ECON 546
Assignment 4 – Solution

Q.1 The GLS estimator is

$$\hat{\beta} = [X'(\Sigma \otimes I)^{-1}X]^{-1} X'(\Sigma \otimes I)^{-1} y$$

$$= \left[\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix}(\Sigma^{-1} \otimes I)\begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}\right]^{-1}$$

$$\left[\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix}(\Sigma^{-1} \otimes I)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right]$$

Let $\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{12} & \sigma^{22} \end{pmatrix}$. (It must be symmetric)

So, $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} \sigma^{11} X_1'X_1 & \sigma^{12} X_1'X_2 \\ \sigma^{12} X_2'X_1 & \sigma^{22} X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11} X_1'y_1 + \sigma^{12} X_1'y_2 \\ \sigma^{12} X_2'y_1 + \sigma^{22} X_2'y_2 \end{bmatrix}$

Now let's apply the partitioned inverse formula, (A-74), on p. 966 of Greene (6th. ed.). Focussing on $\hat{\beta}_2$:

$$\hat{\beta}_2 = -F_2 \sigma^{12} X_2'X_1 \frac{1}{\sigma^{11}} (X_1'X_1)^{-1} [\sigma^{11} X_1'y_1 + \sigma^{12} X_1'y_2]$$

$$+ F_2 (\sigma^{12} X_2'y_1 + \sigma^{22} X_2'y_2)$$

where $F_2 = [\sigma^{22} X_2'X_2 - \sigma^{12} X_2'X_1 \frac{1}{\sigma^{11}} (X_1'X_1)^{-1} \sigma^{12} X_1'X_2]^{-1}$

Now, if $X_1(X_1'X_1)^{-1}X_1' = X_2(X_2'X_2)^{-1}X_2'$, then

$$F_2 = [\sigma^{22} X_2'X_2 - (\sigma^{12})^2/\sigma^{11} . X_2'X_2]^{-1}$$

$$= [\sigma^{22} - \frac{(\sigma^{12})^2}{\sigma^{11}}]^{-1} (X_2'X_2)^{-1}$$

Note that

$$\hat{\beta}_2 = F_2 [(\sigma^{12} X_2' y_1 + \sigma^{22} X_2' y_2) - (\sigma^{12} X_2' X_1 (X_1' X_1)^{-1} X_1' y_1)$$
$$- \left(\frac{(\sigma_{12})^2}{\sigma^{11}} X_2' X_1 (X_1' X_1)^{-1} X_1' y_2 \right)]. \quad (1)$$

So, under the stated condition, this collapses to

$$\hat{\beta}_2 = F_2 [\sigma^{12} X_2' y_1 + \sigma^{22} X_2' y_2 - \sigma^{12} X_2' y_1 - \frac{(\sigma_{12})^2}{\sigma^{11}} X_2' y_2]$$
$$= F_2 [(\sigma^{22} - \frac{(\sigma_{12})^2}{\sigma^{11}}) X_2' y_2].$$

where $F_2 = [(\sigma^{22} - \frac{(\sigma_{12})^2}{\sigma^{11}})^{-1} (X_2' X_2)^{-1}]$.

So, $\hat{\beta}_2 = (X_2' X_2)^{-1} X_2' y_2 = b_2$, the OLS estimator. By symmetry, we also get $\hat{\beta}_1 = b_1$.

[This problem is set as exercise 2.14 in V.K. Srivastava & D.E.A. Giles, Seemingly Unrelated Regression Equations Models, Marcel Dekker, 1987. A very compact proof is given in Proposition 1 of T. Kariya, "Tests for the Independence Between Two Seemingly Unrelated Regression Equations", The Annals of Statistics, 1981, 9(2), 381-390.]

* Why is this condition also necessary? If $\hat{\beta}_2$ is to collapse to b_2 in (1), it must not depend on y_1, & this only happens if $X_2' X_1 (X_1' X_1)^{-1} X_1' = 0$. Similarly, if $\hat{\beta}_1 = b_1$, we would need $X_1' X_2 (X_2' X_2)^{-1} X_2' = 0$. They can only both be zero if $X_1 (X_1' X_1)^{-1} X_1' = X_2 (X_2' X_2)^{-1} X_2'$. #.

N.B. You did not have to give the necessity proof.

Q-2

With just one observation, the likelihood fcn is

$$L(\theta|y) = p(y|\theta) = \exp\{-(y-\theta)\}$$

By Bayes' Theorem, the posterior pdf for θ is given by:

$$p(\theta|y) \propto p(\theta)\, L(\theta|y)$$

$$\propto \exp\{-(y-\theta)\}\,[\pi(1+\theta^2)]^{-1}$$

To get the Bayes' estimator of θ under a zero-one loss fcn, we set $\tilde{\theta}$ to the mode of the posterior.

Now, the mode of the posterior will coincide with the mode of the logarithm of the posterior, which is

$$\log p(\theta|y) = \log k - (y-\theta) - \log \pi - \log(1+\theta^2)$$

& $\dfrac{\partial \log p(\theta|y)}{\partial \theta} = \left[1 - \dfrac{2\theta}{1+\theta^2}\right] = 0$; for mode

$$\Rightarrow \quad 1+\tilde{\theta}^2 = 2\tilde{\theta}, \quad \text{or} \quad \tilde{\theta}^2 - 2\tilde{\theta} + 1 = 0$$

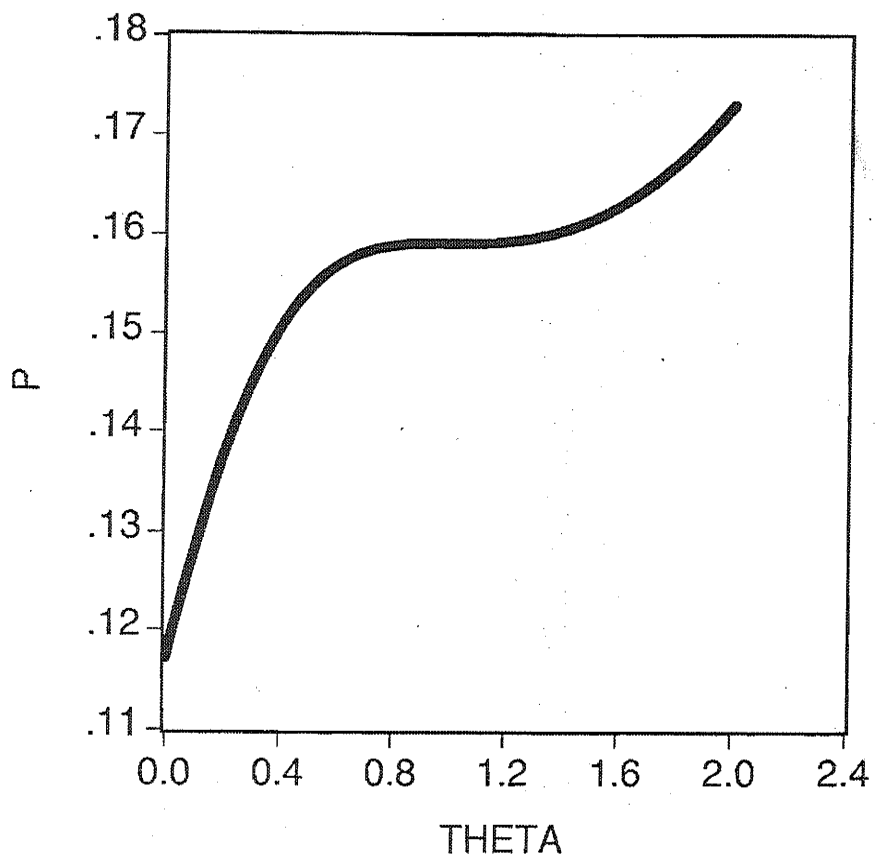
$$\Rightarrow \quad \tilde{\theta} = \frac{2 \pm \sqrt{4-4}}{2} = 1.$$

Now does $\tilde{\theta} = 1$ actually locate the mode?

$$\partial^2 \log p(\theta|y) / \partial\theta^2 = -\left\{ \frac{2}{1+\theta^2} + \frac{2\theta(-1)(2\theta)}{(1+\theta^2)^2} \right.$$

& when $\theta = 1$, this value is 0 !!

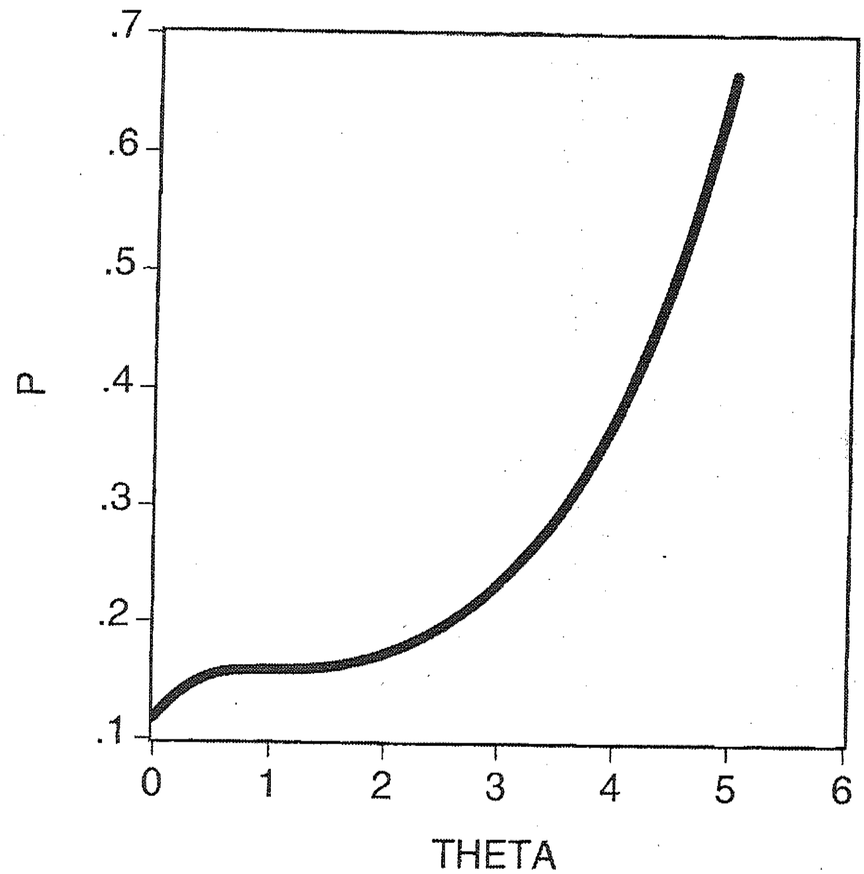
So $\tilde{\theta}$ actually locates a point of inflexion for the posterior, regardless of the value of y:



$(y = 1)$

$p(\theta|y) = \frac{e^{-y+\theta}}{\pi(1+\theta^2)}$, so $p(\theta|y) \to \infty$ as $\theta \to \infty$ (exponential term dominates)

So really, the Bayes' estimator of θ would be any infinitely large value, as this will maximize $p(\theta|y)$.



$(y=1)$

Q 3 (a) The Bayes estimator minimizes average risk : $r(\theta, \hat{\theta}) = \int_{\Omega} R(\hat{\theta}, \theta) p(\theta) d\theta$

where $R(\hat{\theta}, \theta) = \int_{y} L(\theta, \hat{\theta}) p(y|\theta) dy$

is the usual risk.

Suppose the Bayes estimator, θ^*, is <u>inadmissible</u>.

Then there exists an estimator, $\tilde{\theta}$, such that

$R(\tilde{\theta}, \theta) \leq R(\theta^*, \theta) \quad ; \forall \theta$

& $R(\tilde{\theta}, \theta) < R(\theta^*, \theta) \quad ;$ at least some θ.

Then, $r(\tilde{\theta}) = \int_{\Omega} R(\tilde{\theta}, \theta) p(\theta) d\theta \leq \int_{\Omega} R(\theta^*, \theta) p(\theta) d\theta$

for all θ

and, $r(\tilde{\theta}) = \int_{\Omega} R(\tilde{\theta}, \theta) p(\theta) d\theta < \int_{\Omega} R(\theta^*, \theta) p(\theta) d\theta$

for at least some θ,

because $p(\theta) \geq 0 \; ; \forall \theta$.

That is, $r(\hat{\theta}) \leq r(\theta^*) \quad ; \forall \theta$

$< r(\theta^*) \quad ;$ some θ

But this <u>contradicts</u> the fact that θ^* <u>minimizes</u> $r(\hat{\theta})$, as it is the Bayes estimator.

Hence, Bayes estimators are <u>admissible</u>.

#

=.

(b) Diffuse prior + Normal errors $\Rightarrow$ OLS, which, in general, is inadmissible.
The reconciliation is that admissibility of a Bayes estimator generally requires a proper prior, otherwise the Bayes' risk is not well defined. (We can still use the MEL rule in this case).

Q.4 $\quad \mathcal{L} = {}^nC_x\, \theta^x (1-\theta)^{n-x}$

(a) $\quad p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \quad ; a, b > 0; \ 0 \leq \theta \leq 1.$

So, by Bayes' Theorem,

$$\begin{aligned} p(\theta|y) &\propto p(\theta).\mathcal{L} \\ &\propto \theta^x (1-\theta)^{n-x}\, \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{a+x-1}(1-\theta)^{n+b-x-1} \end{aligned}$$

Comparing the posterior pdf with prior pdf, we ~~see~~ see that the former is also a Beta pdf, with parameters $(a+x)$ & $(n+b-x)$. In other words, this is the natural-conjugate prior for a binomial likelihood. When we used the Beta pdf in our consumption fcn. example, we saw that its mean is $(a+x)/[a+x+n+b-x]$, or, $(a+x)/(a+n+b)$, as required for the Bayes' estimator under quadratic loss.

(b) To determine the (unique) mode of $p(\theta)$, we solve $\partial p(\theta)/\partial\theta = 0$, for θ. Equivalently, we can solve $\partial \log p(\theta)/\partial\theta = 0$, for θ.

$$\log p(\theta) = (a-1)\log\theta + (b-1)\log(1-\theta)$$

$$\partial \log p(\theta)/\partial\theta = \frac{a-1}{\theta} - \frac{(b-1)}{1-\theta} = 0$$

$$\Rightarrow (1-\theta)(a-1) - \theta(b-1) = 0$$

$$\Rightarrow a-1 - \theta(a+b-2) = 0$$

$$\Rightarrow \theta = (a-1)/[a+b-2]. \quad \#$$

(c) Obviously, then, the (single) mode of $p(\theta|y)$ is at $(a+x-1)/[a+x+n+b-x-2]$, or at $\theta^* = (a+x-1)/(n+a+b-2)$.

θ^* is the Bayes estimator of θ under a 0-1 loss structure.

(d) X is Binomial, so $E(X) = n\theta$. Let $\hat{\theta}$ denote $E(\theta|y)$ — the Bayes estimator under quadratic loss. Then:

$$E(\hat{\theta}) = E\left[\frac{a+x}{a+n+b}\right] = \frac{a+E(X)}{(a+n+b)}$$

$$= \left(\frac{a+n\theta}{a+n+b}\right)$$

$$\text{Bias}(\theta) = E(\hat{\theta}) - \theta$$

$$= \left(\frac{a+n\theta}{a+n+b}\right) - \theta$$

so $\text{Bias}(\hat{\theta}) = \left[\dfrac{a(1-\theta) - b\theta}{a+n+b}\right]$

Similarly,

$$E(\theta^*) = E\left[\frac{a+x-1}{n+a+b-2}\right] = \frac{(a-1)+E(x)}{(n+a+b-2)}$$

$$= \left[\frac{(a-1)+n\theta}{(n+a+b-2)}\right]$$

& $\text{Bias}(\theta^*) = E(\theta^*) - \theta$

$$= \left[\frac{(a-1)+n\theta}{n+a+b-2}\right] - \theta$$

$$= \left[\frac{a(1-\theta) - b\theta + (2\theta-1)}{(n+a+b-2)}\right].$$

The MLE for θ is (x/n) – the proportion of "successes" in the "n" trials.

So $E(\tilde{\theta}) = E(x)/n = n\theta/n = \theta.$

The MLE is unbiased, whereas the 2 Bayes estimators are both biased, & the extent of their bias depends, in part, on the choice of loss function.

Q.5.

(a) $n = 6$; $\bar{X} = 64$; $\sigma_0^2 = 300$; $\bar{\mu} = 60$; $\bar{V} = 80$

$$\bar{\bar{\mu}} = \left[\frac{1}{\bar{V}} \bar{\mu} + \frac{n}{\sigma_0^2} \bar{X} \right] / \left[\frac{1}{\bar{V}} + \frac{n}{\sigma_0^2} \right]$$
$$= (2.03 / 0.0325)$$
$$= 62.4615$$

& $\frac{1}{\bar{\bar{V}}} = \left[\frac{1}{\bar{V}} + \frac{n}{\sigma_0^2} \right] = 0.0325$,

so, $\bar{\bar{V}} = 30.7692$.

(b) Now, $n = 4$; $\bar{X} = 92.5$; $\bar{\mu} = 62.4615$; $\bar{V} = 30.7694$

& $\bar{\bar{\mu}} = (3.26333 / 0.04583) = 71.199$

(Note: $\sigma_0^2 = 300$, still.)

& $\bar{\bar{V}} = \left\{ \frac{1}{\bar{V}} + \frac{n}{\sigma_0^2} \right\}^{-1} = (1 / 0.04583) = 21.8198$.

Notice that as more information is gathered & this information is taken into account, the posterior variance falls (the precision increases).

(c) $p(\mu \mid \text{Data}) \sim N[71.119, \text{var.} = 21.8198]$

$$\Pr.[50 < \mu < 80] = \Pr.\left[\frac{50 - 71.199}{\sqrt{21.8198}} < Z < \frac{80 - 71.199}{\sqrt{21.8198}} \right]$$
$$= \Pr[-4.5383 < Z < 1.884] \approx \Pr[Z < 1.884]$$
$$= 96.99\%.$$

(d) Use the mode (= mean, because Normal)
$= 71.199$