ENSAE: Bayesian Inference Report

Ali HAIDAR, Yunhao CHEN

January 15, 2023

Abstract

In this project, several methods proposed in [1] have been reproduced.[1]

1 Introduction

In many areas of economics and other sciences, the seemingly unrelated regression (SUR) model is used as a tool to study a wide range of phenomena. Many studies have contributed to the development of estimation, testing, prediction and other inference techniques for the analysis of SUR models. One of the most popular approaches for estimating the SUR model in a Bayesian framework involves the use of a Markov Chain Monte Carlo (MCMC) simulation approach. However, MCMC methods involve many decisions to be made by users: for example, the number of burn-in steps, choice of an appropriate proposal density so that the MCMC algorithm will have a high acceptance rate, determination of the number of MCMC samples.

In view of these problems involved in the use of MCMC techniques, a direct Monte Carlo (DMC) approach has been proposed in [2]. It is shown theoretically and in applications that the DMC approach is easily applicable and provides computational results directly without many of the concerns associated with MCMC approaches mentioned above. Unlike MCMC which is directly applied on standard SUR model, DMC needs to be applied on transformed SUR model. In [1], the authors developed new efficient Bayesian approaches which combines a DMC and an importance sampling procedure. Moreover, a hierarchical prior on the coefficients has been employed.

In this project, we have implemented three methods mentioned above to estimate the parameters in SUR model: MCMC (Markov Chain Monte Carlo) applied on original SUR model; DMC (direct Monte Carlo) on transformed SUR model; DMC-IS (direct Monte Carlo-importance sampling) on transformed SUR model. During the implementation of the last method, we encountered with one difficulty. We decided to switch to DMC-IS in another paper [3], which makes different hypothesis.

2 Problem description

The linear SUR model involves a set of regression equations with cross-equation parameter restrictions and correlated error terms having differing variances. Algebraically, the SUR model is given by:

$$y_j = X_j\beta_j + u_j, j = 1, ...m, \quad \text{with} \quad E[u_i u_j^T] = \begin{cases} \omega_{ij} I, & (i \neq j) \\ \omega_i^2 I, & (i = j) \end{cases}$$

we can model the problem with:

$$y = X\beta + u$$

where,

- $u \sim \mathcal{N}(0, \Omega \otimes I)$

[1]Codes is provided here: https://github.com/yunhao-tech/Bayesian-Inference-project

- Ω is m $\times$ m symmetric matrix with diagonal elements $\{\omega_1^2, ...\omega_m^2\}$
- $X = \text{diag}\{X_1, ...X_m\}$

we can write the likehood function as:

$$L(y|\beta,\Omega) = \frac{1}{(2\pi)^{\frac{nm}{2}}|\Omega|^{\frac{n}{2}}}\exp[-\frac{1}{2}tr\{R\Omega^{-1}\}]$$

where $R = (r_{ij})$, $r_{ij} = (y_i - X_i\beta_i)^T(y_j - X_j\beta_j)$.

2.1 MCMC Approach

In the absence of prior knowledge, a widely used noninformative priors (Jeffrey's invariant prior) has been employed here:

$$\pi(\beta,\Omega) = \pi(\beta)\pi(\Omega) \propto |\Omega|^{-\frac{m+1}{2}}$$

where m is the number of equations in SUR model.

Using this prior, we can obtain the joint posterior density function :

$$\pi(\beta,\Omega|D) \propto |\Omega|^{-\frac{n+m+1}{2}}\exp[(-\frac{1}{2}\{\text{tr}(R\Omega^{-1})\}]$$

We can deduce the conditional posteriors $\pi(\beta|\Omega, D)$ and $\pi(\Omega|\beta, D)$:

$$\beta|\Omega, D \sim \mathcal{N}(\hat{\beta}, \hat{\Omega}_\beta) \text{ and } \Omega|\beta, D \sim IW(R, n) \tag{1}$$

where,

$\hat{\beta} = \{X^T(\Omega^{-1}\otimes I)X\}^{-1}X^T(\Omega^{-1}\otimes I)y$ and $\hat{\Omega}_\beta = (X^T(\Omega^{-1}\otimes I)X)^{-1}$

The conditional posteriors of β and Σ depend upon each other and the densities $\beta|\Omega, D$ and $\Omega|\beta, D$ are available in formula (1). Therefore, the parameters in standard SUR model can be simulated by a 2-block Gibbs sampling.

2.2 DMC Approach

First of all, the standard SUR model can be reformulated as follows:

$$\begin{cases} y_1 = X_1\beta_1 + e_1 \equiv Z_1 b_1 + e_1 \\ y_j = X_j\beta_j + \sum_{l=1}^{j-1}\rho_{jl}(y_l - X_l\beta_l)e_j \equiv Z_j b_j + e_j \end{cases}$$

where the $n \times (p_j + j - 1)$ matrices Z_j are functions of $\beta_{j-1}, ...\beta_1$ and

$$E[e_i e_j^T] = \begin{cases} O, (i \neq j) \\ \sigma_i^2 I, (i = j) \end{cases}$$

where O is a zero matrix and $\Sigma = \text{diag}\{\sigma_1^2, ...\sigma_m^2\}$

The joint posterior density can be deduced as:

$$\pi(b_j|b_{j-1}, ...b_1, \sigma_j^2, D) = \mathcal{N}(\hat{b}_j, \sigma_j^2(Z_j^T Z_j)^{-1}) \tag{2}$$

$$\pi(\sigma_j^2|b_{j-1}, ...b_1, D) = IG(\frac{\hat{\gamma}_j}{2}, \frac{\hat{\nu}_j}{2}) \tag{3}$$

where,

- $\hat{b}_j = (Z_j^T Z_j)^{-1} Z_j^T y_j$
- $\hat{\gamma}_j = (y_j - Z_j \hat{b}_j)^T (y_j - Z_j \hat{b}_j)$
- $\hat{\nu}_j = n - m - p_j + j + 1$

We have to notice that, there is a one to one relation between the parameters of the standard SUR and the transformed SUR, mentioned in [2]. To simplify the formula, here we point out this one to one relation in case of $m = 2$:

- $b_1 = \beta_1$, $b_2 = (\beta_2, \rho_{21})$
- $Z_1 = X_1$, $Z_2 = (X_2, y_1 - X_1 b_1)$
- $\omega_1^2 = \sigma_1^2$, $\omega_2^2 = \rho_{21}\sigma_1^2 + \sigma_2^2$, $\omega_{12} = \rho_{21}\sigma_1^2$
- $\rho_{21} = \frac{\omega_{12}}{\omega_1^2}$

According to formulas (2) and (3), we can simulate the parameters σ and b in an iterative way. This is different from Gibbs sampling. Here, the conditional posteriors of b and σ do not depend upon each other. More precisely, in case of $m = 2$, the DMC algorithm is as follows:

First $j = 1$, sample σ_1^2 from (3); then, inject these simulations into (2) and sample b_1; After that, we arrive at $j = 2$: inject simulations of b_1 into (3) and sample σ_2^2; inject simulations of σ_2^2 into (2) and sample b_2. Finally, we make use of the one to one relation mentioned above, in order to obtain the simulations of parameters $\beta_1, \beta_2, \omega_1^2, \omega_2^2, \omega_{12}$ in standard SUR model.

2.3 DMC-IS Approach

We use again the transformed SUR model but with hierarchical prior structure.

$$\pi(\beta, \Omega, \lambda) = \pi(\beta|\lambda)\pi(\Omega)\pi(\lambda)$$

- $\pi(\beta|\lambda) \propto \exp\{-\frac{1}{2}(\beta - \beta_0)^T \Gamma(\lambda, A)(\beta - \beta_0)\}$
- $\pi(\Omega) \propto |\Omega|^{-\frac{m+1}{2}}$
- $\pi(\lambda) = \prod_{j=1}^m \pi(\lambda_j) \propto \prod_{j=1}^m (\lambda_j)^{a_0 - 1} \exp\{-\lambda_j / b_0\}$

The conditional posterior probabilities:

$$\pi(b_j | b_{j-1}..., b_1, \sigma_j^2, \lambda, D) = \mathcal{N}(\hat{b}_j, \sigma_j^2 (Z_j^T Z_j + \gamma_j D_j)^{-1}) \tag{4}$$

$$\pi(\sigma_j^2 | b_{j-1}, ...b_1, \lambda, D) = IG(\hat{a}_j/2, \hat{h}_j/2) \tag{5}$$

With,

- $\hat{b}_j = (Z_j^T Z_j + \gamma_j D_j)^{-1} [(Z_j^T Z_j)^{-1} \tilde{b} + \lambda_j D_j b_0]$
- $\tilde{b}_j = (Z_j^T Z_j)^{-1} Z_j^T y_j$
- $\hat{a}_j = n - m - 2a_0 + 2j + 1$
- $\hat{h}_j = \gamma_j / b_0 + (n - p + j - 1)\hat{s}_j^2 + (\hat{b}_j - b_{0j})^T (Z_j^T Z_j + \gamma_j D_j)^{-1} (\hat{b}_j - b_{0j})$
- $\hat{s}_j^2 = (y_j - Z_j \tilde{b}_j)^T (y_j - Z_j \tilde{b}_j)/(n - p_j + j - 1)$

Here, $\gamma_j = \lambda_j \sigma_j^2$. We'd like to follow the same procedures as DMC approach. But **we encountered with one difficulty**: in formula (5), the parameters in the law of σ_j^2 depend on γ_j, thus depend on σ_j^2 itself. That makes simulations impossible.

After consulting the paper [3] which discuss DMC-IS on SUR model, with the error terms being assumed to have a heavy-tailed distribution (Student-t distribution). We decided to implement with this hypothesis.

By assuming a multivariate Student-t density with degrees of freedom ν for $u_j, j = 1...m$, the conditional posterior density of the $\{b, \Sigma\}$ is equivalent to the conditional multivariate Student-t and beta prime (BP):

$$\pi(b_j|b_{j-1}..., b_1, \sigma_j^2, \nu, D) = St(\hat{b}_j, \tilde{\sigma_j}^2(Z_j^T Z_j)^{-1}, \hat{\nu_j}) \tag{6}$$

$$\pi((\nu_j\sigma_j^2)/(\tilde{\nu_j}s^2)|b_{j-1}, ...b_1, \nu, D) = BP((\nu_j - q_j + 2)/2, (n + q_j - \tilde{p_j} - 2)/2) \tag{7}$$

With,

- $\tilde{\sigma_j}^2 = (\nu\sigma_j^2 + (\hat{\nu}_j - \nu)s_j^2)/\hat{\nu}_j$
- $\hat{\nu}_j = n - \tilde{p_j} + \nu$
- $\hat{b}_j = (Z_j^T Z_j)^{-1} Z_j^T y_j$
- $s_j^2 = (y_j - Z_j\hat{b}_j)^T(y_j - Z_j\hat{b}_j)/(n - \tilde{p_j})$
- $\tilde{p_j} = p_j + j - 1$

3 Implementation

We tested the methods in the previous section with the simulated data. Without loss of generality in the model structure, we set $m = 2$ and $p_j = 2, j = 1, 2$ in the standard SUR model. The model can be written as follows:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

For $j = 1, 2$, y_j and u_j are $n \times 1$ vectors, X_j is an $n \times 2$ matrix, generated from a uniform density over the interval $(-5, 5)$. The coefficient vector was set to be $\beta_1 = (3, -2)^T$ and $\beta_2 = (2, 1)^T$. Note: n is the number of observations, which is set to 100.

Each element of Ω is set to be:

$$\Omega = \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_1^2 \end{pmatrix} = \begin{pmatrix} 0.1 & -0.05 \\ -0.05 & 0.2 \end{pmatrix}$$

In our implementation, we generated 10000 samples for each approach. The total number of MCMC iterations is chosen to be 11000, of which the first 1000 iterations are discarded (burn-in steps equal to 1000).

3.1 Results

For estimations of *beta*, we obtain the satisfactory results, shown in Fig.1 and Fig.2. More precisely, for all three approaches, the average value of four *beta* parameters is close to the true values, with a small standard deviation. If we compare the results with the results in original paper [1] (c.f. Fig.4 and Fig.5 in Appendix), the error in average value is similar; while the standard deviation in our results is generally smaller than results in [1].

However, the estimations of *omega* do not arrive at the same performance as the original paper, especially for ω_2^2 (cf. Fig.3). The average value of ω_2^2 in all three methods is far away from the true value 0.2. As for ω_1^2 and ω_{12}, the results looks a little bit better: the average is much closer to their true value, with a more reasonable standard deviation. For this collapse of performance, in the first time, we guessed that it was due to coding bugs or due to some parameters setting mistakes. But by developing programs separately and independently, our two members in the team got the same simulation results. Thus, we guess now the fail of task may be due to different versions of mathematical libraries and programming languages used by us and in paper [1].

	beta11				beta12			
	mean	std	CI_95_low	CI_95_high	mean	std	CI_95_low	CI_95_high
method								
DMC	2.993962	0.019547	2.955653	3.032272	-2.004602	0.021647	-2.047027	-1.962177
DMC-IS	3.005906	0.016328	2.973904	3.037908	-1.976421	0.015397	-2.006598	-1.946244
MCMC	2.961451	0.017299	2.927546	2.995356	-1.975476	0.018204	-2.011155	-1.939798
True_value	3.000000	-	-	-	-2.000000	-	-	-

Figure 1: Beta1

	beta21				beta22			
	mean	std	CI_95_low	CI_95_high	mean	std	CI_95_low	CI_95_high
method								
DMC	1.944519	0.034404	1.877091	2.011948	1.058309	0.035809	0.988128	1.12849
DMC-IS	2.018471	0.03924	1.941564	2.095378	0.928812	0.039078	0.852222	1.005401
MCMC	1.991471	0.035653	1.921594	2.061347	0.990957	0.041277	0.910058	1.071857
True_value	2.000000	-	-	-	1.000000	-	-	-

Figure 2: Beta2

	omega11				omega12				omega22			
	mean	std	CI_95_low	CI_95_high	mean	std	CI_95_low	CI_95_high	mean	std	CI_95_low	CI_95_high
method												
DMC	0.107297	0.01554	0.076839	0.137755	-0.161437	0.041592	-0.242954	-0.079921	0.898778	0.161595	0.582067	1.215489
DMC-IS	0.136463	0.094046	-0.047858	0.320783	-0.153008	0.117556	-0.383406	0.077391	1.141213	0.87518	-0.574054	2.85648
MCMC	0.106693	0.015418	0.076476	0.136911	-0.127667	0.039334	-0.204758	-0.050577	1.225412	0.182722	0.867295	1.583529
True_value	0.100000	-	-	-	-0.150000	-	-	-	0.200000	-	-	-

Figure 3: Omega

References

[1] Tomohiro Ando and Arnold Zellner. "Hierarchical Bayesian analysis of the seemingly unrelated regression and simultaneous equations models using a combination of direct Monte Carlo and importance sampling techniques". In: *Bayesian Analysis* 5.1 (2010), pp. 65–95.

[2] Arnold Zellner and Tomohiro Ando. "A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model". In: *Journal of Econometrics* 159.1 (2010), pp. 33–45. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2010.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0304407610001119.

[3] Arnold Zellner and Tomohiro Ando. "Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with Student-t errors, and its application for forecasting". In: *International Journal of Forecasting* 26.2 (2010). Special Issue: Bayesian Forecasting in Economics, pp. 413–434. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2009.12.012. URL: https://www.sciencedirect.com/science/article/pii/S0169207009002131.

Appendix

DMC-IS approach

	TV	Mean	Mode	SDs	95%CIs	
β_{11}	3.00	2.9869	2.9833	0.0507	2.8827	3.0839
β_{12}	-2.00	-2.0318	-2.0333	0.0516	-2.1369	-1.9332
β_{21}	2.00	2.0063	2.0233	0.0749	1.8767	2.1684
β_{22}	1.00	1.0121	1.0276	0.0668	0.8951	1.1622
ω_1^2	0.10	0.1073	0.1101	0.0164	0.0824	0.1467
ω_{12}	-0.05	-0.0566	-0.0618	0.0181	-0.1009	-0.0297
ω_2^2	0.20	0.2206	0.2282	0.0352	0.1684	0.3048

Figure 4: Result of DMC-IS in paper [1]

MCMC algorithm

	TV	Mean	Mode	SDs	95%CIs	
β_{11}	3.00	2.9873	2.9819	0.0494	2.8849	3.0792
β_{12}	-2.00	-2.0353	-2.0307	0.0502	-2.1281	-1.9309
β_{21}	2.00	2.0266	2.0235	0.0707	1.8815	2.1613
β_{22}	1.00	1.0108	1.0284	0.0633	0.9037	1.1531
ω_1^2	0.10	0.1049	0.1055	0.0155	0.0793	0.1397
ω_{12}	-0.05	-0.0554	-0.0592	0.0167	-0.0957	-0.0295
ω_2^2	0.20	0.2059	0.2068	0.0301	0.1565	0.2740

Figure 5: Result of MCMC in paper [1]

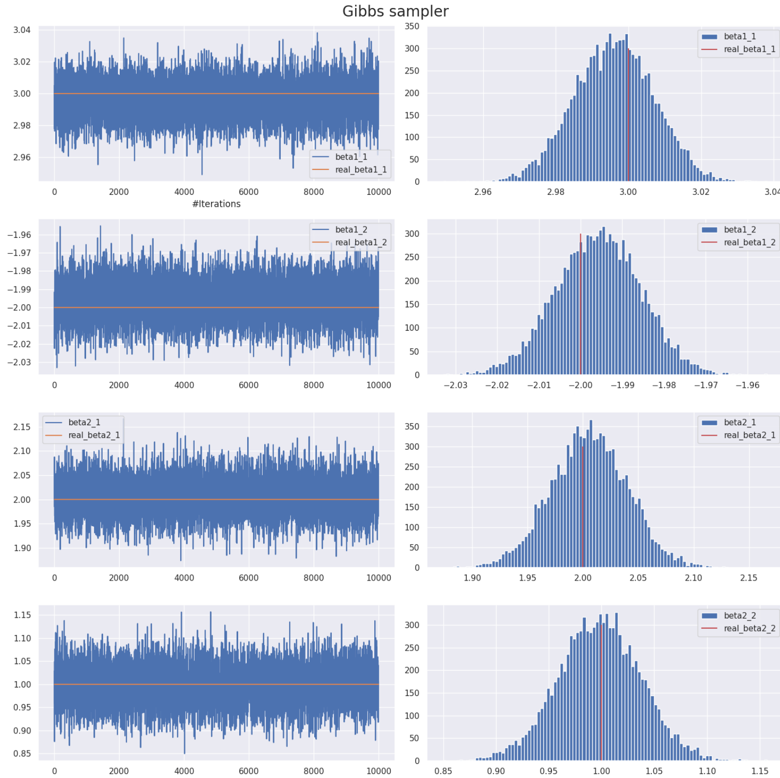

Figure 6: Sampling details of beta in Gibbs MCMC

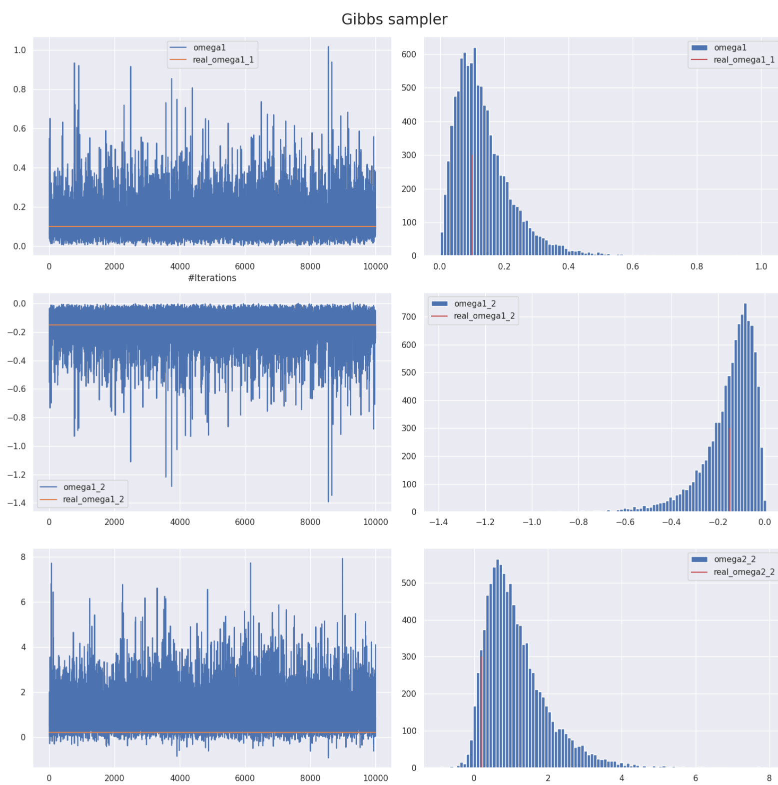

Figure 7: Sampling details of omega in Gibbs MCMC

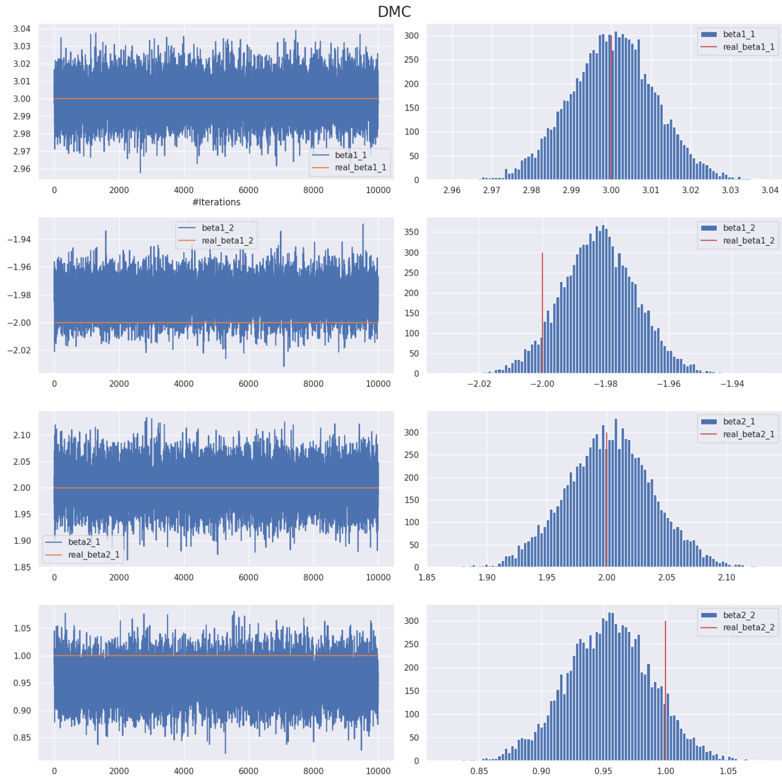

Figure 8: Sampling details of beta in DMC

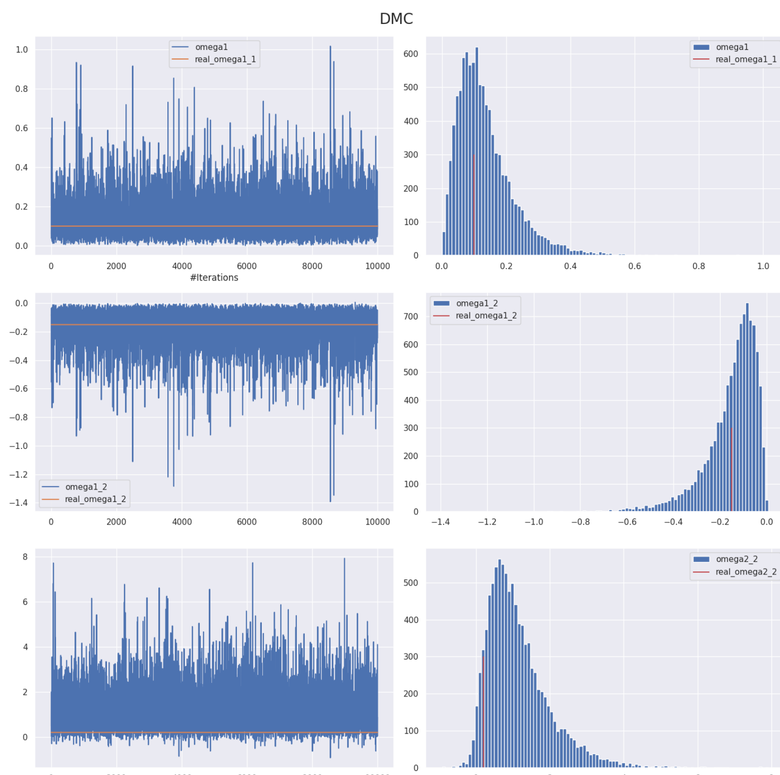

Figure 9: Sampling details of omega in DMC

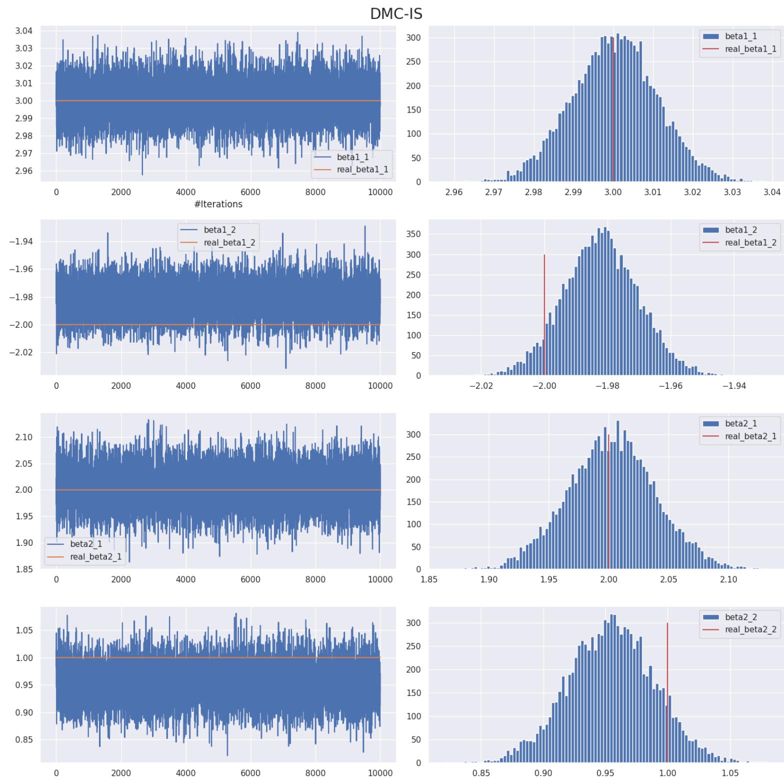

Figure 10: Sampling details of beta in DMC-IS

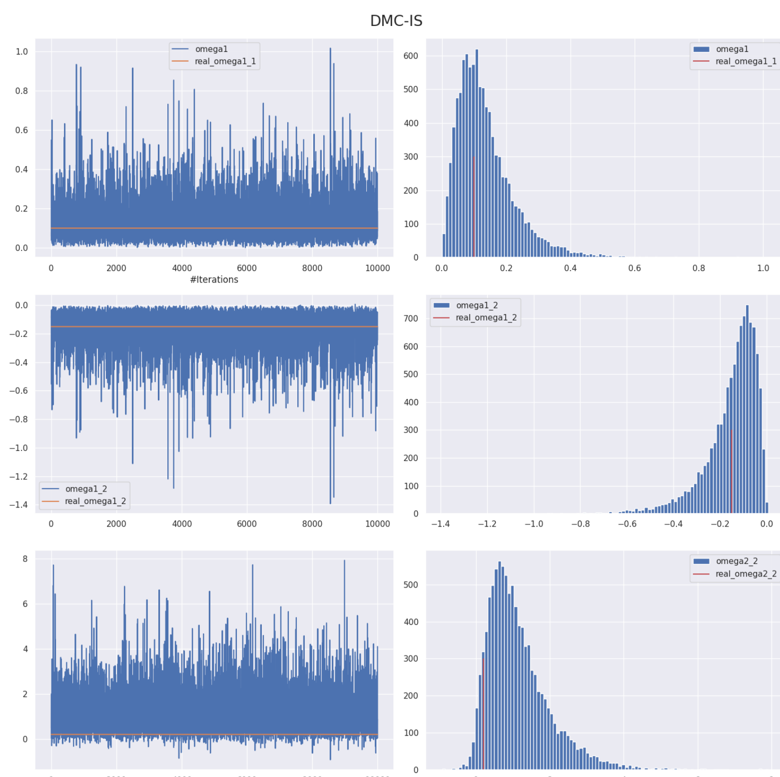

Figure 11: Sampling details of omega in DMC-IS