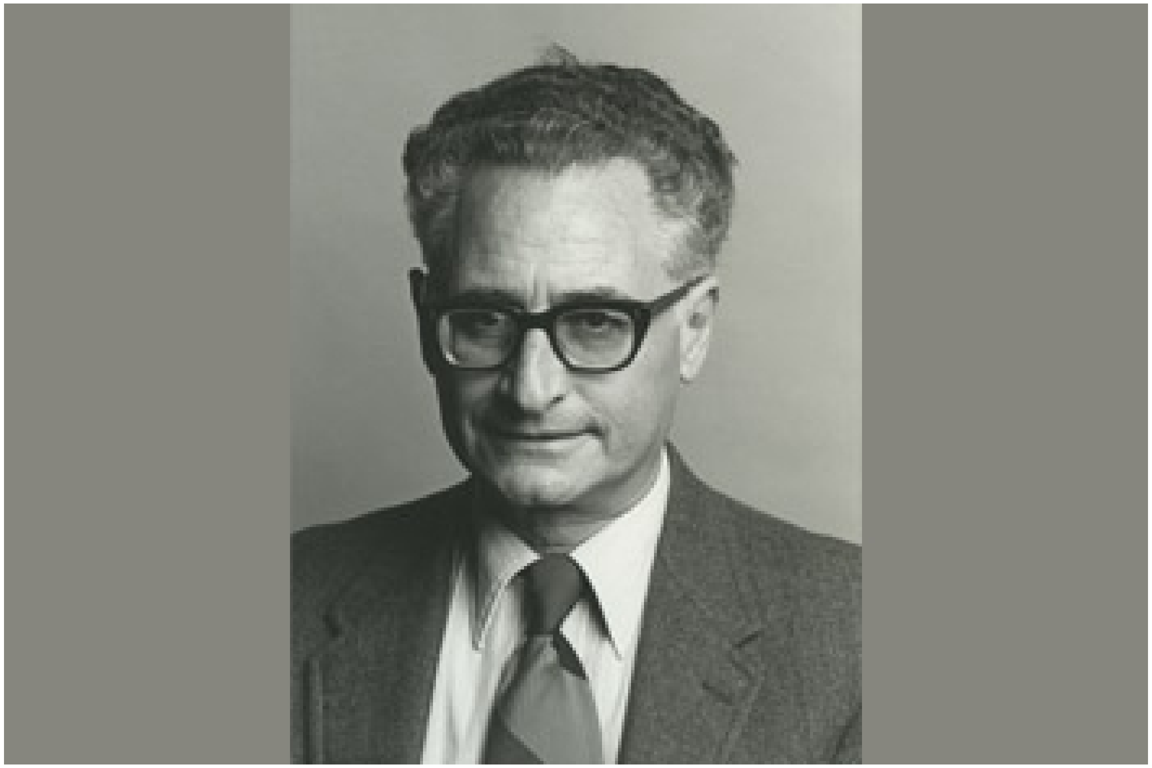
Topic 3

Bayesian Econometrics

4. Bayesian Inference for the Linear Regression Model



Arnold Zellner (1927 -2010)

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad ; \qquad \boldsymbol{\varepsilon} \sim N[0\,,\sigma^2 I_k]$$

$$rank(X) = k \quad ; \quad X \text{ is non-random}$$

- Consider two situations:

  (i) "Diffuse" Prior

  (ii) Natural-Conjugate Prior

- Likelihood Function:

$$L(\boldsymbol{\beta},\sigma\,|\boldsymbol{y}) = p(\boldsymbol{y}\,|\boldsymbol{\beta},\sigma) \propto \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})\}$$

- Re-write the Likelihood function –

Let $\quad \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$.

Then:

$$(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta}) = [(\boldsymbol{y}-X\widehat{\boldsymbol{\beta}}) + (X\widehat{\boldsymbol{\beta}}-X\boldsymbol{\beta})]'[(\boldsymbol{y}-X\widehat{\boldsymbol{\beta}}) + (X\widehat{\boldsymbol{\beta}}-X\boldsymbol{\beta})]$$

Note that $(\boldsymbol{y}-X\widehat{\boldsymbol{\beta}})'(X\widehat{\boldsymbol{\beta}}-X\boldsymbol{\beta}) = 0$

- So,

$$(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta}) = (\boldsymbol{y}-X\widehat{\boldsymbol{\beta}})'(\boldsymbol{y}-X\widehat{\boldsymbol{\beta}}) + (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})$$
$$= vs^2 + (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})$$

- Here, $v = (n-k)$ and $s^2 = (\boldsymbol{e}'\boldsymbol{e})/(n-k)$; where $\boldsymbol{e} = \boldsymbol{y}-X\widehat{\boldsymbol{\beta}}$.
- So, the Likelihood function becomes:

$$L(\boldsymbol{\beta},\sigma\,|\boldsymbol{y}) \propto \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}[vs^2 + (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})]\}$$

(i) Diffuse Prior:

- $p(\boldsymbol{\beta},\sigma) = p(\boldsymbol{\beta})p(\sigma)$

- $p(\beta_i) \propto constant \quad ; \quad -\infty < \beta_i < \infty \quad ; \quad i = 1, 2, \ldots \ldots, k$
- $p(\sigma) \propto 1/\sigma \quad ; \quad 0 < \sigma < \infty$
- So, $p(\boldsymbol{\beta}, \sigma) \propto 1/\sigma$
- Apply Bayes' Theorem –

$$p(\boldsymbol{\beta}, \sigma \,|\boldsymbol{y}) \propto \sigma^{-(n+1)} \exp\{-\frac{1}{2\sigma^2}[vs^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})]\}$$

- First, let's *condition* on σ, and see what we can observe.
- Then we'll *marginalize* and focus just on the $\boldsymbol{\beta}$ vector.
- ***Conditioning***:

$$p(\boldsymbol{\beta}|\sigma, \boldsymbol{y}) \propto \exp\{-\frac{1}{2\sigma^2}[vs^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})]\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}[vs^2]\right\} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})' [\sigma^2 (X'X)^{-1}]^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})]\right\}$$

- So, the *Conditional* Posterior is multivariate Normal with a mean of $\widehat{\boldsymbol{\beta}}$ and a covariance matrix of $\sigma^2 (X'X)^{-1}$.
- Recall that $\widehat{\boldsymbol{\beta}}$ is the MLE for the coefficient vector (= OLS).
- So, $\widehat{\boldsymbol{\beta}}$ is the Bayes estimator under various symmetric Loss Functions.
- ***Marginalizing***:

$$p(\boldsymbol{\beta}|\boldsymbol{y}) = \int_0^\infty p(\boldsymbol{\beta}, \sigma\,|\boldsymbol{y})d\sigma$$

$$\propto \int_0^\infty \sigma^{-(n+1)} \exp\{-\frac{1}{2\sigma^2}[\upsilon s^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})]\} d\sigma$$

- Change of variables: $c = \left[\upsilon s^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right]$

$$z = c/\sigma^2$$

- So:

$$dz = -\left(\frac{2c}{\sigma^3}\right) d\sigma \;\; ; \qquad d\sigma = -(\sigma^3/2c)dz \;\; ; \;\; \sigma = (\frac{c}{z})^{1/2}$$

- $p(\boldsymbol{\beta}\,|\boldsymbol{y}) \propto \int_0^\infty \sigma^{-(n+1)} \exp\{-\frac{1}{2\sigma^2}[\upsilon s^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})]\} d\sigma$

$$\propto \int_\infty^0 (\frac{c}{z})^{-(n+1)/2} e^{-z/2} (-\sigma^3/2c) dz$$

$$\propto \int_0^{\infty} \left(\frac{c}{z}\right)^{-\frac{n+1}{2}} e^{-\frac{z}{2}} \left(\frac{1}{2}\right) \left(\frac{1}{z}\right)^{\frac{3}{2}} c^{\frac{3}{2}} \left(\frac{1}{c}\right) dz$$

$$\propto c^{-n/2} \int_0^{\infty} z^{\left(\frac{n}{2}-1\right)} e^{-z/2} dz$$

Kernel of a Gamma density with shape = (n / 2), scale = ½

The integral = $\Gamma(\frac{n}{2})(\frac{1}{2})^{n/2}$ It's a **finite** constant.

- $p(\boldsymbol{\beta} \mid \boldsymbol{y}) \propto c^{-\frac{n}{2}} = [vs^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^{-n/2}$
- This is the kernel of a **Multivariate Student-t** density.
- Specifically, it's MVST $\left[\hat{\boldsymbol{\beta}}\ ;\ s^2(X'X)^{-1}(\frac{v}{v-2})\right]$.

- Once again, $\widehat{\boldsymbol{\beta}}$ is the Bayes' estimator under various symmetric Loss Functions.

- The marginal densities of a MVST density are univariate Student-t densities, so the *Marginal Posterior* p.d.f. for each β_i is Student-t, with a mean (= median = mode) equal to $\widehat{\beta_\iota}$.

- Similarly, we can obtain the *Marginal Posterior* p.d.f. for σ :

$$p(\sigma \mid \boldsymbol{y}) = \int_{-\infty}^{\infty} p(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) d\boldsymbol{\beta}$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) d\beta_1 \, d\beta_2 \dots \, d\beta_k$$

- It turns out that

$$p(\sigma \mid \boldsymbol{y}) \propto \sigma^{-(v+1)} exp[-vs^2/2\sigma^2]$$

- This is the density for an Inverted Gamma distribution.
- It has the following features –

(i) Single *mode* at $\sigma_m = s\left[\frac{v}{v+1}\right]^{1/2}$.

(ii) $E(\sigma \mid \boldsymbol{y}) = s v^{\frac{1}{2}}\left[\Gamma(\frac{v-1}{2})\right] / \left[\Gamma(\frac{v}{2})\right]$; $v > 1$.

(iii) $Var.(\sigma \mid \boldsymbol{y}) = \left[\frac{vs^2}{v-2}\right] - [E(\sigma \mid \boldsymbol{y})]^2$; $v > 2$.

(iv) $Skew(\sigma \mid \boldsymbol{y}) = s\left\{\frac{\left[\Gamma\left(\frac{v-2}{2}\right)\right]\left(\frac{v}{2}\right)^{\frac{1}{2}}}{\left[\Gamma\left(\frac{v}{2}\right)\right]} - \left(\frac{v}{v+1}\right)^{1/2}\right\} / \sqrt{Var.(\sigma \mid \boldsymbol{y})}$

; $v > 2$.

- Generally, the Skewness is positive.
- Under our standard Loss Functions, the Bayes estimator of σ differs from the OLS – associated estimator (= s).
- However, recall that $v = (n-k)$, so $v \rightarrow \infty$ when $n \rightarrow \infty$.
- So, $\lim_{n\rightarrow\infty} \sigma_m = s$.
- Also, $s\,\frac{v^{\frac{1}{2}}\left[\Gamma\left(\frac{v-1}{2}\right)\right]}{\left[\Gamma\left(\frac{v}{2}\right)\right]} = s\left[\Gamma(\frac{v}{2}+1)\Gamma(\frac{v-1}{2})\right] / \left[\Gamma(\frac{v}{2})\right]^2$,

  So, $\lim_{n\rightarrow\infty} E(\sigma\,|\boldsymbol{y}) = s$.

(ii) Natural Conjugate Prior:

- Given the Normal Likelihood function, the N.C. prior is

$$p(\boldsymbol{\beta}, \sigma) = p(\boldsymbol{\beta} \,|\sigma) p(\sigma)$$

where

$p(\boldsymbol{\beta} \,|\sigma)$ is a (conditional) Multivariate Normal $[\bar{\beta} \; ; \; \sigma^2 A^{-1}]$

$p(\sigma)$ is Inverted Gamma $(w\,,\,c)$

- $p(\boldsymbol{\beta} \,|\sigma) \propto |A|^{1/2} \sigma^{-k} exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\right\}$; A is p.d.s.

$p(\sigma) \propto \sigma^{-(w+1)} exp\{-wc^2/(2\sigma^2)\}$; $w\,, c > 0$

- Choose values of w, c, $\bar{\boldsymbol{\beta}}$ and A to reflect our prior beliefs.

- Recall that the Likelihood Function is

$$L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) = p(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})\}$$

- Apply **Bayes' Theorem**:

$$p(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) \propto p(\boldsymbol{\beta}, \sigma) L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y})$$

- So, $p(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) \propto$

$$\sigma^{-(n+k+w+1)} exp\left[-\frac{1}{2\sigma^2}\left\{wc^2 + \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\right)' A\left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\right) + (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})\right\}\right]$$

- Define: $\gamma = \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\right)' A\left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\right) + (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})$

$$= \boldsymbol{\beta}'(A + X'X)\boldsymbol{\beta} - 2\boldsymbol{\beta}'\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right) + (\boldsymbol{y}'\boldsymbol{y} + \overline{\boldsymbol{\beta}}' A\overline{\boldsymbol{\beta}})$$

- This is *Quadratic* in the $\boldsymbol{\beta}$ vector – we'll "complete the square".

- $q = (ax^2 + bx + c\ ) = a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$

- $q = (\boldsymbol{x}'A^*\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{b} + c\ ) = \left(\boldsymbol{x} + \frac{1}{2}A^{*-1}\boldsymbol{b}\right)'A^*\left(\boldsymbol{x} + \frac{1}{2}A^{*-1}\boldsymbol{b}\right)$

$$+\left(c - \frac{1}{4}\boldsymbol{b}'A^{*-1}\boldsymbol{b}\right)$$

- So,

$$\gamma = \left[\boldsymbol{\beta} - (A + X'X)^{-1}\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)\right]'(A + X'X)\left[\boldsymbol{\beta} - (A + X'X)^{-1}\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)\right]$$

$$+\ \boldsymbol{y}'\boldsymbol{y} + \ \overline{\boldsymbol{\beta}}'A\overline{\boldsymbol{\beta}} - \left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)'(A + X'X)^{-1}\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)$$

- Define: $\breve{\boldsymbol{\beta}} = (A + X'X)^{-1}\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)$.

- So, $p(\boldsymbol{\beta}, \sigma \,|\boldsymbol{y}) \propto \sigma^{-(n+k+w+1)} exp\left\{-\frac{1}{2\sigma^2}\left[wc^2 + \boldsymbol{y}'\boldsymbol{y} + \overline{\boldsymbol{\beta}}' A \overline{\boldsymbol{\beta}} + \right.\right.$

$$\left.\left.(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})'(A + X'X)(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}}) - \breve{\boldsymbol{\beta}}'(A + X'X)\breve{\boldsymbol{\beta}}\right]\right\}$$

$$\propto \sigma^{-(n'+k+1)} exp\left\{-\frac{1}{2\sigma^2}\left[n'c_1^2 + (\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})'(A + X'X)(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})\right]\right\}$$

where: $n' = (n + w)$

$$n'c_1^2 = (wc^2 + \boldsymbol{y}'\boldsymbol{y} + \overline{\boldsymbol{\beta}}' A \overline{\boldsymbol{\beta}} - \breve{\boldsymbol{\beta}}'(A + X'X)\breve{\boldsymbol{\beta}})$$

- As before, let’s first *condition* on $\boldsymbol{\sigma}$, and see what we can observe.

- Then we’ll *marginalize* and focus just on the $\boldsymbol{\beta}$ vector.

***Conditioning*:**

- The *conditional posterior* p.d.f. for $\boldsymbol{\beta}$ should be Multivariate Normal as a consequence of the Natural Conjugacy of the prior.
- $p(\boldsymbol{\beta}\,|\sigma,\boldsymbol{y}) \propto exp\left\{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{\beta}-\breve{\boldsymbol{\beta}}\right)'(A+X'X)\left(\boldsymbol{\beta}-\breve{\boldsymbol{\beta}}\right)\right]\right\}$
- This is MVN $\left[\,\breve{\boldsymbol{\beta}}\,,\sigma^2(A+X'X)^{-1}\,\right]$.

***Marginalizing*:**

- $p(\boldsymbol{\beta}\,|\boldsymbol{y}) = \int_0^\infty p(\boldsymbol{\beta},\sigma|\boldsymbol{y})d\sigma$
- In the case of the "Diffuse" Prior, we evaluated the integral:

$$\int_0^\infty \sigma^{-(n+1)}\text{exp}\{-\frac{1}{2\sigma^2}[vs^2+\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)'X'X\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)]\}d\sigma$$

- We showed that it was *proportional to*

$$[vs^2 + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' X'X(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})]^{-n/2}$$

- Here, we need to integrate

$$\sigma^{-(n'+k+1)} exp\left\{-\frac{1}{2\sigma^2}\left[n'c_1^2 + (\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})'(A + X'X)(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})\right]\right\}$$

with respect to σ .

- Immediately, then:

$$p(\boldsymbol{\beta}\,|\boldsymbol{y}) \propto \left[n'c_1^2 + (\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})'(A + X'X)(\boldsymbol{\beta} - \breve{\boldsymbol{\beta}})\right]^{-(n'+k)/2}$$

- This is the kernel for the density of a Multivariate Student-t distribution.

- $p(\boldsymbol{\beta} \mid \boldsymbol{y})$ is MVST$\left[\breve{\beta}\ ;\ \ n'c_1^2(A+X'X)^{-1}/(n'-2)\right]$

- So, the Bayes estimator of $\boldsymbol{\beta}$, under our standard Loss Functions, is

$$\breve{\boldsymbol{\beta}} = (A+X'X)^{-1}\left(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y}\right)$$

- Recall that the marginal densities of a MVST are univariate Student-t, centered at the $\breve{\beta}_i$ elements.

- If we want to draw inferences about σ, we can obtain the marginal posterior p.d.f. for this parameter. Conjugacy tells us that it will be Inverted Gamma.

- $p(\sigma \mid \boldsymbol{y}) = \int_{-\infty}^{\infty} p(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y})\, d\boldsymbol{\beta}$

$$\propto \sigma^{-(n'+1)} exp\left[-\left(\frac{n'c_1^2}{2\sigma^2}\right)\right]$$

- This is indeed the kernel of an Inverted Gamma density.

- Now, let's note some interesting and important features of the N.C. Bayes estimator of $\boldsymbol{\beta}$:

$$\breve{\boldsymbol{\beta}} = (A + X'X)^{-1}(A\overline{\boldsymbol{\beta}} + X'\boldsymbol{y})$$; <u>doesn't</u> depend on *w* or *c*

(i) *A* is positive-definite, so the estimator is defined even if rank(X) < *k*.

(ii) How does this relate to the "problem" of Multicollinearity?

(iii) We can write

$$\breve{\boldsymbol{\beta}} = \left(\frac{1}{\sigma^2}A + \frac{1}{\sigma^2}X'X\right)^{-1}\left(\frac{1}{\sigma^2}A\overline{\boldsymbol{\beta}} + \frac{1}{\sigma^2}X'X\hat{\boldsymbol{\beta}}\right)$$

(iv) $\breve{\boldsymbol{\beta}}$ is a *matrix* weighted average of $\overline{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ (= OLS = MLE).

(v) Individual elements of $\breve{\boldsymbol{\beta}}$ need not lie "between" the corresponding individual elements of $\bar{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$.

(vi) If $\bar{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ then $\breve{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}$. (This makes intuitive sense.)

(vii) Some non-Bayesian estimators of $\boldsymbol{\beta}$ are can be interpreted as special cases of this N.C. Bayes estimator:

- The "Ridge" estimator:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{R}} = (\kappa I + X'X)^{-1}(X'X)\hat{\boldsymbol{\beta}}$$

$A = \kappa I$ and $\bar{\boldsymbol{\beta}} = 0$. Why "shrink" the estimator towards zero?

- James-Stein & Theil's "mixed" regression estimators are other examples.