Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

A GUIded tour around Bayesian regression

Andrés Ramírez–Hassan
Universidad EAFIT

Mateo Graciano-Londoño
Protección

Abstract

This paper presents a graphical user interface (GUI) to perform Bayesian regression analysis in a very friendly environment without any programming skills. This paper is designed for teaching and applied purposes at introductory level; we present the basic theory underlining 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.

Keywords: Bayesian regression, Graphical user interface, Variable selection.

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 show the basic theoretical framework, and the tutorial for implementing these models. Therefore, practitioners and applied researchers can apply Bayesian regression analysis understanding its theoretical foundation without requiring programming skills. The latter seems to be one relevant impediment to increase the use of the Bayesian framework (Woodward 2005).

Table 1 shows available graphical user interfaces to perform Bayesian regression analysis. ShinyStan is a very flexible open source, but users do require programming skills. BugsXLA is a less flexible open source; However, users do not require programming skills. On the other hand, Matlab toolkit, Stata and BayES are not open sources.

We developed our GUI based on an interactive web application using shiny (Chang 2018), and some libraries in R (R Core Team 2018). In Table 2 can be seen specific libraries and commands that are used in our GUI. It has 9 univariate models, 4 multivariate, 3 hierarchical longitudinal, Bayesian bootstrap and 6 Bayesian model averaging frameworks. In addition, it gives basic summaries and diagnostics of the posterior chains, as well as the posterior

chains, and different plots such as trace, autocorrelation and densities. Regarding flexibility and possibilities is between ShinyStan and BugsXLA, users do not require any programming skills. Our GUI, BEsmarter,[1] is freely available at www.besmarter-team.org; see Appendix for instructions on how to download and install the software.

After this brief introduction, we present in Section 2 some general theoretical and computational aspects of the Bayesian framework. Section 3 shows specific technical details of the models implemented in our GUI. Section 4 introduces our GUI, and how to use it. Finally, Section 5 presents some concluding remarks and future developments.

2. Bayesian framework: A brief summary of theory and computation

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

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

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

The Bayesian framework allows to report the full posterior distribution. In case that just one value of the posterior distribution should be reported, decision theory offers an elegant framework, for instance, a quadratic loss function implies reporting posterior means. 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), $PO_{12} = \frac{p(H_1|\boldsymbol{y})}{p(H_2|\boldsymbol{y})} = \frac{p(\boldsymbol{y}|H_1)}{p(\boldsymbol{y}|H_2)} \times \frac{\pi(H_1)}{\pi(H_2)}$, 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 $p(\boldsymbol{y}|H_j) = \int_{\boldsymbol{\Theta}_j} p(\boldsymbol{y}|\boldsymbol{\theta}_j, H_j)\pi(\boldsymbol{\theta}_j|H_j)d\boldsymbol{\theta}_j$ (Kass and Raftery 1995). Under this framework is also easy to account for model uncertainty based on posterior model probabilities, $p(M_j|\boldsymbol{y}) = \frac{\pi(M_j)p(\boldsymbol{y}|M_j)}{\sum_{m=1}^{M}\pi(M_m)p(\boldsymbol{y}|M_m)}$, where $\pi(M_j)$ is the prior model probability, and $p(\boldsymbol{y}|M_j) = \int_{\boldsymbol{\Theta}_j} p(\boldsymbol{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$ is the marginal likelihood under model j, $j = 1, 2, \ldots, M$. In addition, it is also easy to perform 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 perform inference of functions of these parameters that are well identified. Finally, predictive distributions acknowledge parameter uncertainty, $\pi(\boldsymbol{y}^{New}|\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{y}^{New}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{y}^{New}|\boldsymbol{\theta}^{(s)})$, $\boldsymbol{\theta}^{(s)} \sim \pi(\boldsymbol{\theta}|\boldsymbol{y})$.

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 perform this task (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953; Hastings 1970). Given a target distribution

[1] Bayesian Econometrics: simulations, models and applications to research, teaching and encoding with responsibility.

$\pi(\boldsymbol{\theta}|\boldsymbol{y})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathcal{R}^k$, such that $\pi(d\boldsymbol{\theta}^c|\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}, d\boldsymbol{\theta}^c)\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$ 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

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

where $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ is a proposal density function to draw candidates $(\boldsymbol{\theta}^c)$, $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = min\left\{\frac{\pi(\boldsymbol{\theta}^c|\boldsymbol{y})q(\boldsymbol{\theta}^c, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)}, 1\right\} = min\left\{\frac{p(\boldsymbol{y}|\boldsymbol{\theta}^c)\pi(\boldsymbol{\theta}^c)q(\boldsymbol{\theta}^c, \boldsymbol{\theta})}{p(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)}, 1\right\}$ is the probability to move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^c$,[2] $\delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}^c)$ is equal to 1 if $\boldsymbol{\theta} \in d\boldsymbol{\theta}^c$ and 0 otherwise, and $\left[1 - \int_{\boldsymbol{\Theta}} q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)d\boldsymbol{\theta}^c\right]$ is the probability that the chain remains at $\boldsymbol{\theta}$ (Chib and Greenberg 1995).[3] 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.[4] 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 $q(\boldsymbol{\theta}^{(s-1)}, \cdot)$, $s = 1, 2, \ldots, S$, 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, $s = 1, 2, \ldots, S$ iterations.
5: Return the values $\left\{\boldsymbol{\theta}^{(b)}, \boldsymbol{\theta}^{(b+d)}, \ldots, \boldsymbol{\theta}^{(S)}\right\}$ 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 computational demanding. This algorithm requires to define the proposal density, generates autocorrelated chains by construction, the chain remains at $\boldsymbol{\theta}^{(s-1)}$ if $\boldsymbol{\theta}^c$ is rejected, and it is necessary to discard some initial draws due to avoiding dependence to the initial value, and enough iterations to achieve convergence to the target distribution.

Regarding the proposal density, we implement in this GUI the random-walk and the tailored proposals. 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 $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = q(\boldsymbol{\theta}^c, \boldsymbol{\theta})$. 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, $\left(-\frac{\partial^2 ln\pi(\boldsymbol{\theta}|\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\hat{\boldsymbol{\theta}}}\right)^{-1}$.

A crucial point associated with the proposal densities is the acceptance rate. A 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 chain will converge too slowly. A sensible value depends on the dimension of the parameter

[2] $\pi(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) \neq 0$ is usually satisfied in practice.
[3] Observe that $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ depends on $\boldsymbol{y}$, and the proposal $q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)$ can depend or cannot depend on $\boldsymbol{y}$.
[4] Intuitively, the reversibility condition, $\pi(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = \pi(\boldsymbol{\theta}^c|\boldsymbol{y})q(\boldsymbol{\theta}^c, \boldsymbol{\theta})$, establishes that the probability to move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^c$ is equal that the probability to move from $\boldsymbol{\theta}^c$ to $\boldsymbol{\theta}$.

space. A rule of thumb in the case that the dimension is lower or equal to 2 is 0.50. If the dimension is grater than 2, the acceptance rate should be around 0.25 (Roberts, Gelman, and Gilks 1997).

Regarding convergence issues, we implement some diagnostics in order to check the adequacy of the posterior chains (Plummer, Best, Cowles, Vines, Sarkar, Bates, Almond, and Magnusson 2016). In particular, trace plots, which should look stable, and autocorrelation plots that should decrease very fast. In addition, 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 the target distribution converged. Raftery and Lewis' test (Raftery and Lewis 1992) is designed to calculate the approximate number of iterations (S), burn–in (b) and thinning parameter (d) required to estimate $p\left[H(\boldsymbol{\theta}) \leq h\right]$, $H(\boldsymbol{\theta}) : \mathcal{R}^k \to \mathcal{R}$, given a specific quantile of interest (q), precision (r) and probability (p). Their diagnostic is based on the dependence factor, $I = \frac{S \times b}{S_{Min}}$, $S_{Min} = \Phi^{-1}\left(\frac{1}{2}(p+1)\right)^2 q(1-q)/r^2$, $\Phi(\cdot)$ 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, $\text{CVM}(B_S) = \int_0^1 B_S(t)^2 dt$, $B_S(t) = (S_{[St]} - [St]\,\bar{\boldsymbol{\theta}}^S)/(Sp(0))^{1/2}$, $S_S = \sum_{s=1}^{S} \boldsymbol{\theta}^{(s)}$, $\bar{\boldsymbol{\theta}}^S = S_S/S$, $p(0)$ 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),[5] 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 conditional posterior distributions are not standard. In this case, Metropolis-within-Gibbs algorithm can be adopted. Rather than simulating $\boldsymbol{\theta}_l^{(s)} \sim \pi(\boldsymbol{\theta}_l|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{l-1}, \boldsymbol{\theta}_{l+1}, \ldots, \boldsymbol{\theta}_k)$, 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]This statement requires the positivity condition, if $\pi(\boldsymbol{\theta}_l|\boldsymbol{y}) > 0$, then $\pi(\boldsymbol{\theta}|\boldsymbol{y}) > 0$, $l = 1, 2, \ldots, k$.
[6]Note that $\boldsymbol{\theta}_l$, $l = 1, 2, \ldots, k$ can be either scalars or vectors in Algorithm A2.

Algorithm A2 The Gibbs sampler algorithm

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

3. Models

3.1. Univariate models

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

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

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\beta}^*, \sigma^2 \boldsymbol{B}),^{7}$$

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X} \sim \mathcal{IG}(\alpha^*/2, \delta^*/2),^{8}$$

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

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

Logit: In the logit model the dependent variable is binary, then it follows a Bernoulli distribution, $y_i \overset{i.n.d}{\sim} \mathcal{B}(\pi_i)$,[9] that is $p(y_i = 1) = \pi_i$, such that $\pi_i = \frac{exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}{1+exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}$. We specify a Gaussian distribution as prior $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$.

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, Quinn, and Park 2011, 2018). In particular, the proposal is multivariate Gaussian centered

[7] $\mathcal{N}$ denotes a normal density.
[8] $\mathcal{IG}$ denotes an inverse gamma density.
[9] B denotes a Bernoulli density.

at the current value, and covariance matrix $\boldsymbol{S}(\boldsymbol{B}_0^{-1} + \widehat{\boldsymbol{\Sigma}}^{-1})^{-1}\boldsymbol{S}$, where $\boldsymbol{S} = s\boldsymbol{I}_{dim\boldsymbol{\beta}}$, $s > 0$ is a tune parameter,[10] 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, $y_i = \begin{Bmatrix} 0, & y_i^* \leq 0 \\ 1, & y_i^* > 0 \end{Bmatrix}$ such that $y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \mu_i$, $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$.[11] This implies $p(y_i = 1) = \pi_i = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$.

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.

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

$$y_i^*|\boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X} \sim \begin{Bmatrix} \mathcal{TN}_{(-\infty,0]}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1) & , \ y_i = 0 \\ \mathcal{TN}_{(0,\infty)}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1) & , \ y_i = 1 \end{Bmatrix},$$ [12]

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

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

Multinomial probit: The multinomial probit model is used to model mutually exclusive discrete outcomes or qualitative response variables. We observe $y_{il} = \begin{Bmatrix} 1, & y_{il}^* \geq max(\boldsymbol{y}_i^*) \\ 0, & otherwise \end{Bmatrix}$ such that $\boldsymbol{y}_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{y}_i^*$ is an unobserved latent L dimensional vector, $\boldsymbol{X}_i$ is a $L \times k$ matrix of regressors for each alternative, $l = 1, 2, \ldots, L$. We take into account simultaneously alternative-varying regressors (alternative attributes) and alternative-invariant regressors (individual characteristics).[13] $\boldsymbol{y}_i^*$ can be stacked up into a multiple regression with correlated stochastic errors, $\boldsymbol{y}^* = \boldsymbol{X\beta} + \boldsymbol{\mu}$, where $\boldsymbol{y}^{*\prime} = [\boldsymbol{y}_1^{*\prime}, \boldsymbol{y}_2^{*\prime}, \ldots, \boldsymbol{y}_n^{*\prime}]$, $\boldsymbol{X}' = [\boldsymbol{X}_1', \boldsymbol{X}_2', \ldots, \boldsymbol{X}_n']$, and $\boldsymbol{\mu}' = [\boldsymbol{\mu}_1', \boldsymbol{\mu}_2', \ldots, \boldsymbol{\mu}_n']$.

Following the practice of expressing $y_{il}^{*\prime}$ relative to ${y_{iL}^*}'$ by letting $\boldsymbol{w}_i' = [w_{i1}, w_{i2}, \ldots, w_{iL-1}]$, $w_{il} = y_{il}^* - y_{iL}^*$, we can write $\boldsymbol{w}_i = \boldsymbol{R}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{R}_i$ denotes the $(L-1) \times k$ matrix of observation vectors obtained from $\boldsymbol{X}_i$ by subtracting the L-th row from the first $(L-1)$ rows.

[10] Tune parameters should be set such that achieve reasonable diagnostic criteria.

[11] The 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.

[12] $\mathcal{TN}$ denotes a truncated normal density.

[13] Note 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.

The likelihood function in this model is $p(\boldsymbol{\beta}, \boldsymbol{\Omega}|\boldsymbol{y}, \boldsymbol{R}) = \prod_{i=1}^{n}\prod_{l=1}^{L} p_{il}^{y_{il}}$ where $p_{il} = p(y_{il}^* \geq max(\boldsymbol{y}_i^*))$. We assume independent priors, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\boldsymbol{\Omega}^{-1} \sim \mathcal{W}(\alpha_0, \boldsymbol{\Sigma}_0)$.[14] We can apply Gibbs sampling in this model because this is a standard Bayesian linear regression model when data augmentation in $\boldsymbol{w}$ is used. The posterior conditional distributions are

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

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

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

We can collapse the multinomial vector $\boldsymbol{y}_i$ into the indicator variable $d_i = \sum_{l=1}^{L-1} l \times I_{max(\boldsymbol{w}_l)=w_{il}}$,[15] then the distribution of $\boldsymbol{w}_i|\boldsymbol{\beta}, \boldsymbol{\Omega}^{-1}, d_i$ is a $L-1$ dimensional Gaussian distribution truncated over the appropriate cone in $\mathcal{R}^{L-1}$. McCulloch and Rossi (1994) propose to draw from the univariate conditional distributions $w_{il}|\boldsymbol{w}_{i,-l}, \boldsymbol{\beta}, \boldsymbol{\Omega}^{-1}, d_i \sim \mathcal{TN}_{I_{il}}(m_{il}, \tau_{ll}^2)$, where $I_{il} = \begin{Bmatrix} w_{il} > max(\boldsymbol{w}_{i,-l}, 0), & d_i = l \\ w_{il} < max(\boldsymbol{w}_{i,-l}, 0), & d_i \neq l \end{Bmatrix}$, and permuting the columns and rows of $\boldsymbol{\Omega}^{-1}$ such that the l-th column and row is the last,

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

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

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 $\frac{1}{\omega_{1,1}}\boldsymbol{\Omega}$.

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).[16]

In this setting there are L mutually exclusive alternatives, and the dependent variable y_i is equal to l if the lth alternative is chosen by individual i, $l = \{1, 2, \ldots, L\}$. The likelihood function is $p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} p(y_i = l|\boldsymbol{\beta}, \boldsymbol{X})$, where $p(y_i = l|\boldsymbol{\beta}, \boldsymbol{X}) = \frac{exp\{\boldsymbol{x}_{il}'\boldsymbol{\beta}\}}{\sum_{j=1}^{L} exp\{\boldsymbol{x}_{ij}'\boldsymbol{\beta}\}}$, $\boldsymbol{X}$ is a $nL \times k$ matrix, $k = k_1 + k_2$ is the total number of regressors, k_1 and k_2 are the number of

[14] $\mathcal{W}$ denotes the wishart density.

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

[16] The multinomial mixed logit model can be implemented as a conditional logit model.

alternative-varying and alternative-invariant regressors, respectively. In addition, we assume $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ as prior distribution.

As the multinomial logit model does not have a standard posterior distribution, we propose a "tailored" Metropolis-Hastings algorithm as proposal distribution. In particular, a multivariate student-t distribution with v degrees of freedom (tune 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 $y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \mu_i$, $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, such that $y_i = l$ if and only if $\alpha_{l-1} < y_i^* \leq \alpha_l$, $l = \{1, 2, \ldots, L\}$, where $\alpha_0 = -\infty$, $\alpha_1 = 0$ and $\alpha_L = \infty$.[17] Then, $p(y_i = l) = \Phi(\alpha_l - \boldsymbol{x}_i'\boldsymbol{\beta}) - \Phi(\alpha_{l-1} - \boldsymbol{x}_i'\boldsymbol{\beta})$, and the likelihood function is $p(\boldsymbol{\beta}, \boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n} p(y_i = l|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{X})$. The independent priors of this model are $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\gamma}_0, \boldsymbol{\Gamma}_0)$, $\boldsymbol{\gamma} = [\gamma_2, \gamma_3, \ldots, \gamma_{L-1}]'$, such that $\boldsymbol{\alpha} = \left[exp\{\gamma_2\}, \sum_{l=2}^{3} exp\{\gamma_l\}, \ldots, \sum_{l=2}^{L-1} exp\{\gamma_l\}\right]'$.

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 $\boldsymbol{y}^*$,

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

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

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

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

We assume that $y_i \overset{i.n.d}{\sim} \mathcal{NB}(\gamma, \theta_i)$, 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 $\theta_i = \frac{\lambda_i}{\lambda_i+\gamma}$, $\lambda_i = exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}$ and $\gamma = exp\{\alpha\}$. The independent priors for this model are $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\alpha \sim \mathcal{G}(\alpha_0, \delta_0)$.[18]

This model does not have standard conditional posterior distributions, then we use a random-walk Metropolis-Hastings algorithm where the proposal distribution for $\boldsymbol{\beta}$ is Gaussian centered at the current stage and covariance matrix $s_{\boldsymbol{\beta}}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ where $s_{\boldsymbol{\beta}}$ is a tune parameter and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ is the

[17] Identification issues imply 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.

[18] $\mathcal{G}$ denotes a gamma density.

maximum likelihood covariance estimator. In addition, the proposal for α is normal centered at the current value, and variance $s_\alpha^2\hat{\sigma}_\alpha^2$ where s_α is a tune 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 regressors are completely observed. In particular,

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

where $y_i^* \overset{i.n.d}{\sim} \mathcal{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2)$.[19]

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

The posteriors distributions are

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

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

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

Quantile: In quantile regression the location parameters vary according to the quantile of the dependent variable. Let $q_\tau(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau$ denote the τ-th $(0 < \tau < 1)$ quantile regression function of y_i given $\boldsymbol{x}_i$ such that $y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau + \mu_i$ where $\int_{-\infty}^0 f_\tau(\mu_i)d\mu_i = \tau$. In particular, $f_\tau(\mu_i) = \tau(1-\tau)exp\left\{\mu_i(\tau - I_{\mu_i<0})\right\}$ (asymmetric Laplace distribution). Kozumi and Kobayashi (2011) propose the location-scale mixture of normals 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 $z_i \sim \mathcal{N}(0,1)$, $e_i \perp z_i$.[20] As a consequence of this representation and $i.i.d$ sample, $p(\boldsymbol{y}|\boldsymbol{\beta}_\tau, \boldsymbol{e}, \boldsymbol{X}) \propto \left(\prod_{i=1}^n e_i^{-1/2}\right) exp\left\{-\sum_{i=1}^n \frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\tau - \theta e_i)^2}{2\psi^2 e_i}\right\}$.

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

[19] We can set L and U equal to $-\infty$ and ∞ to model censored data in just one side.
[20] $\mathcal{E}$ denotes an exponential density.

$$\boldsymbol{\beta}_\tau|\boldsymbol{e},\boldsymbol{y},\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\beta}_\tau^*,\boldsymbol{B}_\tau),$$
$$e_i|\boldsymbol{\beta}_\tau,\boldsymbol{y},\boldsymbol{X}\sim\mathcal{GIG}(1/2,\alpha_i^*,\delta_i^*),\text{ [21]}$$

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

3.2. Multivariate models

Multivariate regression: This model is used when there are m multiple dependent variables which share the same set of regressors, and their stochastic errors are contemporaneously correlated. In particular, $\boldsymbol{Y}=[\mathbf{y_1},\mathbf{y_2},\ldots,\mathbf{y_m}]$ is a $n\times m$ matrix that is generated as $\boldsymbol{Y}=\boldsymbol{XB}+\boldsymbol{U}$ where $\boldsymbol{X}$ is a $n\times k$ matrix, $\boldsymbol{B}=[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_m]$ is a $k\times m$ matrix of parameters, and $\boldsymbol{U}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m]$ is a matrix of stochastic random errors such that $\boldsymbol{u}_i\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$, $i=1,2,\ldots,n$ is each row of $\boldsymbol{U}$. Then, $p(\boldsymbol{B},\boldsymbol{\Sigma}|\boldsymbol{Y},\boldsymbol{X})\propto|\boldsymbol{\Sigma}|^{-n/2}\exp\left[-\frac{1}{2}tr(\boldsymbol{Y}-\boldsymbol{XB})'(\boldsymbol{Y}-\boldsymbol{XB})\boldsymbol{\Sigma}^{-1}\right]$, tr denotes the trace operator.

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

$$\boldsymbol{\beta}|\boldsymbol{\Sigma},\boldsymbol{Y},\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\beta}^*,\boldsymbol{\Sigma}\otimes\boldsymbol{\Delta}^*),$$
$$\boldsymbol{\Sigma}|\boldsymbol{Y},\boldsymbol{X}\sim\mathcal{IW}(\alpha^*,\boldsymbol{\Sigma}^*),\text{ [22]}$$

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

We can apply a Gibbs sampling algorithm in this model due to having standard conditional posterior distributions.

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

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

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

[21] $\mathcal{GIG}$ denotes a generalized inverse Gaussian density.
[22] $\mathcal{IW}$ denotes an inverse wishart density.

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

We can apply a Gibbs sampling algorithm in this model due to having standard conditional posterior distributions.

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

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

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

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

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

where $\boldsymbol{B}^* = (\omega_1^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{B}_0^{-1})^{-1}$, $\boldsymbol{\beta}^* = \boldsymbol{B}^*\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \omega_1^{-1}\sum_{i=1}^n \left[\boldsymbol{x}_i\left(y_i - \frac{\sigma_{12}(x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma})}{\sigma_{22}}\right)\right]\right)$,
$\omega_1 = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, $\boldsymbol{G}^* = (\omega_2^{-1}\sum_{i=1}^n \boldsymbol{w}_i\boldsymbol{w}_i' + \boldsymbol{G}_0^{-1})^{-1}$, $\boldsymbol{\gamma}^* = \boldsymbol{G}^*\left(\boldsymbol{G}_0^{-1}\boldsymbol{\gamma}_0 + \omega_2^{-1}\sum_{i=1}^n \left[\boldsymbol{w}_i\left(x_{si} - \frac{\sigma_{12}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})}{\sigma_{11}}\right)\right]\right)$,
$\omega_2 = \sigma_{22} - \sigma_{12}^2/\sigma_{11}$, $\alpha^* = \alpha_0 + n$ and $\boldsymbol{\Sigma}^* = \left[\boldsymbol{\Sigma}_0^{-1} + \sum_{i=1}^n \begin{pmatrix} y_i - \boldsymbol{x}_i'\boldsymbol{\beta} \\ x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma}\end{pmatrix}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma})\right]^{-1}$.

We also use a Gibbs sampling algorithm in this model given that we obtain 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 alternative l, $i = 1, 2, \ldots, n$, $l = 1, 2, \ldots, L$.[23] In particular, $y_{il} = \begin{Bmatrix} 0, & y_{il}^* \leq 0 \\ 1, & y_{il}^* > 0 \end{Bmatrix}$ such that $\boldsymbol{y}_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\mu}_i \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{y}_i^*$ is an unobserved latent L dimensional vector, $\boldsymbol{X}_i$ is a $L \times K$ design matrix of regressors, $K = L \times k$, k is the number of regressors, and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_k']'$, $\boldsymbol{\beta}_j$ are a L dimensional vector of coefficients, $j = 1, 2, \ldots, k$. We take into account simultaneously alternative-varying regressors (alternative attributes) and alternative-invariant regressors (individual characteristics).

[23] These alternatives are not mutually exclusive.

The likelihood function in this model is $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{n}\prod_{l=1}^{L} p_{il}^{y_{il}}$ where $p_{il} = p(y_{il}^* \geq 0)$. Observe that $p(\boldsymbol{y}_i^* \geq 0) = p(\boldsymbol{\Lambda}\boldsymbol{y}_i^* \geq 0)$, where $\boldsymbol{\Lambda} = diag\{\lambda_{ll}\}$, $\lambda_{ll} > 0$. This generates identification issues. We follow the post processing strategy proposed by Edwards and Allenby (2003) to get identified parameters, that is, $\tilde{\boldsymbol{\beta}} = vec\{\boldsymbol{D}\boldsymbol{\Gamma}\}$ and the correlation matrix $\boldsymbol{R} = \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}$, where $\boldsymbol{D} = diag\{\sigma_{ll}\}^{-1/2}$ and $\boldsymbol{\Gamma} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k]$.[24]

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

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

where $\boldsymbol{B} = (\boldsymbol{B}_0^{-1} + \boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}$, $\boldsymbol{\beta}^* = \boldsymbol{B}(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}^{*\prime}\boldsymbol{y}^{**})$, $\boldsymbol{\Sigma}^{-1} = \boldsymbol{C}'\boldsymbol{C}$, $\boldsymbol{X}_i^{*\prime} = \boldsymbol{C}'\boldsymbol{X}_i$, $\boldsymbol{y}_i^{**} = \boldsymbol{C}'\boldsymbol{y}_i^*$, $\boldsymbol{X}^* \begin{bmatrix} \boldsymbol{X}_1^* \\ \boldsymbol{X}_2^* \\ \vdots \\ \boldsymbol{X}_n^* \end{bmatrix}$, $\alpha^* = \alpha_0 + n$, $\boldsymbol{\Sigma}^* = (\boldsymbol{\Sigma}_0 + \sum_{i=1}^{n}(\boldsymbol{y}_i^* - \boldsymbol{R}_i\boldsymbol{\beta})'(\boldsymbol{y}_i^* - \boldsymbol{X}_i\boldsymbol{\beta}))^{-1}$, $I_{il} = \begin{cases} y_{il}^* > 0, & y_{il} = 1 \\ y_{il}^* \leq 0, & y_{il} = 0 \end{cases}$, $m_{il} = \boldsymbol{x}_{il}'\boldsymbol{\beta} + \boldsymbol{f}'(\boldsymbol{y}_{i,-l}^* - \boldsymbol{X}_{i,-l}\boldsymbol{\beta})$, $\boldsymbol{y}_{i,-l}^*$ is a $L-1$ dimensional vector of all components of $\boldsymbol{y}_i^*$ excluding y_{il}^*, $\boldsymbol{x}_{il}$ is the l-th row of $\boldsymbol{X}_i$, $\boldsymbol{X}_{i,-l}$ is $\boldsymbol{X}_i$ deleting l-th row, $\tau_{ll}^2 = 1/\sigma^{ll}$, and $\boldsymbol{f} = -\sigma^{ll}\boldsymbol{\omega}_{l,-l}$, σ^{jl} is the jl-th element of $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\omega}_1' \\ \boldsymbol{\omega}_2 \\ \vdots \\ \boldsymbol{\omega}_L' \end{bmatrix}$, $\boldsymbol{\omega}_{l,-l}'$ is the l-th row of $\boldsymbol{\Sigma}^{-1}$ extracting the l-th element.

3.3. Hierarchical longitudinal models

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

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

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

[24] In a Bayesian setting, we can have a no identified model; however, the posterior of the model parameters exists given a proper prior distribution (Edwards and Allenby 2003).

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

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

Logit: The hierarchical longitudinal logit model establishes $y_{ij}\sim\mathcal{B}(\pi_{ij})$ where $logit(\pi_{ij})=log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)\equiv y_{ij}^*=\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\boldsymbol{w}_{ij}'\boldsymbol{b}_i+\mu_{ij}$, $i=1,2,\ldots,m$ and $j=1,2,\ldots,n_i$ where $\boldsymbol{x}_{ij}$ are k dimensional vectors of regressors, $\boldsymbol{\beta}$ is a k dimensional vector of "fixed" effects, $\boldsymbol{w}_{ij}$ are q dimensional vectors of regressors associated with random effects (q typically less than k), $\boldsymbol{b}_i$ is a q dimensional vector of unit-specific random effects such that $\boldsymbol{b}_i\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{D})$, 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, $\boldsymbol{\beta}\sim\mathcal{N}(\boldsymbol{\beta}_0,\boldsymbol{B}_0)$, $\sigma^2\sim\mathcal{IG}(\alpha_0,1/\delta_0)$ and $\boldsymbol{D}\sim\mathcal{IW}(d_0,d_0\boldsymbol{D}_0)$, and taking into account that $\boldsymbol{y}_i^*|\boldsymbol{\beta},\boldsymbol{D},\sigma^2\sim\mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta},\boldsymbol{V}_i)$ where $\boldsymbol{V_i}=\sigma^2\boldsymbol{I}+\boldsymbol{W}_i\boldsymbol{D}\boldsymbol{W}_i'$, then

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

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

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 $\boldsymbol{y}^*=\left[y_{ij}^*\right]$ 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 $\pi_{ij}=\frac{1}{1+e^{-y_{ij}^*}}$ and $\pi_{ij}^c=\frac{1}{1+e^{-y_{ij}^{*c}}}$.

Poisson: The hierarchical longitudinal poisson model establishes $y_{ij}\sim\mathcal{P}(\lambda_{ij})$,[25] where $log(\lambda_{ij})\equiv y_{ij}^*=\boldsymbol{x}_{ij}'\boldsymbol{\beta}+\boldsymbol{w}_{ij}'\boldsymbol{b}_i+\mu_{ij}$, $i=1,2,\ldots,m$ and $j=1,2,\ldots,n_i$ where $\boldsymbol{x}_{ij}$ are k dimensional vectors of regressors, $\boldsymbol{\beta}$ is a k dimensional vector of "fixed" effects, $\boldsymbol{w}_{ij}$ are q dimensional vectors of regressors associated with random effects (q typically less than k), $\boldsymbol{b}_i$ is a q dimensional vector of unit-specific random effects such that $\boldsymbol{b}_i\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{D})$, and $\mu_{ij}\sim\mathcal{N}(0,\sigma^2)$ are stochastic errors.

[25] $\mathcal{P}$ denotes a poisson density.

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

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

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

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 $\boldsymbol{y}^* = \left[y_{ij}^*\right]$ 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 $\lambda_{ij}^c = e^{y_{ij}^{*c}}$.

3.4. Bayesian bootstrap

We implement the Bayesian bootstrap (Rubin 1981) to linear regression models. In particular, the Bayesian bootstrap simulates posterior distributions assuming that the sample cumulative distribution function (cdf) is the population cdf (this assumption is also implicit in the frequentist bootstrap (Efron 1979)).

Given $y_i \overset{i.n.d}{\sim} \mathcal{F}$ where $\mathcal{F}$ does not define a particular parametric family distribution, $i = 1, 2, \ldots, n$, but sets $E(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$, such that $\boldsymbol{x}_i$ 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 values of Y that are not observed have zero posterior probability.

The algorithm to implement the Bayesian bootstrap is the following:

3.5. Bayesian model averaging

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

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

Algorithm A3 Bayesian bootstrap in linear regression

1: Draw $\boldsymbol{g} \sim \mathcal{D}ir(\alpha_1, \alpha_2, \ldots, \alpha_n)$ such that $\alpha_i = 1\ \forall i$.
2: $\boldsymbol{g} = (g_1, g_2, \ldots, g_n)$ is the vector of probabilities to attach to $(y_1, \boldsymbol{x}_1'), (y_2, \boldsymbol{x}_2'), \ldots, (y_n, \boldsymbol{x}_n')$ en each Bayesian bootstrap replication.
3: Sample S_1 times $(y_i, \boldsymbol{x}_i')$, $i = 1, 2, \ldots, n$ with replacement and probabilities g_i.
4: Estimate $\boldsymbol{\beta}$ using ordinary least squares in the model $E(\boldsymbol{Y}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{y}$ is a S_1 dimensional vector of realizations of $\boldsymbol{Y}$, and $\boldsymbol{X}$ is a $S_1 \times k$ 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.

where $\pi(M_j)$ is the prior model probability,[26] $p(\boldsymbol{y}|M_j) = \int_{\boldsymbol{\Theta}_j} p(\boldsymbol{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$ 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

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

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

The posterior variance highlights how BMA methodology accounts for model uncertainty. The first term is the weighted variance for 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

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

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

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

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

[26] We set equal prior probability 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).

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

BMA allows to incorporate 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 (Barbieri and Berger 2004).

There are two main computational issue implementing BMA. First, the marginal likelihood $p(\boldsymbol{y}|M_j) = \int_{\boldsymbol{\Theta}_j} p(\boldsymbol{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta}_j$, which most of the cases does not have an analytic solution, and second, the amount of models in the model space, 2^k, this some times can be enormous.

The Bayesian information criterion is a possible solution for the first issue. Defining $h(\boldsymbol{\theta}|M_j) = -\frac{log(p(\boldsymbol{y}|\boldsymbol{\theta}_j,M_j)\pi(\boldsymbol{\theta}_j|M_j))}{n}$, then $p(\boldsymbol{y}|M_j) = \int_{\boldsymbol{\Theta}_j} exp\left\{-nh(\boldsymbol{\theta}|M_j)\right\} d\boldsymbol{\theta}_j$. If n is sufficiently large $(n \to \infty)$, we can make the following assumptions (Hoeting, Madigan, Raftery, and Volinsky 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 as $\hat{\boldsymbol{\theta}}_{MLE}$.

We get the following results under these assumptions:

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

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

This implies

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

Taking into account that $\frac{k_j}{2}log(2\pi)$ and $log(\pi(\hat{\boldsymbol{\theta}}_j^{MLE}|M_j))$ are constants as functions of $\boldsymbol{y}$, and $|\boldsymbol{\Sigma}|$ is bounded by a finite constant, then

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

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 from different approaches such as

Occam's window criterion (Madigan and Raftery 1994; Raftery, Madigan, and Hoeting 1997), reversible jump Markov chain Monte Carlo computation (Green 1995), Markov chain Monte Carlo model composition (Madigan, York, and Allard 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.[27]

In Occam's window a model is discarded if its predictive performance is far less well than the best model (Madigan and Raftery 1994; Raftery *et al.* 1997). Thus models not belonging to $\mathcal{M}' = \left\{M_j : \frac{Max_r \pi(M_r|\boldsymbol{y})}{\pi(M_j|\boldsymbol{y})} \leq c\right\}$ 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, $\mathcal{M}'' = \left\{M_j : \exists M_m \in \mathcal{M}', M_m \subset M_j, \frac{\pi(M_m|\boldsymbol{y})}{\pi(M_j|\boldsymbol{y})} > 1\right\}$. 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 fewer than 25.

However, the previous theory framework requires calculating $Max_r \pi(M_r|\boldsymbol{y})$, which implies to calculate all possible models in $\mathcal{M}$. This is computationally burdensome. Then, a heuristic approach is proposed by Raftery, Hoeting, Volinsky, Painter, and Yeung (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. Denote M_0 a model with one regressor less than model M_1. If $log(\pi(M_0|\boldsymbol{y})/\pi(M_1|\boldsymbol{y})) > log(O_R)$ then M_1 is rejected and M_0 is considered, if $log(\pi(M_0|\boldsymbol{y})/\pi(M_1|\boldsymbol{y})) \leq -log(O_L)$ then M_0 is rejected, and M_1 is considered, and if $log(O_L) < log(\pi(M_0|\boldsymbol{y})/\pi(M_1|\boldsymbol{y})) \leq log(O_R)$, M_0 and M_1 are considered. 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 2000) is implemented to improve 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 to ensure that no good models are discarded.[28]

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

[27] Variable selection (model selection) is a related topic 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 to tackle this issue. The former allows to calculate an estimate similar to PIP. However, these frameworks do not offer a clear way to introduce posterior model probabilities.

[28] $O_L = 1/O_R^2$ is defined by default in Raftery *et al.* (2012).

candidate model is accepted with probability

$$\alpha(M_{s-1}, M_c) = Min\left\{\frac{|nbd(M)|p(\boldsymbol{y}|M_c)\pi(M_c)}{|nbd(M^c)|p(\boldsymbol{y}|M_{(s-1)})\pi(M_{(s-1)})}, 1\right\}.$$

Observe that by construction $|nbd(M)| = |nbd(M_c)| = \frac{1}{k}$, except in pathological cases where a model just has one regressor or all regressors.

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

- 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 (MC^3): We implement the Gaussian linear model using MC^3 using the marginal multivariate Student's t distribution.
- Instrumental variables: Let's consider the previous instrumental variable framework (Multivariate models section) assuming $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(3, \boldsymbol{I})$ (Karl and Lenkoski 2012). Then, the posterior distributions are

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

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

  Lenkoski, Karl, and Neudecker (2013) propose an algorithm based on conditional Bayes factors (Dickey and Gunel 1978) that allows embedding MC^3 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 $CBF^{2nd} = \frac{p(\boldsymbol{y}|M_c^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}{p(\boldsymbol{y}|M_{s-1}^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})}$, where $p(\boldsymbol{y}|M_c^{2nd}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = \int_{\mathcal{M}^{2nd}} p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma})\pi(\boldsymbol{\beta}|M_c^{2nd})d\boldsymbol{\beta} \propto |\boldsymbol{B}^*|^{1/2} Exp\left\{\frac{1}{2}\boldsymbol{\beta}^{*\prime}\boldsymbol{B}^{*-1}\boldsymbol{\beta}^*\right\}$. For the first stage, $CBF^{1st} = \frac{p(\boldsymbol{y}|M_c^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma})}{p(\boldsymbol{y}|M_{s-1}^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma})}$, where $p(\boldsymbol{y}|M_c^{1st}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int_{\mathcal{M}^{1st}} p(\boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\pi(\boldsymbol{\gamma}|M_c^{1st})d\boldsymbol{\gamma} \propto |\boldsymbol{G}^*|^{1/2} Exp\left\{\frac{1}{2}\boldsymbol{\gamma}^{*\prime}\boldsymbol{G}^{*-1}\boldsymbol{\gamma}^*\right\}$.[29] 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.[30]

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

[30] Koop, León-Gonzalez, and Strachan (2012) and Lenkoski, Eicher, and Raftery (2014) propose other frameworks for BMA taking into account endogeneity.

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, $p_i(y_i|\theta_i, \phi) = h(y_i, \phi)Exp\{(\theta_i y_i - b(\theta_i))/a(\phi)\}$ (canonical representation), $y_i \overset{i.n.d}{\sim} p_i$, $i = 1, 2, \ldots, n$. It also has a linear predictor $\theta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$, and a link function g such that $E(Y_i|x_i) \equiv \mu_i = b'(\theta_i) = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})$ (g is monotonic and differentiable), and $V(Y_i) = b''(\theta_i)a(\phi)$ (McCullagh and Nelder 1989). The identity function $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ is the canonical link function in the case of the Gaussian model.[31] This statistical framework can help us to characterize:

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

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

4. Using BEsmarter

The top panel in Figure 1 shows the class of models that can be estimated in our GUI. The radio button on the left hand side shows the specific models inside each generic class. In particular, users can see that the normal model is selected inside univariate class models.

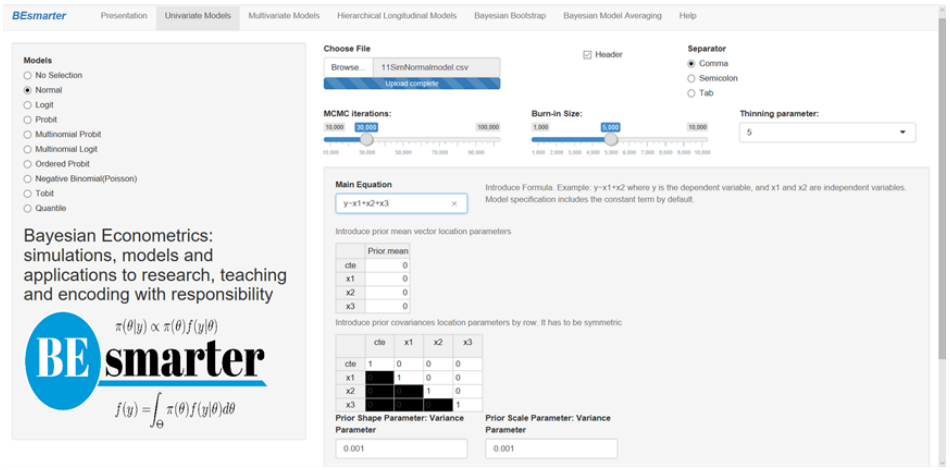

Figure 1: Univariate models: Specification

Then, the right hand side panel displays a widget to upload input data set, which should be a csv file with headers in the first row, this file has to be closed, we also should select what kind

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

of separator is used in input file: comma, semicolon or tab.[32] Range sliders help to set MCMC iterations and burn-in size, and thinning parameter can be selected as well. After this, users should specify the equation, this is set in R software form (see Figure 1, $y \sim x1 + x2 + x3$). Take into account that univariate models include intercept by default, except ordered probit, where specification has to do this explicit, for instance $y \sim x1 + x2 + x3 - 1$. Finally, we should define prior hyperparameters, for instance mean, covariance, shape and scale in the normal-inverse gamma model (see Figure 1).

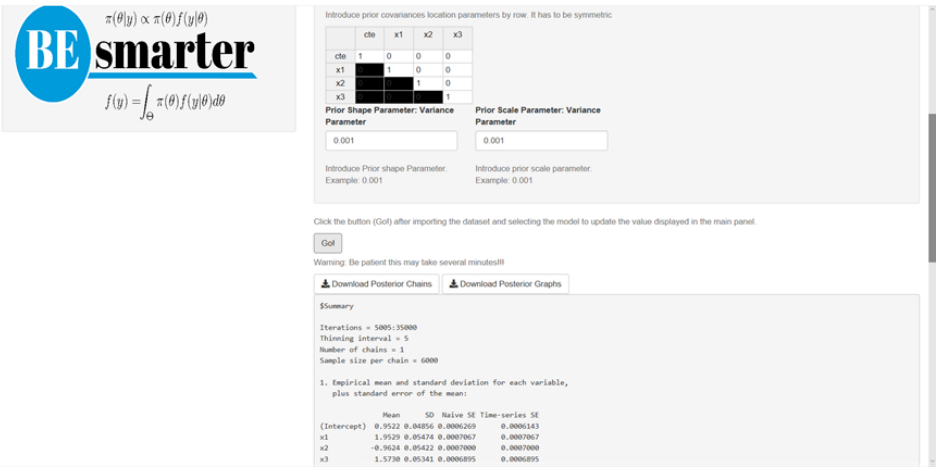

Figure 2: Univariate models: Results

After this specification process, users should click the Go! button to initiate the estimation process. We display summary statistics and convergence diagnostics after this process is finished (see Figure 2). We also have widgets to download posterior chains (csv file) and graphs (pdf and eps files). Take into account that order of coefficients in results (summary, posterior chains and graphs) is first for location parameters, and then, scale parameters.

Multinomial models (probit and logit) require a data set 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 to define 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 3). Multinomial logit also allows to define the tune parameter, degrees of freedom in this case, for the Metropolis-Hastings algorithm. This is a feature in our GUI when estimation of models is based on Metropolis-Hastings algorithm. The order of the coefficients in results of these models is first intercepts, and then, non-alternative specific regressors, these by the order of categories, and finally, coefficients for alternative specific regressors. Take into account that non-alternative specific regressors associated with base category are equal to zero (they do not appear in results). In addition, some coefficients of the main diagonal of the covariance matrix are constant by identification issues in multinomial and multivariate probit models.

[32] The folder called *DataSim* has examples of input data sets for each model that can be estimated using our GUI.

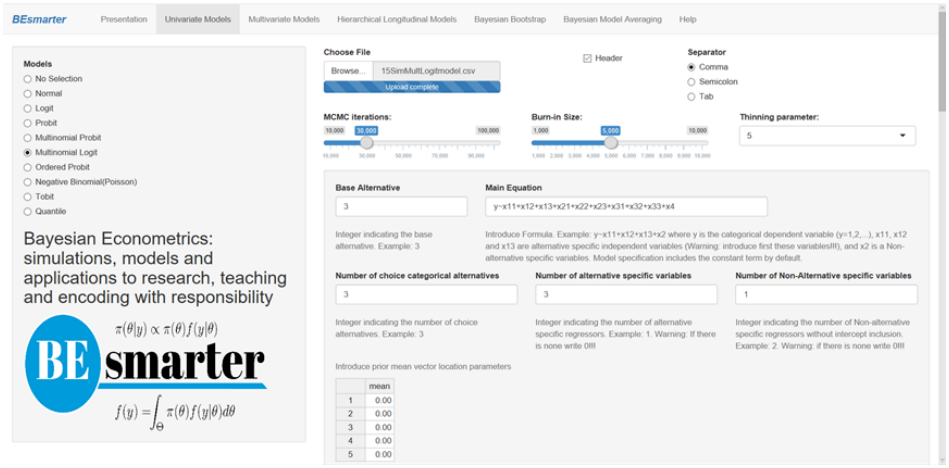

Figure 3: Univariate models: Multinomial

In the case of the negative binomial model, users should define dispersion parameter (α, see negative binomial model). User should define censuring points and quantile in Tobit and quantile models, respectively.

Figure 4 displays the multivariate regression setting. In this case, the input file should have first the dependent variables, and then regressors, if there are intercepts in each equation, there should be a column of 1's after the dependent variables in the input file. Users also has to define the number of dependent variables and the number of regressors, if necessary include the intercept (see Figure 4).

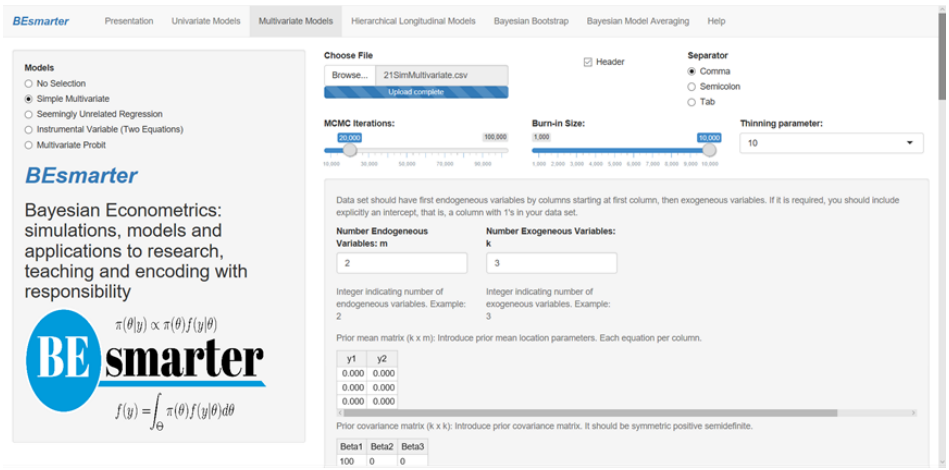

Figure 4: Multivariate models: Simple multivariate

The input file in seemingly unrelated regressions should have first the dependent variables, and then regressors by equation, include 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 equation (if necessary include in-

tercepts, each intercept is an additional regressor), and the number of regressors by equation (if necessary include intercept).

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

In the instrumental variable setting, users should specify the main equation, and the auxiliary equation. This setting includes intercept by default. The first variable on the right hand side in the main equation has to be the variable with endogeneity issues. Users can find first the results associated with the endogeneous regressor, then location parameters of the auxiliary regression, and location parameters of the exogenous regressors. Finally, the posterior covariance matrix.

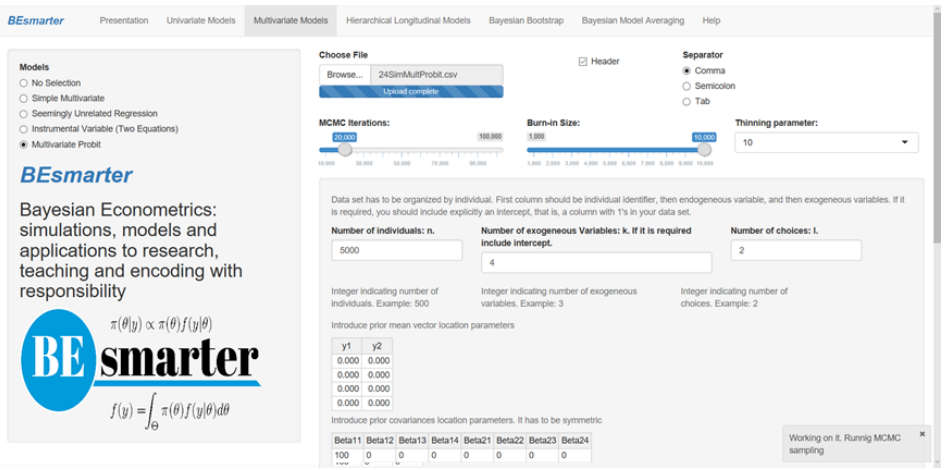

Figure 5: Multivariate models: Multivariate probit

Multivariate probit model requires an input data set ordered by unit, for instance 3 choices implies repeat 3 times each unit, the first column has to be identification of each unit, users should use ordered integers, then the dependent variable, just one vector, composed by 0's and 1's, then regressors, which should include a column of 1's for intercepts. Users should define number of units, number of regressors, and number of choices (see Figure 5). Results display first posterior location parameters by equation, and then the posterior covariance matrix.

Input files of hierarchical longitudinal models should have first the dependent variable, then regressors and a cross sectional identifier ($i = 1, 2, \ldots, m$). It is not a requirement to have a balanced data set, n_i can be different for each i. Users should specify the fixed part equation, and random part equation. Both in R software form. In case of just requiring random intercepts, do not introduce anything in this part (see Figure 6). Users should also type the name of the cross sectional identifier variable. Results include fixed and random effects posteriors, but do not include convergence diagnostics tests.

Bayesian bootstrap just requires upload a data set, specify MCMC iterations, resampling size and equation. Users can see the progress bar after clicking the Go! button (see Figure 7).

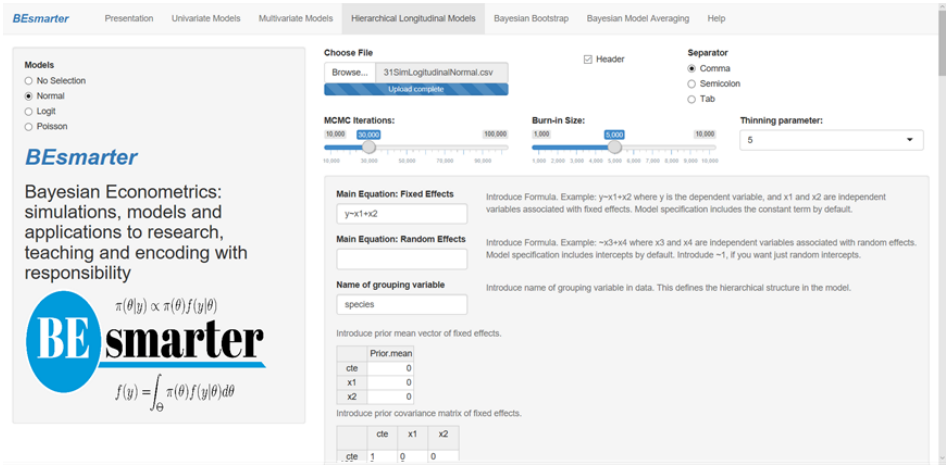

Figure 6: Hierarchical longitudinal models: Specification

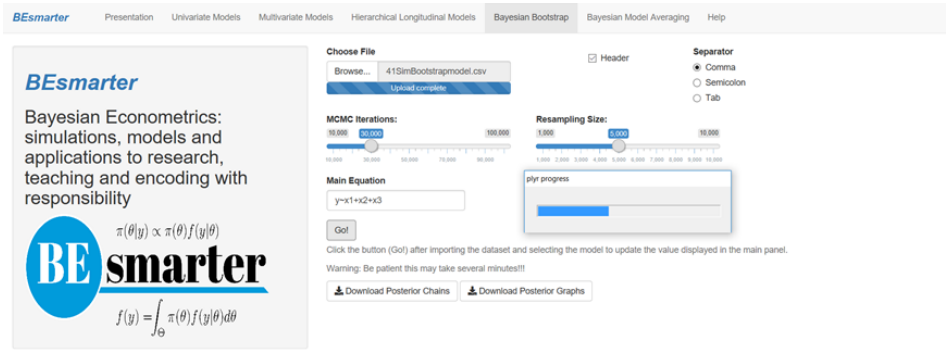

Figure 7: Bayesian bootstrap: Specification

Bayesian model averaging based on a Gaussian distribution can be performed using Bayesian information criterion (BIC) approximation, Markov chain Monte Carlo model composition (MC3) or instrumental variables (see Figure 8). The former two approaches require an input data set where the first column is the dependent variable, and then, the potentially important regressors. Users should define the band width model selection parameter (O_R) and number of iterations for BIC and MC3, respectively. Results include posterior inclusion probability ($p! = 0$), expected value (EV), and standard deviation (SD) of coefficients associated with each regressors. The BIC framework also displays the most relevant models including number of regressors, R^2, BIC, and posterior model probability. Users can download two csv files: *Best models* and *Descriptive statistics coefficients*. The former is a 0-1 matrix such that columns are regressors and rows are models. 1 indicates presence of regressor l in model M_m, 0 otherwise. Take into account that the last column of this file is the posterior model probability. The latter file shows posterior inclusion probabilities, expected values and standard deviations associated with each regressor taking into account the BMA procedure based on the best models.

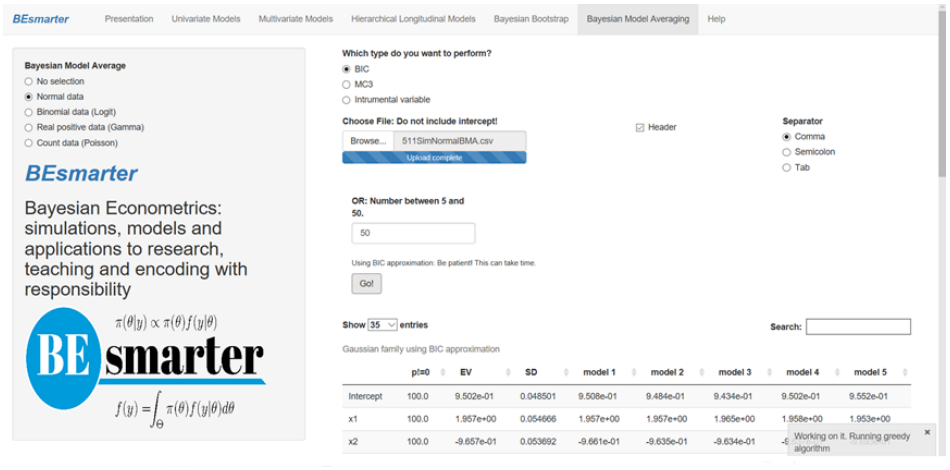

Figure 8: Bayesian model averaging: Specification and results

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 regressors with endogeneity issues, and then, exogeneous regressors, users should include a column of 1's if an intercept is required. The second input file has all instruments. Users should also introduce number of regressors with endogeneity issues (see Figure 9).

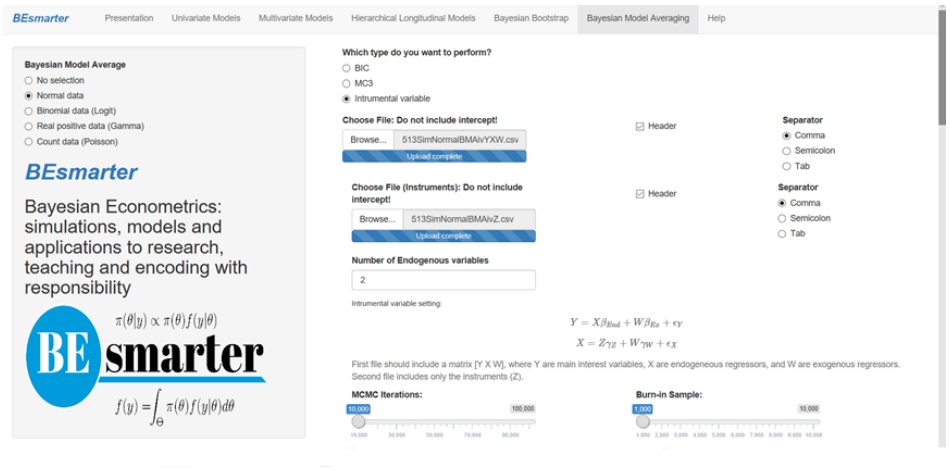

Figure 9: Bayesian model averaging: Instrumental variable specification

Results include posterior inclusion probabilities and expected values for each regressor. Users can find results of the main equation, and then, auxiliary equations. Users can download csv files of BMA results for the second stage (main equation), and first stage (auxiliary equations). In addition, users can download posterior chains of location parameters of the main equation, β_l, $l = 1, 2, \ldots, dim\{\boldsymbol{\beta}\}$, location parameters of auxiliary equations, $\gamma_{j,i}$, $j = 1, 2, \ldots, dim\{\boldsymbol{\beta}_s\}$ where $dim\{\boldsymbol{\beta}_s\}$ is the number of regressors with endogeneous issues, $i = 1, 2, \ldots, dim\{\boldsymbol{\gamma}\}$, where $dim\{\boldsymbol{\gamma}\}$ is the number of regressors in auxiliary regressors (exo-

geneous regressors + instruments), and elements of covariance matrix. $\sigma_{j,k}$.

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

5. Concluding remarks

The Bayesian statistical framework has become very popular among scientists since the computational revolution in 1990's. In particular, computational burdensome procedures such as MCMC algorithms can be easily applied nowadays. However, most of the open source software to apply these procedures require programming skills. This may be a cause why the Bayesian framework is not very popular among applied researchers and practitioners. We introduce in this paper a Graphical User Interface to implement Bayesian regression analysis under different frameworks, explaining the basic theory, such that users can understand the basic principles of Bayesian statistics, and apply them easily. Our objective is to increase popularity of the Bayesian statistical framework among applied researchers and practitioners.

Table 1: Graphical user interfaces to perform Bayesian regression analysis.

Name	**Language**	**Models**	**Open source**
ShinyStan	R+Stan	MCMC Implementation*	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	rnegbinRw	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)

*We slightly modified these functions in our GUI.

References

Albert JH, 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*. URL https://CRAN.R-project.org/package=bayesboot.

Barbieri M, Berger J (2004). "Optimal predictive model selection." *The Annals of Statistics*, **32**(3), 870–897.

Casella G, 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. URL http://shiny.rstudio.com/.

Chib S (1992). "Bayes inference in the Tobit censored regression model." *Journal of Econometrics*, **51**, 79–99.

Chib S, Carlin B (1999). "On MCMC sampling in hierarchical longitudinal models." *Statistics and Computing*, **9**, 17–26.

Chib S, Greenberg E (1995). "Understanding the Metropolis-Hastings Algorithm." *The American Statistician*, **49**(4), 327–335.

Clyde M, George E (2004). "Model Uncertatinty." *Statistical Science*, **19**(1), 81–94.

Dickey JM, Gunel E (1978). "Bayes factors from mixed probabilities." *Journal of the Royal Statistical Society: Series B (Methodology)*, **40**, 43–46.

Edwards YD, Allenby GM (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 GM, Wilson RW (2000). "Regressions by Leaps and Bounds." *Technometrics*, **42**(1), 69–79.

Gelfand AE, Smith AFM (1990). "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*, **85**, 398–409.

Gelman A, Rubin DB (1992). "Inference from iterative simulation using multiple sequences." *Statistical Science*, **7**(4), 457–472.

Geman S, 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, McCulloch R (1993). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, **88**(423), 881–889.

George E, 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 PJ (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, Welch PD (1983). "Simulation run length control in the presence of an initial transient." *Operations Research*, **31**(6), 1109–1144.

Hoeting J, Madigan D, Raftery A, Volinsky C (1999). "Bayesian Model Averaging: A Tutorial." *Statistical Science*, **14**(4), 382–417.

Ishwaran H, Rao JS (2005). "Spike and slab variable selection: Frequentist and Bayesian strategies." *The Annals of Statistics*, **33**(2), 730–773.

Jhonson VE, Rossell D (2012). "Bayesian model selection in high–dimensional settings." *Journal of the American Statistical Association*, **107**(498), 649–660.

Karl A, Lenkoski A (2012). "Instrumental variable bayesian model averaging via conditional Bayes factor." *Technical report*, Heidelberg University.

Kass RE, Raftery AE (1995). "Bayes factors." *Journal of the American Statistical Association*, **90**(430), 773–795.

Koop G, León-Gonzalez R, Strachan R (2012). "Bayesian model averaging in the instrumental variable regression model." *Journal of Econometrics*, **171**, 237–250.

Kozumi H, Kobayashi G (2011). "Gibbs sampling methods for Bayesian quantile regression." *Journal of Statistical Computation and Simulation*, **81**(11), 1565âĂŞ–1578.

Lenkoski A, Eicher TS, Raftery A (2014). "Two-stage Bayesian model averaging in endogeneous variable models." *Econometric Reviews*, **33**.

Lenkoski A, Karl A, Neudecker A (2013). *Package ivbma*. URL https://CRAN.R-project.org/package=ivbma.

Ley E, Steel MF (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, Raftery AE (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, Allard D (1995a). "Bayesian graphical models for discrete data." *International Statistical Review*, **63**, 215–232.

Madigan D, York JC, Allard D (1995b). "Bayesian Graphical Models for Discrete Data." *International Statistical Review*, **63**(2), 215–232.

Martin AD, Quinn KM, Park JH (2011). "MCMCpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software*, **42**(9), 1–21.

Martin AD, Quinn KM, Park JH (2018). *Package MCMCpack*. URL https://CRAN.R-project.org/package=MCMCpack.

McCullagh P, Nelder JA (1989). *Generalized Linear Models*. London: Chapman and Hall.

McCulloch R, 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, Teller E (1953). "Equations of state calculations by fast computing machines." *J. Chem. Phys*, **21**, 1087–1092.

Park T, 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, Magnusson A (2016). *Output Analysis and Diagnostics for MCMC*. URL https://CRAN.R-project.org/package=coda.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raftery A, Hoeting J, Volinsky C, Painter I, Yeung KY (2012). *Package BMA*. URL https://CRAN.R-project.org/package=BMA.

Raftery A, Lewis S (1992). "One Long Run with Diagnostics: Implementation Strategies for Markov chain Monte Carlo." *Statistical Science*, **7**, 493–497.

Raftery AE (1993). "Bayesian model selection in structural equation models." *Sage Focus Editions*, **154**, 163–163.

Raftery AE, Madigan D, Hoeting JA (1997). "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, **92**(437), 179–191.

Robert C, Casella G (2004). *Monte Carlo Statistical Methods*. Second edition. Springer.

Robert C, Casella G (2010). *Introducing Monte Carlo Methods with R*. Second edition. Springer.

Roberts GO, Gelman A, Gilks WR (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*. URL https://CRAN.R-project.org/package=bayesm.

Rossi P, Allenby GM, McCulloch R (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons.

Rubin DB (1981). "The Bayesian Bootstrap." *The Annals of Statistics*, **9**(1), 130–134.

Simmons SJ, Fang F, Fang Q, 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.

Tanner MA, Wong WH (1987). "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association*, **82**(398), 528–540.

Tierney L, Kadane JB (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the american statistical association*, **81**(393), 82–86.

Woodward P (2005). "BugsXLA: Bayes for the common man." *Journal of Statistical Software*, **14**(5), 1–18.

Affiliation:

Andrés Ramírez–Hassan
Department of Economics
School of Economics and Finance
Universidad EAFIT
Cra. 49, Medellín, Antioquia, Colombia
e-mail: aramir21@eafit.edu.co
URL: www.besmarter-team.org

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd