scientific reports

OPEN

Joint distribution properties of fully conditional specification under the normal linear model with normal inverse-gamma priors

Mingyang Cai✉, Stef van Buuren & Gerko Vink

Fully conditional specification (FCS) is a convenient and flexible multiple imputation approach. It specifies a sequence of simple regression models instead of a potential complex joint density for missing variables. However, FCS may not converge to a stationary distribution. Many authors have studied the convergence properties of FCS when priors of conditional models are non-informative. We extend to the case of informative priors. This paper evaluates the convergence properties of the normal linear model with normal-inverse gamma priors. The theoretical and simulation results prove the convergence of FCS and show the equivalence of prior specification under the joint model and a set of conditional models when the analysis model is a linear regression with normal inverse-gamma priors.

Multiple imputation[1] is a widely applied approach for the analysis of incomplete datasets. It involves replacing each missing cell with several plausible imputed values that are drawn from the corresponding posterior predictive distributions. There are two dominant approaches to arrive at those posterior distributions under multivariate missing data: joint modeling (JM) and fully conditional specification (FCS).

Joint modeling requires a specified joint model for the complete data. Schafer[2] illustrated joint modeling imputation under the multivariate normal model, the saturated multinomial model, the log-linear model, and the general location model. However, with an increasing number of variables and different levels of measurement, it can be challenging to formulate the joint distribution of the data.

Fully conditional specification offers a solution to this challenge by allowing a flexible specification of the imputation model for each partially observed variable. The imputation procedure then starts by imputing missing values with a random draw from the marginal distribution. Each incomplete variable is then iteratively imputed with a specified univariate imputation model.

Fully conditional specification has been proposed under a variety of names: chained equations stochastic relaxation, variable-by-variable imputation, switching regression, sequential regressions, ordered pseudo-Gibbs sampler, partially incompatible MCMC and iterated univariate imputation[3]. Fully conditional specification can be of great value in practice because of its flexibility in model specification. FCS has become a standard in practice and has been widely implemented in software (e.g. `mice` and `mi` in R, `IVEWARE` in SAS, `ice` in STATA and module `MVA` in SPSS)[4].

Although many simulation studies demonstrated that fully conditional specification yields plausible imputations in various cases, the theoretical properties of fully conditional specification are not thoroughly understood[5]. A sequence of conditional models may not imply a joint distribution to which the algorithm converges. In such a case, the imputation results may systematically differ according to different visit sequences, which is named "order effects"[6].

Van Buuren[3] stated two cases in which FCS converges to a joint distribution. First, if all imputation models are linear with a homogenous normal distributed response, the implicit joint model would be the multivariate normal distribution. Second, if three incomplete binary variables are imputed with a two-way interactions logistic regression model, FCS would be equivalent to the joint modeling under a zero three-way interaction log-linear model. Liu et al.[7] illustrated a series of sufficient conditions under which the imputation distribution for FCS converges in total variation to the posterior distribution of a joint Bayesian model when the sample size moves to infinity. Complementing the work of Liu et al.[7], Hughes[6] pointed out that, in addition to the compatibility,

Department of Methodology and Statistics, Utrecht University, Utrecht 3584 CH, The Netherlands. ✉email: m.cai@uu.nl

a "non-informative margins" condition is another sufficient condition for the equivalency of FCS and joint modeling for finite samples. Hughes[6] also showed that with multivariate normal distributed data and a non-informative prior, both compatibility and the non-informative margins conditions are satisfied. In that case, fully conditional specification and joint modeling provide imputations from the same posterior distribution. Zhu & Raghunathan[8] discussed conditions for convergence and assessed the properties of FCS. Many authors illustrated convergence properties of FCS when the prior for conditional models is non-informative. However, the case of informative priors has not received much attention. Therefore, we should consider the equivalent prior specification for informative priors under a sequence of conditional and corresponding joint models. This additional investigation allows the imputer to perform imputations under FCS even if they only collect the prior joint information for the incomplete dataset.

For the initial step to evaluate convergence properties of FCS with informative priors, it is sensible to focus on the Bayesian normal linear models and the typical informative prior: normal inverse-gamma prior. This paper will briefly overview joint modeling, fully conditional specification, compatibility, and non-informative margins. Then, we derive a theoretical result and perform a simulation study to evaluate the non-informative margins condition. We also consider the prior for the target joint density of a sequence of normal linear models with normal inverse-gamma priors. Finally, some remarks are concluded.

Background

Joint modeling. Let Y^{obs} and Y^{mis} denote the observed and missing data in the dataset Y. Joint modeling involves specifying a parametric joint model $p(Y^{obs}, Y^{mis}|\theta)$ for the complete data and an appropriate prior distribution $p(\theta)$ for the parameter θ. Incomplete cases are partitioned into groups according to various missing patterns and then imputed with different sub-models. Under the assumption of ignorability, the imputation model for each group is the corresponding conditional distribution derived from the assumed joint model

$$p(Y^{mis}|Y^{obs}) = \int p(Y^{mis}|Y^{obs}, \theta)p(\theta|Y^{obs})d\theta.$$

Since the joint modeling algorithm converges to the specified multivariate distribution, once the joint imputation model is correctly specified, results will be valid and theoretical properties are satisfactory.

Fully conditional specification. Fully conditional specification attempts to define the joint distribution $p(Y^{obs}, Y^{mis}|\theta)$ by positing a univariate imputation model for each partially observed variable. The imputation model is typically a generalized linear model selected based on the nature of the missing variable (e.g. continuous, semi-continuous, categorical and count). Starting from some simple imputation methods, such as mean imputation or a random draw from the sampled values, FCS algorithms iteratively repeat imputations over all missing variables. Precisely, the tth iteration for the incomplete variable Y_j^{mis} consists of the following draws:

$$\theta_j^t \sim f(\theta_j)f(Y_j^{obs}|Y_{-j}^{t-1}, \theta_j)$$
$$Y_j^{mis(t)} \sim f(Y_j^{mis}|Y_{-j}^{obs}, Y_{-j}^t, \theta_j^t),$$

where $f(\theta_j)$ is generally specified with a non-informative prior. After a sufficient number of iterations, typically ranging from 5 to 10 iterations[3,9], the stationary distribution is achieved. The final iteration generates a single imputed dataset, and the multiple imputations are created by applying FCS in parallel m times with different seeds. If the underlying joint distribution defined by separate conditional models exists, the algorithm is equivalent to a Gibbs sampler.

The attractive feature of fully conditional specification is the flexibility of model specification, which allows models to preserve features in the data, such as skip patterns, incorporating constraints and logical, and consistent bounds[5]. Such restrictions would be difficult to formulate when applying joint modeling. One could conveniently construct a sequence of conditional models and avoid the specification of a parametric multivariate distribution, which may not be appropriate for the data in practice.

Compatibility. The definition of compatibility is given by Liu et al.[7]: let $Y = (Y_1, Y_2, \ldots, Y_p)$ be a vector of random variables and $Y_{-j} = (Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p)$. A set of conditional models $\{f_j(Y_j|Y_{-j}, \theta_j) : \theta_j \in \Theta_j, j = 1, 2, \ldots, p\}$ is said to be compatible if there exists a joint model $\{f(Y|\theta) : \theta \in \Theta\}$ and a collection of surjective maps $\{t_j : \Theta \to \Theta_j\}$ such that for each j, $\theta_j \in \Theta_j$ and $\theta \in t_j^{-1}(\theta_j) = \{\theta : t_j(\theta) = \theta_j\}$. In that case

$$f_j(Y_j|Y_{-j}, \theta_j) = f(Y_j|Y_{-j}, \theta).$$

Otherwise, $\{f_j, j = 1, 2, \ldots, p\}$ is said to be incompatible. A simple example of compatible models is a set of normal linear models for a vector of continuous data:

$$Y_j = N((1, Y_{-j})\beta_j, \sigma_j^2),$$

where β_j is the vector of coefficients and 1 is a vector of ones. In such a case, the joint model of $(Y_1, Y_2, \ldots, Y_p)$ would be a multivariate normal distribution and the map t_j is derived by conditional multivariate normal formula. On the other hand, the typical example of an incompatible model would be the linear model with squared terms[7,10].

Incompatibility is a theoretical weakness of fully conditional specification since, in some cases, it is unclear whether the algorithm indeed converges to the desired multivariate distribution[11–14]. Consideration of compatibility is significant when the multivariate density is of scientific interest. Both Hughes et al.[6] and Liu et al.[7] stated the necessity of model compatibility for the algorithm to converge to a joint distribution. Several papers introduced some cases in which FCS models are compatible with joint distributions[3,15]. Van Buuren[14] also performed some simulation studies of fully conditional specification with strongly incompatible models and concluded the effects of incompatibility are negligible. However, further work is necessary to investigate the adverse effects of incompatibility in more general scenarios.

Non-informative margins.   Hughes et al.[6] showed that the non-informative margins condition is sufficient for fully conditional specification to converge to a multivariate distribution. Suppose $\pi(\theta_j)$ is the prior distribution of the conditional model $p(Y_j|Y_{-j}, \theta_j)$ and $\pi(\theta_{-j})$ is the prior distribution of the marginal model $p(Y_{-j}|\theta_{-j})$, then the non-informative margins condition is satisfied if the joint prior could be factorized into independent priors $\pi(\theta_j, \theta_{-j}) = \pi(\theta_j)\pi(\theta_{-j})$. It is worthwhile to note that the non-informative margin condition does not hold if $p(Y_j|Y_{-j}, \theta_j)$ and $p(Y_{-j}|\theta_{-j})$ have the same parameter space. When the non-informative margins condition is violated, an order effect appears. In such a case, the inference of parameters would have systematic differences depending on the sequence of the variables in FCS algorithm. Simulations performed by Hughes et al.[6] demonstrated that such an order effect is subtle. However, more research is needed to verify such claims, and it is necessary to be aware of the existence of the order effect.

Theoretical results

This section proves the convergence of fully conditional specification under the normal linear model with normal inverse-gamma priors to a joint distribution. Since the compatibility of the normal linear model is well understood, we will check the satisfaction of the non-informative margins condition.

Starting with the problem of Bayesian inference for $\theta = (\mu, \Sigma)$ under a multivariate normal model, let us apply the following prior distribution. Suppose that, given Σ, the prior distribution of μ is assumed to be the conditionally multivariate normal,

$$\mu|\Sigma \sim N(\mu_0, \tau^{-1}\Sigma), \tag{1}$$

where the hyperparameters $\mu_0 \in \mathcal{R}^p$ and $\tau > 0$ are fixed and known and where p denotes the number of variables. Moreover, suppose that the prior distribution of Σ is an inverse-Wishart,

$$\Sigma \sim W^{-1}(m, \Lambda) \tag{2}$$

for fixed hyperparameters $m \geq p$ and Λ. The prior density for θ can then be written as

$$\pi(\theta) \propto |\Sigma|^{-(\frac{m+p+2}{2})} \exp\{-\tfrac{1}{2}tr(\Lambda^{-1}\Sigma^{-1})\} \\ \times \exp\{-\tfrac{\tau}{2}(\mu - \mu_0)^T\Sigma^{-1}(\mu - \mu_0)\}. \tag{3}$$

For each variable Y_j, we partition the mean vector μ as $(\mu_j, \mu_{-j})^T$ and the covariance matrix Σ as

$$\begin{pmatrix} \omega_j & \xi_j^T \\ \xi_j & \Sigma_{-j} \end{pmatrix},$$

such that $Y_j \sim \mathcal{N}(\mu_j, \omega_j)$ and $Y_{-j} \sim \mathcal{N}(\mu_{-j}, \Sigma_{-j})$. Similarly, we partition the scale parameter μ_0 as $(\mu_{0j}, \mu_{0-j})^T$ and Λ as

$$\begin{pmatrix} \Lambda_j & \psi_j^T \\ \psi_j & \Lambda_{-j} \end{pmatrix}.$$

The conditional model of Y_j given Y_{-j} is the normal linear regression $Y_j = \alpha_j + \beta_j^T Y_{-j} + \sigma_j$, where $\beta_j^T = \xi_j^T \Sigma_{-j}^{-1}$, $\alpha_j = \mu_j - \xi_j^T \Sigma_{-j}^{-1}\mu_{-j}$ and $\sigma_j = \omega_j - \xi_j^T \Sigma_{-j}^{-1}\xi_j$. The corresponding vectors of parameters θ_j and θ_{-j} would be

$$\begin{aligned} \theta_j &= (\alpha_j, \beta_j, \sigma_j) \\ \theta_{-j} &= (\mu_{-j}, \Sigma_{-j}). \end{aligned} \tag{4}$$

By applying the partition function[16] and by block diagonalization of a partitioned matrix, the joint prior for θ_j and θ_{-j} can be derived from $\pi(\theta)$ as :

$$\begin{aligned} \pi(\theta_j, \theta_{-j}) &= p(\sigma_j)p(\beta_j|\sigma_j)p(\Sigma_{-j}) \\ &\quad \times \exp\{-\frac{\tau}{2}(\alpha_j + \beta_j\mu_{0-j} - \mu_{0j})^T(\sigma_j)^{-1}(\alpha_j + \beta_j\mu_{0-j} - \mu_{0j})\} \\ &\quad \times \exp\{-\frac{\tau}{2}(\mu_{-j} - \mu_{0-j})^T\Sigma_{-j}^{-1}(\mu_{-j} - \mu_{0-j})\} \times |\Sigma_{-j}| \\ &= \pi(\theta_j)\pi(\theta_{-j}), \end{aligned} \tag{5}$$

where

$$\pi(\theta_j) = p(\sigma_j)p(\beta_j|\sigma_j)$$
$$\times \exp\{-\frac{\tau}{2}(\alpha_j + \beta_j\mu_{0-j} - \mu_{0j})^T(\sigma_j)^{-1}(\alpha_j + \beta_j\mu_{0-j} - \mu_{0j})\}, \tag{6}$$

$$\pi(\theta_{-j}) = p(\Sigma_{-j}) \times \exp\{-\frac{\tau}{2}(\mu_{-j} - \mu_{0-j})^T\Sigma_{-j}^{-1}(\mu_{-j} - \mu_{0-j})\} \times |\Sigma_{-j}| \tag{7}$$

and
$$p(\sigma_j) \sim W^{-1}(m, \lambda_j), \, p(\beta_j|\sigma_j) \sim \mathcal{N}(\psi_j^T\Lambda_{-j}^{-1}, \lambda_j\Lambda_{-j}^{-1}), \, p(\Sigma_{-j}) \sim W^{-1}(m-1, \Lambda_{-j}), \, \lambda_j = \Lambda_j - \psi_j^T\Lambda_{-j}^{-1}\psi_j \text{[16]}.$$

Since the joint prior distribution factorizes into independent priors, the "non-informative" margins condition is satisfied. Based on equations (6) and (7), we could derive the prior for the conditional linear model from the prior for the multivariate distribution:

$$p(\sigma_j) \sim W^{-1}(m, \lambda_j)$$
$$p(\beta_j|\sigma_j) \sim \mathcal{N}(\psi_j^T\Lambda_{-j}^{-1}, \lambda_j\Lambda_{-j}^{-1}) \tag{8}$$
$$p(\alpha_j|\sigma_j) \sim \mathcal{N}(\mu_{0j} - \psi_j^T\Lambda_{-j}^{-1}\mu_{0-j}, \tau^{-1}\sigma_j - (\mu_{0-j})^2\lambda_j\Lambda_{-j}^{-1}).$$

Since the conditional $\beta_j|\sigma_j$ follows a normal distribution, the marginal distribution β_j would be a student's t-distribution $\beta_j \sim t(\psi_j^T\Lambda_{-j}^{-1},$
$m\Lambda_{-j}^{-1}\lambda_j^{-1}, 2m - p + 1)$. When the sample size increases, β_j tends to the normal distribution $N(\psi_j^T\Lambda_{-j}^{-1}, \frac{\lambda_j\Lambda_{-j}}{m-1})$. Similarly, the marginal distribution α_j would be $t(\mu_{0j} - \psi_j^T\Lambda_{-j}^{-1}\mu_{0-j}, m(\tau^{-1} - (\mu_{0-j})^2\Lambda_{-j}^{-1})\Lambda_j^{-1}, 2m - p + 1)$. When the sample size increases, α_j tends to the normal distribution $N(\mu_{0j} - \psi_j^T\Lambda_{-j}^{-1}\mu_{0-j},$
$\frac{1}{(\tau^{-1} - (\mu_{0-j})^2\Lambda_{-j}^{-1})(m-1)}\Lambda_j)$. Usually, when the sample size is over 30, the difference between the student's t-distribution and the corresponding normally distributed approximation is negligible. With the prior transformation formula, one could apply Bayesian imputation under the normal linear model with normal inverse-gamma priors. This holds for both the prior information about the distribution of the data (e.g. location and scale of variables) and the scientific model (e.g. regression coefficients).

Simulation

We perform a simulation study to demonstrate the validity and the convergence of fully conditional specification when the conditional models are simple linear regressions with an inverse gamma prior for the error term and a multivariate normal prior for regression weights. In addition, we look for the disappearance of order effects, which is evident in the convergence of fully conditional specification to a multivariate distribution.

We repeat the simulation 500 times and generate a dataset with 200 cases for every simulation according to the following multivariate distribution :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 1 \\ 4 \\ 9 \end{pmatrix}, \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 9 \end{pmatrix}\right]$$

Fifty percent missingness is induced on either variable x, y or z. The proportion of the three missing patterns is equal. When evaluating whether it is appropriate to specify a normal inverse gamma prior, we consider both missing completely at random (MCAR) mechanisms and right-tailed missing at random (MARr) mechanisms where higher values have a larger probability of being unobserved. When investigating the existence of order effects, we only conduct the simulation under MCAR missingness mechanism to ensure that the missingness does not attribute to any order effects. We specify a weak informative prior for two reasons. First, with a weak informative prior, the frequentist inference is still plausible by applying Rubin's rules[1]. Second, Goodrich et al.[17] suggested that compared with flat non-informative priors, weak informative priors places warranted weight to extreme parameter values. In such a case, The prior under the joint model is specified as: $\mu_0 = (0, 0, 0)^T, \tau = 1, m = 3$ and

$$\Lambda = \begin{pmatrix} 60 & 0 & 0 \\ 0 & 60 & 0 \\ 0 & 0 & 60 \end{pmatrix}$$

and the corresponding prior for separated linear regression model would be the same, with $\pi(\sigma) \sim W^{-1}(3, 60)$ and

$$(\alpha, \beta)^T \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 60 & 0 & 0 \\ 0 & 3600 & 0 \\ 0 & 0 & 3600 \end{pmatrix}\right].$$

Scalar inference for the mean of variable Y. The aim is to assess whether Bayesian imputation under a normal linear model with normal inverse gamma priors would yield unbiased estimates and exact coverage of the nominal 95% confidence intervals. Table 1 shows that with weak informative prior, fully conditional specification also provides valid imputations. The estimates are unbiased, and the coverage of the nominal 95% confidence intervals is correct under both MCAR and MARr. Without the validity of a normal inverse gamma prior

4

		Bias	Cov	Ciw
FCS	MCAR	0	0.95	0.74
	MARr	−0.01	0.97	0.73
CCA	MCAR	0	0.94	0.68
	MARr	−0.18	0.81	0.68

Table 1. Bias of the estimates ($E(Y)$), coverage of nominal 95% confidence intervals (Cov) and the corresponding confidence intervals widths (Ciw) under MCAR and MARr.

specification, further investigations into the convergence would be redundant. Complete case analysis (CCA) gives a biased estimate and reduces the coverage of confidence intervals, demonstrating poor performance in analyzing the incomplete data set without addressing missingness.

Order effect evaluation. The visit sequence laid upon the simulation is z, x and y. To identify the presence of any systematic order effect, we estimate the regression coefficient directly after updating variable z and after updating variable x. Specifically, the ith iteration of fully conditional specification would be augmented as:

1. Impute z given x^{i-1} and y^{i-1}.
2. Build the linear regression $y = \alpha + \beta_1 x + \beta_2 z + \epsilon$ and collect the coefficient β_1, denoted as $\hat{\beta_1}^z$.
3. Impute x given z^i and y^{i-1}.
4. Build the linear regression $y = \alpha + \beta_1 x + \beta_2 z + \epsilon$ and collect the coefficient β_1, denoted as $\hat{\beta_1}^x$.
5. Impute y given z^i and x^i.

After a burn-in period with 10 iterations, the fully conditional specification algorithm was performed with an additional 1000 iterations, in which differences between the estimates $\hat{\beta_1}^z - \hat{\beta_1}^x$ are recorded. The estimates from the first 10 iterations are omitted since the FCS algorithms commonly reach convergence around 5 to 10 iterations. Estimates from the additional 1000 iterations would be partitioned into subsequences with equal size, which are used for variance calculation. We calculate the nominal 95% confidence interval of the difference. The standard error of the difference is estimated with batch-means methods[18]. The mean of $\hat{\beta_1}^z - \hat{\beta_1}^x$ is set to zero. Since only three 95% confidence intervals derived from 500 repetitions do not cross the zero, there is no indication of any order effects. We also monitor the posterior distribution of the coefficient under both joint modeling and fully conditional specification. Figure 1 shows a quantile-quantile plot demonstrating the closeness of the posterior distribution for β_1 derived from both joint modeling and fully conditional specification. Since the posterior distributions for β_1 under joint modeling and FCS are very similar, any differences may be considered negligible in practice.

All these results confirm that under the normal inverse gamma prior, Bayesian imputation under normal linear model converges to the corresponding multivariate normal distribution.

Conclusion

Based on the theory of the non-informative margins condition proposed by Hughes et al.[6], we prove the convergence of fully conditional specification under the normal linear model with normal-inverse-gamma prior distributions. Since it has been shown that a sequence of normal linear models is compatible with a multivariate normal density, we only focus on the non-informative margins condition for the prior. The transformation of
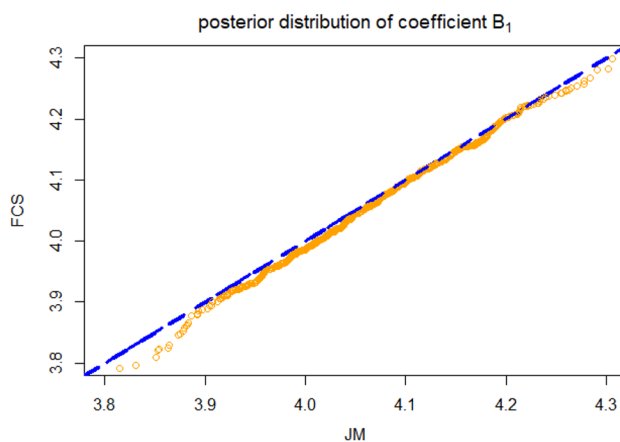
Figure 1. qqplot demonstrating the closeness of the posterior distribution of JM and FCS for β_1.

the prior between a normal inverse gamma for fully conditional specification and a normal inverse Wishart for joint modeling is useful. With transformation, one could apply fully conditional specification when having prior information about statistical moments (e.g., mean and variance of some variables) rather than prior information about parameters of fully conditional models.

The prior reflects the analyst's pre-data knowledge about the data or the model. The analyst specifies the prior when only a small sample size is available, for instance, patients in clinical research. Generally, prior distributions are determined by location and variance parameters. The location parameters [(for example, μ_0 in (1) and m in (2)] are commonly based on the results of previous studies. The variance parameters [(for example, $\tau^{-1}\Sigma$ in (1) and Λ in (2)] are specified based on the exchangeability of the prior and current study[19]. Exchangeability indicates the same population for the prior and current studies. Hence, lower variance parameters can be applied. Otherwise, higher variance parameters can be used to include large support of parameters.

We perform simulations under the case when the number of variables is larger than the sample size. However, based on Bayesian theories, the result is valid when the number of variables is smaller than the sample size. For example, Huang et al.[20,21] proposed to generate "synthetic data" under a simpler prior distribution to augment the sample size. In this case, the statistical inference heavily depends on the prior specification.

Fully conditional specification is an appealing imputation method because it allows one to specify a sequence of flexible and simple conditional models and bypass the difficulty of multivariate modeling in practice. The default prior for normal linear regression is Jeffreys prior, which satisfies the non-informative margin condition. However, it is worth developing other types of priors for fully conditional specification such that one could select the prior that suits the description of prior knowledge best. Many researchers have discussed the convergence condition of FCS. However, there is no conclusion for the family of posterior distributions that satisfies the condition of convergence. In such a case, when including new kinds of priors in fully conditional specification algorithms, it is necessary to investigate the convergence of the algorithm with new posterior distributions. Specifically, one should study the non-informative margin conditions for new priors. Compatibility should also be considered if the imputation model is novel. Our work takes steps in this direction.

Although a series of investigations have shown that the adverse effects of violating compatibility and the non-informative margin conditions may be small, all of these investigations rely on pre-defined simulation settings. More research is needed to verify conditions under which the fully conditional specification algorithm converges to a multivariate distribution and cases in which the violation of compatibility and non-informative margin has negligible adverse impacts on the result.

There are several directions for future research. From one direction, it is possible to develop a prior setting to eliminate order effects of the fully conditional specification algorithm under the general location model since the compatibility and non-informative margins conditions are satisfied under the saturated multinomial distribution. Moreover, various types of priors of the generalized linear model (e.g., non-linear normal regression) for the fully conditional specification and corresponding joint modeling rationales could be developed. Another open problem is the convergence condition and properties of block imputation, which partitions missing variables into several blocks and iteratively imputes blocks[3]. Block imputation is a more flexible and user-friendly method. However, its properties have yet to be studied. Finally, it is necessary to investigate the implementation of prior specifications in software.

Data availibility
The data used in the article is simulation data. The details are available from the GitHub repository: https://github.com/Mingyang-Cai/informative_prior.

References
1. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*, vol. 81 (John Wiley & Sons, 2004).
2. Schafer, J. L. *Analysis of Incomplete Multivariate Data* (Chapman and Hall/CRC, 1997).
3. Van Buuren, S. *Flexible Imputation of Missing Data* (CRC press, 2018).
4. Van Buuren, S. & Groothuis-Oudshoorn, K. Mice: Multivariate imputation by chained equations in r. *J. Stat. Softw.* **45**, 1–68 (2010).
5. Van Buuren, S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **16**, 219–242 (2007).
6. Hughes, R. A. *et al.* Joint modelling rationale for chained equations. *BMC Med. Res. Methodol.* **14**, 28 (2014).
7. Liu, J., Gelman, A., Hill, J., Su, Y.-S. & Kropko, J. On the stationary distribution of iterative imputations. *Biometrika* **101**, 155–173 (2014).
8. Zhu, J. & Raghunathan, T. E. Convergence properties of a sequential regression multiple imputation algorithm. *J. Am. Stat. Assoc.* **110**, 1112–1124 (2015).
9. Oberman, H. I., van Buuren, S. & Vink, G. Missing the point: Non-convergence in iterative imputation algorithms. arXiv preprint arXiv:2110.11951 (2021).
10. Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R. & Initiative, A. D. N. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat. Methods Med. Res.* **24**, 462–487 (2015).
11. Arnold, B. C. & Press, S. J. Compatible conditional distributions. *J. Am. Stat. Assoc.* **84**, 152–156 (1989).
12. Arnold, B. C., Castillo, E. & Sarabia, J. M. Compatibility of partial or complete conditional probability specifications. *J. Stat. Plan. Inference* **123**, 133–159 (2004).
13. Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R. & Kadie, C. Dependency networks for inference, collaborative filtering, and data visualization. *J. Mach. Learn. Res.* **1**, 49–75 (2000).
14. Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G. & Rubin, D. B. Fully conditional specification in multivariate imputation. *J. Stat. Comput. Simul.* **76**, 1049–1064 (2006).
15. Raghunathan, T. E. *et al.* A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **27**, 85–96 (2001).

16. Eaton, M. L. Multivariate statistics: A vector space approach. *Lecture Notes-Monograph Series* **53**, i–512 (2007).
17. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. rstanarm: Bayesian applied regression modeling via Stan. (2019). R package version 2.19.2.
18. Albert, J. *Bayesian Computation with R* (Springer, 2009).
19. Van de Schoot, R. & Miocević, M. *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners* (Taylor & Francis, 2020).
20. Huang, D., Stein, N., Rubin, D. B. & Kou, S. Catalytic prior distributions with application to generalized linear models. *Proc. Natl. Acad. Sci.* **117**, 12004–12010 (2020).
21. Huang, D., Wang, F., Rubin, D. B. & Kou, S. Catalytic priors: Using synthetic data to specify prior distributions in bayesian analysis. arXiv preprint arXiv:2208.14123 (2022).

Author contributions

M.C. wrote the main manuscript. G.V. and S.vB. reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.