THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ECON 7320 Advanced Microeconometrics

Assignment IV

Student name: Chi-Sheng Lo
Student ID number: 44046918

Q1. Bayesian Updating

Given y|p, x ~ Bernoulli ($p_0(1-x)+\ p_1 x$) where p_0, p_1 are an unknown parameters and x is a binary random variable.

(i) Show that Prob[y = 1|x = 0] = p_0 and Prob[y=1|x=1] = p_1

Prob[y = 1|x = 0] ∝ L(x = 0|y = 1)P(y = 1) = p_0

Prob[y = 1|x = 1] ∝ L(x = 1|y = 1)P(y = 1) = p_1

Where: L(x = 0) = $(\text{x} = 0)^S(1 - (\text{x} = 0))^{N-S}$

Where: L(x = 1|) = $(\text{x} = 1)^S(1 - (\text{x} = 1))^{N-S}$

(ii) Write down a suitable parameter space for p_0, p_1.

P(y|x) = $\begin{cases} 0 \le p_1 \le 1 & \text{if } x = 1 \\ 0 \le p_0 \le 1 & \text{if } x = 0 \end{cases}$

(iii) Suppose that your prior $\pi(p_0, p_1)$ satisfies $\pi(p_0, p_1) \propto p_1 1(0 \le p_0 \le 1,\ 0 \le p_1 \le 1)$ where 1(.) is the indicator function. Find $\pi(p_0, p_1)$. Use $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\pi(p_0, p_1)\mathrm{d}p_0\mathrm{d}p_1$ = 1.

$\pi(p_0, p_1) \propto p_1 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)$

$\qquad = cp_1 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)$

Where c is the normalizing constant which does not depend on p_0, p_1.

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\pi(p_0, p_1)\mathrm{d}p_0\mathrm{d}p_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} cp_1 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)\,\mathrm{d}p_0\mathrm{d}p_1$

$\int_0^1 c\, p_1 1 dp_1$ = 1 and $c\frac{p_1^2}{2}|_0^1 = \frac{1}{2}c$ = 1 and c = 2

Therefore: $\pi\ (p_0, p_1) \propto 2p_1 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)$

(iv) Is the prior $\pi(p_0, p_1)$ consistent with the state "Having seen no data, I think all values of p_0, p_1 are equally likely", Justify your answer.

No, since we start with y|p, x ~ Bernoulli ($p_0(1-x)+\ p_1 x$) which does not guarantee that p_0, p_1 are equally likely at the first place. Recall that in part (iii), I showed that $\int_0^1 2\,1 dp_1 = 1 = \pi(p_0)$ and subsequently, $\int_0^1 2\ p_1\ 1 dp_0 = 2p_1 = \pi(p_1)$.

(v) Suppose one observation with y = x = 1.

Show that $\boldsymbol{\pi(p_0, p_1|}$y = 1, x = 1) = 3$\boldsymbol{p_1^2}$1(0 ≤ $\boldsymbol{p_0}$ ≤ 1, 0 ≤ $\boldsymbol{p_1}$ ≤ 1).

$\pi(p_0, p_1|$y =1, x = 1) $\propto p_1^2 1(0 \leq p_0 \leq 1, 0 \leq p_1 \leq 1)$.

$= cp_1^2 1(0 \leq p_0 \leq 1, 0 \leq p_1 \leq 1)$

Where c is the normalizing constant which does not depend on p_0, p_1.

Since it must integrate to 1, then

$\int_0^1 cp_1^2 \mathrm{d}p_1 = 1$

$c\frac{p_1^3}{3}|_0^1 = \frac{1}{3}c = 1$

$c = 3$

Therefore: $\pi(p_0, p_1|$y = 1, x = 1) = $3p_1^2 1(0 \leq p_0 \leq 1, 0 \leq p_1 \leq 1)$

(vi) Show that an equal tailed 0.95 credible interval for $\boldsymbol{p_0}$ is (0.025, 0.975).

Start by finding $\boldsymbol{\pi(p_0|}$y = 1, x = 1) = $\int_{-\infty}^{\infty}\boldsymbol{\pi(p_0, p_1|}$y = 1, x = 1)d$\boldsymbol{p_1}$.

Since: $\pi(p_0|$y = 1, x = 1) = $\int_{-\infty}^{\infty}\pi(p_0, p_1|$y = 1, x = 1)d$p_1$

$= \left[\int_{-\infty}^{\infty} f(y = 1,\ x = 1|\, p_0, p_1)\pi(p_1)\pi(p_0|p_1)dp_1\right] \div$ f(y)

$= \left[\int_{-\infty}^{\infty} f(y = 1,\ x = 1|\, p_1)\pi(p_1)\pi(p_0|p_1)dp_1\right] \div$ f(y)

$= \int_0^1 3p_1^2 \mathrm{d}p_1 = 1$

Thereafter:

$\int_0^a \pi(p_0|$y = 1, x = 1) $\pi((p_0|p_1)$dp_0 = (a – 0) = a = $\frac{0.05}{2}$ = 0.025

b = 1 − 0.025 = 0.075

The 95% credible interval for p_0 is:

$\left[\widehat{(p_0)}_B - 1.96\sqrt{V(p_0|y = 1, x = 1}, \widehat{(p_0)}_B\ +\ 1.96\sqrt{V(p_0|y = 1, x = 1}\ \right]$ = [0.025, 0.975].

(vii) Suppose observe another 99 independent observations verifying

	y = 0	**y = 1**
x = 0	30	20
x = 1	40	9

Giving you a total sample size of 100, compute the posterior $\pi(p_0, p_1|100$ observations) up to its constant proportionality.

Final answer should take the form: $\pi(p_0, p_1|100$ observations) ∝ …

$\pi(p_0, p_1|100 \text{ observations}) \propto \pi(100 \text{ observations}|\,(p_0, p_1))\, \pi(p_0, p_1)$

$$\propto p_0^{20} p_1^{10}(1-p_0)^{30}(1-p_1)^{40} 2p_1 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)$$

$$\propto p_0^{20} p_1^{(20-9)}(1-p_0)^{30}(1-p_1)^{40} 1(0 \le p_0 \le 1, 0 \le p_1 \le 1)$$

Q2. Non-Parametric Density Estimation

Consider Kernel density estimation of f(y), which is known to be flat for – 0.1 ≤ y ≤ 0.1. You have a sample of size N = 100 with sample standard deviation 1.3628, and sample interquartile range 2.5. Exactly fifty of the observations fall in the interval (– 1, 1).

(i) Given the Kernel: K(z) = $\frac{1}{2}$1(|z| < 1), explain why a good choice of the bandwidth is h = 1.

By using Silver's rule which can be illustrated as:

$\widetilde{h^*} = 1.3643\delta N^{-1/5}\min\left\{S,\ \frac{\widehat{q_{.75}} - \widehat{q_{.25}}}{1.349}\right\} = 0.7338 \times \min\left\{S,\ \frac{\widehat{q_{.75}} - \widehat{q_{.25}}}{1.349}\right\} \approx 1$

Where N = 100, S = 1.3628, δ = 1.3510, $\frac{\widehat{q_{.75}} - \widehat{q_{.25}}}{1.349}$ = 1.8532.

This optimal h = 1 strikes the balance between variance and bias.

(ii) Does the bandwidth chosen by Silverman's rule satisfy the asymptotic criteria for pointwise consistency of the Kernel density estimator (KDE)?

It will satisfy the asymptotic criteria for pointwise consistency of this uniform Kernel density estimator as N gets larger. Firstly, the bias is elevating in h and disappears asymptotically if h → 0 as N → ∞. Secondly, the variance is falling in h and disappears asymptotically if Nh → ∞ as N → ∞. Altogether, the KDE is said to be pointwise consistent as $\hat{f}(y) \xrightarrow{P} f(y)$. Moreover, to achieve uniform consistency, we require $\sup_y \left|\hat{f}(y) - f(y)\right| \xrightarrow{P} 0$ which occurs when Nh/lnN → ∞.

(iii) Using the Kernel and bandwidth from part (i), show that the Kernel density estimator is $\hat{\boldsymbol{f}}$(0) = 0.25.

$\hat{f}(0) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}\mathrm{K}\left(\frac{y-y_i}{h}\right) = \frac{1}{100} \times \frac{1}{1} \times 50 \times \frac{1}{2} \times 1 = 0.25$

The KDE smoothes each data point y_i into a small density bumps; subsequently, the KDE sum up all of these small bumps together to obtain the final density estimate. K decides the shape of the bump and h indicates the width of the bump.

(iv) Explain why $\hat{\boldsymbol{f}}$(0) is unbiased.

$\text{Bias(y)} = \mathrm{E}\big(\hat{f}(\mathrm{y})\big) - \mathrm{f(y)} = \frac{1}{2}h^2 f''(\mathrm{y}) \int z^2 K(z)dz + 0h^2 = 0$

Where $f''(\mathrm{y}) = f''(0) = 0$; therefore, unbiased.

Theoretically, the bias of the KDE is caused by the second derivative of the density function. Moreover, as h → 0, the bias contracts at a rate of $0h^2$.

(v) Use the asymptotic distribution of $\hat{\boldsymbol{f}}$(0) to show that $\widehat{\boldsymbol{V}}[\hat{\boldsymbol{f}}(0)]$ = 0.00125.

$\hat{V}[\hat{f}(\mathrm{y})] = \frac{1}{Nh}h^2\mathrm{R(K)f(y)} + 0\frac{1}{Nh} = \frac{1}{Nh}\mathrm{f(y)}\int K(z)^2 dz = 0.00125$

Where $\mathrm{R(K)} = \int K(z)^2 dz = 0.5$, f(0) = 0.25, and Nh = 100

Since point wise consistent: $\hat{f}(\mathrm{y}) \xrightarrow{P} f(\mathrm{y})$

To obtain uniform consistency: $\sup_{y}\left|\hat{f}(\mathrm{y}) - f(\mathrm{y})\right| \xrightarrow{P} 0$

The limit distribution: $\sqrt{NH}\big(\hat{f}(\mathrm{y}) - f(\mathrm{y}) - \mathrm{Bias(y)}\big) \xrightarrow{d} \mathrm{N}(0, f(y)\int K(z)^2 dz)$

(vi) Construct a 0.95 confidence interval for f(0).

$$\text{95\% CI for f(0)} = \left[\hat{f}(0) - 1.96\sqrt{\hat{V}[\hat{f}(0)]}, \hat{f}(0) + 1.96\sqrt{\hat{V}[\hat{f}(0)]}\right]$$

$$= [0.25 - 1.96(0.03536), 0.25 + 1.96(0.03536)]$$

$$= [0.25 - 1.96(0.03536), 0.25 + 1.96(0.03536)]$$

$$= [0.1807, \ 0.3193]$$

Q3. Data Analysis on Bayesian Regression

(i) Obtain the least squares estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma^2}$ and report your results.

Source	SS	df	MS
Model	9041.60857	1	9041.60857
Residual	1025.13493	998	1.02718931
Total	10066.7435	999	10.0768203

Number of obs	=	1,000
F(1, 998)	=	8802.28
Prob > F	=	0.0000
R-squared	=	0.8982
Adj R-squared	=	0.8981
Root MSE	=	1.0135

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	2.993721	.031909	93.82	0.000	2.931105	3.056338
_cons	10.00444	.0320792	311.87	0.000	9.941491	10.06739

Comment:

$\widehat{\beta_1}$ = 2.9937, $\widehat{\beta_0}$ = 10.0044, $\widehat{\sigma^2}$ = 1.0272. In this regression, we reject the hypothesis of H_0: β_1 = 0 at five percent significance level based on the p-value of 0.000. Moreover, the 95 percent confidence interval for the coefficient for x is [2.93, 3.06] which does not contain the value of zero indicating that x is a significant predictor of y.

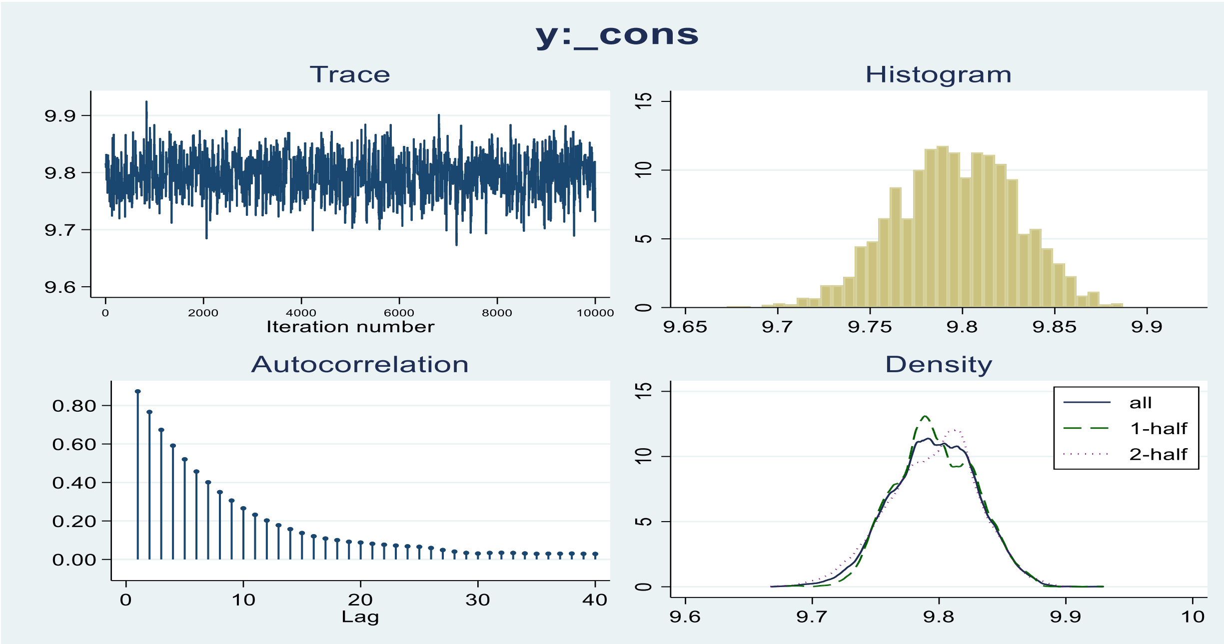
(ii) Perform a Bayesian regression to estimate the posterior distribution for $\boldsymbol{\theta} = (\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\sigma^2})$. Report and explain how it relates to the priors used and to part (i).

Bayesian normal regression
Random-walk Metropolis-Hastings sampling

MCMC iterations	=	12,500
Burn-in	=	2,500
MCMC sample size	=	10,000
Number of obs	=	1,000
Acceptance rate	=	.2218
Efficiency: min	=	.05949
avg	=	.07816
max	=	.1135

Log marginal likelihood = -2507.0241

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
y						
x	2.940083	.031821	.000944	2.939559	2.878487	3.001799
_cons	9.796056	.0331826	.001338	9.796849	9.729622	9.85743
var	1.075796	.050838	.002084	1.076214	.9804819	1.180232



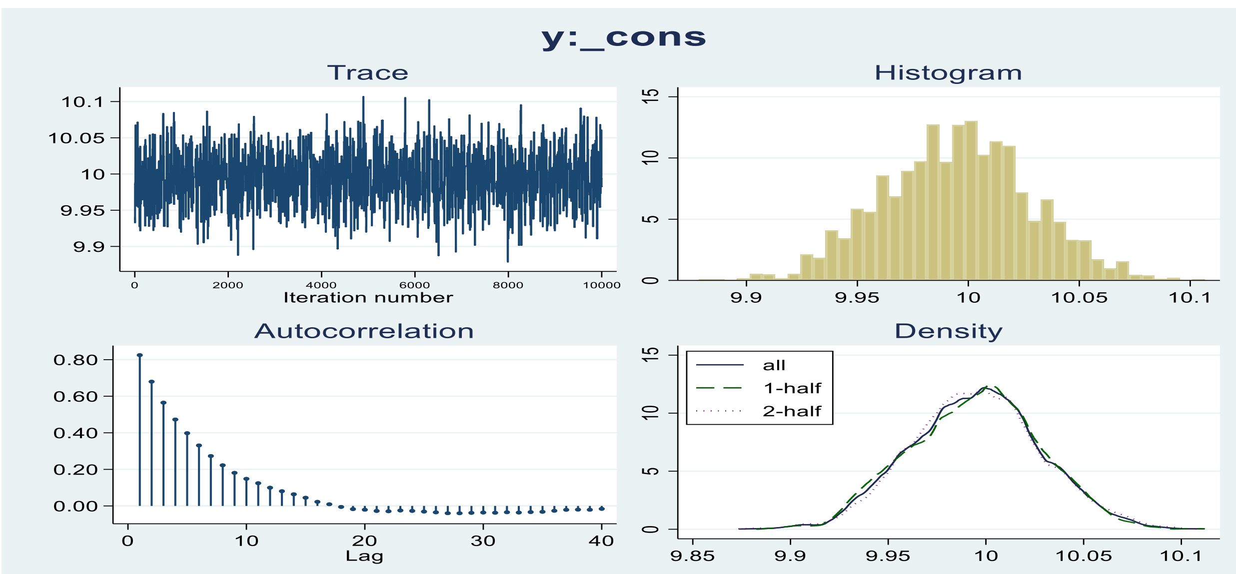
Comment:

This Bayesian estimate has posterior distribution that is described by the prior distribution of parameter and the likelihood function of the data provided in part (i). The mean column of the table above shows the means of the marginal posterior distributions of the parameters. In this setting, both priors for β_0 and β_1 are N(0, 0.05); consequently, the posterior mean estimate for x is slightly different from the coefficient of OLS estimation in part (i). Furthermore, an acceptance rate of 22.18% suggests that convergence problem should not be a concern. The average efficiency is moderately good. From the diagnostic plots for {y:_cons}, the trace plot exhibits stationary process. Additionally, both histogram and kernel density represent normal priors. The autocorrelation can be nearly phased out after lag 30.

(iii) Repeat part (ii) replacing the priors for β_0 and β_1 with N(0, 1). Comment on any differences in the posterior distribution of θ relative to parts (i) – (iii).

```
Bayesian normal regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      1,000
                                                 Acceptance rate  =      .1954
                                                 Efficiency:  min =      .0745
                                                              avg =     .08498
Log marginal likelihood = -1495.8829                          max =     .09936
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
y						
x	2.990714	.031502	.001106	2.990677	2.929507	3.046792
_cons	9.994692	.0331959	.001053	9.99524	9.93155	10.05873
var	1.029765	.045624	.001672	1.027269	.9461358	1.124587



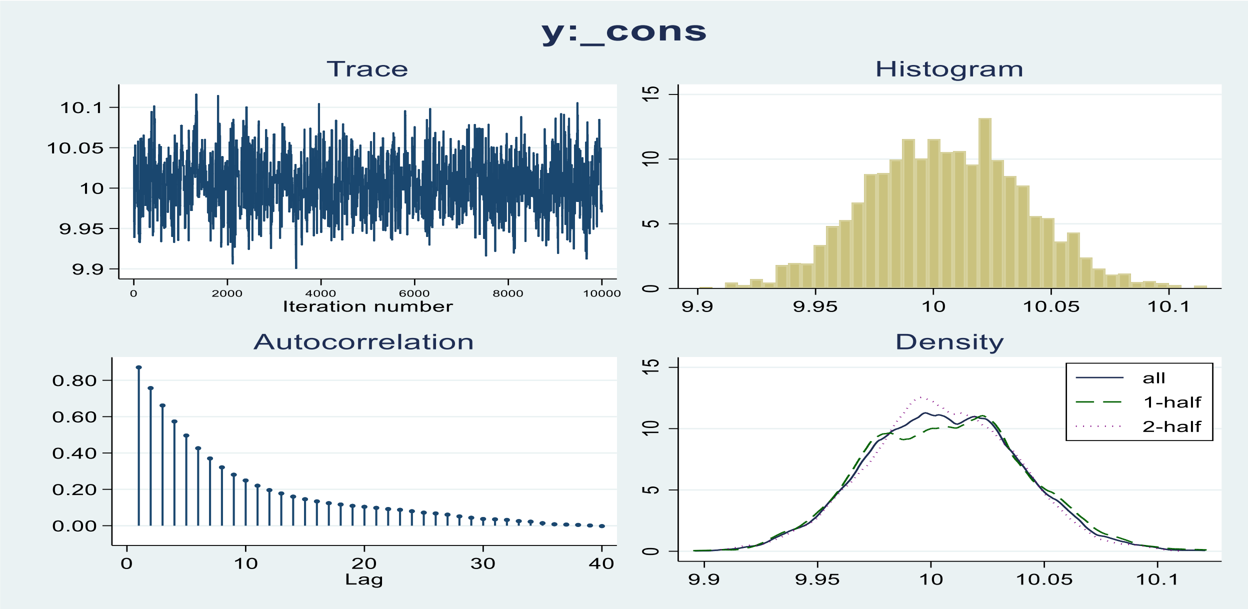
Comment:

The posterior mean estimates for x is very close to the OLS estimation in part (i) and a little far from that in part (ii). This phenomenon suggests that the prior is less informative. In addition, the Monte Carlo standard error (MCSE) that indicates the precision of the posterior mean estimate for x is slightly larger (less accurate) than that in part (ii). The acceptance rate increased slightly but still indicates no convergence problem since it is above 10%. The efficiency has improved compared to part (ii). From the diagnostic plots for {y:_cons}, the trace plot exhibits Gaussian white noise. Both histogram and kernel density resemble normal priors. The autocorrelation will be gone after lag 20.

(iv) Repeat part (ii) replacing the priors for $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ with uninformative (flat) priors. Comment on any differences in the posterior distribution of $\boldsymbol{\theta}$ relative to parts (i) – (iii).

```
Bayesian normal regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      1,000
                                                 Acceptance rate  =      .2088
                                                 Efficiency:  min =     .06398
                                                              avg =      .0824
Log marginal likelihood = -1439.5107                          max =      .1002

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
y            |
           x |  2.994356   .0326351   .001133   2.993478   2.930909   3.061675
       _cons |  10.00564   .0333214   .001317   10.00489   9.940405   10.07149
-------------+----------------------------------------------------------------
         var |  1.028363   .0470476   .001486   1.026529   .9388758   1.121281
------------------------------------------------------------------------------
```



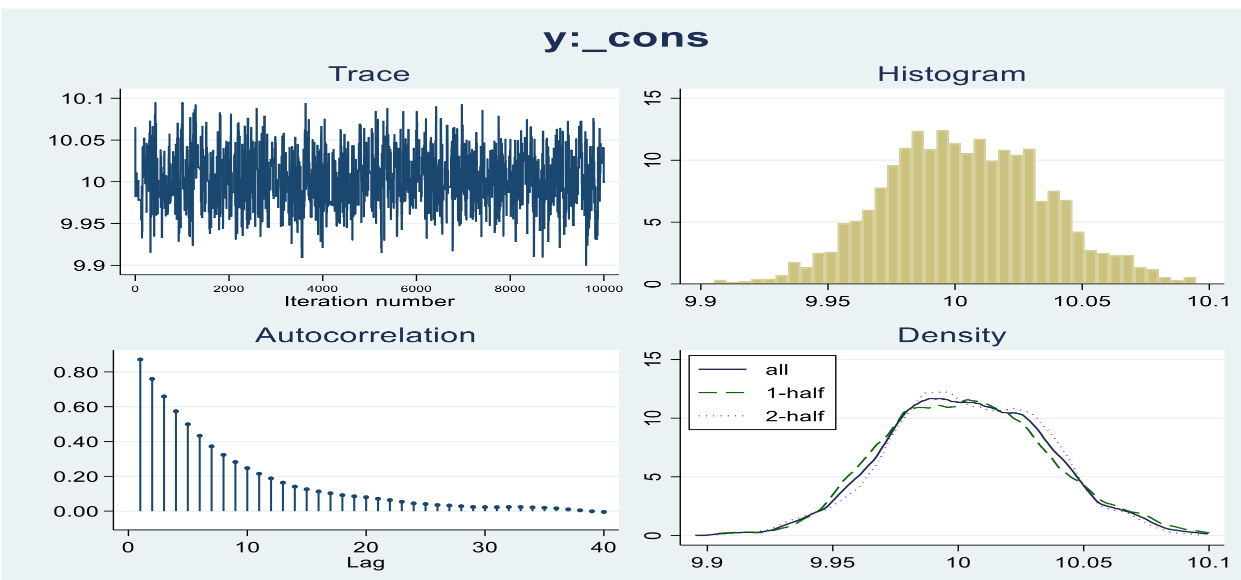
Comment:

Since part (iv) uses uninformative (flat) priors for β_0 and β_1, this posterior mean estimate of Bayesian regression returns highly closed result with the coefficients generated in OLS regression and is still statistically significant. Meanwhile, the MCSE for x is similar to that in part (iii). The acceptance rate is the best among all compared to parts (ii) and (iii). The efficiency is slightly lower than that in part (iii) and higher than that in part (ii). From the diagnostic plots for {y:_cons}, the trace plot still shows mean reversion but more volatile than previous parts. Both histogram and kernel density reflect flat priors. The autocorrelation can be removed after lag 40.

(v) Repeat part (ii) replacing the prior for β_1 with Uniform (3, 10) and using a flat prior for β_0. Comment on any differences in the posterior distribution of θ relative to parts (i) – (iv).

```
Bayesian normal regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      1,000
                                                 Acceptance rate  =      .1979
                                                 Efficiency:  min =     .05938
                                                              avg =     .06353
Log marginal likelihood = -1442.1373                          max =      .0677
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
y						
x	3.022638	.0179178	.000735	3.018846	3.000501	3.065718
_cons	10.00368	.0317602	.001221	10.00307	9.943369	10.06742
var	1.028522	.0464207	.001842	1.027493	.9434659	1.125309



Comment:

$\widehat{\beta_0}$ tracks closely to that in part (i) since part (v)'s prior for β_0 is flat. Compare with other previous parts, this posterior mean estimate for x is the largest among all. Also, it has the smallest MCSE (best accuracy) and results in the thinnest 95 percent credible interval among all. The acceptance rate is still similar to those in previous parts. Nonetheless, the efficiency can be a concern which is noticeably low and is the lowest among all. From the diagnostic plots for {y:_cons}, the trace plot still shows no clear trend. Both histogram and kernel density reflect flat priors. The autocorrelation can be nearly terminated after lag 40.

(vi) Explain why uniform priors such as that in part (v) might lead to posterior inconsistency, whereas normal priors such as those in part (ii) would not.

The posterior inconsistency can be caused by misspecification due to nonconglomerability or improper priors on β_1 and β_0. Having said that, the Bayes estimators of particularly the uniform priors, in part (v), are still within the boundary of the true parameter; therefore, there is no problem in posterior inconsistency. In the contrast, part (ii) uses normal priors for both β_0 and β_1; as a result, the Bayes estimators are always admissible due to this proper priors and can remain consistent when the sample size gets larger.

Appendix: STATA code of Q3

```
set more off
log using ECON7320A4Q3CSLO.log, replace
cd "D:\UQ\2020 S1\ECON7320 Advanced MicroMetrics\ECON7320 Assignment 4"
use data.dta
//Q3-1//
reg y x
//Q3-2//
set seed 7339
bayesmh y x, likelihood(normal({var})) ///
prior({y:x}, normal(0,0.05)) ///
prior({y:_cons},normal(0,0.05)) ///
prior({var},igamma(1,1))
bayesgraph diagnostics {y:_cons}
//Q3-3//
set seed 7339
bayesmh y x, likelihood(normal({var})) ///
prior({y:x}, normal(0,1)) ///
prior({y:_cons},normal(0,1)) ///
```

```
prior({var},igamma(1,1))
bayesgraph diagnostics {y:_cons}
//Q3-4//
set seed 7379
bayesmh y x, likelihood(normal({var})) ///
prior({y:x}, flat) ///
prior({y:_cons},flat) ///
prior({var}, igamma(1,1))
bayesgraph diagnostics {y:_cons}
//Q3-5//
set seed 7379
bayesmh y x, likelihood(normal({var})) ///
prior({y:x}, uniform(3,10)) ///
prior({y:_cons},flat) ///
prior({var},igamma(1,1))
bayesgraph diagnostics {y:_cons}
log cl
```