Prior Choice for the Variance Parameter in the Multilevel Regression and Poststratification Approach for Highly Selective Data: A Monte Carlo Simulation Study

Bruch, Christian; Felderer, Barbara

Austrian Journal of Statistics
July 2022, Volume 51, 76–95.
http://www.ajs.or.at/
doi:10.17713/ajs.v51i4.1361

Prior Choice for the Variance Parameter in the Multilevel Regression and Poststratification Approach for Highly Selective Data. A Monte Carlo Simulation Study.

Christian Bruch
GESIS

Barbara Felderer
GESIS

Abstract

The multilevel and poststratification approach is commonly used to draw valid inference from (non-probabilistic) surveys. This Bayesian approach includes varying regression coefficients for which prior distributions of their variance parameter must be specified. The choice of the distribution is far from being trivial and many contradicting recommendations exist in the literature. The prior choice may be even more challenging when data results from a highly selective inclusion mechanism, such as applied by volunteer panels. We conduct a Monte Carlo simulation study to evaluate the effect of different distribution choices on bias in the estimation of a proportion based on a sample that is subject to a highly selective inclusion mechanism.

Keywords: Bayesian MRP, prior for the variance parameter, self-selection, selective data, simulation study.

1. Introduction

Most surveys are affected by selectivity in the sampling and/or the response process. For probability surveys that do not suffer from sample-selection error, varying willingness between population subgroups to participate in the survey can introduce severe self-selection or nonresponse bias. Non-probability surveys (for theory on non-probability samples see, for example, Little, West, Boonstra, and Hu (2019)) are likely to suffer from sample-selection bias due to non-random sampling and self-selection bias due to selective participation that usually can not be distinguished in practice. The extend of bias differs largely between surveys, while non-probability online surveys have been found to be more selective than probability ones (Cornesse, Blom, Dutwin, Krosnick, De Leeuw, Legleye, Pasek, Pennay, Phillips, Sakshaug, Struminskaya, and Wenz 2020).

In the recent decades, online panels have become a prominent tool to survey the general population. However, due to budget constraints, the recruitment of such panels is often not based on random sampling procedures. Very common examples of non-probabilistic online panels are volunteer panels for which advertisements are fielded over websites, and everybody who comes across the advertisement is invited to participate. The problem hereby is that

different units of the population have different and unknown propensities to see the advertisement and certain units do not have the chance to see the advertisement at all, for examples persons who do not use the Internet or do not have Internet access. In addition, it is very likely for such panels that some population groups are more likely to respond than others. Because the sampling mechanism is unknown, inclusion probabilities are not known in many applications in practice as well and design weights cannot be computed. In addition to the selectivity in the initial recruitment, panel nonresponse and drop-out throughout the panel waves might introduce further selectivity to the respondent sample and therefore severely bias survey estimation.

In order to achieve valid inference based on highly selective samples like those generated by volunteer panels, researchers commonly use weighting procedures to account for selectivity, e.g., by raking the sample such that distributions of the sample match the population distributions known from official statistics. For very selective respondent samples, however, it is very likely that some population groups are very scarce in the sample or even non-existent. This leads to sparse or empty weighting cells that standard weighting procedures can not deal with. During the last decade, more complex weighting and estimation procedures have been developed and applied by social researchers, the most famous being the multilevel regression and poststratification (MRP) approach by Gelman and colleagues (see, for example, Gelman and Little (1997) and Wang, Rothschild, Goel, and Gelman (2015)). This Bayesian approach can successfully be used to stabilize estimation when the sample is highly selective and weighting cells are sparse or empty by borrowing strength from other cells. The weighting and estimation approach includes varying regression coefficients for which prior distributions and, more specifically, the variance parameter of these distributions need to be specified. The choice of the prior distribution of the variance parameters is of great importance and there are contradicting recommendations in the literature. The inconclusive recommendations for the prior distributions make it very hard for practitioners to decide which prior to use in their applications. This is even more challenging when the data stem from a highly selective inclusion mechanism. In this paper, we aim at giving practical advise to inform the choice of the prior distribution for highly selective data. We conduct a Monte Carlo simulation study to evaluate the effect of the choice of the prior distribution of the variance parameter on the estimation in two (highly) selective samples. We thereby use several different distributions that are commonly used in the literature.

In the next section, we give an overview of the MRP approach including a review of the recommendations for the choice of the distribution of the variance of the regression coefficients. Section 3 describes the simulation study. Results are shown in Section 4. We end with a conclusion and discussion in Section 5.

2. Multilevel and poststratification approach

2.1. Notation

We assume that a sample S of size n is drawn from the population P of size N. The final participants set of respondents is described by R. We assume this sample to be very selective, for example, by highly systematic response processes or by highly systematic sampling. These samples lead to a sample composition that differs strongly from the target population and that is often associated with so-called volunteer panels. Thus, the distributions of the relevant variables in the sample might strongly differ from their distributions in the population.

To perform weighting procedures, we assume that weighting variables $x_1 \dots x_d \dots x_D$ are available in the survey and the population, where D is the number of weighting variables. Each weighting variable d consists of C_d categories or groups. The cross-classification of these variables form the weighting cells, whereby a certain cell is indicated by l. For highly selective inclusion processes many sparse or empty cells may appear.

The number of sample elements in a weighting cell l is indicated by n_l, the corresponding number of elements in the population is denoted by N_l.

The variable of interest, for which the estimation is to be performed, is described by y with sample realizations $y_{l[1]} \dots y_{l[i]} \dots y_{l[n]}$. In this paper, we assume we have four weighting variables x_1, x_2, x_3, x_4 and a binary variable y that takes values one and zero. We aim to estimate the proportion p for $y = 1$ in the population based on the sample data. In a typical application, the weighting variables would be socio-demographics like age (in categories), gender, education, household size and marital status which are available for the general population from official statistics.

In the next section, we will present the estimation and weighting procedure that is applied in this paper in more detail.

2.2. Multilevel regression and poststratification approach

Whenever cell population sizes N_l are available, e.g., from official statistics, a poststratification weighting procedure can be applied. In practice, the poststratification procedure bears some challenges when it is applied on survey data that suffer from a selective self-selection process. For example, in case of many weighting cells, the estimation may be highly variable and unstable (Kalton and Flores-Cervantes 2003), particularly when having a lot of sparse or empty weighting cells. In addition to the estimation in sparse cells, also the overall estimation can be deteriorated when many cells with only low numbers of sample elements exist. For weighting cells without elements in the sample, an ordinary poststratification cannot be applied, even when population sizes for this cell are available. In practice, the problem of empty cells is often solved by combining weighting cells resulting in information loss.

One possibility to deal with the problem of sparse and empty sample cells is to use the multilevel logistic and poststratification (MRP) approach, as described by Gelman and colleagues (see, for example, Gelman and Little (1997) and Wang *et al.* (2015)). The MRP procedure can be described as follows (see for the following explanations Wang *et al.* (2015)):

1. First, a Bayesian multilevel logistic regression model is defined by :

$$P(y_i = 1) = logit^{-1}(\alpha + \beta_{l[i]}^{x_1} + \beta_{l[i]}^{x_2} + \beta_{l[i]}^{x_3} + \beta_{l[i]}^{x_4} + \dots)$$

$$\beta_{l[i]}^{var} \sim N(0, \sigma_{var}^2); \sigma_{var}^2 \sim \Psi(\dots)$$

   where $\beta_{l[i]}^{var}$ equals $\beta_{l[i]}^{x_1}$ to $\beta_{l[i]}^{x_4}$ which are the varying coefficients of the weighting variables x_1 to x_4 and σ_{var}^2 are the corresponding variance parameters. The Bayesian part of the model is related to the varying coefficients for which prior distributions need to be specified. These prior distributions are given by normal distributions with a mean of zero and variance parameters σ_{var}^2. The choice of the prior distribution $\Psi(\dots)$ of the variance parameters σ_{var}^2 is not as straightforward as for the varying coefficients and for the variance parameters various recommendations exist in the literature. Because the parameters σ_{var}^2 may have a strong influence on the posterior draws of the varying coefficient and thus on its estimation and its variability, the choice of their prior distributions can have a strong influence on the overall estimation, and thus the priors have to be selected very carefully.

2. In the next step, cell probabilities $\hat{p}_l$ are estimated by using the estimated parameters $\hat{\alpha}, \hat{\sigma}_{var}^2$ and $\hat{\beta}_{l[i]}^{var}$ from the multilevel logistic regression model.

3. Afterwards, the overall proportion p is estimated by:

$$\hat{p} = \frac{\sum_{l=1}^{L} N_l \cdot \hat{p}_l}{\sum_{l=1}^{L} N_l}. \tag{1}$$

Beside the estimated propensities from the multilevel logistic regression model, the population sizes N_l in each cell l are required, but the sample sizes n_l in each cell l are not a part of the equation.

The advantage of MRP compared to an ordinary poststratification is the stabilization of the estimation by including the multilevel logistic regression model. The estimation in sparse weighting cells is improved by using information of weighting cells with many units. This procedure is called *borrowing strength*, a common procedure applied in small area statistics. By improving the estimation in sparse cell, the overall estimation can be improved. Furthermore, it is even possible to obtain estimations for cells without sample elements.

Choosing an appropriate prior distribution for the variance parameters σ^2_{var} can have a large impact on the estimation. We discuss different prior distributions for these parameters in the next section.

2.3. Prior distributions for the variance parameter

In Bayesian statistics, the parameters of interest, such as the variance parameter, are not fixed but underlie some kind of uncertainty. Parameters are drawn from posterior distributions that result from the selected prior distribution and the likelihood that describes the sample data. Priors can be chosen so that the data have a larger or a small influence on the posterior distribution. In general, one can choose between informative and non-informative priors. Non-informative priors have little influence on the posterior distribution as compared to the dominating contribution of the data. Using non-informative priors let's "'the data speak for themselves"' (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin 2013). Using informative priors increases the impact of the prior information on the posterior distribution.

Numerous proposals exist in the literature on which prior distributions should be used for the variance parameters σ^2_{var}. In particular, many non-informative priors are proposed in the literature (Gelman 2006). Depending on the distribution that is used or on the implementation of the distribution in the statistical software (for example rstan (Stan Development Team 2018)), the prior is either applied to the variance parameters σ^2_{var} or σ_{var} (see, for example, Gelman (2006)). In this study, we use several priors proposed by Gelman (2006), Gelman *et al.* (2013) or Gelman (2020).

A first important distribution suggested for the prior of the variance parameter is the inverse gamma distribution. For example, Spiegelhalter, Thomas, Best, Gilks, and Lunn (2003) apply an inverse gamma distribution $IG(0.001, 0.001)$ for their hierarchical model (see also Gelman (2006)). The aim of choosing such small values for the parameter η of the inverse gamma distribution is to obtain a non-informative prior as the distribution becomes more flat and diffuse. However, as mention in Gelman (2006), the inverse gamma distribution may result in an improper posterior density for such parameter values. Inferences are sensitive to the choice of η and the prior distribution may become not non-informative. Gelman (2006) illustrates this using an example with educational testing and experimental school data for which the inverse gamma $IG(1, 1)$ prior distribution strongly constrains the posterior inferences. This constraint was even stronger when using an inverse gamma distribution $IG(0.001, 0.001)$ as prior distribution which was peaked strongly close to zero.

Another distribution that is used as prior distribution for the variance parameter is the scaled inverse chi-squared distribution Scale-Inv- $\chi(\nu, \hat{\sigma}^2_{var})$ which is for example applied or discussed in Wang *et al.* (2015), Gelman *et al.* (2013) and Browne and Draper (2006). The parameter

ν denotes the degrees of freedom and $\hat{\sigma}^2_{var}$ is a prior estimate of σ^2_{var}. The scale inverse chi-square distribution can be expressed by the inverse gamma distribution $IG(\frac{\nu}{2}, \frac{\nu}{2}\hat{\sigma}^2_{var})$ (Browne and Draper 2006). For the scale-inverse chi-squared distribution, the choice of the degrees of freedom has a strong influence on whether the prior is informative or rather non-informative. Higher numbers of degrees of freedom lead to more tight and thus informative prior distributions.

Also, the uniform distribution $U(a_1, a_2)$ is often used as an non-informative prior distribution $\Psi(\ldots)$ for the variance parameter σ_{var} (see for example Gelman (2006), Gelman *et al.* (2013) or Park, Gelman, and Bafumi (2004)). However, using the uniform distribution as prior distribution has some disadvantages: first, as mentioned in Woodward (2011), an upper limit has to be defined and values that are higher than the upper limit have a probability density of zero. Gelman (2020) recommends to apply the uniform distribution only in cases when the bounds are the actual constraints. Second, according to Gelman (2006), the uniform distribution $U(0, a_2)$ shows some problems with respect to miscalibration when the number of groups of a variable is small (lower than 3) and when $a_2 \to \infty$ due to the infinite prior mass $\sigma_{var} \to \infty$.

However, also for three, four or five groups, the right tail of the posterior may be too heavy which may lead to an overestimation of σ_{var}. Gelman (2006) shows this in a school example for a variable with three groups and proposes to use prior distributions that are more suited to constrain the posterior distribution with respect to unrealistic values. In his three-group school example Gelman therefore also applied a Half-Cauchy distribution (for a discussion on the use of the Half-Cauchy distribution see also Gelman (2020)) with scale parameter set to 25, which he describes as weakly informative. The value of the scale parameter was chosen to ensure that σ_{var} is constrained only weakly. By applying this prior, unrealistic high values of σ_{var} could be avoided in this study (Gelman 2006).

Even though the Half-Cauchy distribution performs well in the example above, Gelman (2020) argued that using such a prior might be too weak for many other applications. This is explained again for situations in which the number of groups is small and the data does not contain much information about group-level variance. In such situations, the Half-Cauchy distribution may be too broad and Gelman (2020) recommends to apply stronger priors with not too large scale parameters, such as a Half-normal distribution $N(0, 1)$ or a $StudentT(4, 0, 1)$ distribution in the case that large values of σ_{var} are not plausible. Both a Half-normal distribution with a small variance parameter and the Half-t distribution with higher degrees of freedom lead to a more tight and concentrated distribution.

For highly selective inclusion mechanism, the distribution of the sample data might greatly differ from the population distribution and might have many sparse or empty cells. This selectivity also challenges the choice of the prior for the variance parameter. Having in mind that the posterior distribution of the variance parameter and the varying coefficients result from the combination of the selected prior distribution and the likelihood, using a prior that lets the highly selective data dominate the posterior inference may lead to invalid or unreliable posterior inferences and estimations. For the variance parameter this means that too little or too large variability might occur in posterior inferences of the varying coefficients.

To overcome this problem it may be reasonable to use prior distributions that constrain both the influence of the highly selective data and the posterior inferences of the variance parameters of the varying coefficients in a way that the overall estimation can be stabilized.

Such a procedure, however, might constrain posterior inferences too much, so that important ranges of the parameter space are excluded. This may also have a negative effect on the variability. According to Gelman (2020), hard constraints should only be used when the bounds are actual constraints. For that reason, it may be a thin line between these two extremes (too less or too much variability when letting the data dominate or when excluding ranges of the parameter space) when applying the multilevel regression and poststratification approach on a sample that results from the highly selective inclusion mechanism.

In the next section, we describe the Monte Carlo simulation that was conducted to evaluate the effects of different prior distributions of the variance parameter on the point estimation of y in two different scenarios with differently selective inclusion mechanisms.

3. Simulation study

For our simulation, we build on the simulation study presented in Bruch and Felderer (2021). The simulation study is implemented in **R** (R Core Team 2018).

We use two sampling scenarios: In the first scenario, the inclusion mechanism is selective but to a moderate degree. This scenario serves as a benchmark scenario to which the second scenario is compared to. In the second scenario, the selectivity is further increased. This scenario mimics volunteer panel sampling in which the inclusion mechanism is highly selective, i.e., highly correlated with the covariates leading to sample distributions that are very different from population distributions. In this scenario, many sparse or empty cells challenge the weighting procedures.

In the simulation study, we simulate four weighting variables $x_1 \dots x_4$ and one dependent variable of interest y.

To generate the weighting variables, we generate a synthetic population by drawing first four continuous latent variables $x^U_{1_{cont}}$, $x^U_{2_{cont}}$, $x^U_{3_{cont}}$ and $x^U_{4_{cont}}$ from a multivariate normal distribution with $\zeta = (x^U_{1_{cont}}, x^U_{2_{cont}}, x^U_{3_{cont}}, x^U_{4_{cont}})$ and parameters

$$\zeta \sim N(\mu, \Sigma),$$

where the vector of expectations μ and the covariance matrix Σ are defined by

$$\mu = (200, 50, 80, 0.1)$$

and

$$\Sigma = \begin{pmatrix} 1,000 & 170 & 35 & 10 \\ 170 & 100 & 5 & 10 \\ 35 & 5 & 75 & 10 \\ 10 & 10 & 10 & 10 \end{pmatrix}.$$

To draw the four variables from a multivariate normal distribution, we use the R-Package mvtnorm (Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn 2019). In a next step, the continuous variables are categorized to mimic more realistic data sets usually available for social sciences which often contain many categorical variables. For the first variable x^U_1 the underlying continuous variable $x^U_{1_{cont}}$ is split into five categories (the categorization of the continuous variables can be found in Table 6 in Appendix A). The second variable x^U_2 is generated by building five categories using the continuous $x^U_{2_{cont}}$ variable's 0.25, 0.5, 0.75 and 0.9 quantiles. Similarly, the third variable x^U_3 results from the quartiles of the variable $x^U_{3_{cont}}$. The fourth variable x^U_4 is based on $x^U_{4_{cont}}$ following the same categorization scheme as variable $x^U_{2_{cont}}$.

We simulate a binary survey variable of interest y^U to be depending on the weighting variables. The variable y^U is drawn form a Bernoulli-distribution $y_i \sim B(p_{y,i})$ using the **R**-package LaplacesDemon (Statisticat and LLC. 2018). Each sample element's propensity $p^U_{y,i}$ for element i to choose category $y_i = 1$ depending on the characteristics $x^U_1 \dots x^U_4$ is modeled by

using a logistic model that is the same for both simulation scenarios (the model parameters can be found in Table 10 in Appendix A.):

$$p_{y,i}^U = \frac{\exp(\delta + \gamma_{j[i]}^{x_1^U} + \gamma_{k[i]}^{x_2^U} + \gamma_{o[i]}^{x_3^U} + \gamma_{v[i]}^{x_4^U})}{1 + \exp(\delta + \gamma_{j[i]}^{x_1^U} + \gamma_{k[i]}^{x_2^U} + \gamma_{o[i]}^{x_3^U} + \gamma_{v[i]}^{x_4^U})}. \tag{2}$$

The vectors $\gamma^{x_1^U} \dots \gamma^{x_4^U}$ encompasses values for each variable category and δ is the intercept. These parameters determine how strong the relationships between the weighting variables and the outcome of interest are.

Simulated this way, all joint distributions of weighting variables in the population are set and known for the entire population as well as the benchmark information of the survey variable of interest.

The inclusion mechanism is simulated using the following procedure (the modeling of y^U and the inclusion mechanism to create the volunteer sample is done on the basis of the procedure to model the nonresponse mechanism in the simulation study of Enderle, Münnich, and Bruch (2013)):

At first, we model the propensity ω_i^U to be included in the survey to be depending on $x_1^U \dots x_4^U$. This means that respondents with certain characteristics are more likely to participate than others. The inclusion process is modeled using a logistic model:

$$\omega_i^U = \frac{\exp(\lambda + \xi_{j[i]}^{x_1^U} + \xi_{k[i]}^{x_2^U} + \xi_{o[i]}^{x_3^U} + \xi_{v[i]}^{x_4^U})}{1 + \exp(\lambda + \xi_{j[i]}^{x_1^U} + \xi_{k[i]}^{x_2^U} + \xi_{o[i]}^{x_3^U} + \xi_{v[i]}^{x_4^U})}. \tag{3}$$

The vectors $\xi^{x_1^U} \dots \xi^{x_4^U}$ encompasses values for each variable category, where the value of each reference category is set to zero and λ is the intercept. Different parameters are chosen for the two simulation scenarios depending on the desired correlation between the inclusion mechanism and the weighting variables. The higher the correlations are, the more selective the inclusion process is. Inclusion propensities are in practical applications not known. Thus, in the simulation, they are only used to model the inclusion process but not included in the subsequent weighting and estimation.

In order to create a participation indicator that is either 0 or 1 from the inclusion probability ω_i^U, we draw random numbers from a uniform distribution. A unit participates in the survey if $\omega_i^U > u_i$ and refuses to participate if $\omega_i^U < u_i$. We simulate non-probability samples of size n = 1,000.

However, the Monte Carlo simulation study consists of random procedures which can be repeated a certain number of times. For example, when applying a design-based Monte Carlo simulation study, probability samples are drawn repeatedly from the population of interest by using a certain sampling design. In case of non-probability samples, the inclusion process consists of (mainly) non-random elements. This prevents a meaningful application of a Monte Carlo simulation study with respect to a repeated drawing of samples from the population. Thus, we rather propose to repeat the variable generation process in each simulation run which starts with draws from the multivariate normal distribution. This is often done in so-called (pure) model-based Monte Carlo simulation studies (for an explanation of a model-based or a pure model-based simulation study see Burgard (2015)). As a result, the generation process of the variables y, x_1, x_2, x_3 and x_4 is repeated in each simulation run by applying the non-probabilistic sampling scheme on each generated data set with variables y^U, x_1^U, x_2^U, x_3^U and x_4^U. In doing so, we obtain non-probability samples for each simulation run for which the weighting and complex estimation strategies are applied. In total, we generated 1,000 non-probability samples for each scenario.

In the simulation study, we consider a highly selective and moderately selective scenario varying the concrete numbers for the parameters in the inclusion model. The values for the

parameters can be found in Table 9 in Appendix A. We use rstan (Stan Development Team 2018) to implement the Bayesian approach, to fit the model and to draw the parameters from the posterior distribution. In detail, we use 2000 iteration, 4 chains and 500 warmups for each estimator and in each simulation run. In total, we use 1,000 simulations runs.

Both inclusion models result in sample distributions of the weighting variables that differ from the population benchmark (see Table 6 in Appendix A. Tables 7 and 8 (Appendix A) show the correlation structure of all variables in the realized sample after applying the inclusion mechanism in both scenarios). The difference is larger for scenario 2 with the highly selective inclusion mechanism than for the more moderate scenario 1.

We consider prior distributions for the variance parameter using several parameter constellations that are proposed in the literature as well as prior distributions that are constructed to examine and to improve the performance of the estimation on high selective data. The priors included in the simulation study are presented in the next section. Like Wang *et al.* (2015), we do not consider a fully Bayesian framework in our research and solely draw the main parameters of interest from a posterior distribution, i.e., the parameters of the multilevel approach.

To study how our findings are affected by the sample size and covariance structure, parts of the analysis are repeated using samples with reduced sample size and reduced covariances (see Appendix B).

3.1. Priors used in the simulation study

In the simulation study, we compare several prior distributions for the variance parameter (see 2.3). These are the inverse gamma distribution, the uniform distribution, the scale inverse chi-squared distribution, the Half-Cauchy distribution, the Half-normal distribution and the Half-t-distribution. For most of the distributions, we apply the prior on the standard deviation σ_{var} rather that σ^2_{var} except for the inverse gamma distribution for which the prior is applied on σ^2_{var}. This corresponds to the procedure described in Gelman (2006).

We apply four different strategies to select the scale parameter values for the different prior distributions.

Priors that lead to data-driven posterior inferences First, we use strategies that we call *priors that lead to data-driven posterior inferences*. These strategies include priors that lead to posterior distributions in which the data have a strong influence.

The strategies include the nonproper and non-informative uniform distribution $U(0, \infty)$ that is also applied in Gelman (2006) and that is the default in rstan (Stan Development Team 2018). Because the uniform distribution $U(0, \infty)$ may suffer from an additional overestimation when the number of groups is small (see the explanations in Gelman (2006) and Section 2.3) we further include prior distributions that take the estimated variance parameter for the scale parameter. We include the Half-Student-t distribution and the Half-normal distribution that are, for example, described in Gelman (2020) and Woodward (2011). For these distributions we use the estimated variances from the sample as scale parameters. The variance parameters $\hat{\sigma}_{var}$ are estimated using the glmer function of the lme4 package (Bates, Mächler, Bolker, and Walker 2015). We also apply the estimated variance parameter for the scale parameter of the scaled inverse chi-square distribution as described in Browne and Draper (2006) and applied in Wang *et al.* (2015) (see Table 1 for the concrete parameter constellations in the prior distributions we use in this simulation study).

Selection of scale parameters strongly constraining posterior inferences In a second strategy, we include different priors that constrain the posterior inferences. In this case, the choice of scale parameters is not data-driven but parameters are chosen in a way to limit the influence of the (selective) data. To constrain the variability of the posterior inferences, we use priors

for which a high proportion of their probability mass is shifted to an area close to a variance parameter of zero, for example, the Half-Student-t distribution $StudentT(4, 0, 1)$ and Half-normal distribution $N(0, 1)$ proposed by Gelman (2020) for variables x_1 and x_2. Also for the scaled inverse chi-square distribution we set rather small values for the scale parameter and lower upper bounds of the uniform distribution.

Selection of parameters weakly constraining posterior inferences As a third strategy, the posterior inference of the different variables is constrained to different degrees based on expectations about the variance parameters for the different variables. Comparing the findings from this strategy to those from the second strategy, we analyze the effect of different constraints on the estimation, e.g., whether a weak constraint performs better than a strong constraint that might restrict the estimation too much.

Mixed strategy As a fourth strategy, we use a combination of the data-driven strategy and the strong constraints strategy. The idea is that researchers can put constraints on the posterior inference of certain variables based on their experience from prior research while not using constraints for other variables. Using this strategy, parameters to constrain posterior inferences are only used for variables for which the estimated variance parameters from the sample strongly differ from the range that is usually found for these variables.

Other priors commonly suggested in the literature For the sake of completeness and for comparison reasons, we consider the priors described in Section 2.3 that are further proposed in the literature by Gelman (2006), Gelman (2006) and Spiegelhalter *et al.* (2003):

- $Inv - Gamma(0.001, 0.001)$ (iv.gam.001)
- $Inv - Gamma(1, 1)$. (iv.gam.1)
- $Cauchy(0, 5)$ (cau.5)
- $Cauchy(0, 25)$ (cau.25)

Table 1: Distributions used in the four strategies

strategy	variables	prior distributions			
		uniform	Half-Student-t	Half-normal	Scale-Inv-χ^2
strategy 1 (data-driven)	x_1, x_2	$U(0, \text{inf})$	$StudentT(4, 0, \hat{\sigma}_{var})$	$N(0, \hat{\sigma}_{var})$	Scale-Inv.-$\chi^2(100, \hat{\sigma}_{var})$
	x_3, x_4	$U(0, \text{inf})$	$StudentT(4, 0, \hat{\sigma}_{var})$	$N(0, \hat{\sigma}_{var})$	Scale-Inv.-$\chi^2(100, \hat{\sigma}_{var})$
		(uni.S)	(stu.S)	(norm.S)	(sc.iv.chi.sq.S)
strategy 2 (strong constraint)	x_1, x_2	$U(0, 1)$	$StudentT(4, 0, 1)$	$N(0, 1)$	Scale-Inv.-$\chi^2(100, 0.001)$
	x_3, x_4	$U(0, 1)$	$StudentT(4, 0, 0.1)$	$N(0, 0.1)$	Scale-Inv.-$\chi^2(100, 0.001)$
		(uni.strong)	(stu.strong)	(norm.strong)	(sc.iv.chi.sq.strong)
strategy 3 (weak constraint)	x_1, x_2	$U(0, 10)$	$StudentT(4, 0, 5)$	$N(0, 5)$	Scale-Inv.-$\chi^2(100, 5)$
	x_3, x_4	$U(0, 1)$	$StudentT(4, 0, 0.1)$	$N(0, 0.1)$	Scale-Inv.-$\chi^2(100, 0.001)$
		(uni.weak)	(stu.weak)	(norm.weak)	(sc.iv.chi.sq.weak)
strategy 4 (mixed)	x_1, x_2	$U(0, \text{inf})$	$StudentT(4, 0, \hat{\sigma}_{var})$	$N(0, \hat{\sigma}_{var})$	Scale-Inv.-$\chi^2(100, \hat{\sigma}_{var})$
	x_3, x_4	$U(0, 1)$	$StudentT(4, 0, 0.1)$	$N(0, 0.1)$	Scale-Inv.-$\chi^2(100, 0.001)$
		(uni.mix)	(stu.mix)	(norm.mix)	(sc.iv.chi.sq.mix)

3.2. Estimated variance parameters in both scenarios

To have a first impression of the effect of the sample selectivity on the variance parameters, we estimate the variance parameters $\hat{\sigma}_{var}$ applying a simple maximum likelihood estimation using the glmer function of the lme4 package (Bates *et al.* 2015). We use a multilevel logistic regression with y as dependent variable and x_1, x_2, x_3 and x_4 as independent variables. Summary statistics of the variance parameter estimations for the variables x_1, x_2, x_3 and x_4 over all samples of the simulation study for both scenarios can be found in Table 2. In a

real data application, the information in this table can be seen as the collected knowledge a researcher has on the variables under study based on his prior research. The estimated variance parameters $\hat{\sigma}_{var}$ in each sample are used as prior information in the Half-Student-t distribution, the Half-normal distribution and the the scaled inverse chi-square distribution within the first strategy that leads to data driven posterior inferences.

The estimations of the variance parameters for scenario 2 are higher on average (means and medians) than for scenario 1. Looking at the distribution of the variance parameters more closely, we find large differences for the two scenarios that are caused by the different selectivity of the sampling process: scenario 2 shows much larger outliers than scenario 1 which can be seen by comparing the maximum values for each x-Variable for the two scenarios and by comparing the maximum value to the third quartile (which is given by the ratio in the last column). The outliers are specifically strong for variables x_3 and x_4 in scenario 2 where they are completely out of the range of the simulation study's sample estimations of the variance parameters described by the means and medians. The estimations of the variance parameter for variables x_1 and x_2 also have some large outliers in scenario 2. However, these outliers are not out of the range of the sample estimations to the same extend as those for variables x_3 and x_4.

For the samples that contain very large outliers, choosing priors for x_3 and x_4 that lead to data driven posterior inferences - our first strategy - will likely result in a large variability in the posterior inference and may destabilize the overall estimation of y. It might thus be beneficial to use priors that constrain the effect of the selective data on the posterior inference. We use priors that strongly constrain the estimation for all variables in our second strategy by using priors for which a high proportion of their probability mass is shifted to an area close to a variance parameter of zero. Note that the estimation of the variance parameter for the variables x_3 and x_4 is close to zero on average. Thus, the constraints have a much stronger impact on variables x_1 and x_2, which can be seen by the higher average of their respective estimations of the variance parameter. Putting the same constrains on all variables might not be well suited as well. For the third strategy we thus use the priors for which a high proportion of their probability mass is shifted to an area close to a variance parameter of zero for variables x_3 and x_4, since they have means and medians closer to zero and since they have outliers that are more out of the range compared to x_1 and x_2. Variables x_1 and x_2 have weaker outliers but higher means and medians and thus, using priors with a probability mass close to zero may constrain posterior inferences too much. We thus apply weaker constraints to x_1 and x_2. The constraints are chosen to be close to the medians of the estimated variance parameters in Table 2 and we set the scale parameter of the Half-normal, half t- student and the scale inverse chi square distribution on 5 and the upper bound of the uniform distribution on 10. For the mixed strategy, we put the same constraints as for strategy 2 on variables x_3 and x_4 that show very large outliers in the variance parameter estimation and use the data-driven priors from strategy one for x_1 and x_2 that show weaker outliers. Strategy 2, 3 and 4 only differ in their prior distributions for variables x_1 and x_2 while the same prior distributions are used for variables x_3 and x_4.

Table 2: Distribution of the estimated variance parameters $\hat{\sigma}_{var}$

scenario	Variable	Min.	first Quartile	Median	Mean	third Quartile	Max.	$\frac{Max.}{thirdQuartile}$
scenario 1	x_1	3.16	4.24	4.73	4.73	5.08	6.38	1.26
	x_2	2.99	4.11	4.50	4.64	4.91	7.04	1.43
	x_3	0.33	0.76	0.88	0.89	0.97	1.48	1.52
	x_4	0.00	0.00	0.00	0.17	0.17	0.87	5.19
scenario 2	x_1	3.06	4.64	5.37	5.76	6.21	18.46	2.97
	x_2	3.12	4.86	5.78	6.47	7.06	24.56	3.48
	x_3	0.00	0.64	0.98	1.39	1.34	21.40	15.96
	x_4	0.00	0.00	0.00	1.79	0.00	14.68	inf

4. Results

The results for the two scenarios are presented separately. The several tables show the distribution (minimum, first quartile, median, mean third quartile and maximum) of the different estimators across all samples of the simulation study. The last column informs on the precision of the estimation in terms of the Monte Carlo variance (MCVar) that describes the variance of the point estimation of an estimator across the simulation rounds.

Scenario 1 The results for scenario 1 (moderately selective inclusion mechanism) can be found in Table 3. Most estimations are close to the benchmark of 0.5 and most Monte Carlo variances are about 0.0001. The estimators perform equally well and the results do not depend on the choice of the prior distribution or on the chosen scale parameters. Exceptions are the half-normal distribution within the data-driven strategy (norm.S) that has some large outliers indicating that the estimation is less stable for the half-normal prior than for other priors of the data-driven strategy.

Table 3: Results of the simulation study: estimated p for the different prior distribution of scenario 1. The true value equals $p = 0.5$.

strategy	prior	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.	NA.s	MCVar
strategy 1 (data-driven)	uni.S	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	stu.S	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	norm.S	0.26	0.49	0.50	0.50	0.51	0.75	0.00	0.0003
	sc.iv.chi.sq.S	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
strategy 2 (strong constraint)	uni.strong	0.45	0.48	0.49	0.49	0.50	0.53	0.00	0.0001
	stu.strong	0.46	0.49	0.50	0.50	0.51	0.54	0.00	0.0001
	norm.strong	0.45	0.49	0.50	0.50	0.50	0.54	0.00	0.0001
	sc.iv.chi.sq.strong	0.23	0.26	0.27	0.27	0.28	0.31	0.00	0.0002
strategy 3 (weak constraint)	uni.weak	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	stu.weak	0.46	0.49	0.50	0.50	0.51	0.54	0.00	0.0001
	norm.weak	0.46	0.49	0.50	0.50	0.51	0.54	0.00	0.0001
	sc.iv.chi.sq.weak	0.45	0.49	0.50	0.50	0.50	0.54	0.00	0.0001
strategy 4 (mixed)	uni.mix	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	stu.mix	0.46	0.49	0.50	0.50	0.51	0.54	0.00	0.0001
	norm.mix	0.46	0.49	0.50	0.50	0.51	0.54	0.00	0.0001
	sc.iv.chi.sq.mix	0.45	0.49	0.50	0.50	0.50	0.54	0.00	0.0001
other commonly used prior	iv.gam.001	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	iv.gam.1	0.45	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	cau.5	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001
	cau.25	0.46	0.49	0.50	0.50	0.51	0.55	0.00	0.0001

Within the constraining priors, the inverse chi-squared prior (sc.iv.chi.sq.strong) is found to produce strongly biased estimation, leading to an underestimation of the proportion of y in every simulation round. This means that the posterior inference is restricted too much in this case, particularly for the variance parameters of variables x_1 and x_2. For the uniform distribution uni.strong, the initialization failed for seven samples, for which we have to repeat the estimation process. The mixed strategies and the priors of strategy 4 perform equally well.

In the case of a moderate inclusion mechanism and moderately selective data, almost all considered prior distributions for the variance parameter lead to acceptable results.

Scenario 2 Table 4 shows the results for scenario 2 in which the selectivity of the inclusion mechanism is higher than for scenario 1. In scenario 2, the estimated variances show higher variation and larger outliers than in scenario 1 as described in Section 3.2. For all considered distributions larger Monte Carlo biases of the overall estimation of p is found for scenario 2 than for scenario 1. Also, the estimation is much less stable in scenario 2 due to the highly selective inclusion mechanism and highly selective data. The proportion could not be computed for some simulation runs for the uniform distribution (uni.S) in the data-driven strategies, for the uniform distribution (uni.mix) in the mixed strategy and the inverse gamma distribution (iv.gam.1) leads to missing information (see last column of Table 4). As in

Table 4: Results of the simulation study: estimated p for the different prior distribution of scenario 2. The true value equals $p = 0.5$.

strategy	prior	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.	NA.s	MCVar
strategy 1 (data-driven)	uni.S	0.03	0.11	0.19	0.23	0.32	0.66	112	0.0209
	stu.S	0.03	0.36	0.43	0.42	0.49	0.64	0.00	0.0126
	norm.S	0.06	0.36	0.43	0.42	0.49	0.64	0.00	0.0115
	sc.iv.chi.sq.S	0.08	0.36	0.43	0.42	0.49	0.63	0.00	0.0106
strategy 2 (strong constraint)	uni.strong	0.13	0.25	0.27	0.27	0.30	0.39	0.00	0.0017
	stu.strong	0.19	0.37	0.42	0.42	0.47	0.62	0.00	0.0056
	norm.strong	0.19	0.36	0.40	0.40	0.44	0.58	0.00	0.0033
	sc.iv.chi.sq.strong	0.02	0.03	0.03	0.03	0.04	0.05	0.00	0.0000
strategy 3 (weak constraint)	uni.weak	0.21	0.37	0.44	0.44	0.50	0.63	0.00	0.0086
	stu.weak	0.20	0.37	0.44	0.44	0.50	0.64	0.00	0.0091
	norm.weak	0.21	0.38	0.44	0.44	0.50	0.63	0.00	0.0081
	sc.iv.chi.sq.weak	0.25	0.38	0.44	0.44	0.50	0.63	0.00	0.0068
strategy 4 (mixed)	uni.mix	0.19	0.35	0.42	0.42	0.52	0.64	4.00	0.0150
	stu.mix	0.20	0.37	0.44	0.43	0.50	0.64	0.00	0.0103
	norm.mix	0.20	0.38	0.44	0.44	0.50	0.64	0.00	0.0094
	sc.iv.chi.sq.mix	0.20	0.38	0.44	0.44	0.50	0.64	0.00	0.0086
other commonly used prior	iv.gam.001	0.03	0.22	0.37	0.35	0.46	0.64	1.00	0.0218
	iv.gam.1	0.15	0.33	0.39	0.39	0.45	0.62	0.00	0.0083
	cau.5	0.03	0.26	0.36	0.35	0.44	0.63	0.00	0.0172
	cau.25	0.03	0.19	0.29	0.30	0.41	0.64	0.00	0.0208

scenario 1, the initialization failed for eight samples of the uniform distribution (uni.strong) for which we have to repeat the estimation process. The data-driven strategy does not perform well which can be seen by some very low estimations for p that are close to zero. The high variability of some samples caused by the variance parameters that is completely out of the range of the simulation study (particularly for variables x_3 and x_4) affect the varying coefficients and the overall estimation, leading to such low estimates. The estimators of this strategy also show higher Monte Carlo variances (and thus smaller precision) than the estimators of strategies 2 to 4. In comparison to the data-driven strategy, using strong constrains (strategy 2) leads to less very low estimates except for sc.iv.chi.sq.strong that again constrains posterior inferences too much. On average, the estimation is more biased for strategy 2 than for strategy 1 for each prior distribution except for the non-informative and non-proper uniform distribution uni.S (the problems of this distribution are described in Gelman (2006) and Section 2.3). Thus, posterior inferences may be restricted too much, in particular, for variables x_1 and x_2. Monte Carlo variances, however, are rather low for this strategy.

Strategies 3 and 4 perform about equally well, being less biased than strategies 1 and 2 and showing less very low outliers close to zero. Monte Carlo variances for strategy 3 are slightly lower than for strategy 4. Other priors, that are commonly suggested in the literature do not perform well: they show comparatively large Monte Carlo bias and very small minimum values.

Applying constraining priors for the variance parameters of variable x_3 and x_4 and priors that lead to data-driven posterior inferences for the variance parameters of variable x_1 and x_2 in strategy 4 leads to very good results. Monte Carlo bias is smaller than for estimators of strategy 2. Unrealistic small values below 0.1 can also be avoided.

The constraining priors of strategy 3 lead to good results as well. Applying the scale inverse chi-square distribution results in the largest minimum value in comparison to the other priors and other strategies. At the same time, using the scale inverse chi-square distribution in strategy 3 leads to a lower Monte Carlo bias in comparison to other estimators. In comparison to the strategy 2 that only differs from scenario 3 for the scale parameters for the prior distributions of variables x_1 and x_2, our findings show again that constraining the influence of highly selective data may be helpful to improve the estimation but constraining posterior inferences too much may have an opposite effect. To study this effect, we vary the scale parameters of the scale inverse chi-square distribution for variables x_1 and x_2 starting from parameter constellation of the scale inverse chi-square distribution in strategy 3 that lead to

good estimation results. The results are shown in Table 5.

Comparing the minimum values, we find that decreasing the scale parameters leads to larger minimum values, i.e., unrealistic small values, up to a certain point. After that, the minimum value decreases. In our case, this point is reached when choosing a scale parameter of approximately 3 for variables x_1 and x_2, a value that is out of the range of the estimated variance (Table 2) for both variables. The Monte Carlo bias is highest for $Scale - inv - \chi^2(2,2)$ and lowest for $Scale - inv - \chi^2(7,7)$ and $Scale - inv - \chi^2(6,6)$. An opposite effect can be observed for the Monte Carlo variances. These are lowest for the estimator $Scale - inv - \chi^2(2,2)$ and highest for the estimator $Scale - inv - \chi^2(7,7)$.

Table 5: Different choices of scale parameters for the scale inverse chi-square distribution for variables x_1 and x_2. The prior for variables x_3 and x_4 is $Scale - inv - \chi^2(100, 0.001)$ for each alternative.

Prior	Min.	first Qu.	Median	Mean	third Qu.	Max.	NAs	MCVar
sc.iv.chi.sq.7.7	0.22	0.39	0.45	0.45	0.52	0.64	0.00	0.0086
sc.iv.chi.sq.6.6	0.23	0.39	0.45	0.45	0.51	0.64	0.00	0.0077
sc.iv.chi.sq.5.5	0.25	0.38	0.44	0.44	0.50	0.63	0.00	0.0068
sc.iv.chi.sq.4.4	0.27	0.38	0.43	0.43	0.49	0.62	0.00	0.0056
sc.iv.chi.sq.3.3	0.26	0.37	0.42	0.42	0.46	0.60	0.00	0.0042
sc.iv.chi.sq.2.2	0.23	0.36	0.39	0.40	0.43	0.55	0.00	0.0026

5. Summary and conclusion

In this paper, we compare different prior distributions for the variance parameters of the varying coefficients in the multilevel regression and poststratification approach with respect to their ability to stabilize the estimation that is based on highly selective survey data that result from a highly selective inclusion mechanism. Under the conditions of our simulation study, we find two strategies to perform very well: a strategy combining priors that leads to data-driven posterior inferences and priors that constrain posterior inferences and a strategy applying priors that constrain posterior inferences not too much. Particularly, using the scale inverse chi-square distribution that does not constrain posterior inferences too much leads to the best results.

In scenario 1, in which the inclusion mechanism was less selective than in scenario 2, most of the priors perform well. Notably, strategies that lead to a data-driven posterior inferences, particularly the non-proper and non-informative uniform prior, perform well. In scenario 2, as a result of the highly selective inclusion mechanism and the highly selective data, a large variability arises through the variance parameter for some variables in some samples. The simulation study shows that using priors that result in data-driven posterior inferences in such cases leads to an inefficient estimation and in some samples leads to estimations that are by far too small. Using priors that constrain the influence of the highly selective data may be useful. However, attention should be paid to posterior inferences being not constrained too much. At a certain point, the restriction can lead to an opposite effect and the biases may increase.

The choice of the prior for the variance parameter of the variable's varying coefficient in general depends on the reliability of the prior information and sample information. For each application, the analyst needs to decide which kind of prior strategy to take. In case the analyst does not have prior information, an intuitive choice could be to apply priors that lead to data-driven posterior inferences, particularly, to use non-informative priors. However, our simulation study shows that this strategy is very problematic if the data is heavily selective due to a very selective inclusion mechanism. In such cases, it is preferable to use priors that somehow constrain the influence of the highly selective data and the posterior inference.

The choice of the constraint might be informed by previous surveys or applications or the expert knowledge of the analyst. The analyst may also evaluate the posterior inferences of the variance parameter or the maximum likelihood estimation of the parameter with respect to their plausibility, particularly, when the data result from a highly selective inclusion mechanism. For example, the distributions of posterior simulations of the variance parameter can be evaluated graphically, as for example done in (Gelman 2006). Unrealistically high values may be detected by comparing the variance parameters of different variables or in comparison to previous surveys or comparable applications. Comparing the weighting variables' distribution in the sample with external benchmarks (e.g., from official statistics) can serve as a first evaluation of the sample's selectivity (see, for example, Felderer, Kirchner, and Kreuter (2019)) and inform the decision on which prior strategy to use. In addition, measures such as described in Little *et al.* (2019) might be applied.

The Monte Carlo simulation study shows that the Bayesian multilevel regression and post-stratification approach needs to be carefully applied when using data that result from a selective inclusion mechanism. Up to a certain point of selectivity (in scenario 1), we still receive good estimations for many prior choices. For very selective inclusion mechanisms, however, reliable estimations cannot be ensured as seen for scenario 2. This is even more important in practical applications, when the nonresponse mechanism cannot be compensated completely via a weighting procedure and no perfect prior information is available. Thus, more research is needed for estimation procedures that can be applied on highly selective data, particularly, when this procedures are based on a Bayesian framework.

Our findings are limited to the chosen prior distributions and scale parameters. For example, for the Half-Student-t distribution and the scaled inverse chi-square distribution the degrees of freedom can further be varied. Also, one could think of applying different prior distributions for different variables. We compare a moderately skewed sample to a heavily skewed sample and find the choice of prior only to strongly affect the estimation for the heavily skewed sample. More research is needed on the degree of selectivity for which the choice of the prior is crucial for the outcome and for which arbitrary choices are possible. The simulation study focuses on the usefulness of different prior strategies for samples with different degrees of selectivity. Although selectivity is arguable one of the main problems when conducting surveys, there are other issues that might affect the performance of the different strategies and are thus worth to be studied. For example, further research is needed to evaluate the effect of (non-) coverage on the performance of the strategies. The results of the simulation study are further limited to the parameter constellations and the data sets of the simulation study. As a single Monte Carlo simulation study can only be conducted manipulating certain parameter constellations and data sets, further research should be conducted to simulate different parameter constellations.

A. Variables generated in the simulation study

The following tables are taken from or based on Bruch and Felderer (2021), in which the simulation study used for this article is introduced.

Table 6: Distribution of the weighting variables in the population (benchmark) and the samples of the simulation scenarios

	benchmark	scenario 1	scenario 2
x_1	0.215	0.408	0.574
	0.161	0.155	0.165
	0.124	0.096	0.094
	0.236	0.256	0.150
	0.264	0.085	0.016
x_2	0.250	0.482	0.854
	0.250	0.216	0.062
	0.250	0.175	0.025
	0.150	0.084	0.047
	0.100	0.042	0.012
x_3	0.250	0.266	0.318
	0.250	0.289	0.269
	0.250	0.246	0.230
	0.250	0.199	0.183
x_4	0.250	0.412	0.520
	0.250	0.229	0.220
	0.250	0.200	0.168
	0.150	0.129	0.081
	0.100	0.031	0.011

Table 7: Correlation structure (Spearman correlation; averaged over all simulation runs) in the realized sample in scenario 1

	y	x_1	x_2	x_3	x_4
y	1.00	0.59	0.58	0.02	0.08
x_1	0.59	1.00	0.41	0.09	0.02
x_2	0.58	0.41	1.00	0.03	0.21
x_3	0.02	0.09	0.03	1.00	0.30
x_4	0.08	0.02	0.21	0.30	1.00

Table 8: Correlation structure (Spearman correlation; averaged over all simulation runs) in the realized sample in scenario 2

	y	x_1	x_2	x_3	x_4
y	1.00	0.25	0.36	-0.07	-0.15
x_1	0.25	1.00	-0.10	0.06	-0.15
x_2	0.36	-0.10	1.00	-0.14	-0.24
x_3	-0.07	0.06	-0.14	1.00	0.30
x_4	-0.15	-0.15	-0.24	0.30	1.00

Table 9: Values for the parameters in the inclusion model generating the inclusion propensities ω depending on the weighting variables $x_1^U \dots x_4^U$ including the numbers of resulting weighting cells and amount of empty weighting cells averaged over all simulation rounds

Parameter	category	scenario 1	scenario 2
λ		-1	1e-07
$\xi_{j[i]}^{x_1^U}$	$j=1$	1	50
	$j=2$	0	0
	$j=3$	-0.2	-0.2
	$j=4$	0.4	0.4
	$j=5$	-1	-50
$\xi_{k[i]}^{x_2^U}$	$k=1$	0	0
	$k=2$	-1	-55
	$k=3$	-1	-55
	$k=4$	-1	-50
	$k=5$	-1	-50
$\xi_{o[i]}^{x_3^U}$	$o=1$	0	0
	$o=2$	0.4	0.4
	$o=3$	0.3	0.3
	$o=4$	0.2	0.2
$\xi_{v[i]}^{x_4^U}$	$v=1$	1	50
	$v=2$	0.1	0.1
	$v=3$	0	0
	$v=4$	0.2	0.2
	$v=5$	-1	-50
weighting cells		500	500
empty cells		217.82 (43.6 %)	390.97 (78.2 %)

B. Varying further dimensions of the simulation study

In the simulation study, we mainly focus on the usefulness of the different prior strategies on samples with different degrees of selectivity. In an additional analysis, we repeat some of the analysis varying two more dimensions: the sample size and the correlation between the weighting and dependent variables. Both aspects are analyzed separately due to computational constraints. Since the simulation study is computationally very intensive as result of the Bayesian components of the different estimators, we need to further restrict the additional analysis. Since our main study showed that the differences between the different prior distributions within a certain strategy are lower than the differences between the four strategies, we chose to only use one prior distribution for each strategy. We use the scale inverse chi-squared distribution for all strategies because it showed very good results throughout the strategies.

B.1. Sample size variation

In the first additional analysis, we reduce the sample size from $n = 1,000$ to $n = 500$.

As can be seen in Tables 11 and 12, the variability of most of the estimators is increased and the precision reduced (as compared to Tables 3 and 4).

For the bias analysis we find - like for the main study with $n = 1,000$ - bias to be smaller for the moderately selective inclusion mechanism than for the highly skewed inclusion mechanism. Results for the moderately selective inclusion mechanism are very similar for the reduced and larger sample sizes: the estimators sc.iv.chi.sq.S, sc.iv.chi.sq.weak and sc.iv.chi.sq.mix lead to similar results that are closed to the benchmark. The prior of estimator sc.iv.chi.sq.strong constraints the posterior inference too much and leads to strong underestimations.

For the highly skewed inclusion mechanism, we find bias to be larger for the estimators in the case of $n = 500$ than for a sample size of $n = 1,000$. We find that the relative performance of the estimators is the same for the larger and smaller sample size but the differences are not as strong for the smaller than for the larger sample size. Looking at bias, we find that the estimator sc.iv.chi.sq.strong leads to estimations that are very different from the benchmark. The sc.iv.chi.sq.weak is comparable to estimator sc.iv.chi.sq.mix on average. Both estimators perform slightly better than sc.iv.chi.sq.S. Among these three estimators

Table 10: Values for the parameters in the model that generates the survey variable of interest p_y depending on the weighting variables $x_1^U \dots x_4^U$

Parameter	category	value
δ		-0.5
$\gamma_{j[i]}^{x_1^U}$	$j=1$	-7
	$j=2$	0
	$j=3$	-0.2
	$j=4$	0.4
	$j=5$	7
$\gamma_{k[i]}^{x_2^U}$	$k=1$	-7
	$k=2$	0.4
	$k=3$	0
	$k=4$	-0.2
	$k=5$	7
$\gamma_{o[i]}^{x_3^U}$	$o=1$	1.5
	$o=2$	-0.4
	$o=3$	0
	$o=4$	0.9
$\gamma_{v[i]}^{x_4^U}$	$v=1$	0.1
	$v=2$	0.1
	$v=3$	0.1
	$v=4$	0
	$v=5$	0.1

that show similar biases, the estimator sc.iv.chi.sq.weak leads to the best results in terms of Monte Carlo variance. The results show that for the highly skewed selection mechanism, the performance of all estimators is worse on the smaller than the larger sample size in terms of bias and Monte Carlo variance.

Table 11: Results of the simulation study for estimated p for a reduced sample size of $n = 500$ compared to scenario 1. The true value equals $p = 0.5$.

Prior	Min.	firstQu.	Median	Mean	thirdQu.	Max.	NAs	MCVar
sc.iv.chi.sq.S	0.45	0.49	0.50	0.50	0.51	0.55	0.00	0.0002
sc.iv.chi.sq.strong	0.21	0.25	0.26	0.26	0.28	0.33	0.00	0.0004
sc.iv.chi.sq.weak	0.44	0.48	0.49	0.49	0.51	0.55	0.00	0.0003
sc.iv.chi.sq.mix	0.44	0.48	0.49	0.49	0.51	0.55	0.00	0.0003

Table 12: Results of the simulation study for estimated p for a reduced sample size of 500 compared to scenario 2. The true value equals $p = 0.5$.

Prior	Min.	first Qu.	Median	Mean	third Qu.	Max.	NAs	MCVar
sc.iv.chi.sq.S	0.05	0.32	0.38	0.38	0.45	0.64	0.00	0.0105
sc.iv.chi.sq.strong	0.01	0.03	0.03	0.03	0.04	0.06	0.00	0.0001
sc.iv.chi.sq.weak	0.07	0.33	0.40	0.40	0.46	0.62	0.00	0.0070
sc.iv.chi.sq.mix	0.07	0.32	0.40	0.40	0.47	0.65	0.00	0.0092

B.2. Covariance and correlation structure variation

In the second additional analysis, we reduce the covariances and correlations between the variables. The following population covariance matrix is used:

$$\Sigma = \begin{pmatrix} 1,000 & 50 & 25 & 5 \\ 50 & 100 & 1 & 5 \\ 25 & 1 & 75 & 5 \\ 5 & 5 & 5 & 10 \end{pmatrix}$$

Table 13 shows the resulting correlation structures in the sample after the inclusion mechanism of scenario 1 is applied (for comparison with the main study see tables 7).

Table 13: Sample correlation structure (Spearman correlation; averaged over all simulation runs) for scenario 1 with reduced covariances

	y	x_1	x_2	x_3	x_4
y	1.00	0.51	0.46	0.02	0.02
x_1	0.51	1.00	0.09	0.07	-0.01
x_2	0.46	0.09	1.00	0.00	0.09
x_3	0.02	0.07	0.00	1.00	0.15
x_4	0.02	-0.01	0.09	0.15	1.00

Table 14: Results of the simulation study for estimated p for a reduced covariance structure compared to scenario 1. The true value equals $p = 0.49$.

Prior	Min.	first Qu.	Median	Mean	third Qu.	Max.	NAs	MCVar
sc.iv.chi.sq.S	0.46	0.48	0.49	0.49	0.50	0.52	0.00	0.00
sc.iv.chi.sq.strong	0.24	0.27	0.28	0.28	0.29	0.33	0.00	0.00
sc.iv.chi.sq.weak	0.45	0.48	0.48	0.48	0.49	0.51	0.00	0.00
sc.iv.chi.sq.mix	0.45	0.48	0.48	0.48	0.49	0.52	0.00	0.00

The results are very similar to the results of scenario 1 in the main study. Estimator sc.iv.chi.sq.strong restricts the posterior inferences too much while the estimators sc.iv.chi.sq.S, sc.iv.chi.sq.weak and sc.iv.chi.sq.mix lead to similar results with estimations close to the benchmark.

Acknowledgements

The authors thank Matthias Sand and the anonymous reviewer for their helpful comments on this research.

References

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Browne WJ, Draper D (2006). "A Comparison of Bayesian and Likelihood-Based Methods for Fitting Multilevel Models." *Bayesian Analysis*, **1**(3), 473 – 514. doi:10.1214/06-BA117.

Bruch C, Felderer B (2021). "Applying Multilevel Regression Weighting When Only Population Margins Are Available." *Communications in Statistics - Simulation and Computation (forthcoming)*. doi:10.1080/03610918.2021.1988642.

Burgard JP (2015). *Evaluation of Small Area Techniques for Applications in Official Statistics*. doctoral thesis, Universität Trier. doi:10.25353/ubtr-xxxx-2d32-f845.

Cornesse C, Blom AG, Dutwin D, Krosnick JA, De Leeuw ED, Legleye S, Pasek J, Pennay D, Phillips B, Sakshaug JW, Struminskaya B, Wenz A (2020). "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology*, **8**(1), 4–36. doi:10.1093/jssam/smz041.

Enderle T, Münnich R, Bruch C (2013). "On the Impact of Response Patterns on Survey Estimates from Access Panels." *Survey Research Methods*, **7**(2), 91–101. doi:10.18148/srm/2013.v7i2.5036.

Felderer B, Kirchner A, Kreuter F (2019). "The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias." *Journal of Official Statistics (JOS)*, **35**(1). doi:10.2478/jos-2019-0005.

Gelman A (2006). "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis*, **1**(3), 515–534. doi:10.1214/06-BA117A.

Gelman A (2020). "Prior Choice Recommendations." https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations. Accessed: 2021-02-24.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gelman A, Little TC (1997). "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology*, **23(2)**, 127–135.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2019). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-11, URL https://CRAN.R-project.org/package=mvtnorm.

Kalton G, Flores-Cervantes I (2003). "Weighting Methods." *Journal of Official Statistics*, **19(2)**, 81–97.

Little RJA, West BT, Boonstra PS, Hu J (2019). "Measures of the Degree of Departure from Ignorable Sample Selection." *Journal of Survey Statistics and Methodology*, **8**(5), 932–964. doi:10.1093/jssam/smz023.

Park DK, Gelman A, Bafumi J (2004). "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis*, **12**(4), 375–385. doi:10.1093/pan/mph024.

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

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR, Lunn D (2003). "BUGS: Bayesian Inference Using Gibbs Sampling."

Stan Development Team (2018). "RStan: The R Interface to Stan." R package version 2.18.2, URL http://mc-stan.org/.

Statisticat, LLC (2018). "LaplacesDemon: Complete Environment for Bayesian Inference." R package version 16.1.1.

Wang W, Rothschild D, Goel S, Gelman A (2015). "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting*, **31**, 980–991. doi:10.1016/j.ijforecast.2014.06.001.

Woodward P (2011). *Bayesian Analysis Made Simple: An Excel GUI for WinBUGS*. Taylor & Francis.

Affiliation:

Christian Bruch
GESIS Leibniz Institute for the Social Sciences
Square B2, 1 68159 Mannheim
E-mail: Christian.Bruch@gesis.org
URL: https://www.gesis.org/institut/mitarbeiterverzeichnis/person/christian.bruch