
Seemingly Unrelated Regression	22SimSUR.csv	22SimSUR.R
Instrumental Variable	23SimIV.csv	23SimIV.R
Bivariate Probit	24SimMultProbit.csv	24SimMultProbit.R
Hierarchical longitudinal models		
Model	Data set file	Data set simulation
Normal	31SimLogitudinalNormal.csv	31SimLogitudinalNormal.R
Logit	32SimLogitudinalLogit.csv	32SimLogitudinalLogit.R
Poisson	33SimLogitudinalPoisson.csv	33SimLogitudinalPoisson.R
Bayesian Bootstrap		
Model	Data set file	Data set simulation
Bayesian bootstrap	41SimBootstrapmodel.csv	41SimBootstrapmodel.R
Bayesian model averaging		
Model	Data set file	Data set simulation
Normal (BIC)	511SimNormalBMA.csv	511SimNormalBMA.R
Normal (MC ³)	512SimNormalBMA.csv	512SimNormalBMA.R
Normal (instrumental variables)	513SimNormalBMAivYXW.csv	513SimNormalBMAiv.R
	513SimNormalBMAivZ.csv	
Logit (BIC)	52SimLogitBMA.csv	52SimLogitBMA.R
Gamma (BIC)	53SimGammaBMA.csv	53SimGammaBMA.R
Poisson (BIC)	53SimPoissonBMA.csv	53SimPoissonBMA.R

Table 4: Real datasets in folder *DataApp*.

Univariate models		
Model	Data set file	Dependent variable
Normal	1ValueFootballPlayers.csv	log(Value)
Logit	2HealthMed.csv	Hosp
Probit	2HealthMed.csv	Hosp
Multinomial(Mixed) Probit	Fishing.csv	mode
Multinomial(Mixed) Logit	Fishing.csv	mode
Ordered Probit	2HealthMed.csv	MedVisPrevOr
Negative Binomial(Poisson)	2HealthMed.csv	MedVisPrev
Tobit	1ValueFootballPlayers.csv	log(ValueCens)
Quantile	1ValueFootballPlayers.csv	log(Value)
Multivariate models		
Model	Data set file	Dependent variable
Multivariate	4Institutions.csv	logpcGDP95 and PAER
Seemingly Unrelated Regression	5Institutions.csv	logpcGDP95 and PAER
Instrumental Variable	6Institutions.csv	logpcGDP95 and PAER
Bivariate Probit	7HealthMed.csv	$y = [\text{Hosp SHI}]'$
Hierarchical longitudinal models		
Model	Data set file	Dependent variable
Normal	8PublicCap.csv	log(gsp)
Logit	9VisitDoc.csv	DocVis
Poisson	9VisitDoc.csv	DocNum
Bayesian Bootstrap		
Model	Data set file	Dependent variable
Bayesian bootstrap	1ValueFootballPlayers.csv	log(Value)
Bayesian model averaging		
Model	Data set file	Dependent variable
Normal (BIC)	10ExportDiversificationHHI.csv	avghhi
Normal (MC ³)	10ExportDiversificationHHI.csv	avghhi
Normal (instrumental variables)	11ExportDiversificationHHI.csv	avghhi and avglgdpcap
	12ExportDiversificationHHIInstr.csv	
Logit (BIC)	13InternetMed.csv	internet
Gamma (BIC)	14ValueFootballPlayers.csv	log market value
Poisson (BIC)	15Fertile2.csv	ceb

References

- Acemoglu, D., Johnson, S., and Robinson, J. (2001). The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91(5):1369–1401.
- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.
- Baath, R. (2018). *Package bayesboot*.
- Barbieri, M. and Berger, J. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897.
- Cameron, C. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge.
- Casella, G. and Moreno, E. (2006). Objective Bayesian variable selection. *Journal of the American Statistical Association*, 101(473):157–167.
- Chang, W. (2018). *Web Application Framework for R: Package shiny*. R Studio.
- Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, 51:79–99.
- Chib, S. and Carlin, B. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9:17–26.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.
- Clyde, M. and George, E. (2004). Model uncertainty. *Statistical Science*, 19(1):81–94.
- Dickey, J. M. and Gunel, E. (1978). Bayes factors from mixed probabilities. *Journal of the Royal Statistical Society: Series B (Methodology)*, 40:43–46.
- Edwards, Y. D. and Allenby, G. M. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research*, 40:321–334.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26.
- Furnival, G. M. and Wilson, R. W. (1974). Regressions by leaps and bounds. *Technometrics*, 16(4):499–511.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.
- George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. and McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.
- Geweke, J. (1992). *Bayesian Statistics*, chapter Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Clarendon Press, Oxford, UK.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57:97–109.
- Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417.
- Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, 33(2):730–773.
- Jetter, M. and Ramírez Hassan, A. (2015). Want export diversification? Educate the kids first. *Economic Inquiry*, 53(4):1765–1782.
- Jhonson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660.
- Karabatsos, G. (2016). A menu-driven software package of Bayesian nonparametric (and parametric) mixed models for regression analysis and density estimation. *Behavior Research Methods*, 49:335–362.
- Karl, A. and Lenkoski, A. (2012). Instrumental variable Bayesian model averaging via conditional Bayes factor. Technical report, Heidelberg University.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Koop, G., León-Gonzalez, R., and Strachan, R. (2012). Bayesian model averaging in the instrumental variable regression model. *Journal of Econometrics*, 171:237–250.
- Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578.
- Lenkoski, A., Eicher, T. S., and Raftery, A. (2014). Two-stage Bayesian model averaging in endogeneous variable models. *Econometric Reviews*, 33.

- Lenkoski, A., Karl, A., and Neudecker, A. (2013). *Package ivbma*.
- Ley, E. and Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24(4):651–674.
- Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association*, 89(428):1535–1546.
- Madigan, D., York, J., and Allard, D. (1995a). Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232.
- Madigan, D., York, J. C., and Allard, D. (1995b). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.
- Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42(9):1–21.
- Martin, A. D., Quinn, K. M., and Park, J. H. (2018). *Package MCMCpack*.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- McCulloch, R. and Rossi, P. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64:207–240.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys*, 21:1087–1092.
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., Almond, R., and Magnusson, A. (2016). *Output Analysis and Diagnostics for MCMC*.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raftery, A., Hoeting, J., Volinsky, C., Painter, I., and Yeung, K. Y. (2012). *Package BMA*.
- Raftery, A. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7:493–497.
- Raftery, A. E. (1993). Bayesian model selection in structural equation models. *Sage Focus Editions*, 154:163–163.
- Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.
- Ramírez Hassan, A. (2017). The interplay between the Bayesian and frequentist approaches: a general nesting spatial panel data model. *Spatial Economic Analysis*, 12(1):92–112.

- Ramírez-Hassan, A. (2019). Dynamic variable selection in dynamic logistic regression: An application to Internet subscription. *Empirical Economics*, Forthcoming.
- Ramírez Hassan, A., Cardona Jiménez, J., and Cadavid Montoya, R. (2013). The impact of subsidized health insurance on the poor in Colombia: Evaluating the case of Medellín. *Economía Aplicada*, 17(4):543–556.
- Ramírez-Hassan, A. Guerra-Urzola, R. (2020). Bayesian treatment effects due to a subsidized health program: The case of preventive health care utilization in medellín (Colombia). *Empirical Economics*, Forthcoming.
- Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, second edition.
- Robert, C. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer, second edition.
- Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.
- Rossi, P. (2017). *Package bayesm*.
- Rossi, P., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons.
- Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134.
- Serna Rodríguez, M., Ramírez Hassan, A., and Coad, A. (2018). Uncovering value drivers of high performance soccer players. *Journal of Sport Economics*, Accepted manuscript.
- Simmons, S. J., Fang, F., Fang, Q., and Ricanek, K. (2010). Markov chain Monte Carlo model composition search strategy for quantitative trait loci in a Bayesian hierarchical model. *World Academy of Science, Engineering and Technology*, 63:58–61.
- Stan Development Team (2017). shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. R package version 2.3.0.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.
- Winkelmann, R. (2004). Health care reform and the number of doctor visits - An econometric analysis. *Journal of Applied Econometrics*, 19(4):455–472.
- Woodward, P. (2005). BugsXLA: Bayes for the common man. *Journal of Statistical Software*, 14(5):1–18.
- Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning, Mason, Ohio: South-Western, fifth edition.