UNIVERSITY OF VICTORIA | **EXAMINATIONS, APRIL 2016**

ECONOMICS 546: THEMES IN ECONOMETRICS

TO BE ANSWERED IN BOOKLETS

DURATION: 3 HOURS
INSTRUCTOR: D. Giles

STUDENTS MUST COUNT THE NUMBER OF PAGES IN THIS EXAMINATION PAPER BEFORE BEGINNING TO WRITE, AND REPORT ANY DISCREPANCY IMMEDIATELY TO THE INVIGILATOR.

THIS QUESTION PAPER HAS 10 PAGES.
STATISTICAL TABLES AND A FORMULA SHEET ARE SUPPLIED SEPARATELY.

This is a "closed book/closed notes" examination.
Calculators may be used.

(*Total Marks = 90*)

SECTION A
Answer ANY FOUR QUESTIONS FROM THIS SECTION
(*Extra questions that are attempted will not be graded)*

ALL QUESTIONS ARE WORTH 16 MARKS

Question 1:

(a) Discuss the relative advantages and disadvantages of Maximum Likelihood estimation, and Bayesian estimation.

10 marks

(b) What are the conditions under which Bayes estimators and MLE's will coincide?

6 marks

Question 2:

Suppose that we have a random sample of n observations from a Pareto distribution, with a ***known*** location parameter, y_m, and an ***unknown*** shape parameter, k. That is, the density function for an individual observation is:

$$p(y_i \mid k, y_m) = k\, y_m^k / [y_i^{k+1}] \quad ; \quad y_m < y_i < \infty \quad ; \quad k > 0 .$$

(a) As a Bayesian, suppose that I decide to represent my prior uncertainty about k with a prior density which is Gamma, with a shape parameter, α (> 0), and a scale parameter, θ (> 0). That is:

$$p(k) = k^{\alpha-1} e^{-k/\theta} / [\theta^{\alpha} \Gamma(\alpha)] \quad ; \qquad k > 0 .$$

$\Gamma(\alpha)$ is a Gamma function. (It is just a *constant* once we assign a value to α). The mean of this distribution is $(\alpha\ \theta)$, its variance is $(\alpha\ \theta^2)$. Its mode is at $[(\alpha\ \text{-}1)\theta]$, if $\alpha > 1$.

Show that the posterior density for k is also Gamma, but with a shape parameter which is $(n + \alpha)$, and a scale parameter which is $[\theta^{-1} + \sum_{i=1}^{n} \log_e(y_i / y_m)]^{-1}$.

7 marks

(b) What is the Bayes estimator of k if I have a quadratic loss function? What is the Bayes estimator of k if I have a zero-one loss function?

2 marks

(c) Show that the Maximum Likelihood Estimator (MLE) of k is $\tilde{k} = [n / \sum_{i=1}^{n} \log_e(y_i / y_m)]$.

4 marks

(d) Show that the Bayes estimator you obtained in part (b) above under quadratic loss collapses to the MLE for k if $\theta \to \infty$ and $\alpha \to 0$. Why does this result make sense intuitively? (What is happening to the prior density in this situation?)

3 marks

Question 3:

Consider the Negative Binomial distribution for a random variable, Z, where z denotes the number of failuress before the α'th. success in a sequence of Bernoulli trials. So, the mass function for Z is:

$$p(z \mid \theta, \alpha) = \Gamma(\alpha + z) / [z! \Gamma(\alpha)] \theta^{\alpha} (1-\theta)^{z} \quad ; \quad z = 0, 1, 2, ...$$

where α is <u>known</u>, and $0 \le \theta \le 1$. It can be shown that $E[y_i] = \alpha(1-\theta)/\theta$.

In answering this question you may wish to use the following result. If a random variable, X, follows a Beta distribution with parameters a, $b > 0$, then the kernel of its density function is

$$p(x \mid a, b) \propto x^{a-1}(1-x)^{b-1} \quad ; \quad 0 \le x \le 1$$

and $E(X) = a/(a+b)$. Also, $Mode(X) = (a-1)/(a+b-2)$; if a, $b > 1$

(a) Show that Jeffreys' prior p.d.f. for θ is

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-1/2} \quad ; \quad 0 \le \theta \le 1$$

5 marks

(b) Using Jeffreys' prior, obtain the Bayes estimator of θ, (i) if the loss function is quadratic; and (ii) if the loss function is "zero-one".

4 marks

(c) Show that the Natural Conjugate prior for θ is a Beta density.

3 marks

(d) Using the Natural Conjugate prior, obtain the Bayes estimator of θ, (i) if the loss function is quadratic; and (ii) if the loss function is "zero- one".

4 marks

Question 4:

Suppose that we have a random sample of size '*n*' from an Exponential distribution. The associated density function is:

$$p(y \mid \theta) = \theta^{-1} \exp[-y/\theta] \quad ; \quad 0 < y < \infty \quad ; \quad 0 < \theta < \infty$$

Also, $E[Y] = \theta$; and $Var[Y] = \theta^2$.

(a) Show that the MLE for θ is $\tilde{\theta} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

3 marks

(b) Consider the following alternative estimator of θ:

$$\hat{\theta} = n\bar{y}/(n+1) .$$

Show that the MLE, $\tilde{\theta}$, is ***inadmissible*** if we have a quadratic loss function.

[**Hint:** Recall that if the loss is quadratic, then Risk = MSE.]

7 marks

(c) Using a particular prior for θ, and a quadratic loss function, the following Bayes estimator was obtained:

$$\theta^* = (\beta^{-1} + n\bar{y})/(\alpha + n - 1) ,$$

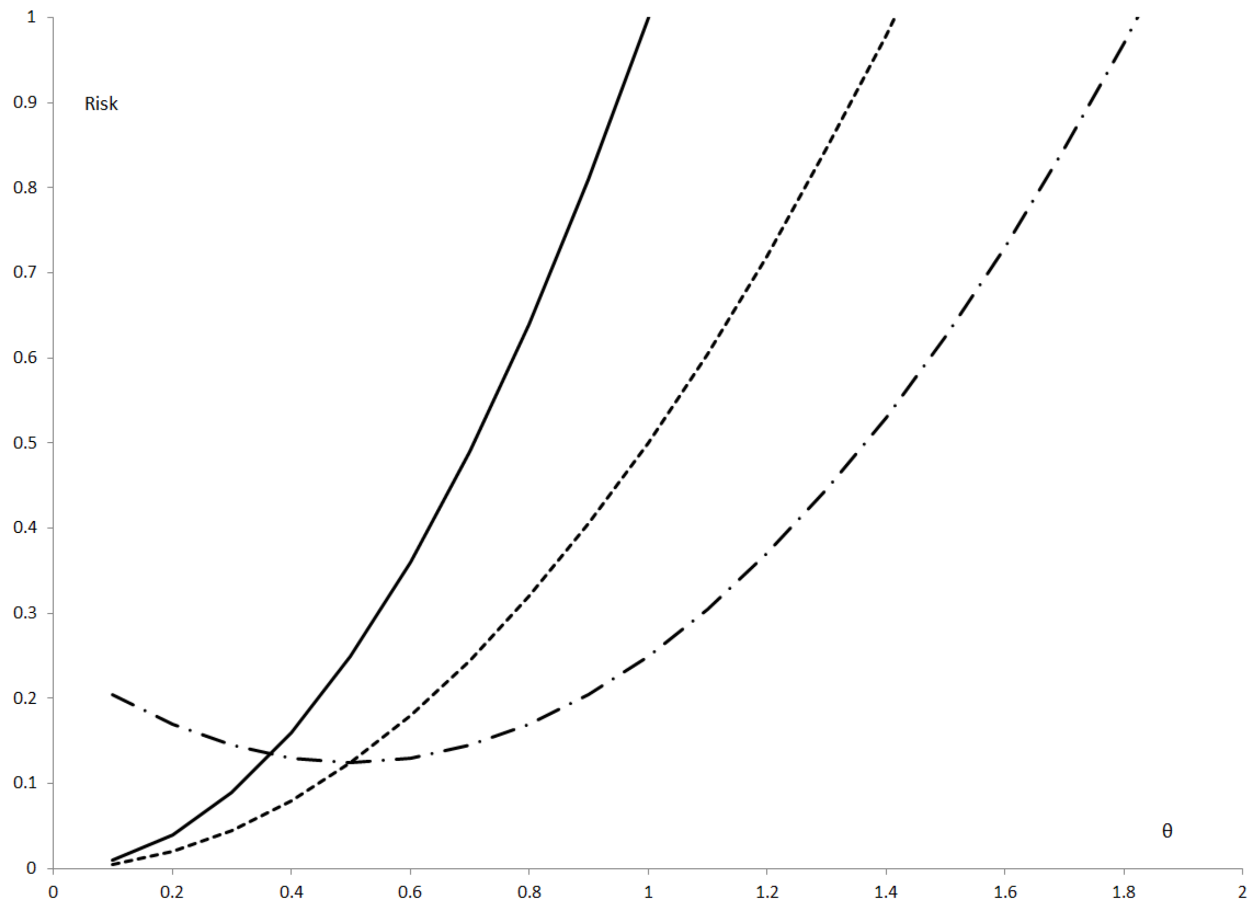
where α and β are the (known) parameters of the prior distribution.

Show that the risk of θ^* (under quadratic loss) is :

$$R(\theta^*) = [(n\theta^2) + (\beta^{-1} + (1-\alpha)\theta)^2]/[\alpha + n - 1]^2 .$$

4 marks

(d) The following diagram shows the risks of $\tilde{\theta}, \hat{\theta},$ and θ^* when n = 1, α = 2, and β = 1:



Briefly discuss what these risk functions indicate. Which risk is associated with which estimator?

2 marks

Question 5: (Answer ANY TWO of the following)

(a) Discuss the following statement: "Model selection is more straightforward if a Bayesian approach is used instead of a frequentist approach".

8 marks

(b) Briefly describe the "Table Lookup" method for generating random values, and explain why it may be useful in Bayesian econometrics.

8 marks

(c) Briefly outline the differences between the Gibbs sampler and the Metropolis-Hastings algorithms.

8 marks

Question 6:

Consider the standard multiple linear regression model, satisfying *all* of the usual assumptions:

$$y = X\beta + \varepsilon \quad ; \qquad \varepsilon \sim N[0, \sigma^2 I_n]$$

The James-Stein (JS) estimator of the coefficient vector is

$$\hat{\beta} = [1 - (ce'e)/(b'X'Xb)]b$$

where 'b' is the ML estimator, 'e' is the ML residual vector, and 'c' is a *positive scalar*.

(a) Explain why the JS estimator is a 'nonlinear' estimator.

1 mark

(b) Explain why we might call the JS estimator a "shrinkage estimator". In what "direction" is the estimator "shrinking" the OLS estimator?

2 marks

(c) In what sense is the JS estimator rather like the natural conjugate Bayes estimator for β?

3 marks

(d) The risk (under quadratic loss) of the JS estimator is approximately

$$Risk(\hat{\beta}) = \sigma^2 tr((X'X)^{-1}) + [nc\sigma^4 / (\beta' X' X\beta)^2][\beta'\beta\{4 + c(n+2)\} - 2(\beta' X' X\beta) tr((X'X)^{-1})]$$

Compare the risks of the JS and OLS estimators when $X'X = I_k$. In particular, for this situation, show that the OLS estimator is <u>inadmissible</u> by proving that it is risk-dominated by the JS estimator for any choice of 'c' such that

$$0 < c < 2(k-2)/(n+2)$$

9 marks

(e) What is the *minimum* number of regressors that must be included in the model for this inadmissibility result to hold?

1 mark

PART B
Answer the ONE QUESTION in this section

THIS QUESTION IS WORTH 26 MARKS

Question 7:

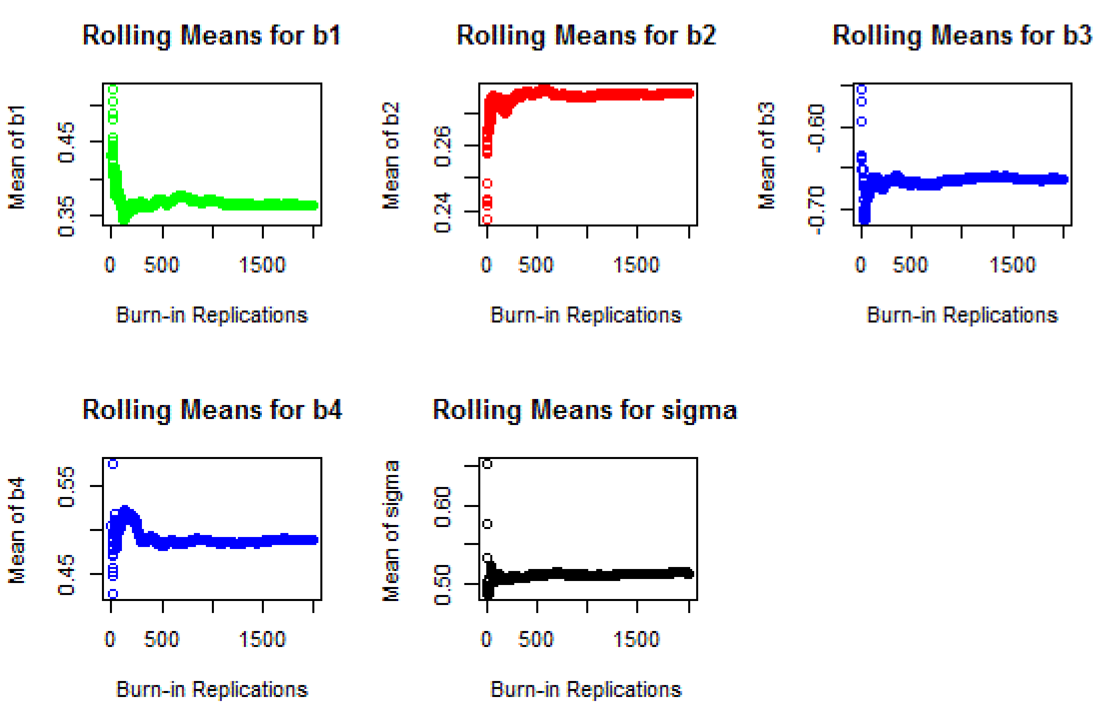
The R code that was used to generate the following results is available in the separate handout that has been supplied.

The model that is being estimated is a linear regression model with an intercept and three other regressors. The errors of the regression model are *independently* distributed according to a Student-t distribution with v degrees of freedom. (Note that this is different from the situation where the error vector follows a *multivariate* Student-t distribution.)

The regression coefficients are called "b1" to "b4", and the scale parameter (standard deviation) for the error distribution is called "sigma".

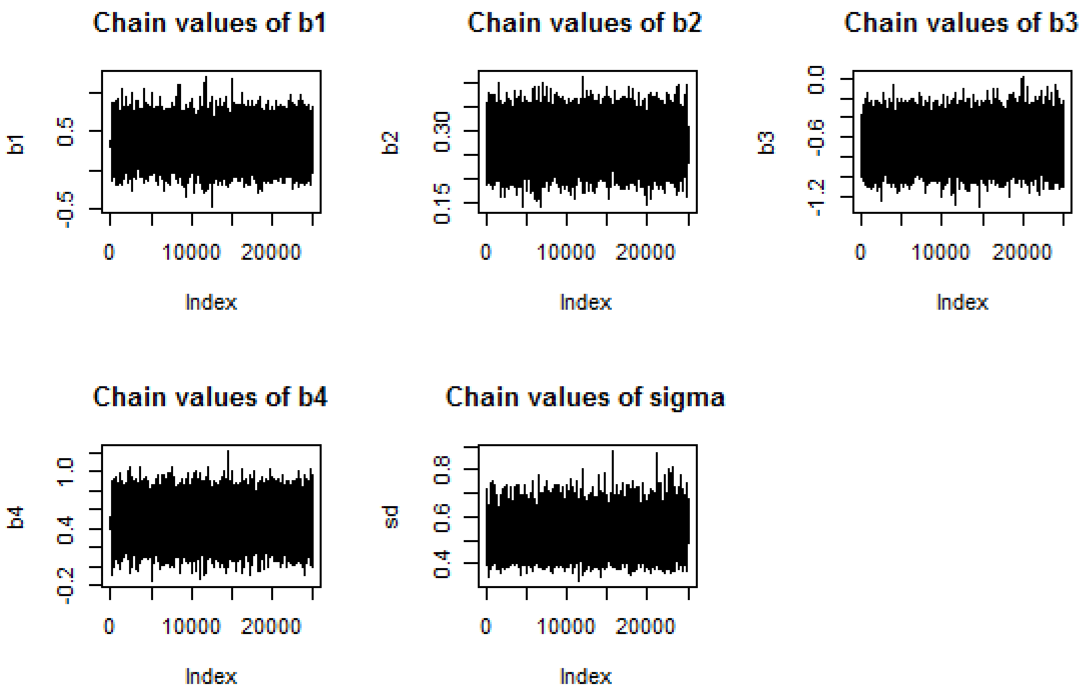
An informative prior distribution is used for the five parameters of the model. The joint posterior density for the parameters is non-standard so a particular MCMC algorithm is used to obtain information about the marginal posterior densities.

(a) **The following results were obtained when v = 4; the mean of the prior density for "sigma" = 1; and the prior means for "b1", "b2", b3", and "b4" are all zero:**



(i) How many MC replications wre used for the "burn-in"? Do you think that enough replications have been used?

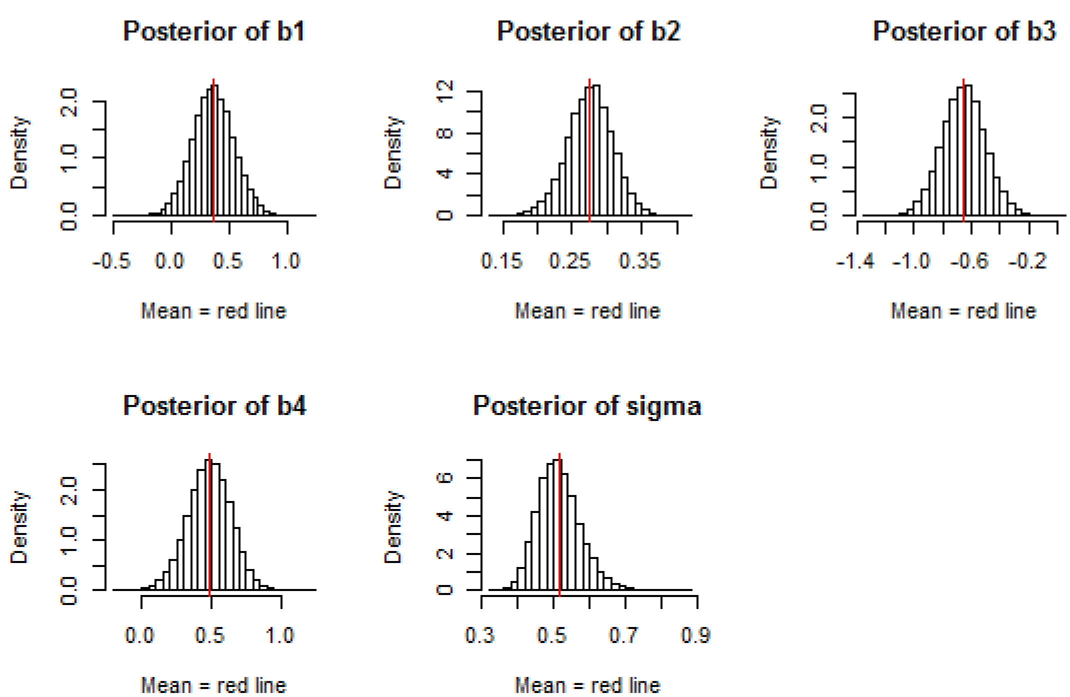
3 marks



(ii) Interpret the graphs immediately above. How many MC replications have been used *after* the "burn-in"?

3 marks

(iii) Discuss the results depicted in the graphs immediately below.



4 marks

```
> # Summarize the marginal posterior p.d.f.'s
> summary(bfit$beta[-(1:burnin),1])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4862  0.2373  0.3554  0.3569  0.4752  1.2160
> summary(bfit$beta[-(1:burnin),2])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1385  0.2542  0.2763  0.2756  0.2975  0.4121
> summary(bfit$beta[-(1:burnin),3])
    Min.   1st Qu.    Median      Mean   3rd Qu.     Max.
-1.30600 -0.75640 -0.65420 -0.65450 -0.55490  0.01361
> summary(bfit$beta[-(1:burnin),4])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.1565  0.3829  0.4855  0.4850  0.5892  1.2190
> summary(bfit$sigma[-(1:burnin)])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3233  0.4735  0.5104  0.5148  0.5503  0.8789
```

(iv) What is the Bayes point estimate of "b1" if we have a quadratic loss function? What is the Bayes point estimate of "sigma" if we have an absolute error loss function?

2 marks

(v) We also have the following results:

```
> #Posterior std. deviations for b1, b2, b3, b4, & sigma:

>   sd(bfit$beta[-(1:burnin),1]);   sd(bfit$beta[-(1:burnin),2]);sd(bfit$beta[-(1:burnin),3]);
sd(bfit$beta[-(1:burnin),4]);sd(bfit$sigma[-(1:burnin)])

[1] 0.1788289
[1] 0.03298178
[1] 0.1514883
[1] 0.1554058
[1] 0.05848021

> # Posterior Quantiles for b1, b2, b3, b4, & sigma:

> quantile(bfit$beta[-(1:burnin),1], probs = c(2.5, 5, 95, 97.5)/100);quantile(bfit$beta[-
(1:burnin),2], probs = c(2.5, 5, 95, 97.5)/100);quantile(bfit$beta[-(1:burnin),3], probs =
c(2.5, 5, 95, 97.5)/100);quantile(bfit$beta[-(1:burnin),4], probs = c(2.5, 5, 95,
97.5)/100);quantile(bfit$sigma[-(1:burnin)], probs = c(2.5, 5, 95, 97.5)/100)

       2.5%          5%         95%       97.5%
0.004459192 0.063116969 0.654733596 0.709129319

     2.5%        5%       95%     97.5%
0.2101656 0.2211769 0.3292389 0.3392983

      2.5%         5%        95%      97.5%
-0.9509986 -0.9032979 -0.4022960 -0.3513928
```

2.5%	5%	95%	97.5%
0.1794102	0.2300648	0.7407045	0.7911273

2.5%	5%	95%	97.5%
0.4133555	0.4273332	0.6184314	0.6420597

Construct a 95% Bayes credible interval for "b1", and a 90% Bayes credible interval for "sigma". Interpret these intervals, and explain how they differ (conceptually) from confidence intervals.

5 marks

(b) **The following results were obtained when *v* = 40 (instead of *v* = 4); the mean of the prior density for "sigma" = 1; and the prior means for "b1", "b2", b3", and "b4" are all zero:**

```
> # Summarize the marginal posterior p.d.f.'s
> summary(bfit$beta[-(1:burnin),1])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4280  0.2621  0.4015  0.4023  0.5434  1.3500
> summary(bfit$beta[-(1:burnin),2])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1160  0.2426  0.2678  0.2675  0.2923  0.4300
> summary(bfit$beta[-(1:burnin),3])
     Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-1.426000 -0.750000 -0.635900 -0.636400 -0.522500 -0.002748
> summary(bfit$beta[-(1:burnin),4])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.2741  0.3687  0.4890  0.4907  0.6130  1.2990
> summary(bfit$sigma[-(1:burnin)])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4604  0.6034  0.6430  0.6479  0.6873  0.9946
```

The Least Squares estimation of the model yields the following results:

```
> summary(lsfit)

Call:
lm(formula = logtime ~ nesting + size + status, data = birdextinct,
    x = TRUE, y = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8410 -0.2932 -0.0709  0.2165  2.5167

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.43087    0.20706   2.081 0.041870 *
nesting      0.26501    0.03679   7.203 1.33e-09 ***
size        -0.65220    0.16667  -3.913 0.000242 ***
status       0.50417    0.18263   2.761 0.007712 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6524 on 58 degrees of freedom
Multiple R-squared:  0.5982,    Adjusted R-squared:  0.5775
F-statistic: 28.79 on 3 and 58 DF,  p-value: 1.577e-11
```

(i) In what sense are these least squares results also Maximum Likelihood estimates?

2 marks

(ii) Use the least squares results and the Bayes results on p.9 to comment on the extent to which the prior information is influencing the results. (You may assume that we have a quadratic loss function.)

4 marks

(c) The following results were obtained when *v* = 4 (as in part (a)); the mean of the prior density for "sigma" = 1; and the prior means for "b1", "b2", b3", and "b4" are all zero. However, instead of setting "g <- 62", as on p.1 of the R code handout, I set "g <- 620":

```
> # Summarize the marginal posterior p.d.f.'s
> summary(bfit$beta[-(1:burnin),1])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.4021  0.2374  0.3573  0.3579  0.4788  1.1650
> summary(bfit$beta[-(1:burnin),2])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1092  0.2554  0.2781  0.2772  0.2995  0.4127
> summary(bfit$beta[-(1:burnin),3])
    Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
-1.28000 -0.76030 -0.65740 -0.65780 -0.55680 -0.07518
> summary(bfit$beta[-(1:burnin),4])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.1990  0.3857  0.4887  0.4893  0.5944  1.3060
> summary(bfit$sigma[-(1:burnin)])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3322  0.4740  0.5103  0.5148  0.5505  0.8612
```

Compare these results with their counterpart on p.8 above. What does this tell you about the sensitivity of the results to the specification of (one aspect) of the prior.

3 marks

END OF EXAMINATION