Bayesian Logistic and Cauchit Regression with Markov Chain Monte Carlo to Address Multicollinearity

1 Objective of Report

This report investigates fitting logitistic and cauchit regression models on a 9-dimensional and 150-observation dataset with very high multicollinearity. Classically, parameter inference of these model using maximum likelihood and Newton-Raphson optimisation gives unstable and unreliable solutions because high multicollinearity causes near-singularity of the Hessian matrix. Bayesian inference is therefore an alternative that imposes regularisation effects by incorporating prior distributions on the parameters, which helps stabilise estimates and reduce the impact of multicollinearity. One challenge is that in this case, the posterior over parameters is intractable because of high dimensionality and non-conjugate likelihood functions. We therefore approximate it with Markov chain Monte Carlo (MCMC) methods, specifically the Random Walk Metropolis (RWM) algorithm.

The report is structured as follows: We firstly experiment with rejection sampling to see the benefit of using heavier-tail distribution as a proposed distribution. We then use this principle to conduct inference on the logistic regression problem. We present 2 different priors and clearly derive the corresponding posterior. We then run MCMC sampling on each posterior, choosing the candidate distribution with best Brier score, and assess estimates and convergence with trace plots and effective sample size. The same procedure applies to the cauchit regression problem.

2 Rejection Sampling

This section experiments with rejection sampling to draw samples from a logistic distribution, using the Cauchy distribution as the proposal.

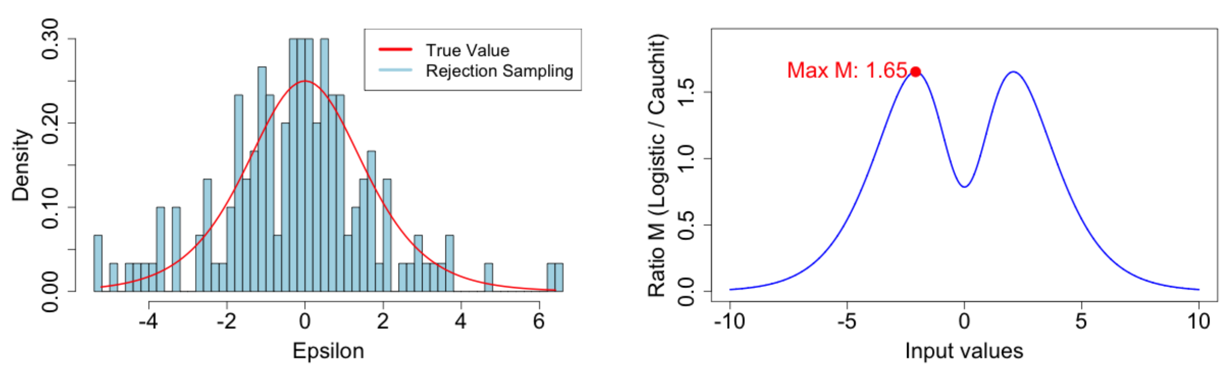

Figure 1: Left: The plot compares the approximated logistic distribution using rejection sampling with the true density. Right: The plot shows the values of M against different values of inputs, the supremum of M is at value 1.65.

Rejection sampling is a technique for generating samples from a target distribution using a proposal distribution and a scaling factor. In this rejection sampling algorithm, the scaling factor M is calculated as the supremum of the ratio of the logistic to cauchy densities. This ensures the proposed samples from the Cauchy distribution can best cover the logistic distribution. Also, cauchy is heavier tail than logistic so M does not grow too large as $|x| \to \infty$. From the right of Figure 1, we see that the largest M is 1.65, and this gives an efficiency of 65.50% when trying to obtain 150 samples, which is reasonable. The approximated density, from the left of Figure 1, shows a good fit to the true logistic distribution, indicating that the sampling process was implemented correctly.

3 Logistic Regression

In this section, we aim to (Goal 1) find the most effective MCMC approximation for two different posterior distributions (Goal 2) compare the two approximated posteriors to determine the better prior choice. The two posteriors are defined in Equation 1 and 2, where the independent and identically distributed normal (IID-N) prior and the Unit Information Prior (UIP) prior are used.

$$\text{Posterior 1} = \prod_{i=1}^{n} \text{Bernoulli}\left(\frac{1}{1+\exp(-\mathbf{x}_i^T \boldsymbol{\beta})}\right) \times \prod_{j=1}^{d} \mathcal{N}(\beta_j; 0, 1^2) \tag{1}$$

$$\text{Posterior 2} = \prod_{i=1}^{n} \text{Bernoulli}\left(\frac{1}{1+\exp(-\mathbf{x}_i^T \boldsymbol{\beta})}\right) \times \mathcal{N}_d\left(\boldsymbol{\beta}; 0, n\left(X^T X\right)^{-1}\right) \tag{2}$$

Candidate Category	Candidate	Posterior 1 (IID-N)	Posterior 2 (UIP)
IID Univariates	Normal	0.01463	0.02484
	Student-t (df=5)	0.01461	0.02478
	Student-t (df=3)	0.01470	0.02589
	Student-t (df=1)	0.01464	0.02457
Multivariate	Normal	0.01513	0.02938
	Student-t (df=5)	0.01526	0.02853
	Student-t (df=3)	0.01579	0.02551
	Student-t (df=1)	0.01469	0.02691
Multivariate + PC	Normal (pc)	0.01481	0.02402
	Student-t (df=5, pc)	**0.01450**	0.02431
	Student-t (df=3, pc)	0.01465	0.02434
	Student-t (df=1, pc)	0.01479	**0.02399**

Table 1: Comparison of **Brier Scores** for Bayesian Logistic regression using parameters' posterior means from the last 10,000 samples (Lower is better - Bold is best of each column)

We begin with the first Goal 1 using the RWM algorithm. The RWM algorithm is a MCMC method that proposes new sample points based on a random walk from the current point using a candidate distribution, and make moves towards regions of higher probability density of the target. Ideally, we want the sampling algorithm to converge to the target distribution with low autocorrelation among samples, yielding a high effective sample size (ESS). To achieve this, we need to propose a candidate that can capture two special characteristics of these two posteriors: they have a heavier tail than the normal distribution and the inputs covariates are highly multicolinear. We therefore experiment with extensive candidate distribution options (1) IID Candidate distributions with increasing heavy tail behaviour: normal and Student-t with degree of freedom (df) 5,3,1 and (2) Multivariate versions of these with and without preconditioning (PC) of the covariance matrix. Regarding the IID candiates, we will conduct component-wise update with separately-tuned step size so that the acceptance rate in each dimension is between 36 and 40%. Regarding PC for multivariate candidates, we repeat the PC process until the chain mixes well by checking the trace plots, leading to 5 times of PC with 10,000 iterations each across both posteriors. The step-size is tuned the same across dimensions and maintained at approximately 23%.

There are three main results for Goal 1. We (1) select the best approximated Poterior 1 and 2 using Brier score, presented in Table 1, and focus on assessing (2) the convergence and estimate quality of them using trace plots presented in Figure 2 and 3 in the Appendix, and (3) checking posterior means, 95% credible intervals (CI) and ESS results calculated using the last 10,000 samples, presented in Table 2 and 3.

The first result is presented in Table 1: we compare Brier scores on each columns of RMW approximation for Posterior 1 and Posterior 2. For Posterior 1, we see that the Student-t with df = 5 and PC achieve the lowest score at 0.01450. For Posterior 2, the Student-t with df = 1 and PC achieve the lowest score at 0.02399.

We then assess convergence and properties of parameter estimates with Posterior 1. From Figure 2, we can see that the all the chain mixes poorly in the first 10,000 iterations, but thanks to PC, they begin to mix well between iteration 10,000 and 50,000. The posterior means also converge well across all parameters, yet there are biases, the level of which is different for each parameter. Visually, biases are relatively reasonable for most parameters except for β_1, β_2 and β_{10}, whose bias is quite high. This comment aligns with the results in Table 2, where the result of β_1 and β_2 are more biased compared to the rest, particularly the true β_1 that lies outside of its estimated 95% CI. The result of β_{10} however appears reasonable and lies inside the CI, and so do the rest of the variables. The ESS stays reasonable between 181 and 293 (the higher the better).

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True Val	-2.00	-1.56	-1.11	-0.67	-0.22	0.22	0.67	1.11	1.56	2.00
Mean	-0.30	-0.36	-0.91	-0.58	-0.13	0.23	-0.16	0.79	1.39	1.44
2.50%	-1.37	-1.88	-2.44	-2.04	-2.13	0.08	-2.00	-0.42	0.84	0.40
97.50%	0.77	1.19	0.67	0.89	1.82	0.39	1.71	2.07	2.10	2.70
ESS	292.73	284.25	228.10	250.57	264.70	180.88	238.77	271.47	263.15	276.63

Table 2: Estimated Bayesian logistic regression parameters (Posterior 1) using RWM, showing true values, posterior means, credible intervals, and effective sample sizes for β_1 to β_{10}.

Similarly, for Posterior 2, as depicted in Figure 3, all chains mixed well, yet displaying significant biases in β_1, β_2, β_5, β_7, and β_{10}. These biases were reflected in the results from Table 3, where only β_4, β_6, and β_9 showed reasonable estimates. The ESS values for this model were lower compared to Posterior 1, ranging from 123 to 189, suggesting higher correlation between sample, meaning poorer convergence and less reliability in the inferences drawn from these samples.

For Goal 2, we can conclude as follows: based Brier score, best approximated Posterior 1 gives better fit than best approximated Posterior 2 given the same logistic dataset. For parameter estimation, Posterior 1 also gives less biased and more reliable results than Posterior 2.

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True Val	-2.00	-1.56	-1.11	-0.67	-0.22	0.22	0.67	1.11	1.56	2.00
Mean	-0.29	1.11	-1.17	-0.60	-5.54	0.26	-3.57	-0.34	1.01	0.07
2.50%	-0.90	-2.08	-4.13	-1.95	-13.38	-0.45	-15.51	-1.72	0.24	-1.02
97.50%	0.33	4.13	1.83	0.68	1.67	0.86	10.95	0.95	1.89	1.35
ESS	167.63	167.99	160.23	197.89	123.60	161.00	161.95	137.21	117.05	189.07

Table 3: Estimated Bayesian logistic regression parameters (Posterior 2) using RWM, showing true values, posterior means, credible intervals, and effective sample sizes for β_1 to β_{10}.

4 Cauchit Regression

We repeat the same procedure with Goal 1 and Goal 2 for the Bayesian cauchit regression model. As the cauchit distribution is as heavy tail as Student-t with df=1, we can explore even heavier tail options using df=0.5 and df=0.1. Regarding the cauchit regression model set-up, the given priors are the same - IID-N and UIP, but now the likelihood are different. The resulting Posteriors are given in 3 and 4.

$$\text{Posterior } 3 = \prod_{i=1}^{n} \text{Bernoulli}\left(\frac{1}{\pi}\arctan(\mathbf{x}_i^T\boldsymbol{\beta}) + \frac{1}{2}\right) \times \prod_{j=1}^{d} \mathcal{N}(\beta_j; 0, 1^2) \tag{3}$$

$$\text{Posterior } 4 = \prod_{i=1}^{n} \text{Bernoulli}\left(\frac{1}{\pi}\arctan(\mathbf{x}_i^T\boldsymbol{\beta}) + \frac{1}{2}\right) \times \mathcal{N}_d\left(\boldsymbol{\beta}; 0, n\left(X^TX\right)^{-1}\right) \tag{4}$$

Candidate Category	Candidate	Posterior 3 (IID-N)	Posterior 4 (UIP)
IID Univariates	Normal	0.01170	0.02653
	Student-t (df=1)	0.01199	0.02667
	Student-t (df=0.5)	0.01202	0.02649
	Student-t (df=0.1)	0.01185	0.02909
Multivariate	Normal	0.01271	0.03401
	Student-t (df=1)	0.01373	0.02877
	Student-t (df=0.5)	0.01169	0.03086
	Student-t (df=0.1)	0.01228	0.02887
Multivariate + PC	Normal (pc)	0.01235	**0.02611**
	Student-t (df=1, pc)	0.01272	0.02659
	Student-t (df=0.5, pc)	0.01107	0.02650
	Student-t (df=0.1, pc)	**0.01054**	0.02613

Table 4: Comparison of **Brier Scores** for Bayesian cauchit regression using parameters' posterior means from the last 10,000 samples (Lower is better - Bold is best of each column)

For Goal 1, we firstly compare Brier scores on each columns of RMW sampling for Posterior 3 and Posterior 4 in the Table 4. For Posterior 3, we see that the multivariate Student-t with df = 0.1 and PC achieve the lowest score at 0.01054. For Posterior 4, the multivariate normal and PC achieve the lowest score at 0.02611.

Next, we assess convergence and properties of estimates for the best approximated Posterior 3. From Figure 4, we can see that the all the chain mixes well and all posterior means shows good convergence. Bias is visually low across all parameters, apart from β_1, β_2 and β_8. This aligns with the results in Table 5, where the result of β_1, β_7 and β_8 are more biased compared to the rest, particularly the true β_1 that also lies outside of its estimated 95% CI interval. The ESS stays relatively low between 120 and 180, indicating the sample did not converge very well.

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True Val	-2.00	-1.56	-1.11	-0.67	-0.22	0.22	0.67	1.11	1.56	2.00
Mean	0.46	-0.72	-1.09	-0.66	0.19	0.33	-0.10	-0.67	1.65	1.80
2.50%	-1.06	-2.17	-2.42	-2.10	-2.20	0.15	-2.12	-1.95	0.97	0.90
97.50%	1.62	0.61	0.21	0.63	2.01	0.55	1.61	0.52	2.46	2.74
ESS	119.92	120.45	127.84	179.82	118.18	137.01	136.54	145.27	121.90	120.34

Table 5: Estimated Bayesian cauchit regression parameters (Posterior 3) using RWM, showing true values, posterior means, credible intervals, and effective sample sizes for β_1 to β_{10}

Similarly, we assess convergence and properties of estimates for the best approximated Posterior 4. In Figure 5, all chains and their posteriors means appear to mix well. Visually, we see that only β_3, β_4 and β_6 converge close to the true value while the rest exhibits significant bias. Cross-checking that with Table 6, the result is consistent with the plots and we see that overall, the posterior means are more poorly estimated compared to that in Posterior 3. However, the ESS values are higher here, ranging between 276 and 309, meaning the samples are less correlated and the chain converged better than Posterior 3.

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True Val	-2.00	-1.56	-1.11	-0.67	-0.22	0.22	0.67	1.11	1.56	2.00
Mean	-0.09	1.52	-1.60	-0.56	-3.32	0.24	-3.19	-0.83	0.79	0.23
2.50%	-0.69	-1.23	-4.58	-1.91	-11.25	-0.43	-17.07	-2.32	0.03	-0.94
97.50%	0.54	4.85	1.04	0.71	4.10	0.95	10.38	0.39	1.66	1.44
ESS	307.29	306.61	308.57	293.48	305.31	279.76	280.47	275.66	292.47	296.24

Table 6: Estimated Bayesian cauchit regression parameters (Posterior 4) using RWM, showing true values, posterior means, credible intervals, and effective sample sizes for β_1 to β_{10}.

For Goal 2, we can conclude as follows: based on Brier score, the best approximated Posterior 3 gives better fit than best approximated Posterior 4 given the same cauchit dataset. For parameter estimation, Posterior 3 also gives less biased result than Posterior 4, although the sample in the latter converged better.

5 Conclusion

In conclusion, this work experimented with and selected the best MCMC approximation for 4 posterior distributions of Bayesian logit and cauchit regression models. Our findings support the claim that using IID Normal priors on the parameters results in better model fits for both logistic and cauchit regressions compared to the Unit Information prior. More exploration on prior choice, e.g with non-zero mean, is also needed as the estimated values of β_1 is poor across all models.

A Appendix

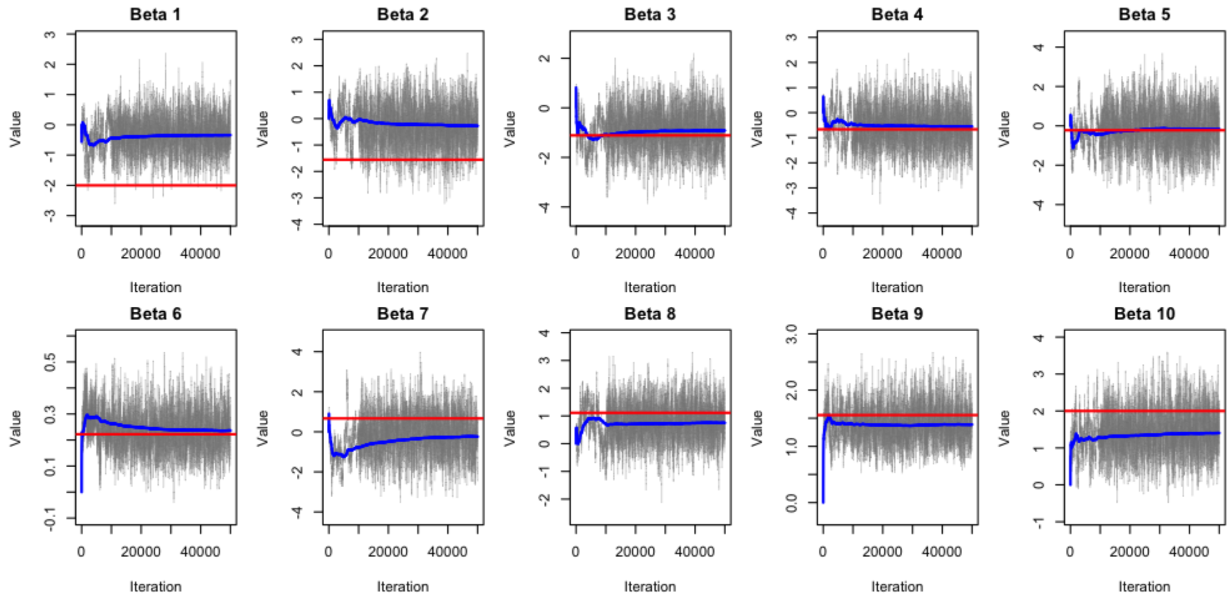

Figure 2: Trace plots for the parameters β_1 to β_{10} in Bayesian logistic regression model with IID-N prior. **Blue line** is **cumulative posterior mean**. **Red line** is **true parameter value**. The selected candidate for RWM sampling is Student-t with df = 5 and preconditioning applied.

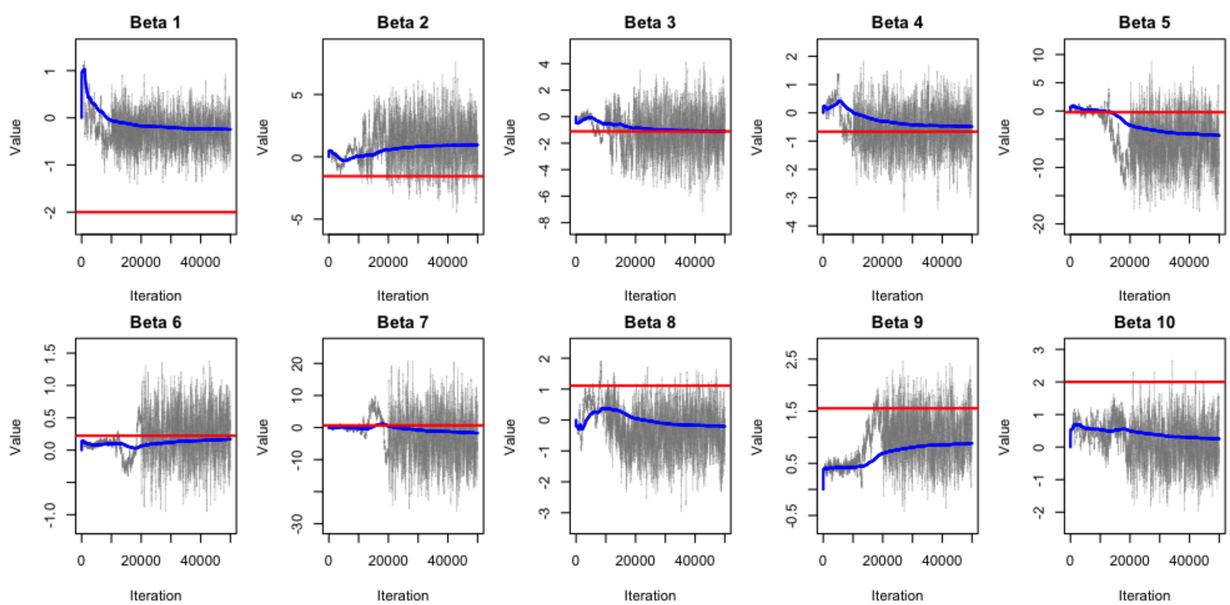

Figure 3: Trace plots for the parameters β_1 to β_{10} in Bayesian logistic regression model with UIP prior. **Blue line** is **cumulative posterior mean**. **Red line** is **true parameter value**. The selected candidate for RWM sampling is Student-t with df = 1 and preconditioning applied.

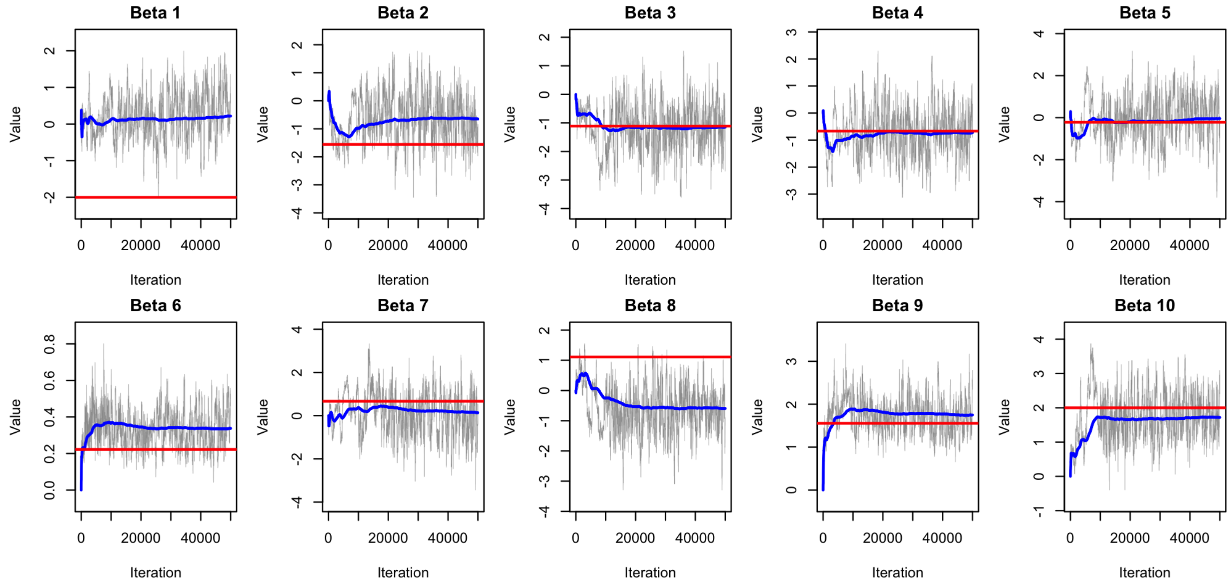

Figure 4: Trace plots for the parameters β_1 to β_{10} in Bayesian cauchit regression model with IID prior. **Blue line** is **cumulative posterior mean**. **Red line** is **true parameter value**. The selected distribution for RWM sampling is Student-t with df $= 0.1$ and preconditioning applied.

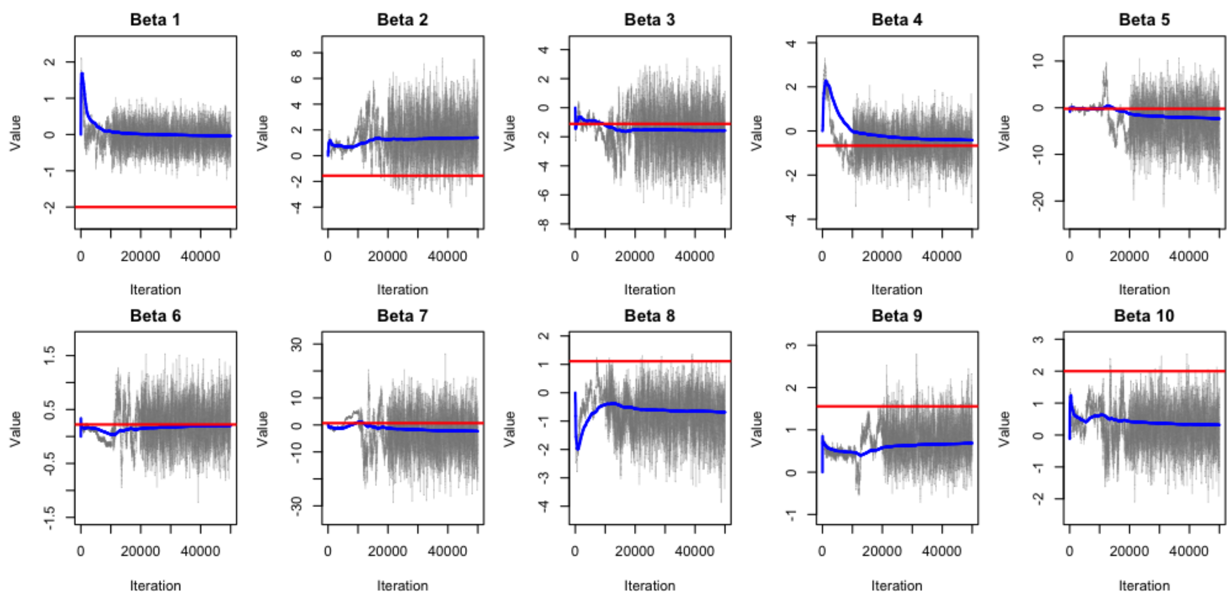

Figure 5: Trace plots for the parameters β_1 to β_{10} in Bayesian cauchit regression model with UIP prior. **Blue line** is **cumulative posterior mean**. **Red line** is **true parameter value**. The selected candidate for RWM sampling is normal and preconditioning applied.