

3 Algorithm Implementation

The algorithm has been implemented in three different simulations, the first two using simulated datasets and the last one using a real dataset.

For each simulated dataset, the covariates are randomly drawn from pre-determined Normal distributions and the true parameters are specified beforehand. Given the matrix of covariates X and the vector of true parameters b , the latent variables are then computed as the matrix product $X'b$, to which random standard Normal errors are added. The dependent variables Y_i 's are then set either equal to 1 or to 0, depending on whether the corresponding latent is above or below 0 respectively. For each generation of the latent, it is then possible to compute the OLS estimators, and, by iterating such process, it is possible to obtain a large sample of OLS estimators for the same true model, from which it is possible to derive the exact distributions of the OLS estimators, which are then used as a reference to evaluate the performance of the algorithms.

For the real dataset, such iteration is obviously not possible, thus the results of the probit regression are used as a reference.

For each dataset, the Metropolis Hastings algorithm, and both the uninformed and informed Gibbs sampler are run using different initial values, to evaluate their robustness, and different priors (when possible), to evaluate their effect on the evolution of the Markov Chain. The criteria for the performance evaluation of the algorithms are the following:

- The speed of convergence of the Markov Chains: the number of burn-in iterations.
- The correctness of the generated distributions: the mean squared errors (MSEs) of their means with respect to the true parameters, and their standard deviation.
- The autocorrelation function of the Markov Chain: if the process converges to stationarity, then the corresponding autocorrelation function should approximately converge to 0 for each parameter.
- The empirical acceptance rate (for the Metropolis Hastings algorithm): to guarantee a proper mixing of the chain, the acceptance rate should stabilize as the process evolves; furthermore, a high acceptance rate would imply a slow convergence.

3.1 Simulations

3.1.1 Simulation 1

The first simulated dataset contains two regressors, and the true parameters are: $[\beta_0 = 2; \beta_1 = 1; \beta_2 = -1]$. Obviously, the distributions of the OLS estimators are Normally distributed and centred around the true values, as it is possible to see from figures 1,2 and 3, along with their respective standard deviations (0.07, 0.032, 0.032). The first simulation (MCMC1 in the table) starts from relatively close values and relies on a multivariate standard Normal as a prior: as it is possible to see, the process stabilizes quickly, and the distributions approximately coincide with the target ones, with larger standard deviations. Changing the initial values (MCMC2) obviously affects the convergence of the process: the further they are, the more burn-in iterations are necessary before convergence is reached (in this case, approximately 3000). However, they do not seem to affect the quality of the final distribution: both the means and the standard deviations coincide with the preceding simulation. The reason for such late convergence can be identified by considering the acceptance rate: for the first 3000 iterations, it is equal to 1, therefore the process does not mix at all, since all the steps are accepted; indeed, as soon as the rate starts to decline, the process stabilizes, as it is also possible to see by the autocorrelation of the estimates converging to 0.

The algorithm is then run multiple times (from MCMC3 to 6) with different starting values and priors, and it is possible to see that the process converges to stationarity relatively well. In this simulation, different priors (more and less informative) do not lead to significant differences in terms of performance.

On the other hand, MCMC7 is a hypothetical simulation, where, instead of substituting the posterior with the product of the likelihood by the prior, the target posterior distribution is directly inserted in the algorithm. As

expected, the generated distributions almost perfectly coincide with the target distributions, both in terms of means and standard deviations.

As it is possible to see from Uninformed Gibbs (UG) 1,2 and 3, the uninformed Gibbs sampler allows for a quicker convergence of the process, even for starting values relatively far from the correct ones (UG 3). The quality of the generated distributions is satisfying: the MSEs are low, while the standard deviations are slightly larger than the true ones.

From Informed Gibbs (IG) 2 and 3, it is possible to identify the effect of a proper prior in the algorithm: given the same starting values ([5,5,5]), IG2 uses a standard Normal distribution as a prior, while IG3 uses exactly the means and the variance-covariance matrix of the target distribution. As it is possible to see, while the speed of convergence is not significantly affected, the distributions generated with the proper prior are almost identical to the correct ones: the means approximately coincide with the true parameters, and the standard deviations are even slightly lower than the true ones.

3.1.2 Simulation 2

The second simulated dataset contains three regressors, and the true parameters are [-1; 2; 3; 0.5]. As it is possible to see from figures 65, 66, 67, and 68, the OLS estimators are still Normally distributed and centred at the true parameters. From MCMC1 and MCMC2, it is possible to see that the algorithm still works properly for relatively close starting values: the MSEs are low, the autocorrelations quickly converge to 0, the acceptance rate quickly stabilizes, and the generated distributions are approximately centred around the true values, with slightly larger standard deviations. MCMC3 and MCMC4 allow to analyse the effect of the prior: given the same starting values, while MCMC3 uses a multivariate standard Normal distribution as a prior, MCMC4 has its prior centred at the true values of the parameters. Compared to MCMC3, MCMC4 process converges more quickly: indeed, the acceptance rate stabilizes at a lower level, implying a more effective mixing as the process evolves. In terms of quality of the generated distributions, there are no significant differences.

The uninformed Gibbs samplers (UG1, UG2 and UG3) still work relatively well: the processes converge relatively quickly, with the autocorrelations converging to 0, and the generated distributions are approximately centred around the true values, with slightly larger standard deviations. Compared to the previous MCMCs, the MSEs are overall slightly higher, however the difference can be considered as negligible.

The informed Gibbs samplers (IG1 and IG2) allow to analyse the effect of a proper prior, given the same starting values: compared to IG1, which relies on a multivariate standard Normal distribution, IG2 uses the means and the variance-covariance matrix of the target distributions. As it is possible to see, IG2 allows for a quicker convergence of the process, and, furthermore, the generated distributions are characterized by a lower standard deviation, which makes them almost identical to the target ones.

3.1.3 Simulation 3: Real Dataset

The real dataset, called “Gaming.csv”, contains information about consumers’ six driving factors behind the choice of a type of computer (“A” or “B”). In particular, the Bernoulli random variable Y_i is set equal to 1 if the corresponding consumer’s choice is “A”, thus the parameters can be imprecisely interpreted as the effect of the regressors on the probability of the consumer choosing “A” over “B”. Given the use of a real dataset, it is not possible to generate the true distribution of the OLS estimators, therefore the STATA output of the probit regression is used as a reference for the true values of the parameters (see figure 111).

As it is possible to see from MCMC1, 2, and 3, neither of the processes converges to stationarity: the autocorrelations do not converge to 0, and the generated distributions are thus obviously wrong. The reason can be easily identified by considering the acceptance rate: it is always equal to 1, thus, all the steps are accepted, and the process never starts mixing. Such lack of convergence arises even when using starting values very close to the probit ones, or when using informative priors close to the probit results. Therefore, one

possible solution may be to find a different proposal transition probability, which may guarantee a proper mixing of the process.

As it is possible to see from UG1, 2, and 3, uninformed Gibbs samplers processes quickly convergence to stationarity, even for starting values which are relatively far from the probit ones (UG3): the MSEs are very low, and the generated distributions are approximately centred around the probit values.

IG1 and IG2 allow to analyse the effect of a proper prior: while IG1 relies on a multivariate standard Normal distribution, IG2 relies on a more informative prior, closer to the probit values. While in both cases the processes quickly converge to stationarity, the MSE of IG1 is higher than the one of IG2 (2 vs 0.12), mainly due to the generated distribution of beta0, which, in IG1, is centred at a wrong value (1 instead of 4). It is also worth mentioning that the generated distributions by both IGs are characterized by lower standard deviations, when compared to the previous UGs.

3.2 Performance Assessment

From the previous simulations, it is possible to conclude that Gibbs sampler is overall a more effective algorithm with respect to the standard Metropolis Hastings (MH), both in terms of convergence and of running time (at least, in our case). In particular, the number of required burn-in iterations before convergence is lower, which does not compromise the quality of the generated distributions. Without an informative prior, uninformed Gibbs sampler proves to be an effective algorithm, provided that the starting values are not too far from the true ones. Instead, given an informative prior, an informed Gibbs sampler or even a Metropolis Hastings algorithm would be a better choice, since they would lead to a faster convergence and more precise distributions, with lower standard deviations and closer to the true ones. However, the choice of the prior needs to be careful, since, as it is possible to see from simulation 3, the insertion of a wrong prior may lead to the generation of wrong distributions. Finally, for all the algorithms, the starting values clearly represent a determinant factor in the convergence of the process: values too far from the true ones would compromise the effectiveness of the algorithm. With respect to this, it is possible to notice that Gibbs sampler still seems to be more robust to distant starting values compared to MH.

Appendix

SIMULATION 1:

For the sake of clarity, charts can be found in the appendix, referenced accordingly in the tables.

Parameters of the Target Distribution

	β_0	β_1	β_2
Pdf	(1)	(2)	(3)
Mean		(4)	
Covariance matrix		(5)	

Simulation 1 Results:

	Distribution	Sim. MC	MSE	Autocorrelation	Acc. rate
MCMC 1	(6)	(7)	(8)	(9)	(10)
MCMC 2	(11)	(12)	(13)	(14)	(15)
MCMC 3	(16)	(17)	(18)	(19)	(20)
MCMC 4	(21)	(22)	(23)	(24)	(25)
MCMC 5	(26)	(27)	(28)	(29)	(30)
MCMC 6	(31) – Target as Prior	(32)	(33)	(34)	(35)
MCMC 7	(36) - Target Posterior	(37)	(38)	(39)	n.d.

	Distribution	Simulated Markov Chain	MSE	Autocorrelation
Uninformed Gibbs 1	(40)	(41)	(42)	(43)
Uninformed Gibbs 2	(44)	(45)	(46)	(47)
Uninformed Gibbs 3	(48)	(49)	(50)	(51)
Uninformed Gibbs 4 (failed convergence)	(52)	(53)	(54)	(55)
Informed Gibbs 1	(56)	(57)	(58)	(59)
Informed Gibbs 2	(60)	(61)	(62)	(63)

SIMULATION 2:

For the sake of clarity, charts can be found in the appendix, referenced accordingly in the tables.

Parameters of the Target Distribution

	β_0	β_1	β_2	β_3
Pdf	(65)	(66)	(67)	(68)
Covariance matrix	(69)			
Mean	(70)			

Simulation 2 Results:

	Distribution	Sim. MC	MSE	Autocorrelation	Acc. rate
MCMC 1	(71)	(72)	(73)	(74)	(75)
MCMC 2	(76)	(77)	(78)	(79)	(80)

<i>MCMC 3</i>	(81)	(82)	(83)	(84)	(85)
<i>MCMC 4</i>	(86)	(87)	(88)	(89)	(90)

	<i>Distribution</i>	<i>Simulated Markov Chain</i>	<i>MSE</i>	<i>Autocorrelation</i>
<i>Uninformed Gibbs 1</i>	(91)	(92)	(93)	(94)
<i>Uninformed Gibbs 2</i>	(95)	(96)	(97)	(98)
<i>Uninformed Gibbs 3</i>	(99)	(100)	(101)	(102)
<i>Informed Gibbs 1</i>	(103)	(104)	(105)	(106)
<i>Informed Gibbs 2</i>	(107)	(108)	(109)	(110)

SIMULATION 3:

Parameters of the Target Distribution

<i>True parameters</i>	(111)
<i>Mean</i>	(112)
<i>Covariance matrix</i>	(113)

Simulation 3 Results:

	<i>Distribution</i>	<i>Simulated Markov Chain</i>	<i>MSE</i>	<i>Autocorrelation</i>	<i>Acc. Rate</i>
<i>MCMC 1</i>	(114)	(115)	(116)	(117)	(118)
<i>MCMC 2</i>	(119)	(120)	(121)	(122)	(123)
<i>MCMC 3</i>	(124)	(125)	(126)	(127)	(128)

	<i>Distribution</i>	<i>Simulated Markov Chain</i>	<i>MSE</i>	<i>Autocorrelation</i>
<i>Uninformed Gibbs 1</i>	(129)	(130)	(131)	(132)
<i>Uninformed Gibbs 2</i>	(133)	(134)	(135)	n.d.
<i>Uniformed Gibbs 3</i>	(136)	(137)	(138)	(139)
<i>Informed Gibbs 1</i>	(140)	(141)	(142)	(143)
<i>Informed Gibbs 2</i>	(144)	(145)	(146)	(147)

SIMULATION 1

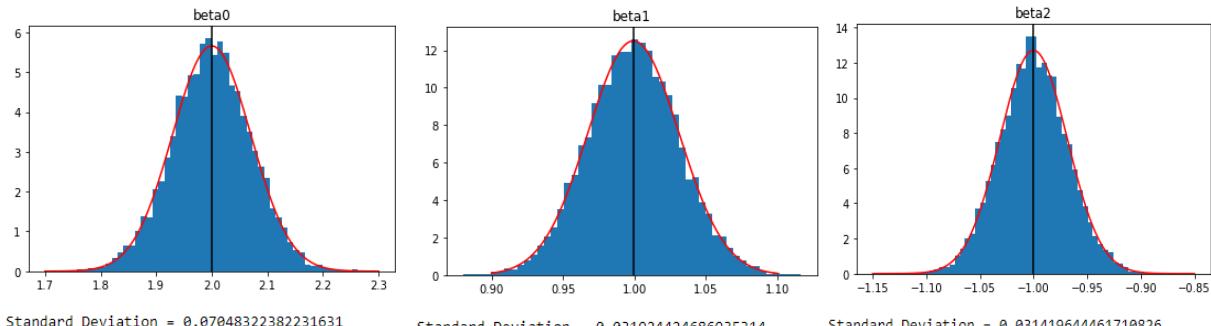


Figure (1) & Figure(2) & Figure (3)

The mean of the target is
[2.00083596 0.99996139 -1.00022078]

(4)

The variance-covariance matrix of the target is
[[5.07399900e-03 -5.89588163e-05 -2.03148306e-03]
[-5.89588163e-05 1.01405867e-03 2.76497760e-05]
[-2.03148306e-03 2.76497760e-05 1.01219678e-03]]

(5)

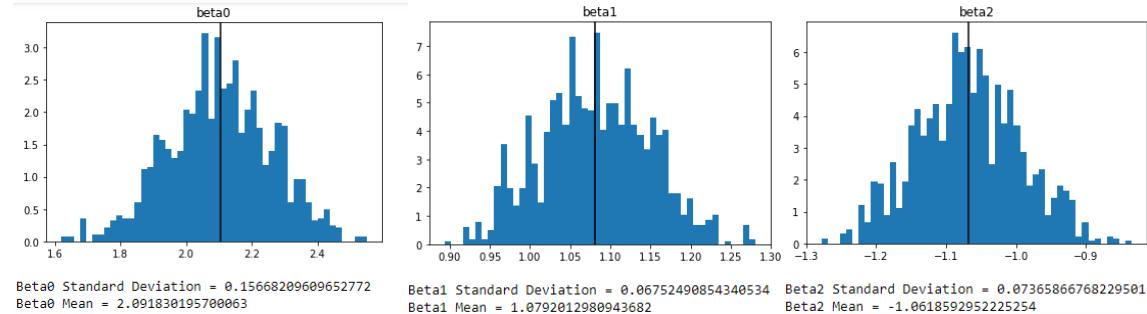


Figure (6)

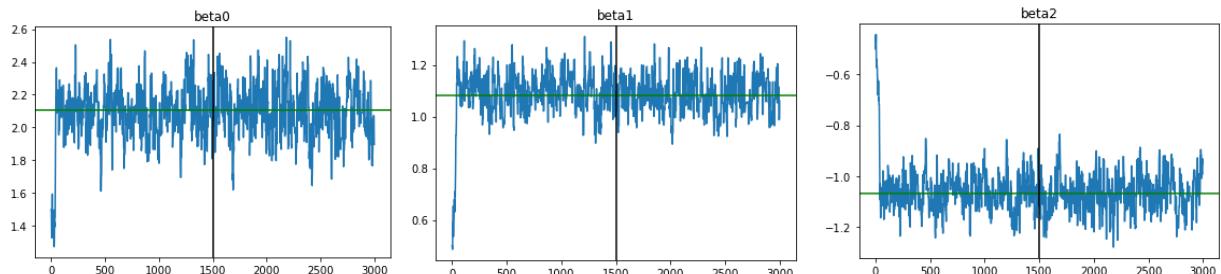


Figure (7)

The MSE is 0.00617740095585747

Figure (8)

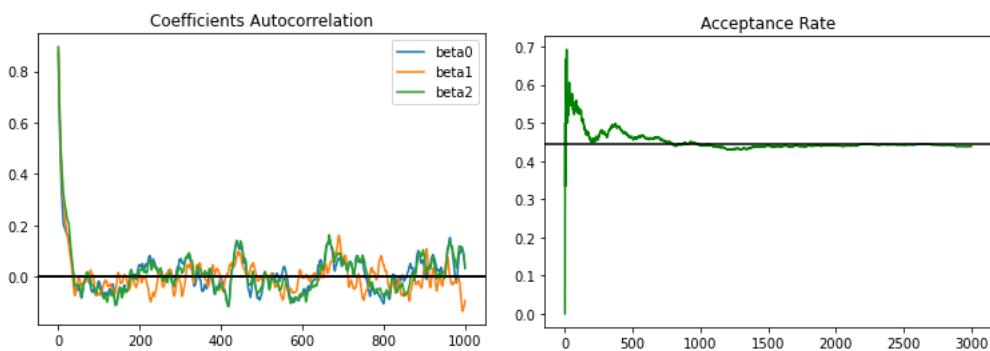


Figure (9) & Figure (10)

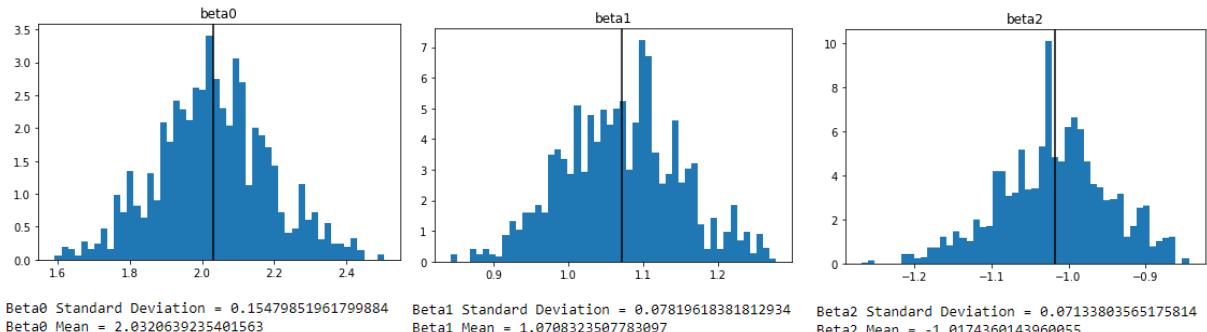


Figure (11)

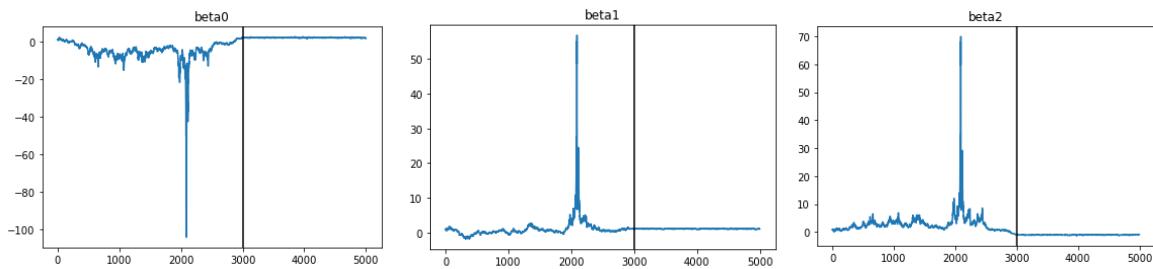


Figure (12)

The MSE is 0.0021164439025294053

Figure (13)

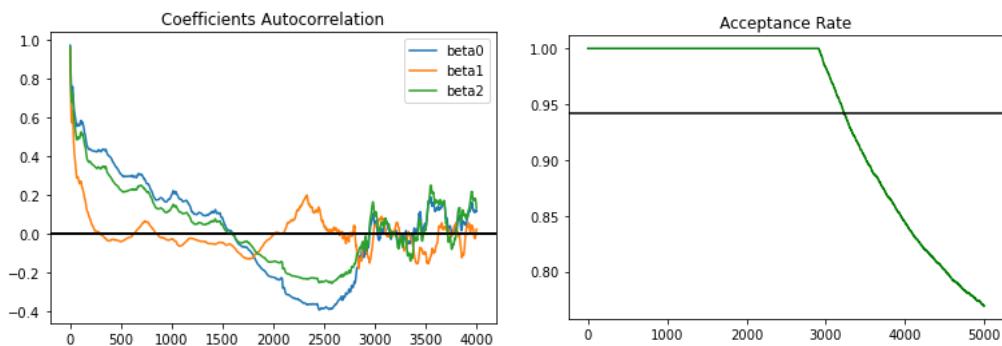


Figure (14) & Figure (15)

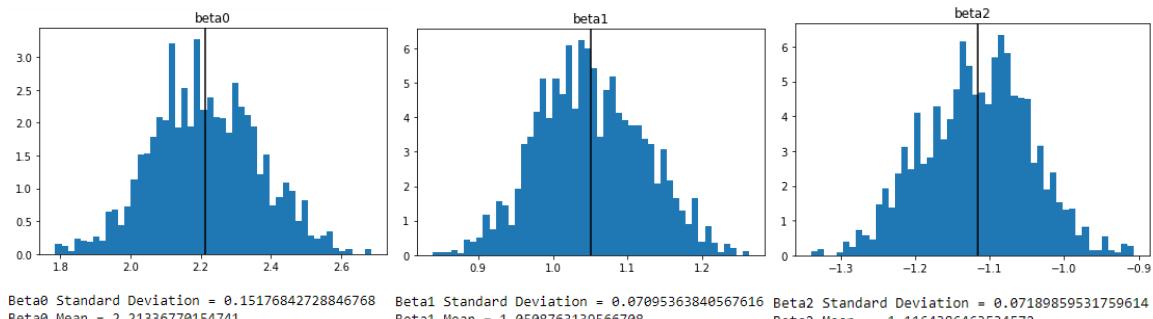


Figure 16

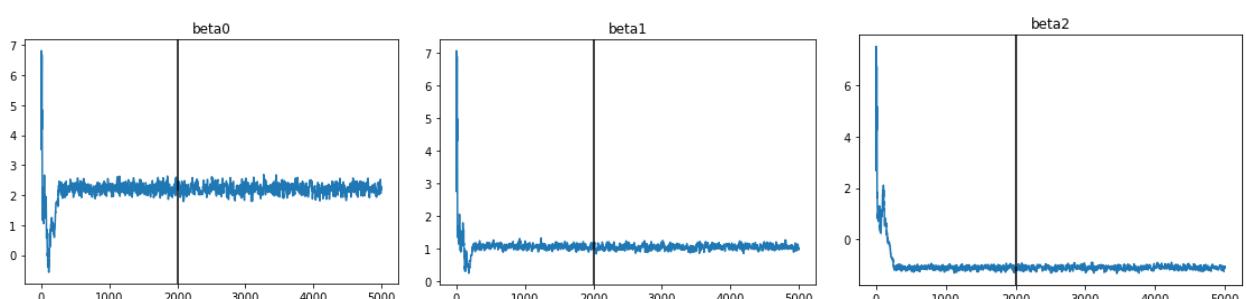


Figure 17

The MSE is 0.020557377916689254

Figure 18

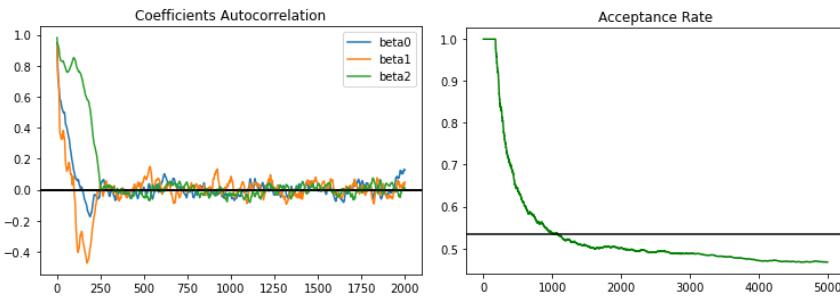


Figure 19 & Figure 20

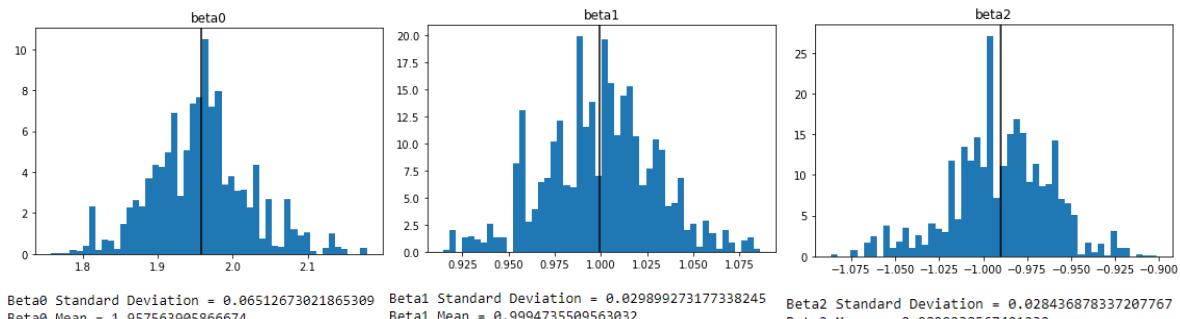


Figure 21

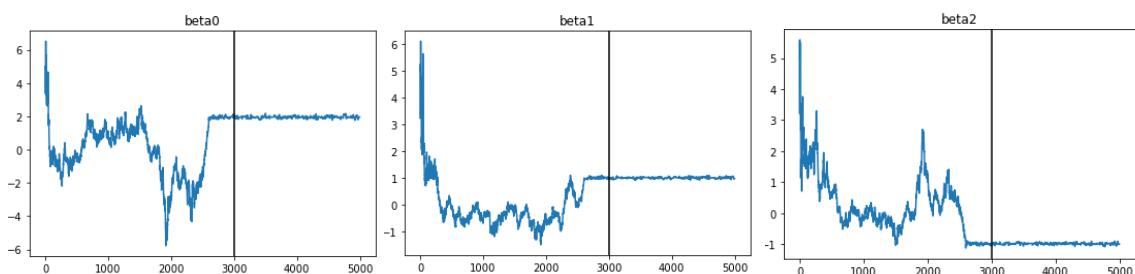


Figure 22

The MSE is 0.0006342092989605569

Figure 23

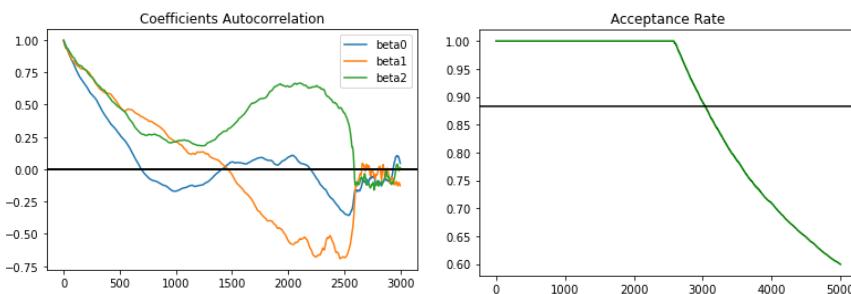


Figure 24 & Figure 25

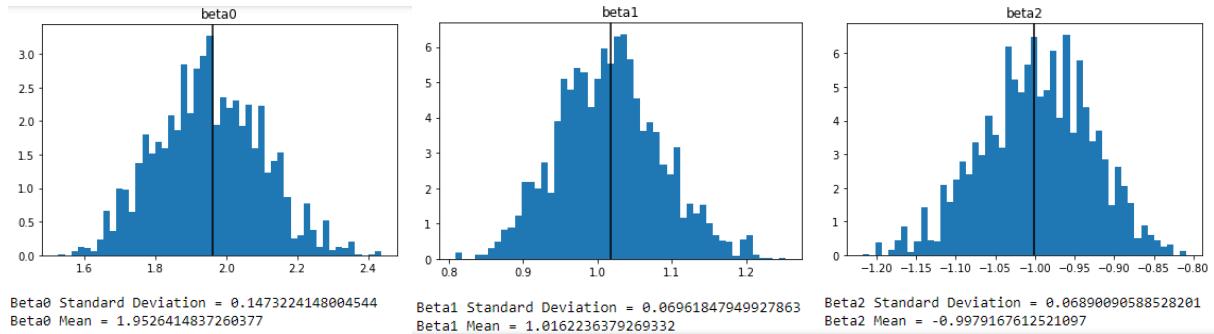


Figure 26

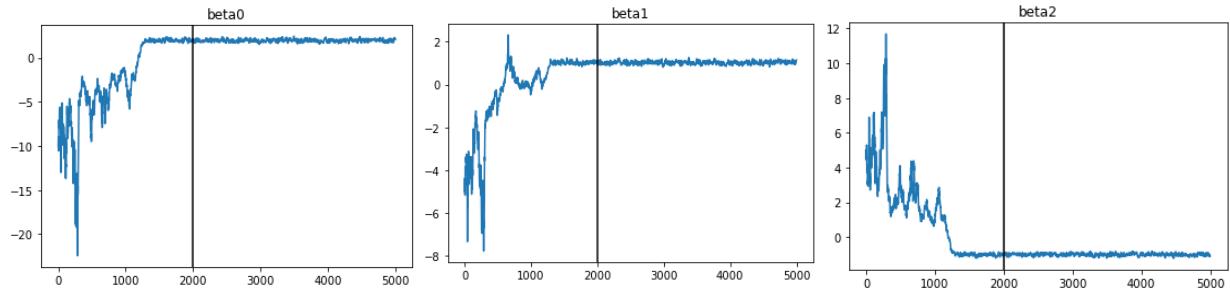


Figure 27

The MSE is 0.0008367917916453612

Figure 28

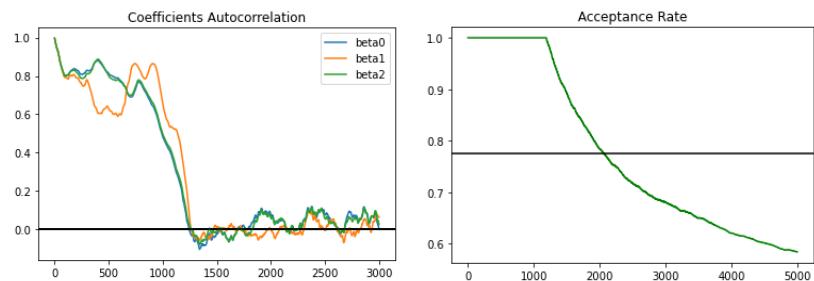


Figure 29 & Figure 30

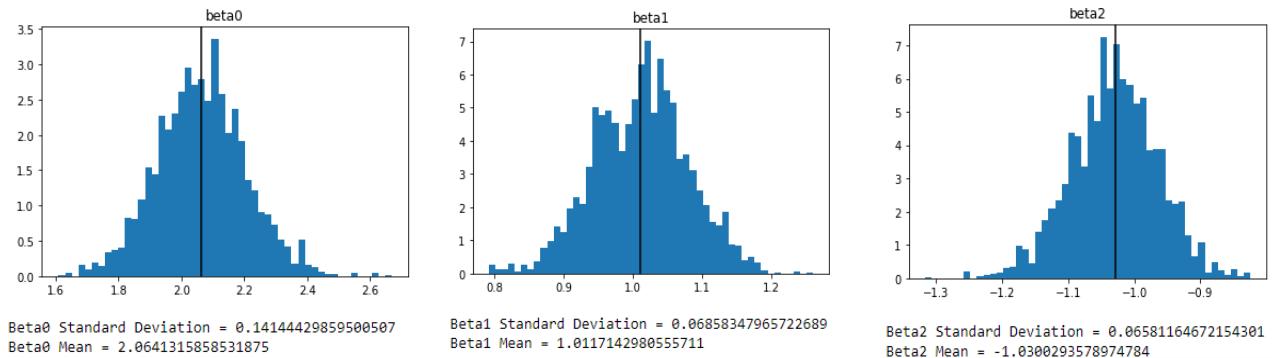


Figure 31

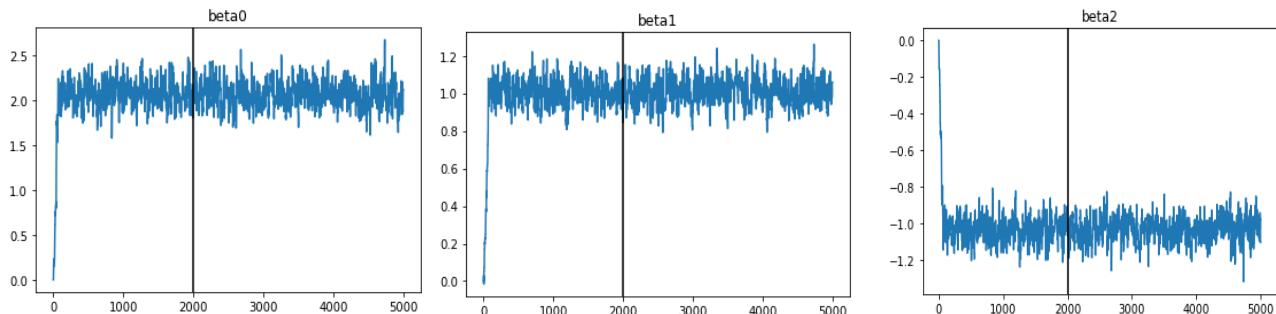


Figure 32

The MSE is 0.0017172824729047862

Figure 33

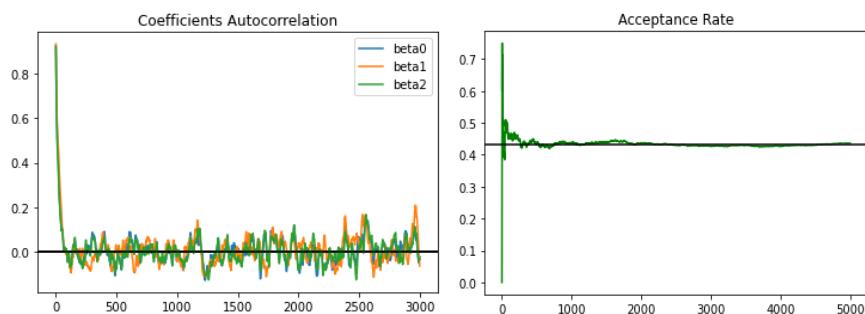


Figure 34 & Figure 35

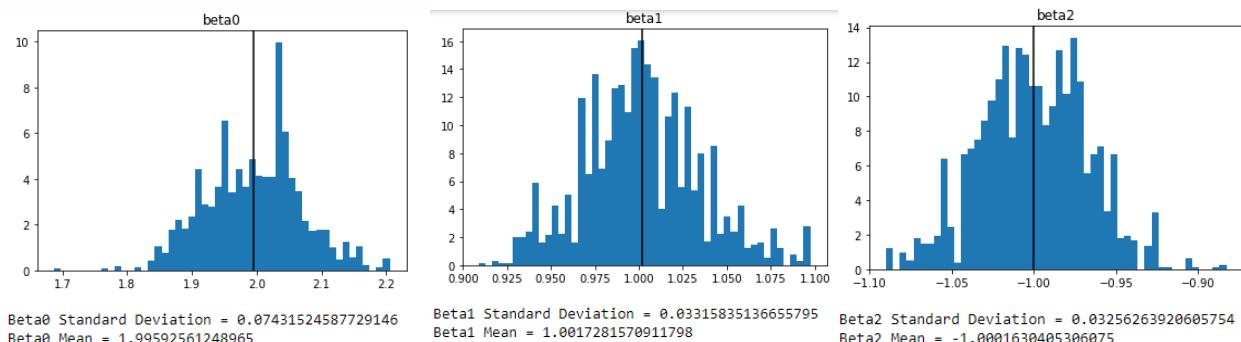


Figure 36

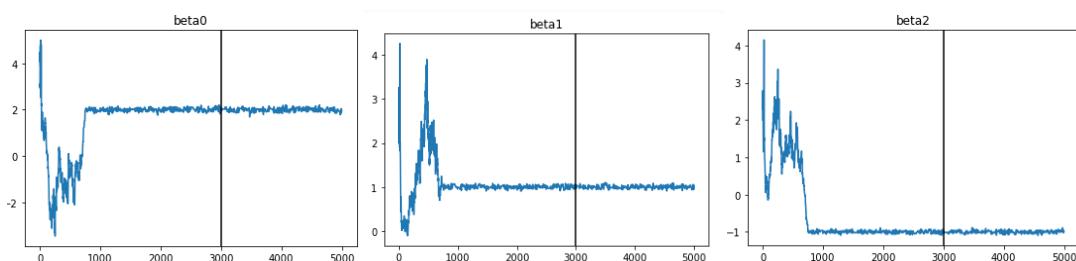


Figure 37

The MSE is 6.53791424363712e-06

Figure 38

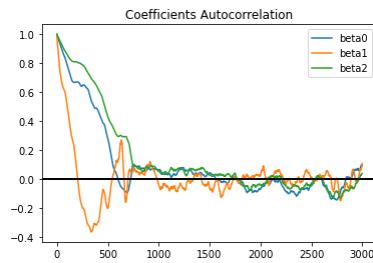


Figure 39

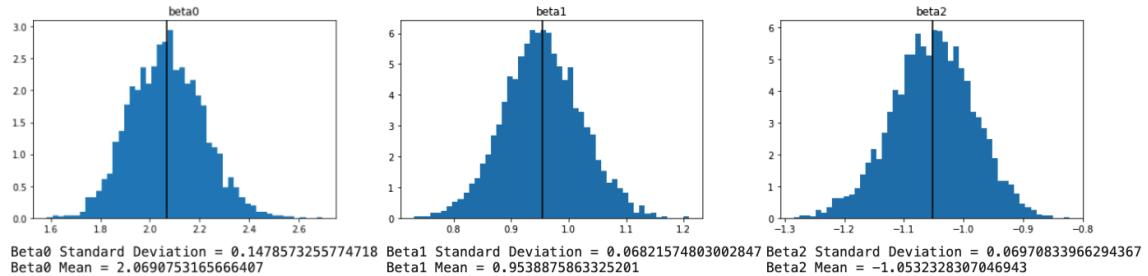


Figure 40

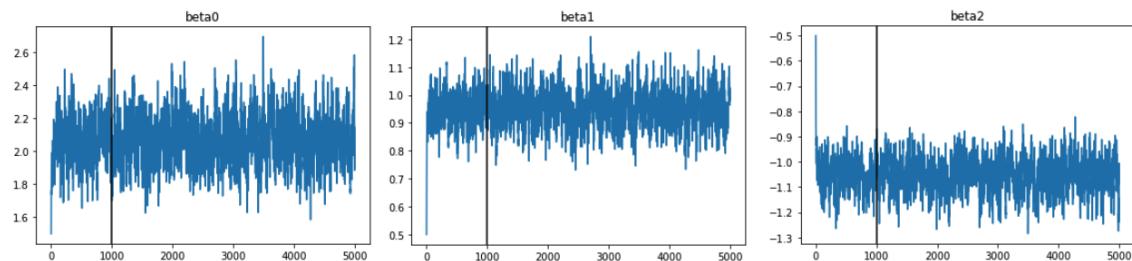


Figure 41

The MSE is 0.0032438294392856844

Figure 42

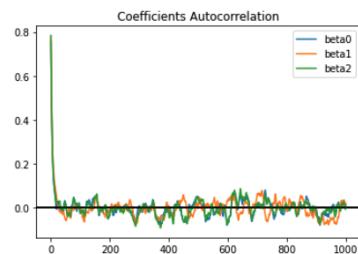


Figure 43

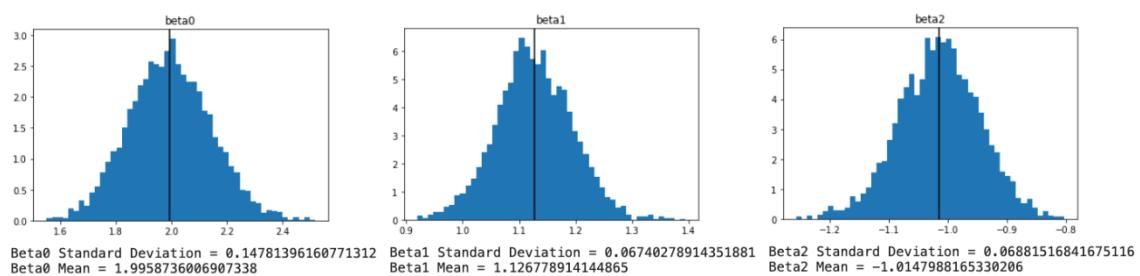


Figure 44

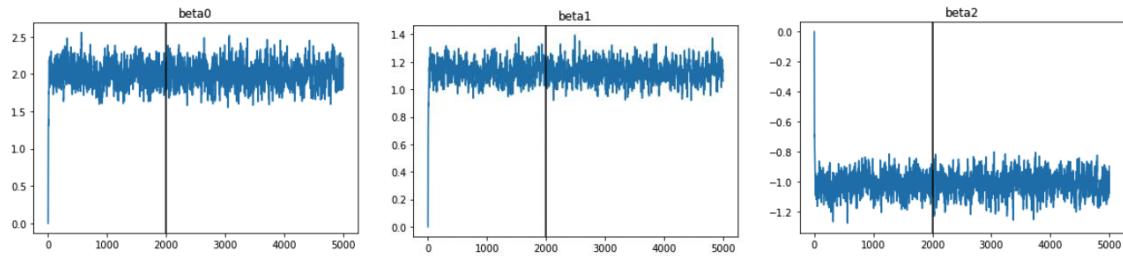


Figure 45

The MSE is 0.005436308404596187

Figure 46

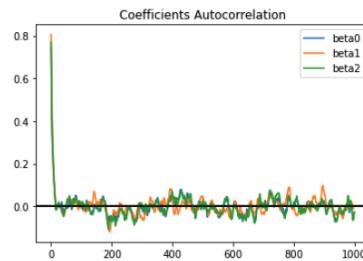


Figure 47

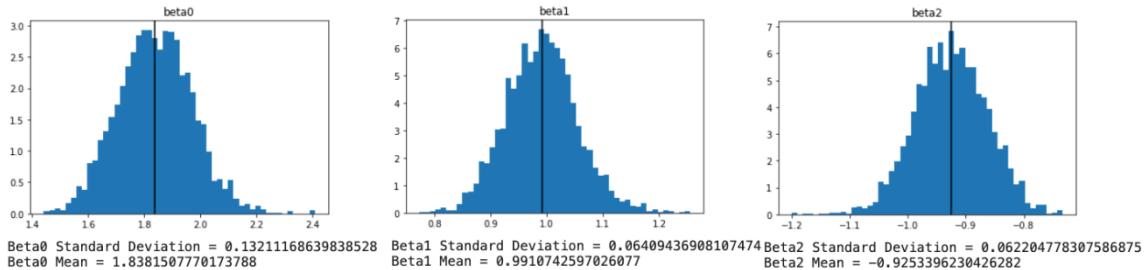


Figure 48

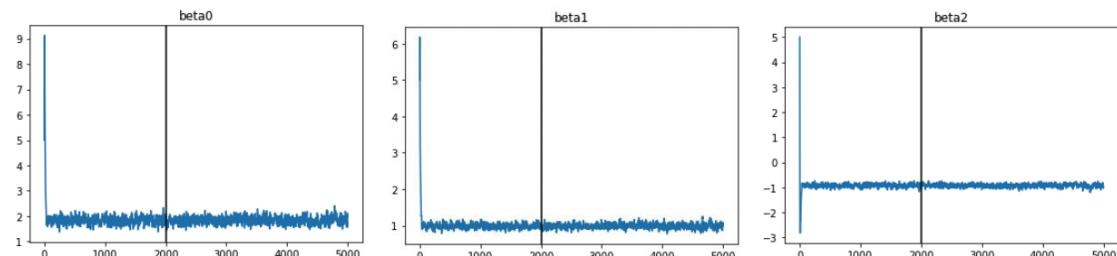


Figure 49

The MSE is 0.01061633723578386

Figure 50

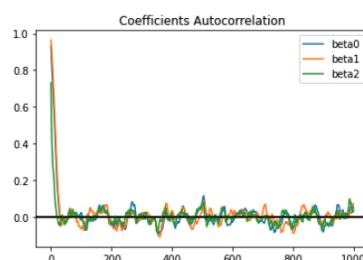


Figure 51

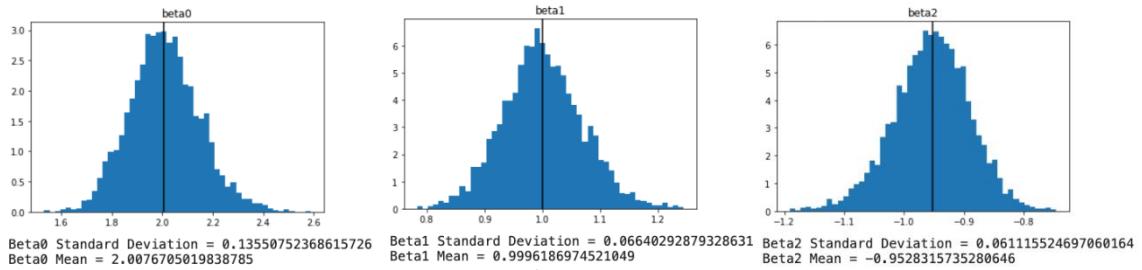


Figure 52

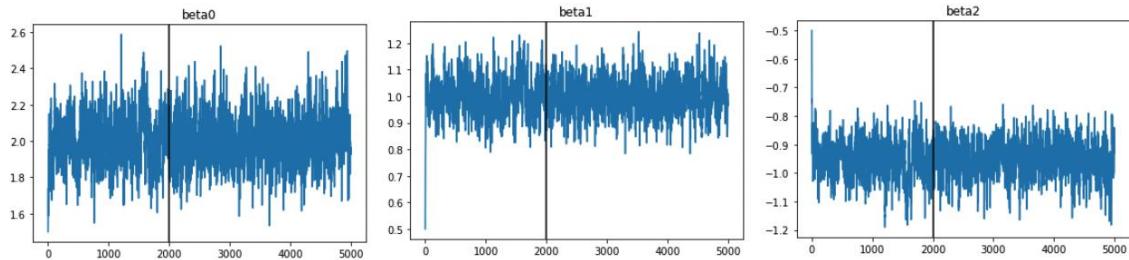


Figure 53

The MSE is 0.0007612808160520314

Figure 54

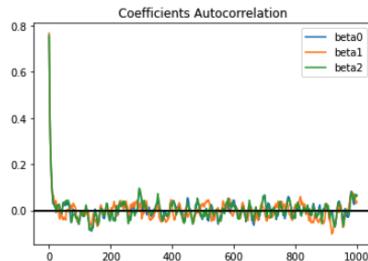


Figure 55

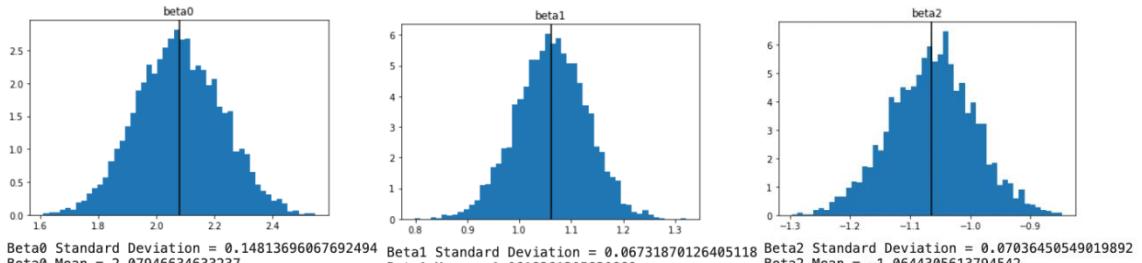


Figure 56

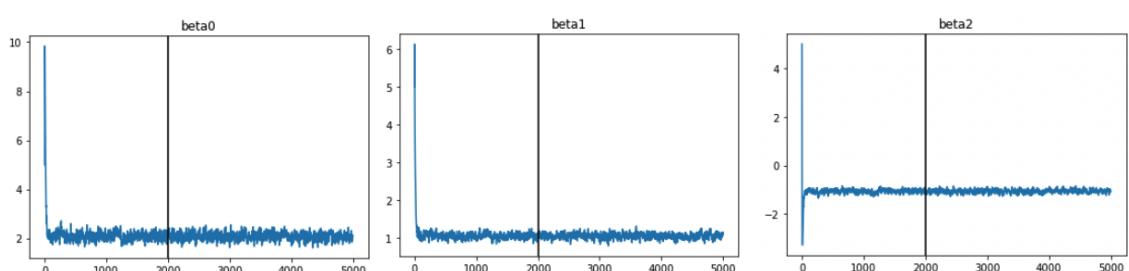


Figure 57

The MSE is 0.004762889287280429

Figure 58

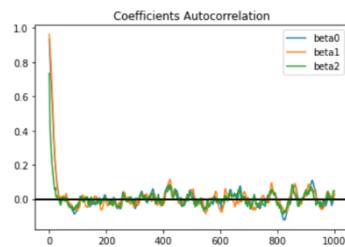


Figure 59

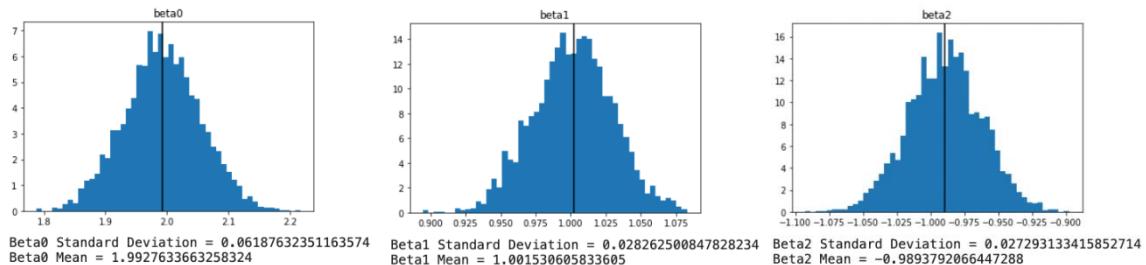


Figure 60

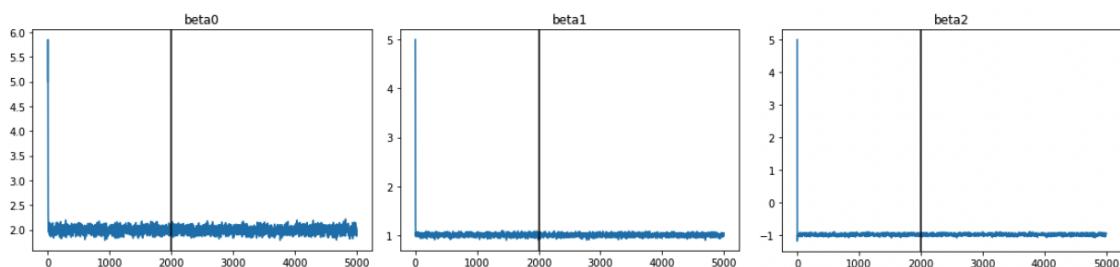


Figure 61

The MSE is $5.5837624215778575e-05$

Figure 62

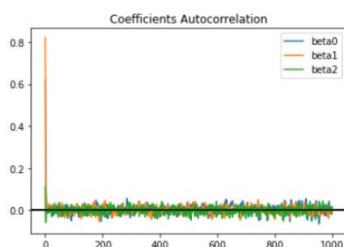


Figure 63

SIMULATION 2

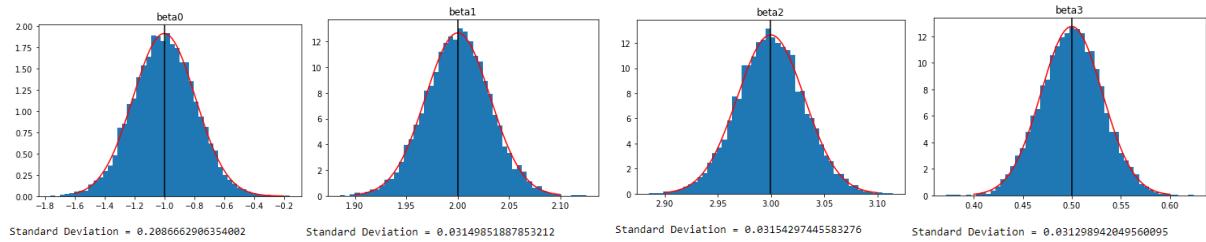


Figure 65-66-67-68

```
The mean of the target is
[-1.00143401  2.00027883  2.99980573  0.49983782]
The variance-covariance matrix of the target is
[[ 4.35459754e-02 -5.00612638e-03  2.98973409e-03 -2.84501362e-03]
 [-5.00612638e-03  9.92255917e-04 -1.41502883e-05 -3.72945259e-06]
 [ 2.98973409e-03 -1.41502883e-05  9.95058743e-04  2.01344582e-05]
 [-2.84501362e-03 -3.72945259e-06  2.01344582e-05  9.79721746e-04]]
```

Figure 69 & Figure 70

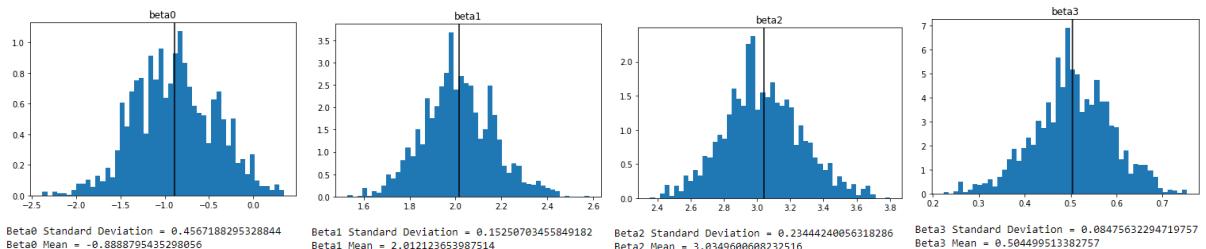


Figure 71

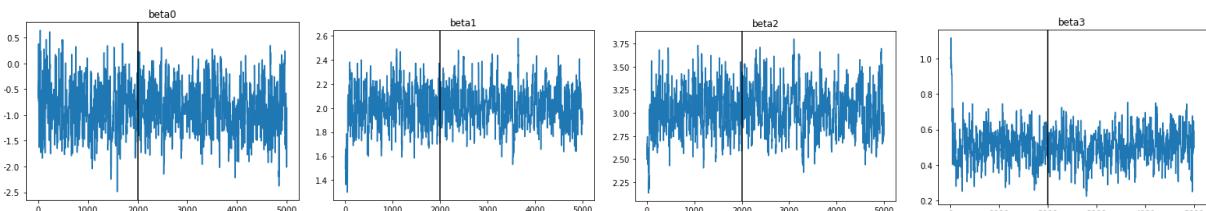


Figure 72

The MSE is 0.0034342975764000984

Figure 73

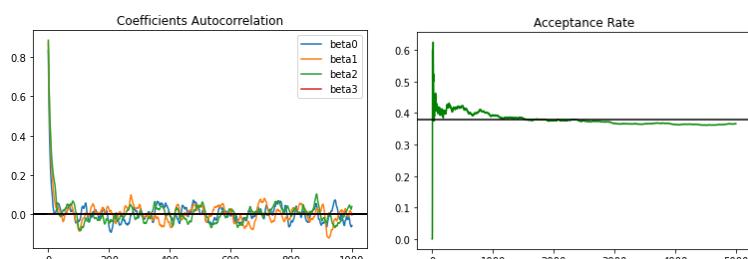


Figure 74 & Figure 75

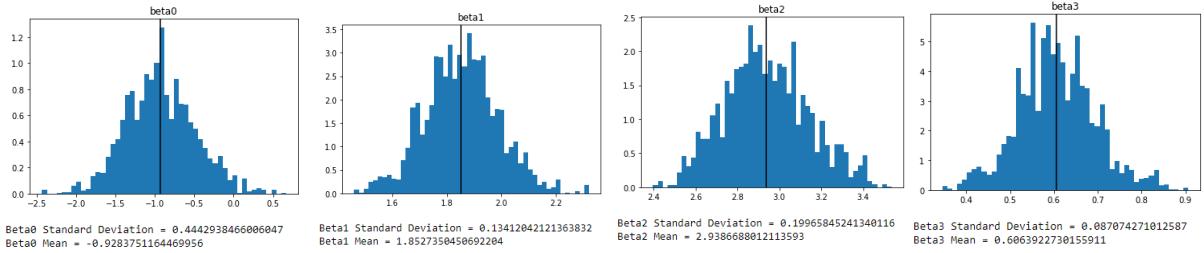


Figure 76

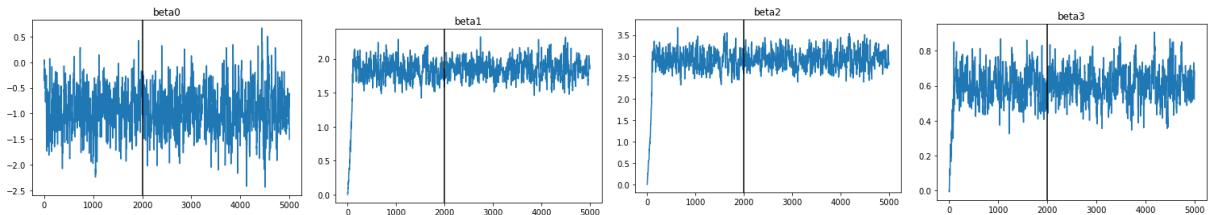


Figure 77

The MSE is 0.010474480649255456

Figure 78

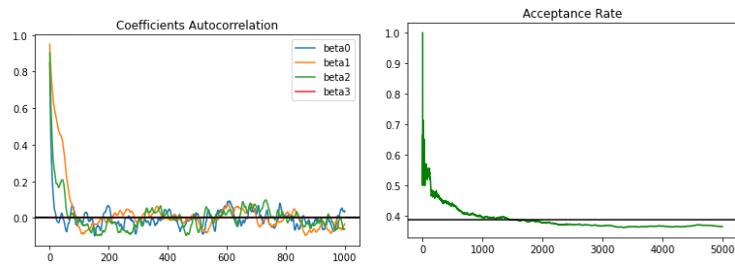


Figure 79 & Figure 80

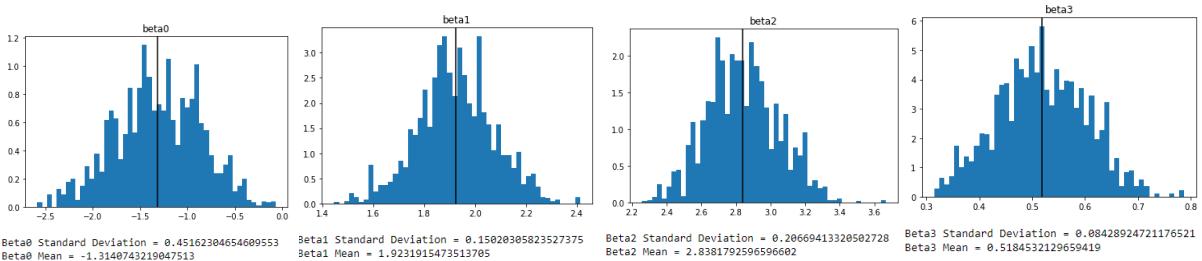


Figure 81

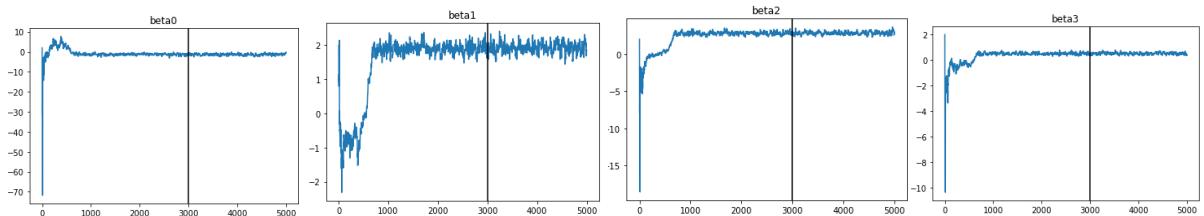


Figure 82

The MSE is 0.03276717278781704

Figure 83

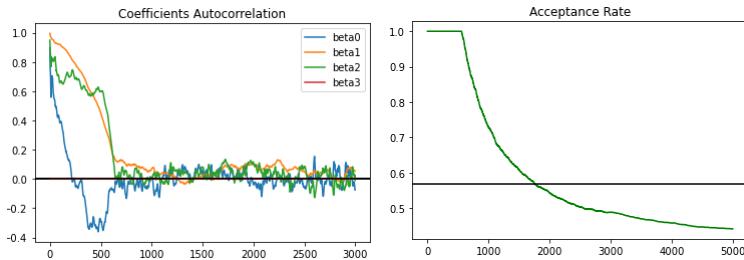


Figure 84 & Figure 85

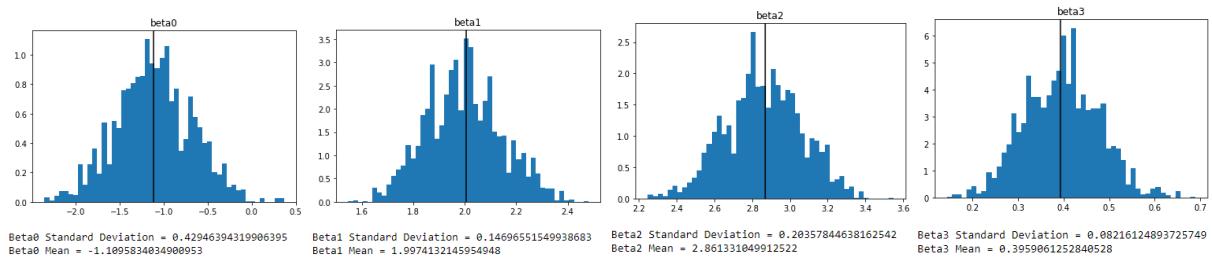


Figure 86

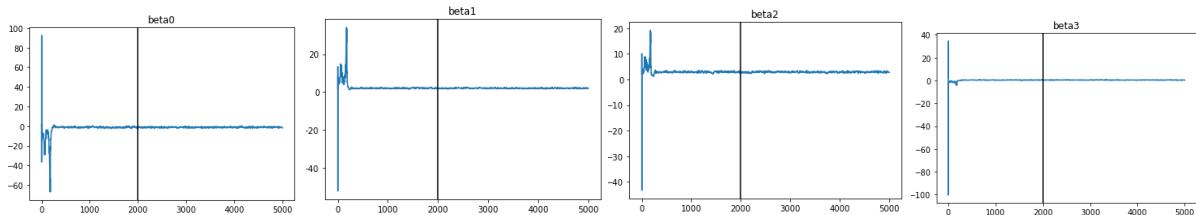


Figure 87

The MSE is 0.010519956562736198

Figure 88

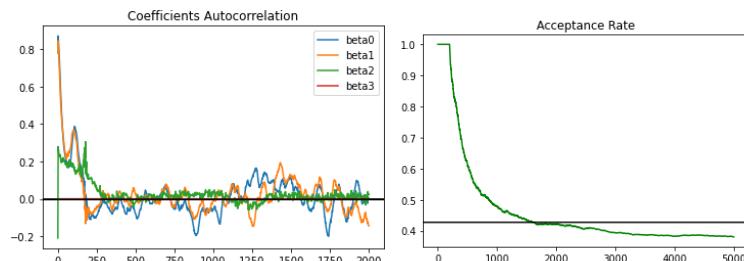


Figure 89 & Figure 90

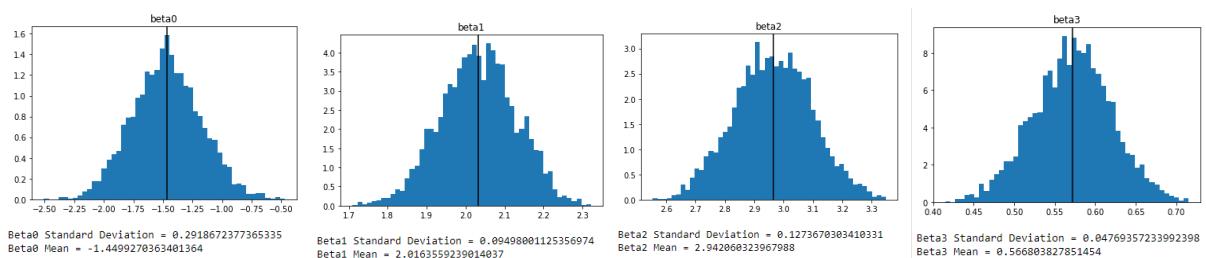


Figure 91

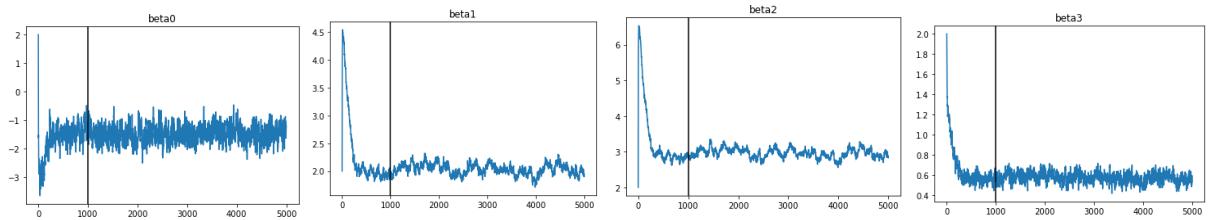


Figure 92

The MSE is 0.05263040293769704

Figure 93

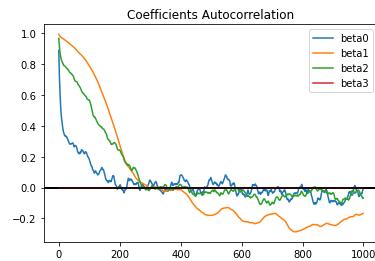


Figure 94

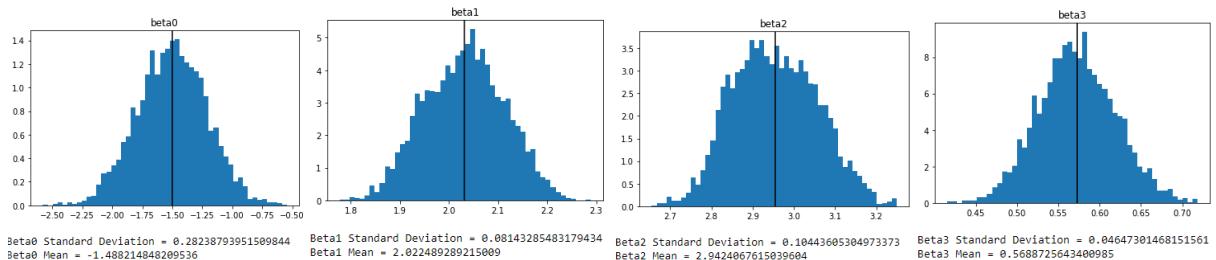


Figure 95

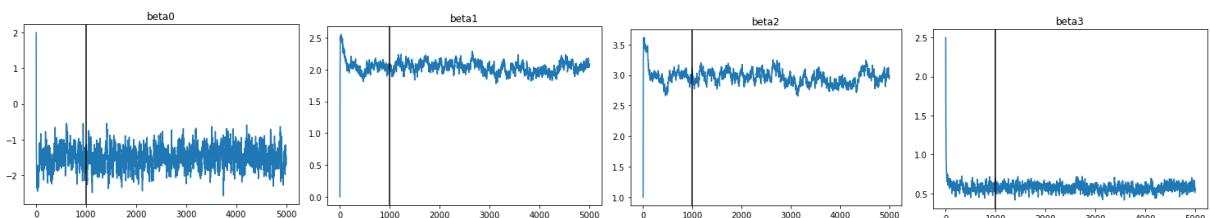


Figure 96

The MSE is 0.061729979345224846

Figure 97

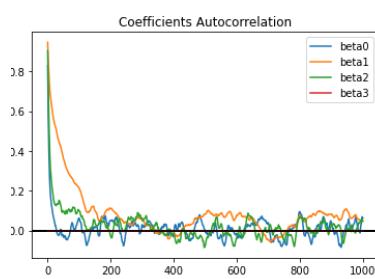


Figure 98

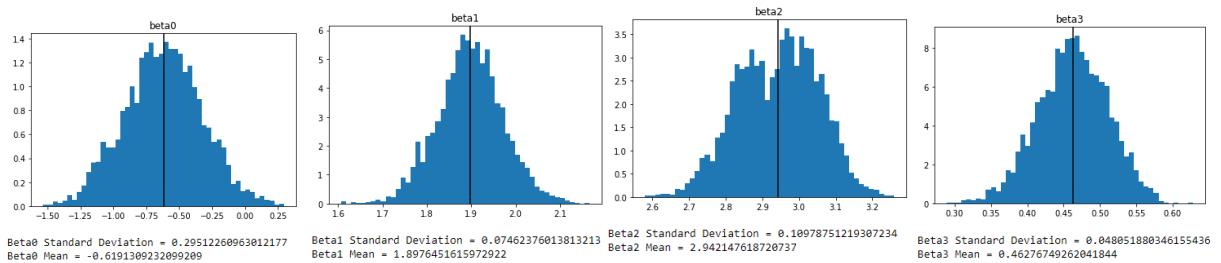


Figure 99

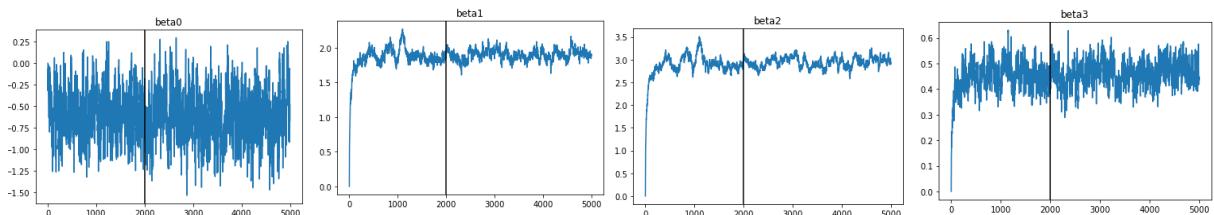


Figure 100

The MSE is 0.04006773105620584

Figure 101

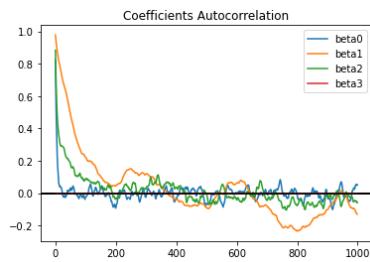


Figure 102

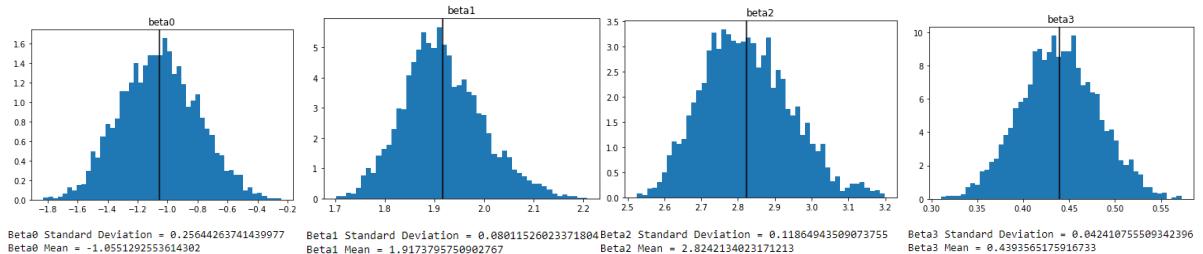


Figure 103

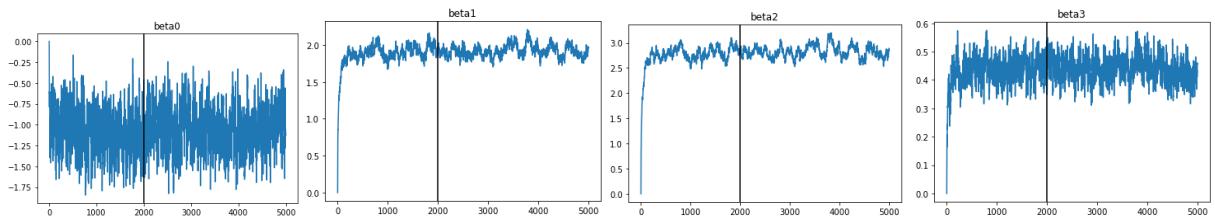


Figure 104

The MSE is 0.011110982323125077

Figure 105

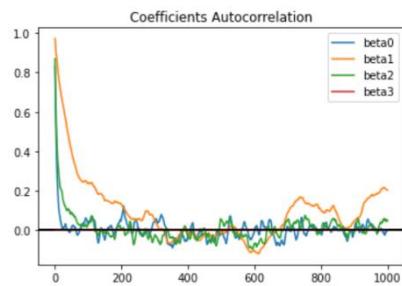


Figure 106

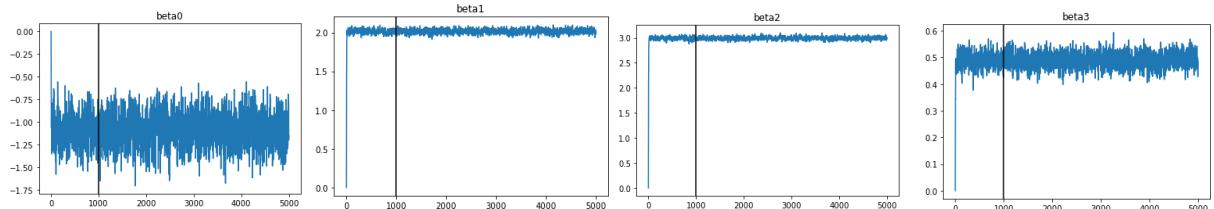


Figure 107

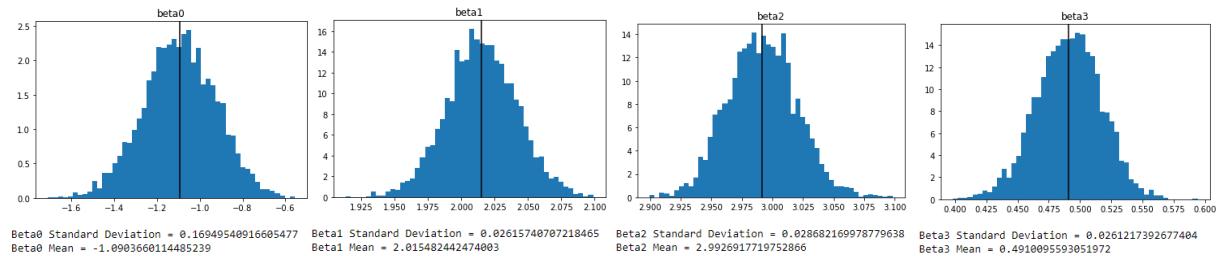


Figure 108

The MSE is 0.00213499006705884

Figure 109

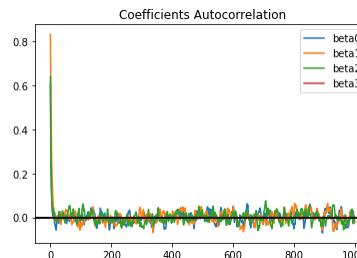


Figure 110

SIMULATION 3

A	Coefficient	Std. err.	z	P> z	[95% conf. interval]
ln_income	-1.14485	.1994539	-5.74	0.000	-1.535773 - .7539278
age	.2783464	.0114136	24.39	0.000	.2559762 .3007166
proportion_gaming_products	-.9335337	.0624765	-14.94	0.000	-1.055986 -.811082
previous_purchase_gaming	.5112991	.0606246	8.43	0.000	.392477 .6301212
male	-.2623994	.0533403	-4.92	0.000	-.3669445 -.1578543
_cons	4.17425	2.091251	2.00	0.046	.0754736 8.273027

Figure 111

The mean of the target is
 $[4.1645595 \quad 0.27840926 \quad -0.93369574 \quad 0.51139309 \quad -1.14414624 \quad -0.26189784]$

Figure 112

The variance-covariance matrix of the target is
 $\begin{bmatrix} 2.19428383e+00 & -6.15020832e-04 & 2.29795032e-03 & -2.79036755e-04 \\ -2.06062551e-01 & -7.34936808e-04 \\ [-6.15020832e-04 & 4.84036027e-05 & -8.59878656e-05 & -3.43881309e-05 \\ -7.86469129e-05 & -5.94031755e-05 \\ [2.29795032e-03 & -8.59878656e-05 & 1.72089515e-03 & 1.36374210e-04 \\ -5.67827690e-05 & 1.32812670e-04 \\ [-2.79036755e-04 & -3.43881309e-05 & 1.36374210e-04 & 1.57185628e-03 \\ 7.64963677e-05 & 3.88976744e-05 \\ [-0.66662551e-01 & -7.86469129e-05 & -5.67827690e-05 & 7.64963677e-05 \\ 1.97496655e-02 & 1.59364126e-04 \\ [-7.34936808e-04 & -5.94031755e-05 & 1.32812670e-04 & 3.88976744e-05 \\ 1.59364126e-04 & 1.42238957e-03 \end{bmatrix}]$

Figure 113

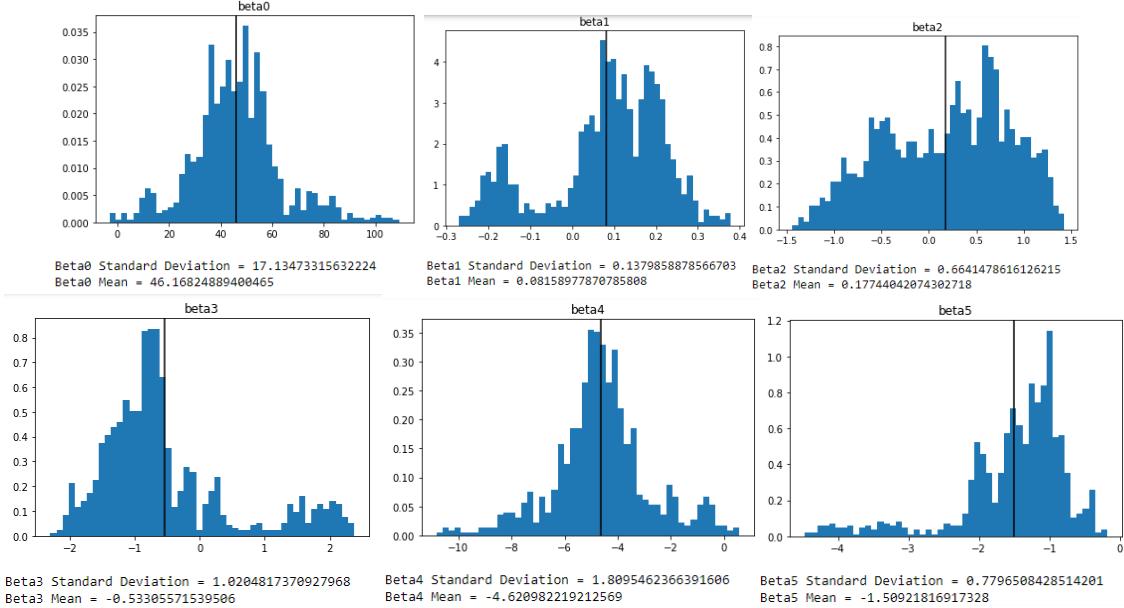


Figure 114

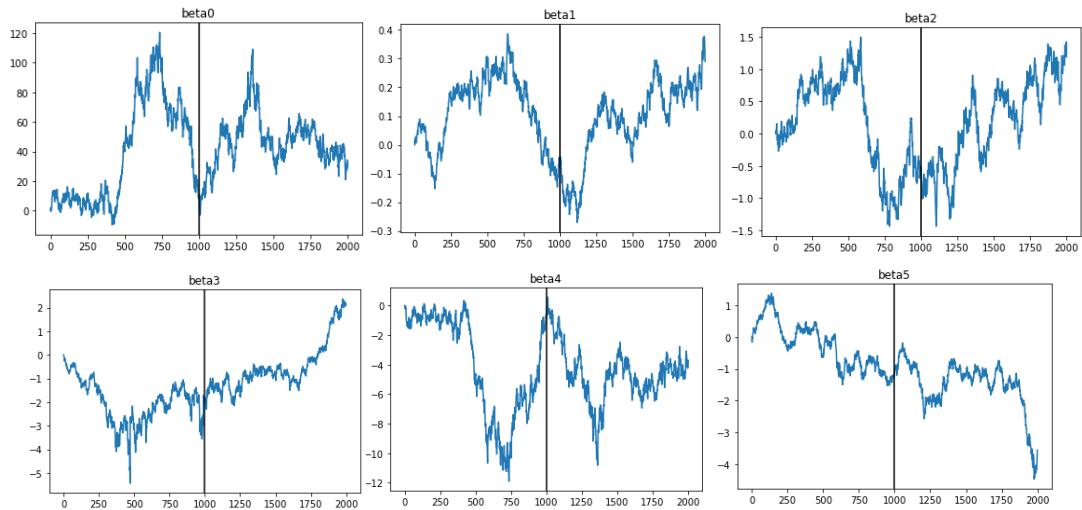


Figure 115

The MSE is 296.58294150061596

Figure 116

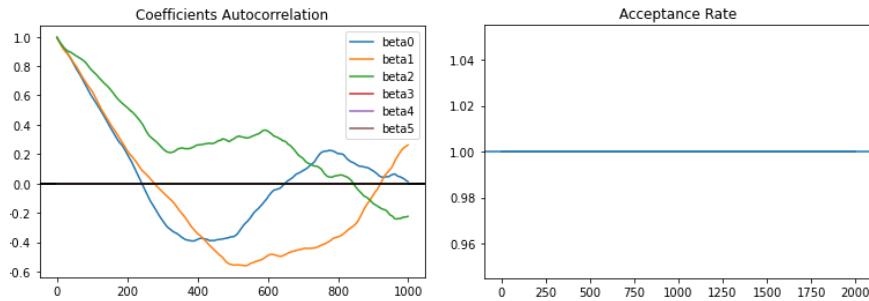


Figure 117 & Figure 118

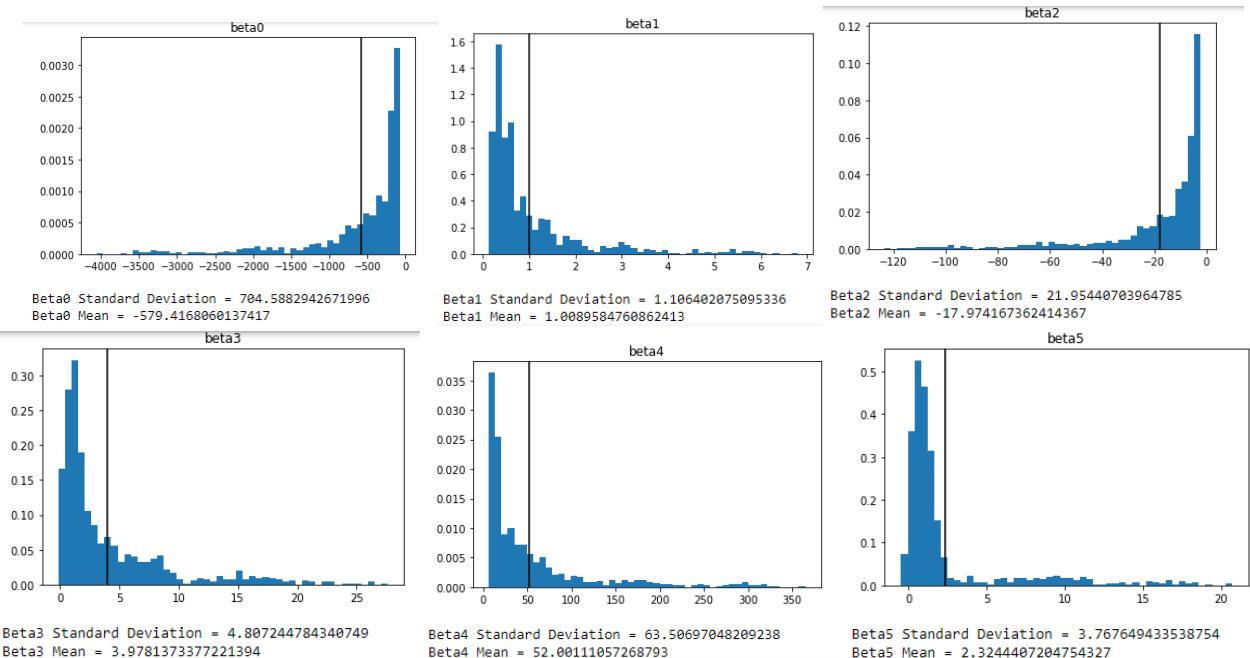


Figure 119

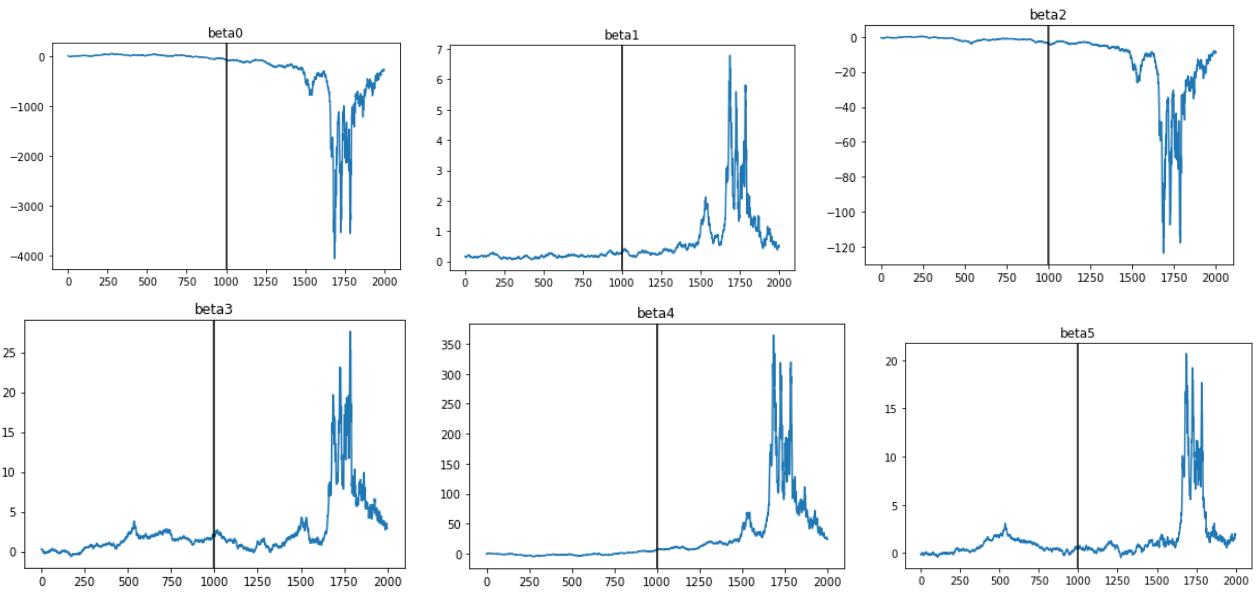


Figure 120

The MSE is 57285.44024720429

Figure 121

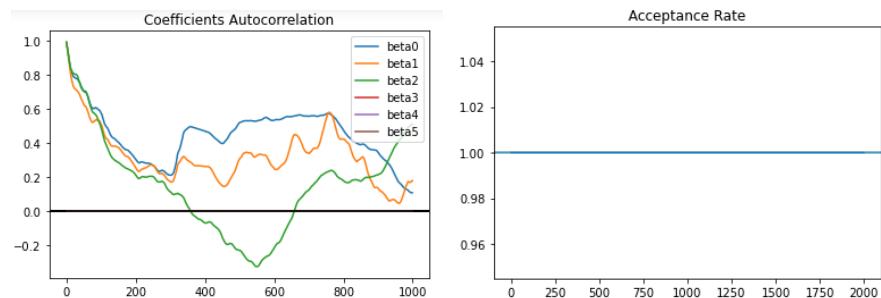


Figure 122 & Figure 123

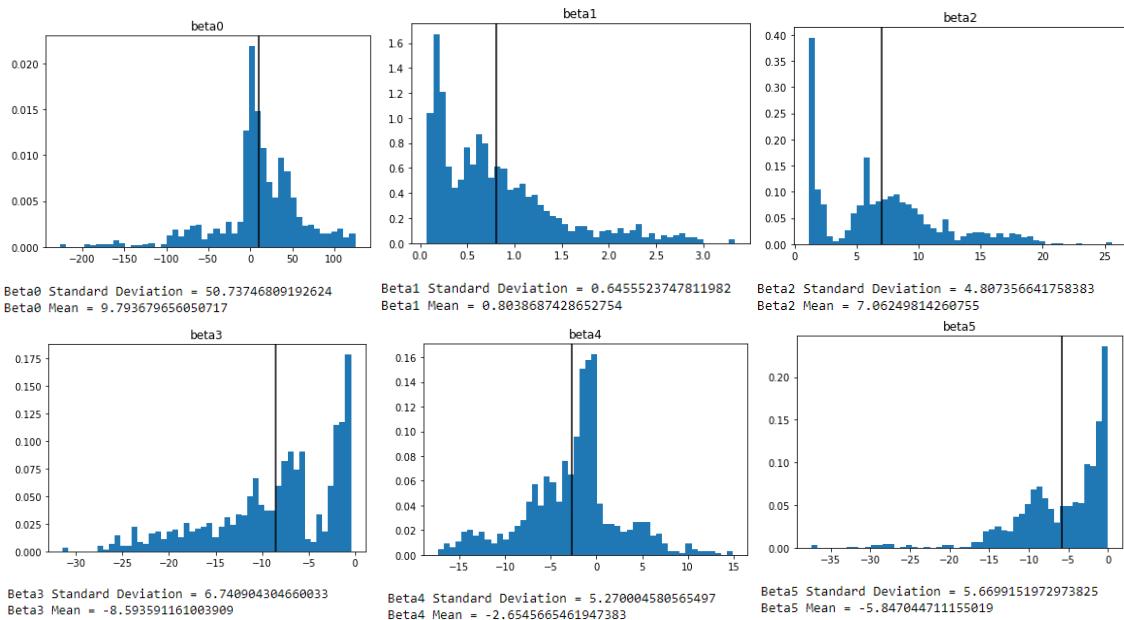


Figure 124

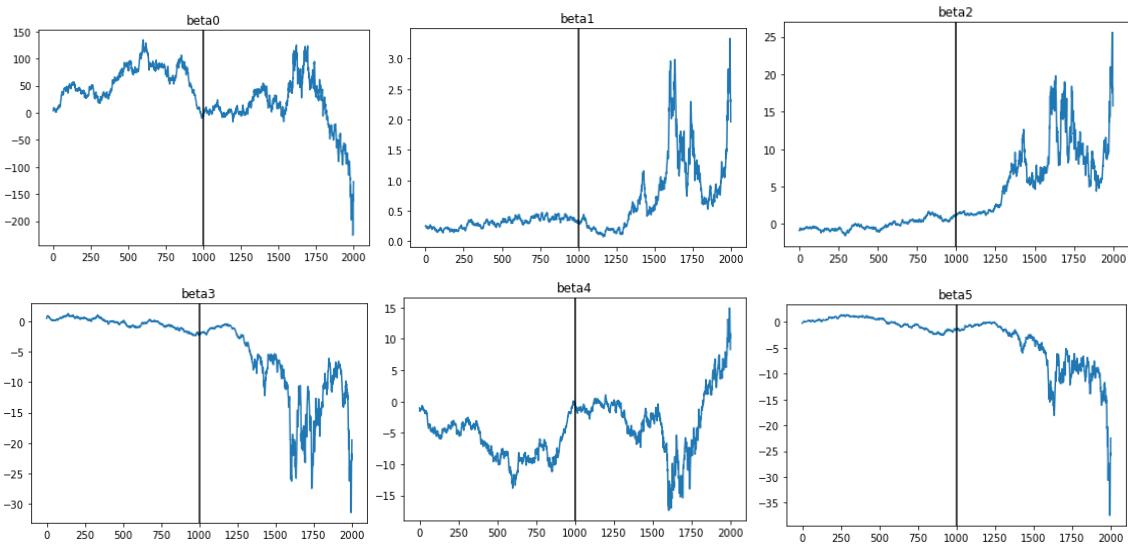


Figure 125

The MSE is 35.359537097721756

Figure 126

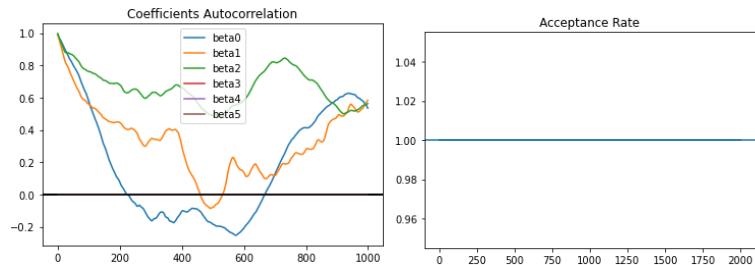


Figure 127 & 128

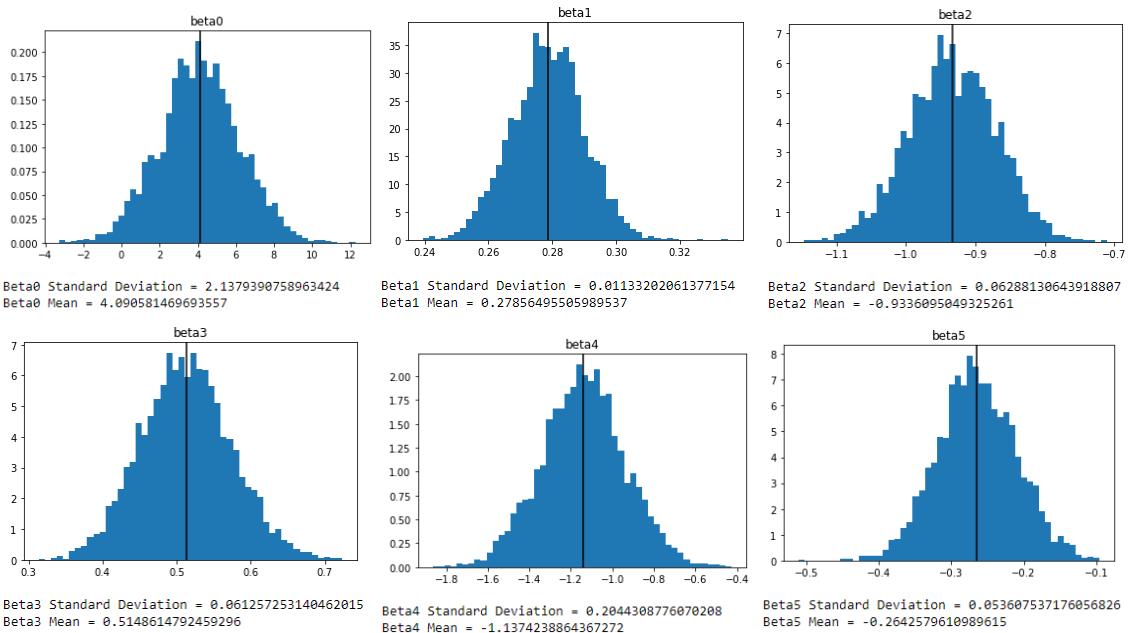


Figure 129

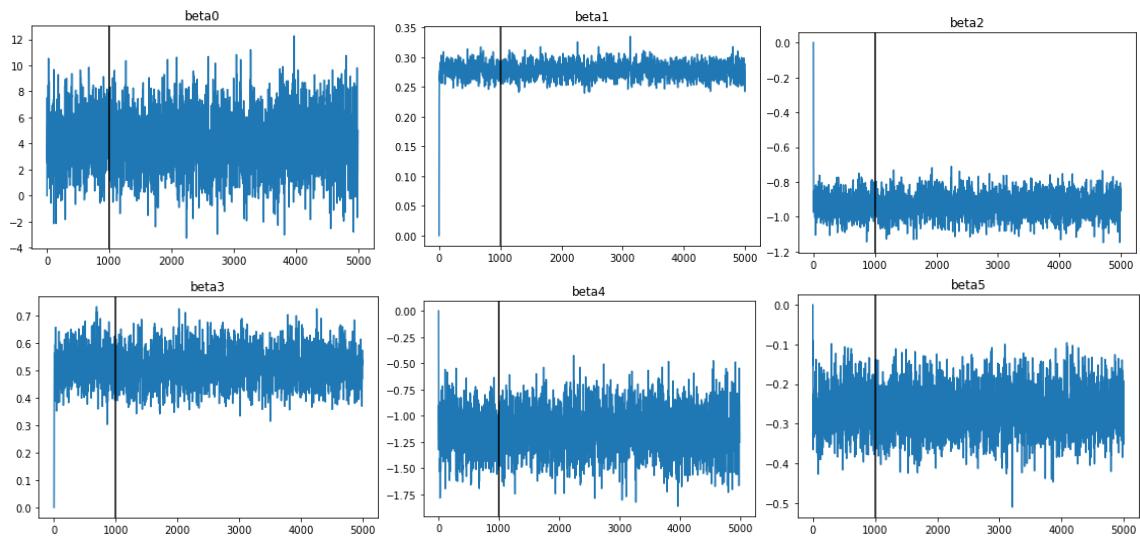


Figure 130

The MSE is 0.001178628072374522

Figure 131

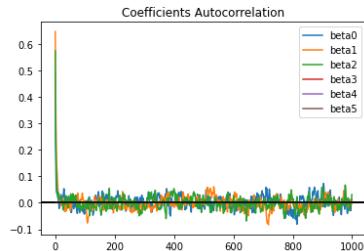


Figure 132

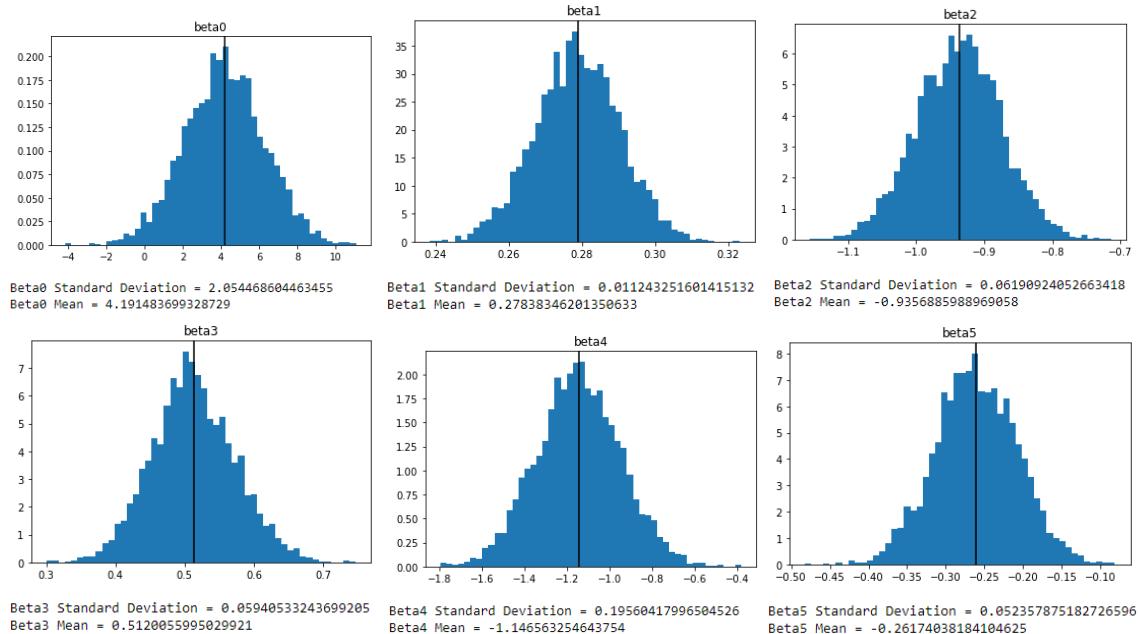


Figure 133

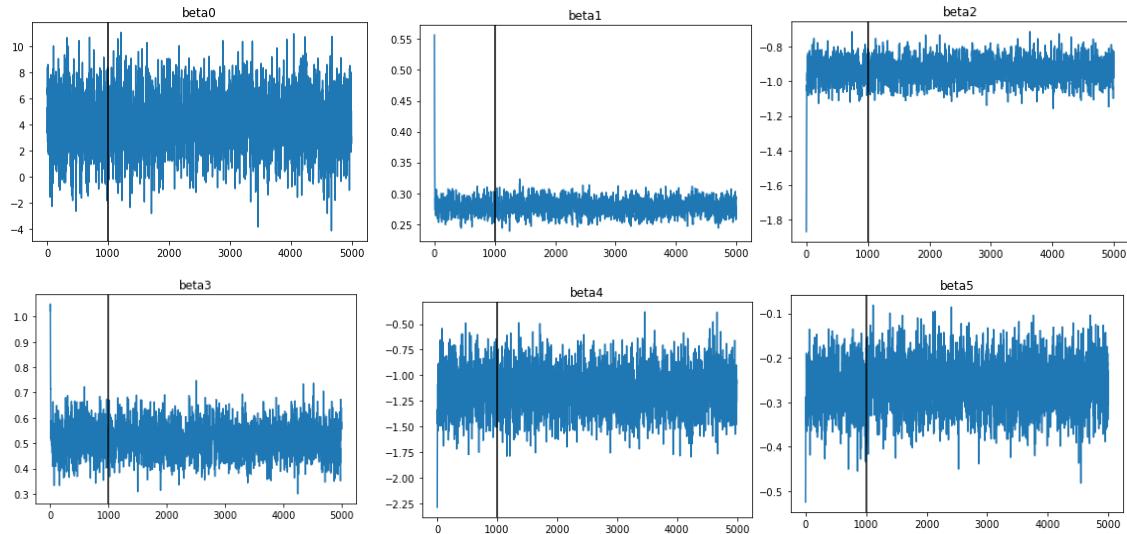


Figure 134

The MSE is 5.091900722627938e-05

Figure 135

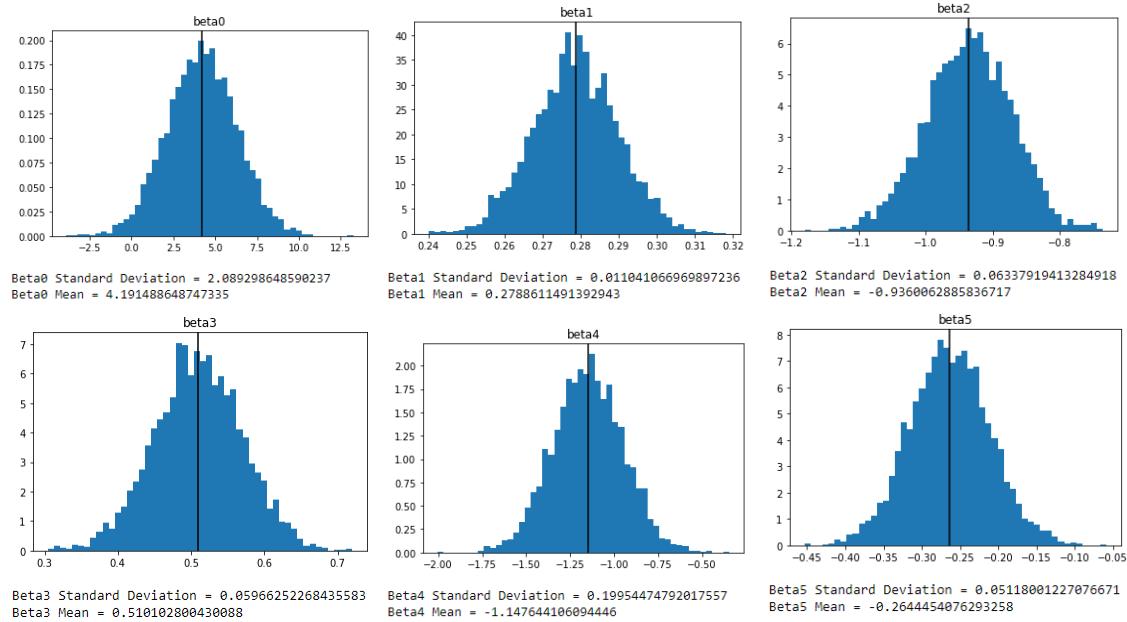


Figure 136

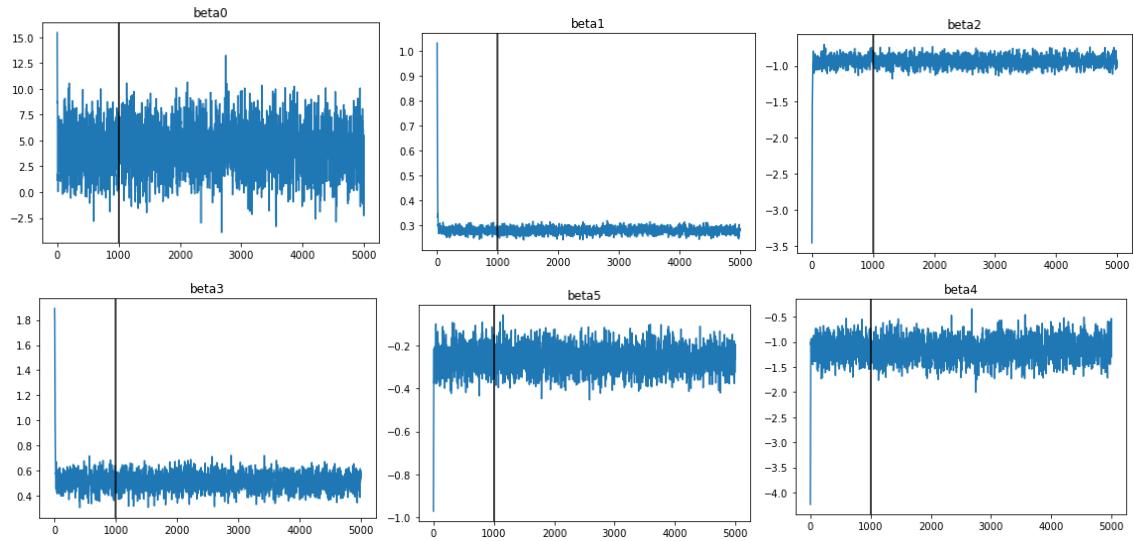


Figure 137

The MSE is 5.282899672695802e-05

Figure 138

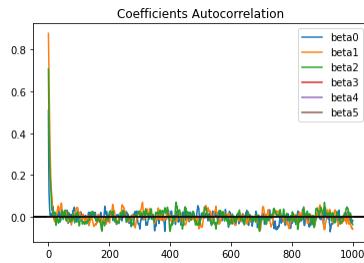


Figure 139

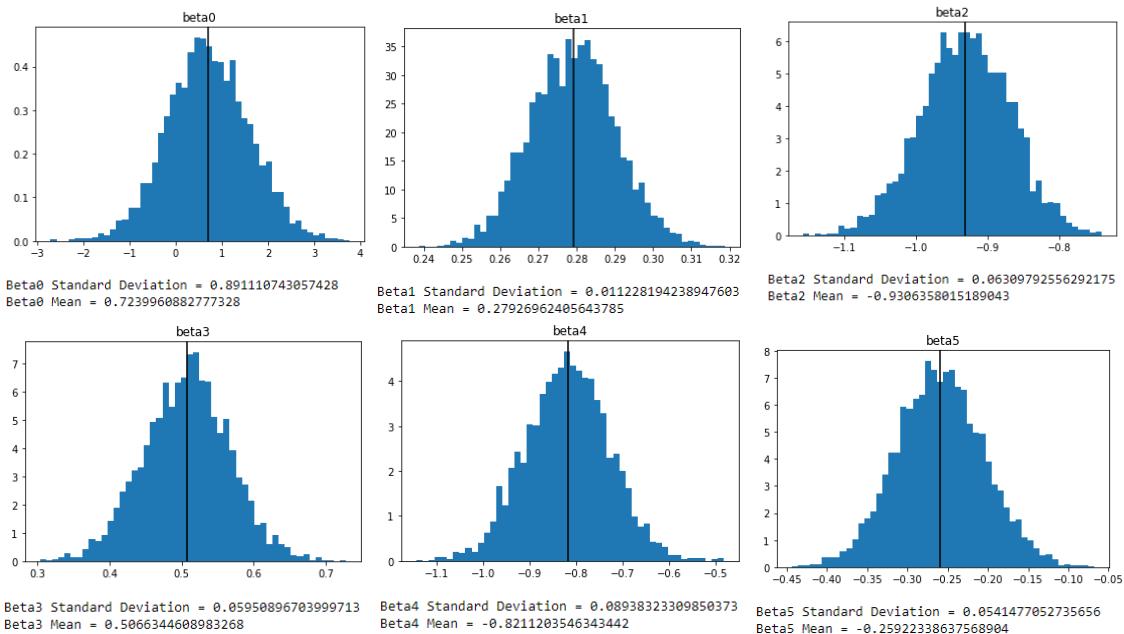


Figure 140

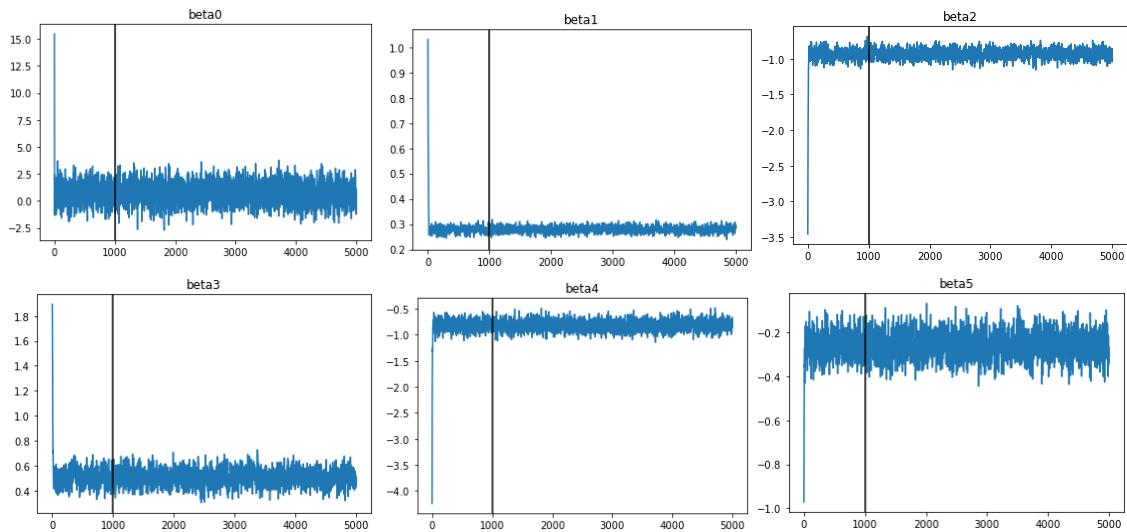


Figure 141

The MSE is 2.001515672453689

Figure 142

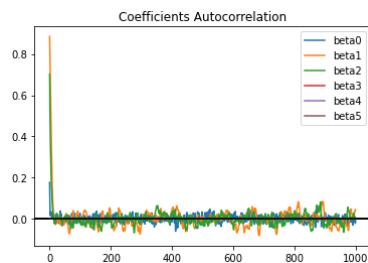


Figure 143

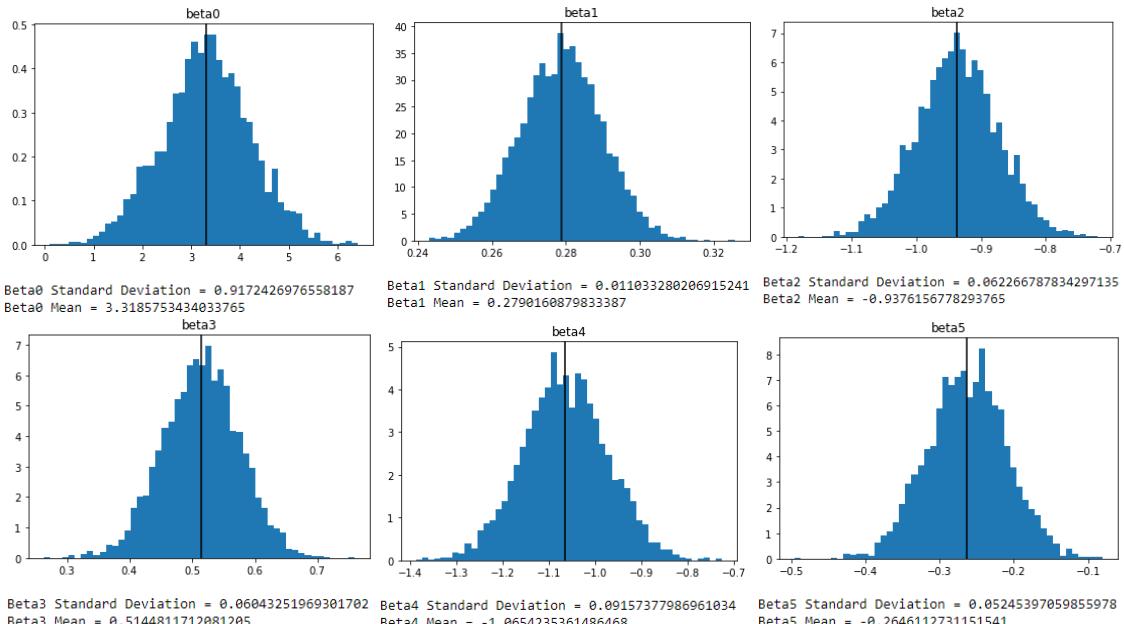


Figure 144

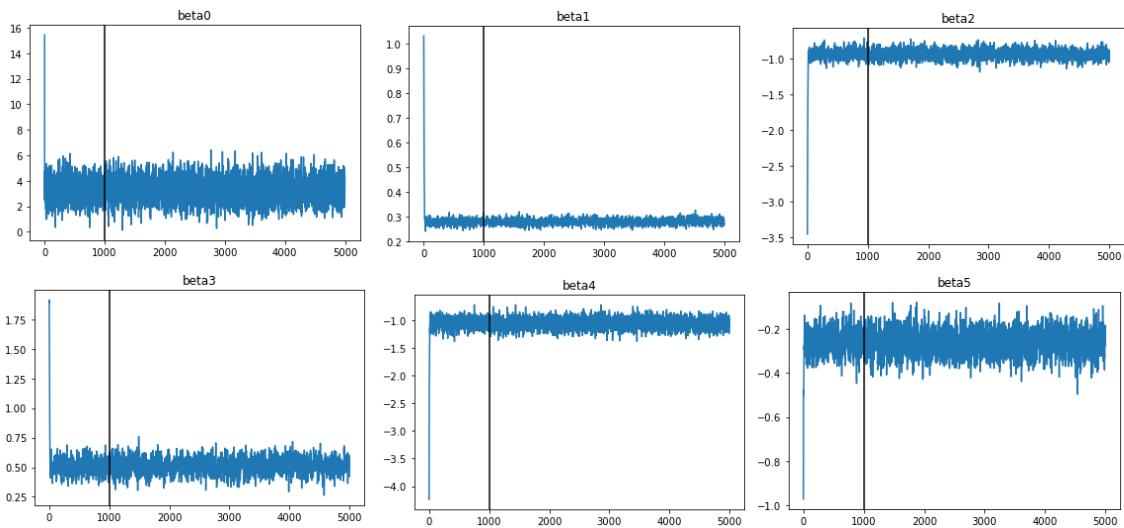


Figure 145

The MSE is 0.12308663501442085

Figure 146

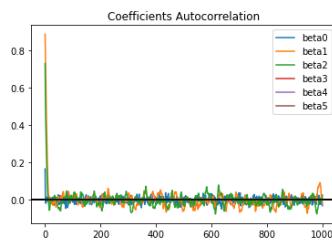


Figure 147