Notes on Arnold Zellner's, *An Introduction to Bayesian Inference in Econometrics*

Prepared by David Giles, 1973

Released, April 2012

dgiles@uvic.ca

Zellner : Chapter 1.

The Unity of Science —

Scientific inferences made about phenomena in economics are not fundamentally different from inferences made in other areas of science.

There are 3 types of inference — deductive, inductive, and reductive (or abductive).

Deductive Inference —

Deductive arguments — proceeds from premises to conclusion.

Not sufficient for all inference because:

(a) Impossibility of complete certainty in knowledge. (Can't be sure that the sun will rise tomorrow).

(b) A given set of data can be generated by an infinity of theories. Deductive logic can't help us to choose between them.

(c) May choose the simplest —
  (i) Desirable that its implications be far-reaching though.
  (ii) Valuable because simple models make strong statements which are readily testable.

Although deductive inference is important, it is insufficient to situations.

Take Jeffreys view of inductive logic — it contains deductive logic within its bounds. Not mutually exclusive.

Inductive Inference —

Fundamental problem is that of learning from experience. Part of this is description of what we have observed already, & part is generalization or induction — future experience is predicted on the basis of past experience.

But induction is more than mere description. Here Jeffreys is in conflict with Mach.
Always an element of uncertainty with induction.

Reductive Inference -

"Reduction" suggests that something *may be*.

It involves studying facts & devising theories to explain them.

The link of reduction with the *unusual fact* is emphasised.

Unusual facts often trigger the reductive process to produce new concepts & generalizations.

We can view the problem of invention or discovery as one of choice among the many possible combinations of ideas.

Preparatory and apparently fruitless work is required, & it helps to be in touch with a variety of fields.

Two possibilities —

(a) A goal being given, & how to reach it?

(b) Discovering a fact & then imagining how it might be useful.

The reductive process involves choosing particular combinations of ideas that seem fruitful.

Jeffrey's Rules for a Theory of Inductive Inference -

(1) The conclusions must follow from the hypotheses.
(2) The theory must be self-consistent.
(3) Any given rule must be applicable in practice.
(4) The theory must allow for the possibility that an inference may be incorrect.
(5) The theory must not deny any empirical proposition *a priori*.

These are the 5 "essential" rules. Also adds three more "inessential" ones as "useful guides."

One of the implications of Jeffreys's rules is that the concept of probability as simply a frequency is ruled out.

Probability is cast in terms of a "reasonable degree of belief", satisfying certain consistency rules which enable it to be expressed in terms of numbers.

Jeffreys is a subjective theory which attempts to provide consistent procedures for behaviour under uncertainty.

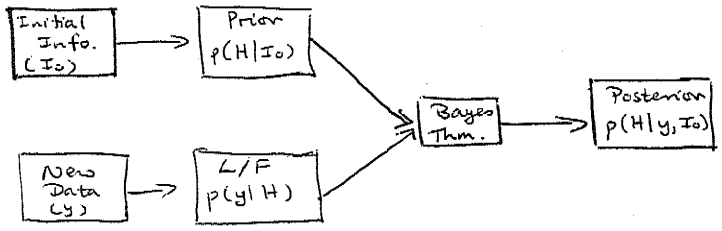
Numerical probabilities can be associated with degrees of confidence that we have in propositions about empirical phenomena.

When I say that something is "probably true", then I mean that on the basis of previous information, studies & experience, I have a high degree of confidence in the explanation that has been offered.

The Bayesian approach (especially Jeffreys') involves a quantification of such phrases as "probably true", by utilizing numerical probabilities to represent degrees of confidence or belief.

The degree of reasonable belief depends upon the state of our current information. So a probability representing a degree of belief is <u>conditional</u> (upon our information). We revise the probability as our information changes. This is the essence of learning from experience.

Bayes's Theorem renders operational the process of revising our probabilities as our information changes.



As the sample information grows it will more & more dominate the posterior density & this will tend to concentrate about the true value of the parameter.

The Bayesian procedure may be applied widely throughout economics & the sciences in general.

It provides a unified & operational approach to a wide variety of inferential problems. The same principles are adopted in each case. c.f. other methods of inference.

Zellner : Chapter 2.

Bayes' Theorem –

The "principle of inverse probability" — here we have a given set of data, & from the information in the data we try to infer what random process generated them.

$$p(y, \theta) = p(y|\theta)p(\theta) = p(\theta|y)p(y)$$

$$\text{So, } p(\theta|y) = \frac{p(\theta)\,p(y|\theta)}{p(y)}$$

$$= \frac{p(\theta)\,p(y|\theta)}{\int_{R_\theta} p(\theta)\,p(y|\theta)\,d\theta}$$

i.e. $p(\theta|y) \propto p(\theta)\,p(y|\theta)$.

where $p(\theta|y)$ is the posterior density,
$p(\theta)$ is the prior "
& $p(y|\theta)$ is the likelihood function, & is <u>not</u> a p.d.f. It is to be viewed as a function of θ.

The latter constitutes the entire evidence of the experiment.

Zellner gives the example of $y_i \sim IN(\mu, \sigma_0^2)$ where σ_0^2 is <u>known</u>, & where $\mu \sim N(\mu_a, \sigma_a^2)$.

Then let $\hat{\mu} = \frac{1}{n}\sum_1^n y_i$; $s^2 = \left(\frac{1}{n-k}\right)\sum_1^n (y_i - \hat{\mu})^2$

$$\text{So, } p(\mu|y, \sigma_0^2) \propto \exp\left\{-\frac{1}{2}\left[\frac{(\mu-\mu_a)^2}{\sigma_a^2} + \left(\frac{n}{\sigma_0^2}\right)(\mu-\hat{\mu})^2\right]\right.$$

Then, as in the <u>Theil Notes</u>, complete the square on μ,
& obtain:

$$p(\mu|y, \sigma_0^2) \propto \exp\left[-\left(\frac{\sigma_a^2 + \sigma_0^2/n}{2\sigma_a^2\sigma_0^2/n}\right)\left(\mu - \frac{\hat{\mu}\sigma_a^2 + \mu_a\sigma_0^2/n}{\sigma_a^2 + \sigma_0^2/n}\right)^2\right]$$

$$\therefore \mu \sim N(E(\mu), var(\mu))$$

$$\text{So, } E(\mu) = \left(\frac{\hat{\mu}\sigma_a^2 + \mu_a\sigma_0^2/n}{\sigma_a^2 + \sigma_0^2/n}\right) = \frac{\hat{\mu}\,(\sigma_0^2/n)^{-1} + \mu_a\,(\sigma_a^2)^{-1}}{(\sigma_0^2/n)^{-1} + (\sigma_a^2)^{-1}}$$

i.e. a weighted average of sample & prior means.

$$\& \; var(\mu) = \left(\frac{\sigma_a^2\sigma_0^2/n}{\sigma_a^2 + \sigma_0^2/n}\right) = \frac{1}{(\sigma_0^2/n)^{-1} + (\sigma_a^2)^{-1}}$$

Let $h_o = (\sigma_o^2/n)^{-1}$; $h_a = (\sigma_a^2)^{-1}$

Then, $E(\mu) = (\hat{\mu} h_o + \mu_a h_a)/(h_o + h_a)$

$var(\mu) = 1/(h_o + h_a)$

So h_o & h_a are "precision parameters" relating to the sample & prior density respectively.

Several Sets of Data -

Suppose that we have a prior $p(\theta)$ & sample y_1

Then, $p(\theta|y_1) \propto p(\theta)\, p(y_1|\theta)$

Then obtain y_2 <u>independent</u> of y_1

So treat $p(\theta|y_1)$ as a prior, relative to y_2.

i.e. $p(\theta|y_1, y_2) \propto p(\theta|y_1)\, p(y_2|\theta)$

$\therefore\ p(\theta|y_1, y_2) \propto p(y_1|\theta)\, p(y_2|\theta)\, p(\theta).$

But, $p(y_1|\theta) . p(y_2|\theta)$ is just the <u>joint</u> L/F. for y_1 & y_2 combined. So it doesn't matter if we handle the data sequentially, or if we pool it & consider the joint sample.

Prior p.d.f.'s —

The prior p.d.f., $p(\theta)$, represents our knowledge about the parameters of the model prior to sampling. When choosing $p(\theta)$ we must take account of the range of θ & ensure that $p(\theta)$ is defined as a p.d.f. exactly over the same range as is θ.

Prior information can come from one (or both) of two types of source —

(i) Data-based (DB) — information in samples of past data.

(ii) Non data-based (NDB) — arising from introspection, casual observation, or theoretical considerations.

Information in the data can be used to make comparisons of alternative prior beliefs or hypotheses. Later in the book Zellner analyzes this problem & develops a framework for making a choice among alternative conflicting beliefs or hypotheses.

If $\exists$ 2 investigators with the same model & with DB prior based on different past data, the posterior densities may differ. If their past samples are pooled, then their priors become identical & they come to the same final solution.

Typically, prior information is fairly scant, so we use a "diffuse" prior density to represent this fact.

If we use a diffuse prior —

$$p(\mu) \propto \text{constant} \quad ; \; -\infty < \mu < \infty$$

Then this is an "improper" p.d.f., since

$$\int_{-\infty}^{\infty} p(\mu)\,d\mu = k\int_{-\infty}^{\infty} d\mu \to \infty.$$

But Bayes' theorem still applies without modification.

Then if $y_i \sim IN(\mu, \sigma_0^2)$; σ_0^2 known.

We have:

$$p(\mu \mid y, \sigma_0^2) \propto k.\left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_1^n (y_i-\mu)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}\sum_1^n\left[(y_i-\hat{\mu})-(\mu-\hat{\mu})\right]^2\right\}$$

$$\therefore \quad p(\mu \mid y, \sigma_0^2) \propto \exp\left\{-\frac{1}{2\sigma_0^2}\left[(n-k)s^2 + n(\mu-\hat{\mu})^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(n-k)s^2\right\} . \exp\left\{\frac{-n}{2\sigma_0^2}(\mu-\hat{\mu})^2\right\}$$

$$\propto \exp\left\{-\frac{n}{2\sigma_0^2}(\mu-\hat{\mu})^2\right\}.$$

where $\hat{\mu} = \frac{1}{n}\sum_i^n \mu_i$.

So, $E(\mu) = \hat{\mu}$; $var.(\mu) = (\sigma_0^2/n)$ and <u>Normal</u>.

c.f. the earlier example with a normal prior. There let $\sigma_a \to \infty$, & get the same result as here.

When we have NDB information to be incorporated via a prior pdf, then we want one that represents our information accurately, yet is mathematically tractable.

The so-called "natural conjugate" prior p.d.f.'s are useful in this respect.

Let $p(y \mid \theta, n)$ be the pdf for y.

If $p(y \mid \theta, n) = p_1(t \mid \theta, n) . p_2(y)$
& $t' = (t_1, ---, t_k)$
where p_2 is independent of θ, then the t_i's are termed <u>sufficient</u> statistics. <u>NB</u> : $t_i = t_i(y)$.

A <u>Natural Conjugate</u> prior p.d.f. for θ is

$$f(\theta \mid .) \propto p_1(t \mid \theta, n)$$

$$= k\, p_1(t \mid \theta, n)$$

where $k = k(t, n)$.

So $f(\theta \mid .)$ has the same functional form as $p_1(t \mid \theta, n)$. Its argument is θ, & its parameters are t & n. So assign values to t & n & get:

$$f(\theta \mid t_0, n_0) \propto p_1(t_0 \mid \theta, n_0)$$

& use this as the <u>informative</u> prior pdf.

<u>NB</u> : If $p(y \mid \theta, n) = p_1(t \mid \theta, n) . p_2(y)$

Then $p(\theta \mid y, n) \propto p(\theta)\, p(y \mid \theta, n)$
$\propto p(\theta)\, p_1(t \mid \theta, n)\, p_2(y)$
$\propto p(\theta)\, p_1(t \mid \theta, n)$.

And as $n \to \infty$, $p(\theta \mid y, n) \propto p_1(t \mid \theta, n)$. also.

Marginal & Conditional Posterior Distributions —

Let $\theta' = (\theta_1' : \theta_2')$

Then, $$p(\theta_1|y) = \int_{R_{\theta_2}} p(\theta_1, \theta_2|y)\,d\theta_2$$
$$= \int_{R_{\theta_2}} p(\theta|y)\,d\theta_2$$
$$= \int_{R_{\theta_2}} p(\theta_1|\theta_2, y)\,p(\theta_2|y)\,d\theta_2.$$

So the marginal posterior pdf for θ_1 is a weighted average of the conditional posterior pdf, the weighting function being the marginal posterior for θ_2.

Let $y' = (y_1, ---, y_n)$; $y_i \sim IN(\mu, \sigma^2)$; both parameters unknown.

& $p(\mu, \sigma|y) \propto p(\mu, \sigma)\,\ell(\mu, \sigma|y)$

Let $p(\mu, \sigma) \propto \frac{1}{\sigma}$; $-\infty < \mu < \infty$; $0 < \sigma < \infty$.

So,
$$p(\mu, \sigma|y) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\sum_1^n (y_i - \mu)^2\right\}.$$

$$\propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu - \hat{\mu})^2\right]\right\}.$$

& $$p(\mu|y) = \int_0^\infty p(\mu, \sigma|y)\,d\sigma$$
$$\propto \int_0^\infty \sigma^{-(n+1)} \exp.\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu - \hat{\mu})^2\right]\right\} d\sigma.$$

And, as noted in the notes from Theil's book, by putting
$$a = \left[(n-1)s^2 + n(\mu - \hat{\mu})^2\right]$$
$$\& \quad x = \left(a/2\sigma^2\right).$$

We get:
$$p(\mu|y) \propto \left\{(n-1)s^2 + n(\mu - \hat{\mu})^2\right\}^{-n/2} = \left\{(n-1)s^2 + n(\mu - \hat{\mu})^2\right\}^{-\left(\frac{\nu+1}{2}\right)}.$$

So $p(\mu|y) \sim u.s.(t)$

& $E(\mu) = \hat{\mu}$.

So the variable $t = \left\{\frac{(\mu-\hat{\mu})}{s/\sqrt{n}}\right\}$ takes a Student t-distribution with $v = (n-1)$ d.f.

Now consider the marginal pdf for σ:

$$p(\sigma|y) = \int_{-\infty}^{\infty} p(\sigma,\mu|y)\,d\mu$$

$$\propto \int_{-\infty}^{\infty} \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[vs^2 + n(\mu-\hat{\mu})^2\right]\right\} d\mu$$

Now, let $z = \frac{(\mu-\hat{\mu})}{\sigma/\sqrt{n}}$

$\therefore dz \propto \frac{d\mu}{\sigma}$

$\therefore d\mu \propto \sigma\, dz$

So, $$p(\sigma|y) \propto \int_{-\infty}^{\infty} \sigma^{-n} \exp\left\{-\frac{vs^2}{2\sigma^2}\right\} . \exp\left\{-z^2/2\right\} dz$$

$$\propto \sigma^{-n} \exp\left\{-\frac{vs^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \exp\left\{-z^2/2\right] dz$$

$$\therefore p(\sigma|y) \propto \sigma^{-n} \exp.\left\{-\frac{vs^2}{2\sigma^2}\right\}. \quad ; 0 < \sigma < \infty.$$

Now, the <u>inverted</u> <u>Gamma</u> distribution takes the form

$$p(y|\gamma,\alpha) = \left(\frac{2}{\Gamma(\alpha)\gamma^{\alpha} y^{2\alpha+1}}\right) e^{-1/\gamma y^2} \quad ; 0 < y < \infty$$

or, put $\sigma = y$, $\alpha = v/2$ & $\gamma = 2/vs^2$

$$\therefore p(\sigma|v,s) = \frac{2}{\Gamma(v/2)}\left(\frac{vs^2}{2}\right)\left(\frac{1}{\sigma^{v+1}}\right) e^{-vs^2/2\sigma^2} \quad ; 0 < \sigma < \infty.$$

So $p(\sigma|v,s) \propto \sigma^{-(v+1)} \exp\left\{-\frac{vs^2}{2\sigma^2}\right\}.$

So our posterior marginal is of the inverted Gamma form.

Point Estimates for Parameters —

Consider an estimate $\hat{\theta} = \hat{\theta}(y)$, & a loss fcn. given as $L(\hat{\theta};\theta)$.

Because θ is random, so is $L(\theta;\hat{\theta})$.

So min. $E[L(\hat{\theta};\theta)]$ in the posterior since —

$$\min_{\hat{\theta}} E[L(\hat{\theta};\theta)] = \min_{\hat{\theta}} \int_{R_\theta} L(\hat{\theta};\theta)\, p(\theta|y)\, d\theta.$$

e.g. if $L = (\theta-\hat{\theta})' C (\theta-\hat{\theta})$

then $E(L) = E(\theta - E\theta)' C (\theta - E\theta) - (\hat{\theta} - E\theta)' C (\hat{\theta} - E\theta)$

So pick $\hat{\theta} = E(\theta)$ & min. loss.

i.e. for the definite quadratic loss fcns, the optimal Bayes estimator is the mean of the posterior density.

A relationship between classical sampling theory & Bayesian estimation —

Let $\tilde{\theta} = \tilde{\theta}(y)$ be the sampling theory estimator.

$$r(\theta) = \int_{R_y} L(\theta;\tilde{\theta})\, p(y|\theta)\, dy$$

& no best $\tilde{\theta}$ $\forall \theta$, so take average —

$$\min_{\tilde{\theta}} \int_{R_\theta}\int_{R_y} L(\theta;\tilde{\theta})\, p(y|\theta)\, p(\theta)\, dy\, d\theta$$

$$= \min_{\tilde{\theta}} \int_{R_y} \left[\int_{R_\theta} L(\theta;\tilde{\theta})\, p(\theta|y)\, d\theta \right] p(y)\, dy$$

$$\therefore \min_{\tilde{\theta}} \int_{R_\theta} L(\theta;\tilde{\theta})\, p(\theta|y)\, d\theta.$$ i.e. <u>pick Bayes' Est.</u>

If the double integral converges, then it & the single integral give rise to the same min.

If it diverges, then ~~[illegible]~~ mins. for double doesn't exist, but $\exists$ min for single — this is a <u>quasi-Bayesian</u> estimator.

On an average risk criterion, the Bayes estimator gives the best performance in repeated sampling.

Bayesian Intervals & Regions —

Given that the posterior $p(\theta|y)$ has been obtained, then we can find the "probability" that θ lies in a given region. (NB that θ is the r.v., not the region.)

i.e. $$P_r(\theta \in \bar{R}|y) = \int_{\bar{R}} p(\theta|y)\,d\theta = \beta$$

This probability measures a degree of belief that $\theta \in \bar{R}$.
Conversely, we can fix the probability, β, & find the region, $\bar{R}$ (not necessarily unique).

Marginal Dist'n. of the Observations —

Sometimes we want the marginal pdf of the observations —

then $$p(y) = \int_{R_\theta} p(y,\theta)\,d\theta$$
$$= \int_{R_\theta} p(y|\theta)p(\theta)\,d\theta$$

This may be useful when evaluating the proportionality constant for our posterior pdf.

Predictive p.d.f. —

Given our sample, y, we want to make inferences about unobserved observations. Denote the latter by $\tilde{y}$.

Then, $$p(\tilde{y}|\theta,y) = p(\tilde{y},\theta|y)/p(\theta|y)$$

i.e. $$p(\tilde{y},\theta|y) = p(\tilde{y}|\theta,y)p(\theta|y)$$

Now, what we want is $p(\tilde{y}|y)$.

Well, $$p(\tilde{y}|y) = \int_{R_\theta} p(\tilde{y},\theta|y)\,d\theta$$

$$\therefore p(\tilde{y}|y) = \int_{R_\theta} p(\tilde{y}|\theta,y)p(\theta|y)\,d\theta.$$

Point Prediction –

We can use the predictive pdf, $p(\tilde{y}|y)$, to make point predictions. Let $\hat{y}$ be a point prediction for $\tilde{y}$. Then $\exists$ $L(\hat{y}, \tilde{y})$. So seek $\hat{y}$ to min. expected loss.

viz:

$$\min_{\hat{y}} \int_{R_{\tilde{y}}} L(\tilde{y}, \hat{y})\, p(\tilde{y}|y)\, d\tilde{y}.$$

If this integral can be solved & minimized, then we have an optimal point prediction. Again, the mean of the predictive pdf is an optimal point prediction if the loss function is quadratic.

Prediction Regions –

Given that we have the predictive pdf, $p(\tilde{y}|y)$, then for a given region $\bar{R}$ we can generally evaluate –

$$\Pr(\tilde{y} \in \bar{R}|y) = \int_{\bar{R}} p(\tilde{y}|y)\, d\tilde{y}.$$

Large-Sample Properties of $p(\theta|y)$ –

We have $p(\theta|y) \propto p(\theta)\, l(\theta|y)$.

$$\propto p(\theta) \exp.[\log l(\theta|y)].$$

The max. of $l(\theta|y)$ & of its log. occur at $\hat{\theta}$, the M.L.E.

Now, $\log[l(\theta|y)]$ is of order n, but $p(\theta)$ is independent of n.

So, as $n \to \infty$, the L/F tends to dominate the posterior. Also, as $n \to \infty$, the L/F tends to become normal, & its centre is at $\hat{\theta}$.

So as $n \to \infty$, $p(\theta|y)$ becomes $N(\hat{\theta}; \sigma^2)$.

To be more explicit –

$$p(\theta) = p(\hat{\theta}) + (\theta - \hat{\theta})p'(\hat{\theta}) + \tfrac{1}{2}(\theta - \hat{\theta})^2 p''(\hat{\theta}) + \cdots$$

$$= p(\hat{\theta})\left[1 + \frac{(\theta - \hat{\theta})p'(\hat{\theta})}{p(\hat{\theta})} + \frac{\frac{1}{2}(\theta - \hat{\theta})^2 p''(\hat{\theta})}{p(\hat{\theta})} + \cdots\right]$$

Now, put $g(\theta) = \log.[\ell(\theta|y)]$.

$$\therefore \exp\{g(\theta)\} = \exp\{g(\hat\theta) + (\theta-\hat\theta)g'(\hat\theta) + \tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta) + \cdots\}$$

But $g'(\hat\theta) = 0$ (at maximum)

So, $\exp\{g(\theta)\} = \exp\{g(\hat\theta)\}.\exp\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}.\exp\{\tfrac{1}{6}(\theta-\hat\theta)^3 g'''(\hat\theta)\}\cdots$

$$\therefore \ell(\theta|y) = \ell(\hat\theta|y).\exp\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}.\exp\{\tfrac{1}{6}(\theta-\hat\theta)^3 g'''(\hat\theta)\}\cdots$$

$$\propto \exp\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}\exp\{\tfrac{1}{6}(\theta-\hat\theta)^3 g'''(\hat\theta)\}\cdots$$

But $e^x = 1 + x + \frac{x^2}{2!} + \cdots$

$$\therefore \ell(\theta|y) \propto \exp.\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}[1 + \tfrac{1}{6}(\theta-\hat\theta)^3 g'''(\hat\theta) + \cdots]$$

$$\therefore p(\theta|y) \propto p(\hat\theta)\left[1 + \frac{(\theta-\hat\theta)p'(\hat\theta)}{p(\hat\theta)} + \cdots\right]$$

$$\circ \exp\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}[1 + \tfrac{1}{6}(\theta-\hat\theta)^3 g'''(\hat\theta) + \cdots]$$

And the leading term gives rise to:

$$p(\theta|y) \propto p(\hat\theta)\exp\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}$$

$$\propto \exp.\{\tfrac{1}{2}(\theta-\hat\theta)^2 g''(\hat\theta)\}$$

$$\therefore E[\theta|y] = \hat\theta \quad ; \quad \text{Var.}(\theta|y) = [-g''(\hat\theta)]^{-1}$$

$$= \left[-\frac{d^2 \log\{\ell(\theta|y)\}}{d\theta^2}\right]^{-1}_{\theta=\hat\theta}.$$

$$\therefore \underset{\lim_{n\to\infty}}{p(\theta|y)} \simeq \frac{|g''(\hat\theta)|^{1/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-\hat\theta)^2|g''(\hat\theta)|}.$$

Appendix : Diffuse Prior Distributions.

If a parameter's value is completely unknown, then Jeffreys suggests 2 rules for choosing a prior distribution —

(a) If $-\infty < \mu < \infty$, then take a uniform distribution:

$$p(\mu)\,d\mu \propto d\mu \quad ; \quad -\infty < \mu < \infty.$$

i.e. $p(\mu) \propto$ constant.

But $\int_{-\infty}^{\infty} p(\mu)\,d\mu = \infty$, so ∞ represents "certainty".

Now, $Pr(a < \mu < b) / Pr.(c < \mu < d) = (0/0)$ i.e. indeterminate so we can make no statement concerning the odds that μ lies in any particular pair of finite intervals. This represents formal ignorance.

However, other distributions other than the uniform one have this property of indeterminacy. Because we want our prior pdf to impart as little knowledge as possible, choose that pdf which minimizes information in some sense.

Measure information by

$$H = \int_{-m}^{m} p(\mu) \log p(\mu)\,d\mu.$$

Then $-H$ = "entropy", or "uncertainty."

Now, min H s.t. $\int_{-m}^{m} p(\mu)\,d\mu = 1$.

So, $\mathcal{L} = \int_{-m}^{m} p \log p\,d\mu + \lambda\left[\int_{-m}^{m} p\,d\mu - 1\right]$

$$\therefore \frac{\partial \mathcal{L}}{\partial p} = \int_{-m}^{m} \frac{\partial}{\partial p}[p \log p]\,d\mu + \lambda \int_{-m}^{m} \frac{\partial}{\partial p}[p]\,d\mu + 0$$

$$= \int_{-m}^{m} [\log p + 1]\,d\mu + \lambda \int_{-m}^{m} d\mu$$

$$= 2m \log p + 2m + \lambda 2m$$

$$= 2m(1 + \lambda + \log p]$$

$$= 0 \quad \text{for } \underline{\text{min}}.$$

So, $1 + \lambda + \log p(\mu) = 0$

$$\therefore p(\mu) = e^{-(1+\lambda)}$$

So set $1 + \lambda = \log 2m$

$$\therefore p(\mu) = \left(\frac{1}{2m}\right)$$

So the <u>uniform</u> dist'n. is the one to pick.

(b) If $0 < \sigma < \infty$, then pick $\log\sigma$ as uniform.
Let $\theta = \log\sigma$.

Then $0 < \sigma < \infty \Rightarrow -\infty < \theta < \infty$.

i.e. pick $p(\theta)d\theta \propto d\theta$; $-\infty < \theta < \infty$.

Now, $\theta = \log\sigma \Rightarrow d\theta = \frac{1}{\sigma}d\sigma$.

So $p(\sigma)d\sigma \propto \frac{d\sigma}{\sigma}$; $0 < \sigma < \infty$.

i.e. $p(\sigma) \propto \frac{1}{\sigma}$; $0 < \sigma < \infty$.

Note the following properties —

(i) Invariance to power transformations —

Let $\phi = \sigma^n$

$\therefore d\phi = n\sigma^{n-1}d\sigma$

$\therefore \frac{d\phi}{\phi} = \left(\frac{n\sigma^{n-1}}{\sigma^n}\right)d\sigma = n\frac{d\sigma}{\sigma}$.

$\therefore \frac{d\phi}{\phi} \propto \frac{d\sigma}{\sigma}$.

So it doesn't matter if we use σ or σ^2.

(ii) $\int_0^\infty p(\sigma)d\sigma \propto \int_0^\infty \left(\frac{d\sigma}{\sigma}\right) = \infty$. ; ∞ = certainty.

(iii) $\int_0^a \left(\frac{d\sigma}{\sigma}\right) = \infty$

(iv) $\int_a^\infty \left(\frac{d\sigma}{\sigma}\right) = \infty$.

$\therefore Pr(0<\sigma<a)/Pr(a<\sigma<\infty)$ is <u>indeter.</u>

So (iii) & (iv) again give us a formal representation of ignorance.

NB If we took $p(\sigma) \propto$ const.
Then $pr(\sigma > c) = \infty$ (for $c > 0$)
So we know $Pr(\sigma < c) = 0$ & that is some information. No good.

Now, if we take $p(\theta) \propto$ const., we min. H.
& this $\Rightarrow p(\sigma) \propto \frac{1}{\sigma}$; if $\theta = \log\sigma$. But is there some other transform of σ to use for θ ?
In general, we need $\theta = f(\sigma)$ when f is diff'ble, $f(0) = -\infty$, $f(\infty) = \infty$.
Further we need $f'(\sigma) = A\sigma^n$, where A & n are constants. Only if we take $n = -1$ will the ratio $Pr[0<\sigma<a]/Pr[a<\sigma<\infty]$ be indeterminate.

Some don't like to use Jeffreys' "improper" pdf's. They introduce a pdf which is ~~prior~~ flat over the dense part of the L/F. Outside of that range, it doesn't matter what the prior pdf does.

To demonstrate this —

$$p(\mu) = p(\mu_0) + (\mu-\mu_0)p'(\mu_0) + \tfrac{1}{2}(\mu-\mu_0)^2 p''(\mu_0) + \cdots$$

so $p(\mu) \simeq p(\mu_0)$ if flat

$$\therefore\ p(\mu|y) \propto p(\mu_0)\,p(y|\mu)$$

$$\therefore\ p(\mu|y) \propto p(y|\mu)$$

We are <u>not</u> satisfying Jeffreys' complete ignorance criterion, & this also implies that we know something about the L/F — this may not be so.

<u>Invariance</u> —

If $p(\theta) \propto |Inf_\theta|^{1/2}$

when Inf_θ = Fisher's Information Matrix

$$= -E\left[\frac{\partial^2 \log p(y|\theta)}{\partial\theta_i\,\partial\theta_j}\right]$$

Then if $\eta = F(\theta)$ is a 1-1 diff'ble transform,

$\Rightarrow p(\eta) \propto |Inf_\eta|^{1/2}$ will lead to consistency.

Formally, ~~$p(\eta)d\eta \propto p(\theta)d\theta$~~ $|Inf_\theta|^{\frac{1}{2}}d\theta = |Inf_\eta|^{\frac{1}{2}}d\eta$

<u>Proof</u>: $Inf_\theta = -E\left[\frac{\partial \log p}{\partial\theta_i}\cdot\frac{\partial \log p}{\partial\theta_j}\right]$.

Now, $(\partial \log p/\partial\theta_i) = \sum_{r=1}^{k}\left(\frac{\partial \log p}{\partial\eta_r}\right)\left(\frac{\partial\eta_r}{\partial\theta_i}\right)$.

$$\therefore\ Inf_{\theta_{ij}} = -E\left(\sum_r \frac{\partial \log p}{\partial\eta_r}\cdot\frac{\partial\eta_r}{\partial\theta_i}\right)\left(\sum_s \frac{\partial \log p}{\partial\eta_s}\cdot\frac{\partial\eta_s}{\partial\theta_j}\right)$$

$$= \sum_r\sum_s\left(\frac{\partial\eta_r}{\partial\theta_i}\right)\left(Inf_{\eta_{rs}}\right)\left(\frac{\partial\eta_s}{\partial\theta_j}\right)$$

where $Inf_{\eta_{rs}}$ is the (r,s)th. element of Inf_η.

$$\therefore\ Inf_\theta = [Inf_{\theta_{ij}}] = J\,Inf_\eta\,J'$$

where J is the Jacobian : $J = \left|\left[\frac{\partial\eta_j}{\partial\theta_i}\right]\right|$

<u>NB</u> $\eta = F(\theta)$

$\therefore\ |J|\,d\theta = d\eta$

So $|\mathrm{Inf}_\theta|^{\frac{1}{2}} = |J|\,|\mathrm{Inf}_\eta|^{\frac{1}{2}}$.

$\therefore |\mathrm{Inf}_\theta|^{\frac{1}{2}} = \left(\frac{d\eta}{d\theta}\right)|\mathrm{Inf}_\eta|^{\frac{1}{2}}$

$\therefore |\mathrm{Inf}_\theta|^{\frac{1}{2}} d\theta = |\mathrm{Inf}_\eta|^{\frac{1}{2}} d\eta$. QED.

Now, note the consistency, regardless of parametization, if the correct prior pdf is chosen –

$p(\theta|y) \propto p(\theta)\ell(\theta|y)$

But $p(\theta)d\theta = p(\eta)d\eta$

So $p(\eta|y) \propto p(\theta|y)$.

Now, let Ω = parameter space; $\mathcal{S}$ = sample space.

Let $\mathcal{F} = \{p(y|\theta)\,;\ \theta \in \Omega \subset R^k;\ y \in S \subset R^n\}$.

Then if $p(\theta|y) \propto p(\theta)\ell(\theta|y)$

& if $p(\theta) \propto |\mathrm{Inf}_\theta|^{\frac{1}{2}}$, the following invariance properties hold –

1) <u>S - labeling invariance:</u>

Let $z = G(y)$ be a 1-1 & diff'ble fctn. for $S \to S^*$. Then $p(\theta|z) \propto p(\theta|y)$.

) <u>Ω - labeling invariance</u>

If $\eta = F(\theta)$ is 1-1 & diff'ble, then

(a) $\exists\ p^*(y|\eta) \equiv p(y|\theta)$.

& (b) $p(\eta|y)d\eta \propto p(\theta|y)d\theta$.

) <u>Ω - restriction Invariance:</u>

Let $\theta \in \Omega^* \subset \Omega$. Then the posterior based on $p^*(y|\theta)$ with $\theta \in \Omega^*$ [viz $p^*(\theta|y) \propto p^*(\theta)p^*(y|\theta)$] will be proportional to $p(\theta|y)$ based on $p(y|\theta)$ for $\theta \in \Omega$

viz $p^*(\theta|y) \propto p(\theta|y)$.

(4) <u>Sufficiency Invariance</u>.

Let $t' = (t_1, \ldots, t_m)$ be sufficient for θ.
& let $p^*(t|\theta)$ be the corres. pdf.

Then $p^*(\theta|t) \propto p(\theta|y)$,

where $p^*(\theta|t) \propto p^*(\theta)\, p^*(t|\theta)$

& $p(\theta|y) \propto p(\theta)\, p(y|\theta)$
& both priors are $\propto |\mathrm{Inf}_\theta|^{1/2}$.

5). <u>Direct Product Invariance.</u>

Let y_1 & y_2 be independent samples.
Take $p_1(y_1|\theta_1)$; $p_2(y_2|\theta_2)$ where $\theta_1 \in \Omega_1$; $\theta_2 \in \Omega_2$.

& $p(y|\theta) \propto p_1(y_1|\theta_1)\, p_2(y_2|\theta_2)$

where $\theta \in \Omega = (\Omega_1 \times \Omega_2)$.

Then $p(\theta|y) \propto p_1(\theta_1|y_1)\, p_2(\theta_2|y_2)$

where $p_i(\theta_i|y_i) \propto p(\theta_i)\, p(y_i|\theta_i)$; $i = 1, 2$.

& all priors are prop. to $|\mathrm{Inf}|^{1/2}$.

6) <u>Repeated Product Invariance</u>.

Take $y_1, y_2, \ldots, y_m$ indep.

Then, $p(y_1, y_2, \ldots, y_m|\theta) = \prod_{i=1}^{m} p(y_i|\theta)$

and $p^*(\theta|y_1, \ldots, y_m) \propto p^*(\theta) \prod_{i=1}^{m} p(y_i|\theta)$

Take $p^*(\theta)$ & $p(\theta) \propto |\mathrm{Inf}_\theta|^{\frac{1}{2}}$.

Then $p^*(\theta|y_1, \ldots, y_m) \propto p(\theta|y_1) \prod_{i=2}^{m} p(y_i|\theta)$.

This is <u>"Repeated Product Invariance."</u>

(<u>NB</u> $p(\theta|y_1) \propto p(\theta)\, p(y_1|\theta)$. Hence other prior.).

Consider Jeffreys' prior: $p(\theta) \propto |\mathrm{Inf}_\theta|^{1/2}$. In what sense can this prior be considered as representing "knowing little"?

Let $p(y|\theta)$ be the pdf for y, given θ.

Define, $I_y(\theta) = \int_{R_y} p(y|\theta) \log. p(y|\theta) dy$

to measure the information in $p(y|\theta)$. Now, let the average prior information be

$$\overline{I}_y = \int_{R_\theta} I_y(\theta) p(\theta) d\theta$$

Let the gain in information in going from the prior to the data be

$$G = \overline{I}_y - \int_{R_\theta} p(\theta) \log p(\theta) d\theta$$

Then a "minimal information" pdf is one which maximizes G for given $p(y|\theta)$.

Example: Suppose that $p(y|\theta) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(y-\theta)^2]$.

~~Then, $I_y(\theta) = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)$~~

Then, $I_y(\theta) = \int_{-\infty}^{\infty} p(y|\theta)\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}(y-\theta)^2\right) dy$

$$= -\frac{1}{2}\log 2\pi \int_{-\infty}^{\infty} p(y|\theta) dy - \frac{1}{2}\int_{-\infty}^{\infty} (y-\theta)^2 p(y|\theta) dy$$

$$= \left(-\frac{1}{2}\log 2\pi\right) - \frac{1}{2}\int_{-\infty}^{\infty} y^2 p(y|\theta) dy + \theta \int_{-\infty}^{\infty} y\, p(y|\theta) dy$$

$$- \frac{1}{2}\theta^2 \int_{-\infty}^{\infty} p(y|\theta) dy$$

$$= \left(-\frac{1}{2}\log 2\pi\right) - \frac{1}{2}[\mathrm{var}\, y + [E(y)]^2]$$

$$+ \theta . E(y) - \frac{1}{2}\theta^2$$

$$= \left(-\frac{1}{2}\log 2\pi\right) - \frac{1}{2} - \frac{1}{2}\theta^2 + \theta^2 - \frac{1}{2}\theta^2$$

$$= -\frac{1}{2}[\log 2\pi + 1].$$

So, $\overline{I}_y = \int_{R_\theta} I_y(\theta) p(\theta) d\theta = -\frac{1}{2}[\log 2\pi + 1] \int_{R_\theta} p(\theta) d\theta$

$$= -\frac{1}{2}[\log 2\pi + 1].$$

$$\therefore G = \overline{I_y} - \int p(\theta) \log p(\theta) d\theta$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \int p(\theta) \log p(\theta) d\theta.$$

& since the 1st. term is independent of θ, we max. G. by min. $\int p(\theta) \log p(\theta) d\theta$.

& this $\Longrightarrow$ <u>uniform dist'n.</u>

i.e. $p(\theta) \propto |\mathrm{Inf}_\theta|^{1/2} \propto$ constant.

So we have a justification for Jeffreys' prior.

Example: $p(y|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{y^2}{2\sigma^2}\right]$; $-\infty < y < \infty$.

$$\therefore I_y(\sigma) = \int p(y|\sigma) \log p(y|\sigma) dy$$

$$= \int_{-\infty}^{\infty} p(y|\sigma)\left[-\log\sigma - \tfrac{1}{2}\log 2\pi - \tfrac{y^2}{2\sigma^2}\right] dy$$

$$= -\log\sigma - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2\sigma^2}\int y^2 p(y|\sigma) dy$$

$$= -\log\sigma - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2\sigma^2}[\mathrm{var}(y) + (E(y))^2]$$

$$= -\log\sigma - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2\sigma^2}[\sigma^2 + 0]$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \log\sigma.$$

$$\therefore \overline{I_y} = \int I_y(\sigma) p(\sigma) d\sigma$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \int \log\sigma \, p(\sigma) d\sigma$$

$$\therefore G = \overline{I_y} - \int p(\sigma) \log p(\sigma) d\sigma$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \int \log\sigma \, p(\sigma) d\sigma - \int p(\sigma) \log p(\sigma) d\sigma$$

Now, maximize G s.t. $\int p(\sigma) d\sigma = 1$.

$$\therefore \mathcal{L} = -\tfrac{1}{2}[\log 2\pi + 1] - \int \log\sigma \, p(\sigma) d\sigma - \int p(\sigma) \log p(\sigma) d\sigma + \lambda\left[\int p(\sigma) d\sigma - 1\right]$$

$$\frac{\partial \mathcal{L}}{\partial p} = 0 - \int \frac{\partial}{\partial p}[p \log\sigma] d\sigma - \int \frac{\partial}{\partial p}[p \log p] d\sigma + \lambda \int \frac{\partial}{\partial p}[p] d\sigma$$

$$= -\int \log\sigma \, d\sigma - \int [\log p + 1] d\sigma + \lambda \int d\sigma$$

$$= -[\sigma\log\sigma - \sigma] - [\log p + 1]\sigma + \lambda\sigma$$

$$= \sigma[1 - \log\sigma - \log p - 1 + \lambda]$$

$$= 0.$$

$$\therefore \quad \lambda - \log p - \log\sigma = 0$$

$$\therefore \lambda = \log p + \log\sigma$$
$$= \log(p\sigma)$$

$$\therefore p\sigma = e^{\lambda}$$

$$\therefore p = \frac{e^{\lambda}}{\sigma} \quad \propto \frac{1}{\sigma}.$$

So this result is also in accordance with

$$p(\sigma) \propto |\mathrm{Inf}_\sigma|^{\frac{1}{2}} \propto 1/\sigma.$$

<u>Example</u>: $\quad p(y|\sigma,\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right] \; ; -\infty < y < \infty.$

$$\therefore I_y(\theta,\sigma) = \int p(y|\theta,\sigma) \log p(y|\theta,\sigma)\, dy.$$

$$= \int p(y|\theta,\sigma)\left[-\log\sigma - \tfrac{1}{2}\log 2\pi - \tfrac{1}{2\sigma^2}(y-\theta)^2\right] dy$$

$$= \left(-\log\sigma - \tfrac{1}{2}\log 2\pi\right) - \frac{1}{2\sigma^2}\int (y-\theta)^2\, p(y|\theta,\sigma)\, dy$$

$$= -\log\sigma - \tfrac{1}{2}\log 2\pi - \frac{1}{2\sigma^2}[\mathrm{var}\, y]$$

$$= -\log\sigma - \tfrac{1}{2}\log 2\pi - \frac{\sigma^2}{2\sigma^2}$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \log\sigma.$$

Then $\quad \bar{I}_y = \iint I_y(\theta,\sigma)\, p(\theta,\sigma)\, d\theta\, d\sigma$

$$= \iint \left\{-\tfrac{1}{2}[\log 2\pi + 1] - \log\sigma\right\} p(\theta,\sigma)\, d\theta\, d\sigma$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] \iint p(\theta,\sigma)\, d\theta\, d\sigma - \iint \log\sigma\, p(\theta,\sigma)\, d\theta\, d\sigma$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \iint \log\sigma\, p(\theta,\sigma)\, d\theta\, d\sigma$$

$$\therefore G = \bar{I}_y - \iint p(\theta,\sigma) \log p(\theta,\sigma)\, d\theta\, d\sigma$$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \iint \log\sigma\, p(\sigma,\theta)\, d\theta\, d\sigma - \iint p(\theta,\sigma) \log p(\theta,\sigma)\, d\theta\, d\sigma$$

Max. G s.t. $\iint p(\theta,\sigma)\, d\theta\, d\sigma = 1.$

$$= -\tfrac{1}{2}[\log 2\pi + 1] - \iint \log\sigma\, p(\sigma,\theta)\, d\theta\, d\sigma - \iint p(\theta,\sigma) \log p(\theta,\sigma)\, d\theta\, d\sigma$$
$$+ \lambda\left[\iint p(\theta,\sigma)\, d\theta\, d\sigma - 1\right]$$

$$\therefore \frac{\partial \mathcal{L}}{\partial p} = 0 - \iint \frac{\partial}{\partial p}[p\log\sigma]\,d\theta\,d\sigma - \iint \frac{\partial}{\partial p}[p\log p]\,d\sigma\,d\theta$$
$$+\lambda \iint \frac{\partial}{\partial p}[p]\,d\theta\,d\sigma + 0$$
$$= -\iint \log\sigma\,d\theta\,d\sigma - \iint [\log p + 1]\,d\sigma\,d\theta + \lambda \iint d\theta\,d\sigma$$
$$= -\int [\sigma\log\sigma - \sigma]\,d\theta - [\log p + 1]\iint d\sigma\,d\theta + \lambda \iint d\sigma\,d\theta$$
$$= [\sigma - \sigma\log\sigma]\theta - \sigma\theta[\log p + 1] + \lambda\theta\sigma$$
$$= \sigma\theta[1 - \log\sigma - \log p + 1 + \lambda]$$
$$= 0$$

$$\therefore \lambda - \log\sigma - \log p = 0$$
$$\therefore \lambda = \log p + \log\sigma = \log(p\sigma)$$
$$\therefore p = (e^{\lambda}/\sigma) \propto \frac{1}{\sigma}, \text{ as required.}$$
$$\text{i.e. } p(\theta,\sigma) \propto \frac{1}{\sigma}.$$

Example: Consider the information matrix for

$$p(y|\theta,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right]$$

$$\therefore \log p(y|\theta,\sigma) = -\log\sigma - \tfrac{1}{2}\log 2\pi - \frac{1}{2\sigma^2}(y-\theta)^2$$

$$\therefore \cancel{\partial \log L}$$

$$\log L = -n\log\sigma - \tfrac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_i (y_i-\theta)^2$$

$$\frac{\partial \log L}{\partial \sigma} = \frac{-n}{\sigma} + 0 + \frac{1}{\sigma^3}\sum (y_i-\theta)^2$$

$$\frac{\partial^2 \log L}{\partial\sigma\,\partial\theta} = \frac{1}{\sigma^3}\left[\sum \frac{\partial}{\partial\theta}(y_i-\theta)^2\right]$$
$$= \left(\frac{1}{\sigma^3}\right)\frac{\partial}{\partial\theta}\left[\sum (y_i^2 + \theta^2 - 2y_i\theta)\right]$$
$$= \left(\frac{1}{\sigma^3}\right)\sum (2\theta - 2y_i)$$
$$= \frac{2}{\sigma^3}[n\theta - \cancel{2}\sum y_i]$$

$$\therefore -E\left[\frac{\partial^2 \log L}{\partial\sigma\,\partial\theta}\right] = -\frac{2}{\sigma^3}[n\theta - \cancel{2}E(\sum y_i)]$$
$$= -\frac{2}{\sigma^3}[n\theta - n\theta] = 0. = -E\left[\frac{\partial^2 \log L}{\partial\theta\,\partial\sigma}\right].$$

& $\dfrac{\partial \log L}{\partial \sigma} = -\dfrac{n}{\sigma} + 0 + \dfrac{1}{\sigma^3}\sum(y_i-\theta)^2$

$\therefore \left(\dfrac{\partial^2 \log L}{\partial\sigma\partial\sigma}\right) = \dfrac{n}{\sigma^2} - \dfrac{3}{\sigma^4}\sum(y_i-\theta)^2$

$\therefore +E\left[\dfrac{\partial^2 \log L}{\partial\sigma\partial\sigma}\right] = \left(n/\sigma^2\right) - \left(\dfrac{3}{\sigma^4}\right)\sum E(y_i-\theta)^2$

$= \left(n/\sigma^2\right) - \left(\dfrac{3}{\sigma^4}\right)\cdot n\sigma^2$

$= n/\sigma^2 - \dfrac{3n}{\sigma^2}$

$= -\dfrac{2n}{\sigma^2}$

$\therefore -E(\ ) = 2n/\sigma^2$

& $\dfrac{\partial \log L}{\partial \theta} = 0 + 0 - \dfrac{1}{2\sigma^2}\sum \dfrac{\partial}{\partial\theta}(y_i-\theta)^2$

$= -\dfrac{1}{2\sigma^2}\sum \dfrac{\partial}{\partial\theta}(y_i^2+\theta^2-2y_i\theta)$

$= -\dfrac{1}{2\sigma^2}\sum(2\theta-2y_i)$

$= -\dfrac{1}{\sigma^2}\sum(\theta-y_i)$

$\therefore \dfrac{\partial^2 \log L}{\partial\theta\partial\theta} = -\dfrac{1}{\sigma^2}\sum(1)$

$= -n/\sigma^2$

$\therefore -E(\ ) = n/\sigma^2$

$\therefore \ \mathrm{Inf}_{\theta,\sigma} = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}.$

$\therefore \ |\mathrm{Inf}_{\theta,\sigma}|^{\frac{1}{2}} = \left(\dfrac{n}{\sigma^2}\cdot\dfrac{2n}{\sigma^2}\right)^{\frac{1}{2}} = \left(\dfrac{2n^2}{\sigma^4}\right)^{\frac{1}{2}} = \sqrt{2}\left(\dfrac{n}{\sigma^2}\right)$

$\propto \left(\dfrac{1}{\sigma^2}\right)$, not $\propto\left(\dfrac{1}{\sigma}\right)$.

Note also that

$$p(\theta,\sigma) \propto \frac{1}{\sigma}$$

is invariant to transformations of the kind $\eta = \theta + k\sigma$.

If we consider an <u>asymptotic</u> form of G and write:

$$G_a = \int p(\theta) \log\sqrt{n\,|\mathrm{Inf}_\theta|}\, d\theta - \int p(\theta)\log p(\theta)\, d\theta$$

Maximize G_a s.t. $\int p(\theta)\,d\theta = 1$

$$\Rightarrow \quad p(\theta) \propto |\mathrm{Inf}_\theta|^{\frac{1}{2}}$$

So, for the asymptotic form of G, Jeffreys' prior is a minimal information prior. However, for the usual form of G, Jeffreys' invariant priors are <u>not always</u> minimal information prior pdf's since they do not always maximize G. [e.g. see third example above.]

If the Jeffreys' prior is <u>not</u> the one which max. G, then use of such a prior introduces more information than need be. When the no. of parameters is large, this discrepency may be important!

Zellner: Chapter 3.

Regression Model.

Let $y_i = \beta_1 + \beta_2 x_i + u_i \quad ; \quad i = 1, 2, \ldots, n.$

Assume:

(1) $u_i \sim IN(0, \sigma^2)$

(2) x_i fixed & non-stochastic

OR

(3) x_i random but distributed indep. of u_i, & with a p.d.f. independent of β_1, β_2 & σ.

Then, $p(y \mid X, \beta_1, \beta_2, \sigma) \propto \left(\frac{1}{\sigma^n}\right) \exp\left[-\frac{1}{2\sigma^2}\sum(y_i - \beta_1 - \beta_2 x_i)^2\right]$

Take a diffuse prior –

$p(\beta_i) \propto \text{const} \quad ; \quad i = 1, 2. \qquad -\infty < \beta_i < \infty$

$p(\sigma) \propto \frac{1}{\sigma} \quad ; \qquad 0 < \sigma < \infty$

& independent

$\therefore \; p(\beta_1, \beta_2, \sigma) \propto 1/\sigma \quad ; \quad -\infty < \beta_i < \infty \; ; \; 0 < \sigma < \infty.$

$\therefore \; p(\beta_1, \beta_2, \sigma \mid y, X) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\sum(y_i - \beta_1 - \beta_2 x_i)^2\right]$

Now, consider $\sum(y_i - \beta_1 - \beta_2 x_i)^2 = \gamma$

$$\therefore \; \gamma = \sum\left\{(y_i - \hat\beta_1 - \hat\beta_2 x_i) - \left[(\beta_1 - \hat\beta_1) + (\beta_2 - \hat\beta_2)x_i\right]\right\}^2$$

$$= \sum(y_i - \hat\beta_1 - \hat\beta_2 x_i)^2 + \sum(\beta_2 - \hat\beta_2)^2 x_i^2 + n(\beta_1 - \hat\beta_1)^2 + 2(\beta_1 - \hat\beta_1)(\beta_2 - \hat\beta_2)\sum x_i$$

$$- 2\sum(y_i - \hat\beta_1 - \hat\beta_2 x_i)\left[(\beta_1 - \hat\beta_1) + (\beta_2 - \hat\beta_2)x_i\right]$$

$$= \sum r_i^2 + (\beta_2 - \hat\beta_2)^2 \sum x_i^2 + \alpha$$

$$= \nu s^2 + (\beta_2 - \hat\beta_2)^2 \sum x_i^2 + \alpha$$

Now, $\alpha = -2\sum(y_i - \hat\beta_1 - \hat\beta_2 x_i)(\beta_1 - \hat\beta_1) - 2\sum(y_i - \hat\beta_1 - \hat\beta_2 x_i)(\beta_2 - \hat\beta_2)x_i$

$$= -2\left(\sum r_i\right)(\beta_1 - \hat\beta_1) - 2(\beta_2 - \hat\beta_2)\sum r_i x_i$$

$$= 0 - 0$$

$$= 0$$

So, $\gamma = \nu s^2 + (\beta_2 - \hat\beta_2)^2 \sum x_i^2 + n(\beta_1 - \hat\beta_1)^2 + 2(\beta_1 - \hat\beta_1)(\beta_2 - \hat\beta_2)\sum x_i$

$$\therefore\ p(\beta_1, \beta_2, \sigma \mid y, X) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta_2-\hat{\beta}_2)^2 \Sigma x_i^2 + n(\beta_1-\hat{\beta}_1)^2 + 2(\beta_1-\hat{\beta}_1)(\beta_2-\hat{\beta}_2)\Sigma x_i\right]\right\}$$

Now, consider the case where σ is <u>given</u>

Then $p(\beta_1, \beta_2, \sigma_0 \mid y, X) \propto \left(\frac{1}{\sigma_0^{n+1}}\right) \exp\left\{-\frac{\nu s^2}{2\sigma_0^2}\right\}$

$$\cdot \exp\left\{-\frac{1}{2\sigma_0^2}\left[(\beta_2-\hat{\beta}_2)^2 \Sigma x_i^2 + n(\beta_1-\hat{\beta}_1)^2 + 2(\beta_1-\hat{\beta}_1)(\beta_2-\hat{\beta}_2)\Sigma x_i\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}\left[(\beta_2-\hat{\beta}_2)^2 \Sigma x_i^2 + n(\beta_1-\hat{\beta}_1)^2 + 2(\beta_1-\hat{\beta}_1)(\beta_2-\hat{\beta}_2)\Sigma x_i\right]\right\}$$

And this is bivariate normal.

Recall the MVN model:

$$p(x \mid \theta, \Sigma) = \frac{1}{(2\pi)^n |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\theta)\Sigma^{-1}(x-\theta)'\right\}$$

Now, let $x = (\beta_1, \beta_2)$

$\theta = (\hat{\beta}_1, \hat{\beta}_2)$

$\Sigma = \sigma^2 \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}^{-1}$

$$\therefore\ (x-\theta)\Sigma^{-1}(x-\theta)' = \frac{1}{\sigma^2}\begin{bmatrix}\beta_1-\hat{\beta}_1 \\ \beta_2-\hat{\beta}_2\end{bmatrix}' \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2\end{bmatrix} \begin{bmatrix}\beta_1-\hat{\beta}_1, & \beta_2-\hat{\beta}_2\end{bmatrix}'$$

$$= \frac{1}{\sigma^2}\left[n(\beta_1-\hat{\beta}_1)^2 + (\beta_2-\hat{\beta}_2)\Sigma x_i,\ (\beta_1-\hat{\beta}_1)\Sigma x_i + (\beta_2-\hat{\beta}_2)\Sigma x_i^2\right]\begin{bmatrix}\beta_1-\hat{\beta}_1 \\ \beta_2-\hat{\beta}_2\end{bmatrix}$$

$$= \frac{1}{\sigma^2}\left[n(\beta_1-\hat{\beta}_1)^2 + (\beta_2-\hat{\beta}_2)^2 \Sigma x_i^2 + 2(\beta_1-\hat{\beta}_1)(\beta_2-\hat{\beta}_2)\Sigma x_i\right]$$

As required.

But σ^2 is rarely known, so this is of little help.

To look at the <u>marginal</u> pdf for β_1 & β_2 only.

$$p(\beta_1, \beta_2 \mid y, X) = \int_0^\infty p(\beta_1, \beta_2, \sigma \mid y, X)\, d\sigma$$

$$\propto \int_0^\infty \left(\frac{1}{\sigma^{n+1}}\right) \exp\left[-\frac{c}{2\sigma^2}\right] d\sigma$$

where $c = [\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)^2 \Sigma x_i^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i]$

Let $z = \left(\frac{c}{\sigma^2}\right)$

$$\therefore dz = -2c\sigma^{-3} d\sigma$$

& $\sigma = (c/z)^{\frac{1}{2}}$

$$\therefore \sigma^{-(n+1)} = (c/z)^{-\frac{1}{2}(n+1)}$$

$$\therefore p(\beta_1, \beta_2 \mid y, X) \propto \int_0^\infty \left(\frac{1}{2c}\right)(c/z)^{3/2}(c/z)^{-\frac{1}{2}(n+1)} e^{-z/2}\, dz$$

$$\propto \tfrac{1}{2} c^{-n/2} \int_0^\infty z^{n/2 - 1} e^{-z/2}\, dz$$

$$\propto c^{-n/2}$$

Since the integrand is $\Gamma(n/2, 2)$.

So, $p(\beta_1, \beta_2 \mid y, X) \propto [\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i + (\beta_2 - \hat{\beta}_2)^2 \Sigma x_i^2]^{-\frac{n}{2}}$

Now, this is a Bivariate t- dist'n, since the general form of the MVT is

$$p(x \mid \theta, V, \nu, m) \propto [\nu + (x - \theta)' V (x - \theta)]^{-(m+\nu)/2}$$

So,

$$p(\beta_1 \mid y, x) \propto \left[\nu + \frac{\Sigma(x_i - \bar{x})^2}{s^2 \Sigma x_i^2 / n} (\beta_1 - \hat{\beta}_1)^2\right]^{-\frac{\nu+1}{2}} ; -\infty < \beta_1 < \infty$$

$$p(\beta_2 \mid y, x) \propto \left[\nu + \frac{\Sigma(x_i - \bar{x})^2}{s^2} (\beta_2 - \hat{\beta}_2)^2\right]^{-\frac{(\nu+1)}{2}} ; -\infty < \beta_2 < \infty.$$

Then

$$\left.\begin{aligned} &\left(\frac{\Sigma(x_i - \bar{x})^2}{s^2 \Sigma x_i^2 / n}\right)^{\frac{1}{2}} (\beta_1 - \hat{\beta}_1) = t_\nu \\ \& \quad &\left(\frac{\Sigma(x_i - \bar{x})^2}{s^2}\right)^{\frac{1}{2}} (\beta_2 - \hat{\beta}_2) = t_\nu \end{aligned}\right\} \text{Student t with } \nu \text{ d.f.}$$

Also,

$$p(\sigma \mid y, x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\beta_1, \beta_2, \sigma \mid y, x)\, d\beta_1\, d\beta_2$$

$$\propto \left(\frac{1}{\sigma^{v+1}}\right) \exp\left[\frac{-vs^2}{2\sigma^2}\right] \quad ; 0 < \sigma < \infty.$$

And this is in the form of an inverted-Γ dist'n.

$$p(\sigma \mid v, s) = \frac{2}{\Gamma(v/2)} \left(\frac{vs^2}{2}\right)^{v/2} \left(\frac{1}{\sigma^{v+1}}\right) \exp\left[\frac{-vs^2}{2\sigma^2}\right]$$

$$\Rightarrow E(\sigma) = s\left(\sqrt{\frac{v}{2}}\right)^{\frac{1}{2}} \frac{\Gamma[(v-1)/2]}{\Gamma(v/2)}$$

$$var(\sigma) = \left(\frac{s^2 v}{v-2}\right) - [E(\sigma)]^2.$$

Transform σ to σ^2:

$$\Rightarrow p(\sigma^2 \mid y, x) \propto [(\sigma^2)^{v/2}]^{-1} \exp\left[\frac{-vs^2}{2\sigma^2}\right] \quad ; 0 < \sigma < \infty$$

Simply by re-writing the expression obtained already.

Or, $p(h \mid y, x) \propto \left(h^{v/2-1}\right) \exp\left(\frac{-vs^2 h}{2}\right) \; ; 0 < h < \infty$

where $h = (1/\sigma^2)$

So since the χ^2 dist'n. is just Γ with $\alpha = v/2, \gamma = 2$,

then $p(x \mid \alpha, \gamma) \propto x^{v/2-1} e^{-x/2}$

$\therefore$ $p(h \mid y, x)$ is just $\chi^2_{(v)}$.

So we can make inferences about β_1, β_2 & $\frac{1}{\sigma^2}$ individually. What about <u>joint</u> inferences on (β_1, β_2)?

Consider the expression

$$\Psi = \frac{1}{2s^2}\left[n(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)\Sigma x_i^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i\right]$$

Then $\Psi \sim F_{2,v}$.

To show this, we must convert $p(\beta_1, \beta_2 | y, X)$ to $p(\psi | y, X.)$ Do this in several steps:

(a) Let $\delta' = (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2) = (\delta_1, \delta_2)$ say.

Then if $A \equiv \frac{1}{2s^2}\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$

$$\Rightarrow \delta' A \delta = (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2) \frac{1}{2s^2}\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{pmatrix} \beta_1 - \hat{\beta}_1 \\ \beta_2 - \hat{\beta}_2 \end{pmatrix}$$

$\therefore \delta' A \delta = \psi.$

NB go from $\beta \Rightarrow \delta$ so $J = \left|\frac{\partial \beta_i}{\partial \delta_j}\right|$.

Now, $p(\delta | y, x) = p(\beta_1, \beta_2 | y, x) |J|$

& $$|J|^{-1} = \begin{vmatrix} \frac{\partial \delta_1}{\partial \beta_1} & \frac{\partial \delta_1}{\partial \beta_2} \\ \frac{\partial \delta_2}{\partial \beta_1} & \frac{\partial \delta_2}{\partial \beta_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

$\therefore |J| = 1.$

So, $p(\delta | y, x) = p(\beta_1, \beta_2 | y, x)$

Now, $\Rightarrow p(\delta | y, x) \propto [\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + \Sigma x_i^2 (\beta_2 - \hat{\beta}_2)^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i]^{-n/2}$

$$\propto \nu s^2 [1 + 2(\delta' A \delta)/\nu]^{-n/2}$$

$$\propto [1 + (\tfrac{2}{\nu})\delta' A \delta]^{-n/2}$$

Now, $A = K'K$

$\therefore \delta' A \delta = \delta' K' K \delta = (K\delta)'(K\delta) = V'V.$

$\therefore V = K\delta = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ & $K\delta = \begin{pmatrix} K\delta_1 \\ K\delta_2 \end{pmatrix}$

So, $\delta = (V/K) = \frac{1}{K}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$

$$\therefore |J| = \begin{vmatrix} \frac{\partial \delta_1}{\partial v_1} & \frac{\partial \delta_1}{\partial v_2} \\ \frac{\partial \delta_2}{\partial v_1} & \frac{\partial \delta_2}{\partial v_2} \end{vmatrix} = \begin{vmatrix} 1/K & 0 \\ 0 & 1/K \end{vmatrix} = 1/K^2$$

$$\therefore p(V | y, x) = (\tfrac{1}{K^2}) p(\delta | y, x)$$

$$\propto (\tfrac{1}{K^2})[1 + (\tfrac{2}{\nu})V'V]^{-n/2}$$

$$\propto [1 + (\tfrac{2}{\nu})V'V]^{-n/2}$$

& let $V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \psi^{\frac{1}{2}} \cos\theta \\ \psi^{\frac{1}{2}} \sin\theta \end{pmatrix}$ ✓

$$\therefore\ p(\Psi|y,x) = |J| \cdot p(V|y,x)$$

$$\& \ |J| = \begin{vmatrix} \frac{\partial v_1}{\partial \Psi} & \frac{\partial v_1}{\partial \theta} \\ \frac{\partial v_2}{\partial \Psi} & \frac{\partial v_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \frac{1}{2}\Psi^{-\frac{1}{2}}\cos\theta & -\Psi^{\frac{1}{2}}\sin\theta \\ \frac{1}{2}\Psi^{-\frac{1}{2}}\sin\theta & \Psi^{\frac{1}{2}}\cos\theta \end{vmatrix}$$

$$= \tfrac{1}{2}(\sin^2\theta + \cos^2\theta$$
$$= \tfrac{1}{2}$$

$$\therefore\ p(\Psi|y,x) \propto \tfrac{1}{2}\left[1 + \left(\tfrac{2}{v}\right)V'V\right]^{-n/2}$$

But, $V'V = \Psi(\cos^2\theta + \sin^2\theta)$
$= \Psi$

$$\therefore\ p(\Psi|y,x) \propto \left[1 + \tfrac{2}{v}\cdot\Psi\right]^{-\left(\frac{v+2}{2}\right)}$$

$$\& \ p(F_{m,v}) \propto F^{\frac{m-2}{2}}\left[1 + F\left(\tfrac{m}{q}\right)\right]^{-\left(\frac{m+q}{2}\right)}$$

So put $m = 2$.

Then $p \propto \left[1 + \left(\tfrac{2}{q}\right)F\right]^{-(q+2)/2}$.

So, $\Psi \sim F_{2,v}$, as required.

Note that if 2 variables are distributed as a Bivariate t-dist'n, then a linear combination of those variables is distributed as a univariate t.

Let $E(y|x = x_0) = \beta_1 + \beta_2 x_0$. (Use for <u>predicting</u>)

Then $\hat{\eta}_0 = \hat{\beta}_1 + x_0\hat{\beta}_2$; i.e. predict with <u>estimated</u> equation.

Change variables β_1 & β_2 to η_0 & β_2 in the joint posterior pdf for β_1 & β_2.

Viz. let $\eta_0 - \hat{\eta}_0 = (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)x_0$

$$\beta_2 - \hat{\beta}_2 = (\beta_2 - \hat{\beta}_2)$$

Then $|J| = \begin{vmatrix} \frac{\partial \eta_0}{\partial \beta_1} & \frac{\partial \eta_0}{\partial \beta_2} \\ \frac{\partial \beta_2}{\partial \beta_1} & \frac{\partial \beta_2}{\partial \beta_2} \end{vmatrix} = \begin{vmatrix} 1 & x_0 \\ 0 & 1 \end{vmatrix} = 1$

$$\therefore \; p(\eta_0, \beta_2 \mid y, x) = |J| \; p(\beta_1, \beta_2 \mid y, x)$$

$$\propto \left[\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + \Sigma x_i^2 (\beta_2 - \hat{\beta}_2)^2 + \Sigma x_i (\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \right]^{-\frac{\eta}{2}}$$

& get $$p(\eta_0 \mid y, x) = \int_{-\infty}^{\infty} p(\eta_0, \beta_2 \mid y, x) \, d\beta_2$$

$$\propto \left[\nu + \frac{n \Sigma (x_i - \bar{x})^2}{s^2 \Sigma (x_i - \bar{x}_0)^2} (\eta_0 - \hat{\eta}_0)^2 \right]^{-\frac{(\nu+1)}{2}}.$$

Then, $$\frac{\eta_0 - \hat{\eta}_0}{s \left[1/n + (x_0 - \bar{x})^2 / \Sigma (x_i - \bar{x})^2 \right]^{\frac{1}{2}}} \sim t_\nu.$$

So use this for forecast intervals.

In practice, how do we make use of the distributions of the parameters to get posterior graphs?

We know that $$\frac{(\beta_2 - \hat{\beta}_2)}{s / [\Sigma (x_i - \bar{x})^2]^{\frac{1}{2}}} \sim t_\nu.$$

Now, $\Rightarrow dt_\nu = k \, d\beta_2$, where $k = \dfrac{\sqrt{\Sigma (x_i - \bar{x})^2}}{s}$.

Now, $p(t_\nu)$ is given by the tables of the t-dist'n.

So $p(t_\nu) \, dt_\nu = p(t_\nu) \, k \, d\beta_2$.

So $p(\beta_2)$ is got by taking $k \, p(t_\nu)$.

Take the multiple regression model.

$$y = X\beta + u$$

$$\therefore\ p(y|X,\beta,\sigma) \propto \left(\frac{1}{\sigma^n}\right) \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

$$\& \left[(y-X\beta)'(y-X\beta)\right] = \left[y - X\hat{\beta} - X(\beta-\hat{\beta})\right]'\left[y - X\hat{\beta} - X(\beta-\hat{\beta})\right]$$

$$= (y-X\hat{\beta})'(y-X\hat{\beta}) + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})$$

$$= \nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})$$

So,

$$p(y|X,\sigma,\beta) \propto \left(\frac{1}{\sigma^n}\right) \exp\left[-\frac{1}{2\sigma^2}\left\{\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right\}\right]$$

Let $p(\beta) \propto$ const ; $-\infty < \beta_i < \infty$

& $p(\sigma) \propto 1/\sigma$; $0 < \sigma < \infty$

So independence $\Rightarrow p(\beta,\sigma) \propto 1/\sigma$.

$$\therefore\ p(\beta,\sigma|y,X) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right]\right\}.$$

$\therefore\ p(\beta|\sigma,y,X)$ is $M.V.N\left(\hat{\beta}, (X'X)^{-1}\sigma^2\right)$.

Now, $$p(\beta|X,y) = \int_0^\infty p(\beta,\sigma|y,X)\,d\sigma$$

Let $c = \left[\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})\right]$

Let $z = c/\sigma^2$

$$\therefore\ dz = -\left(\frac{2c}{\sigma^3}\right)d\sigma$$

$$\therefore\ d\sigma = -(\sigma^3/2c)\,dz$$

& $\sigma = (c/z)^{\frac{1}{2}}$ $\quad\therefore\ \sigma^{-(n+1)} = (c/z)^{-\frac{1}{2}(n+1)}$

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

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

$$\therefore \; I = \int_0^\infty \tfrac{1}{2} c^{-n/2} \left(\tfrac{1}{z}\right)^{-n/2+1} e^{-z/2} dz$$

$$\propto c^{-n/2} \int_0^\infty z^{n/2-1} e^{-z/2} dz$$

$$\propto c^{-n/2} \qquad (\text{'cos } \textstyle\int \text{ is } \Gamma\text{-dist'n})$$

$$\therefore \; p(\beta \mid y, X) \propto \left[\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right]^{-n/2}.$$

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

$$\propto \int_{-\infty}^{\infty} \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{ -\frac{1}{2\sigma^2} \left[\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right] \right\} d\beta$$

$$\propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left[\frac{-\nu s^2}{2\sigma^2} \right] \int \cdots \int \exp\left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right\} d\beta_1 \cdots d\beta_K.$$

Now, let $z_i = \frac{1}{\sigma} x_i (\beta_i - \hat{\beta}_i)$

$$\therefore \; dz_i = \frac{x_i}{\sigma} d\beta_i$$

$$\propto \frac{1}{\sigma} d\beta_i$$

$$\therefore \; d\beta_i \propto \sigma \, dz_i$$

$$\therefore \; p(\sigma \mid y, X) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left[\frac{-\nu s^2}{2\sigma^2} \right] \int \cdots \int \exp\left[-\tfrac{1}{2} \Sigma z_i^2 \right] (\sigma)^K dz_1 \cdots dz_K.$$

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

$$\therefore \; p(\sigma \mid y, X) \propto \left(\frac{1}{\sigma^{\nu+1}}\right) \exp\left[\frac{-\nu s^2}{2\sigma^2} \right] \qquad ; \; 0 < \sigma < \infty.$$

But this is just an inverted Γ-dist'n.

Consider inferences concerning a single element of β, say β_i.

We want $p(\beta_1 | y, X) = \int \cdots \int p(\beta | y, X)\, d\beta_2 \cdots d\beta_K$.

where $p(\beta | y, X) \propto [\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})]^{-n/2}$.

Now, let $\delta = (\beta - \hat{\beta})$; let $H = \frac{1}{s^2}(X'X)$; $\delta_1 = (\beta_1 - \hat{\beta}_1)$;

$\delta_2' = (\beta_2 - \hat{\beta}_2, \beta_3 - \hat{\beta}_3, \cdots, \beta_K - \hat{\beta}_K)$

$$\therefore p(\beta | y, X) \propto [s^2]^{-n/2} [\nu + \tfrac{1}{s^2}(\beta - \hat{\beta})' X' X (\beta - \hat{\beta})]^{-n/2}$$

$$\propto [\nu + \delta' H \delta]^{-n/2}.$$

& $p(\delta | y, X) = p(\beta | y, X)\, |J|$

& $|J| = \left| \frac{\partial \delta_i}{\partial \beta_j} \right| = 1$

So $p(\delta | y, X) \propto [\nu + \delta' H \delta]^{-n/2}$

Let $H = \begin{pmatrix} h_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$

$$\therefore \delta' H \delta = \delta_1^2 (h_{11} - H_{12} H_{22}^{-1} H_{21}) + (\delta_2 + \delta_1 H_{22}^{-1} H_{12})' H_{22} (\delta_2 + \delta_1 H_{22}^{-1} H_{12})$$

[Completing the square on δ_2 :
NB $ax^2 + bx + c = a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$]

~~$\delta' H \delta$~~

$$\therefore p(\delta | y, X) \propto [\nu + \delta_1^2 (h_{11} - H_{12} H_{22}^{-1} H_{21}) + (\delta_2 + \delta_1 H_{22}^{-1} H_{12})' H_{22} \cdot (\delta_2 + \delta_1 H_{22}^{-1} H_{12})]^{-\frac{n}{2}}$$

Let $h^{11} = (h_{11} - H_{12} H_{22}^{-1} H_{21})^{-1}$

Let $C = H_{22} / (\nu + \delta_1^2 / h^{11})$

$$\therefore p(\delta | y, X) \propto [\nu + \delta_1^2 / h^{11}]^{-n/2} \cdot [1 + (\delta_2 + \delta_1 H_{22}^{-1} H_{12}) C (\delta_2 + \delta_1 H_{22}^{-1} H_{12})]^{-\frac{n}{2}}$$

$$(\delta_1 | y, X) \propto [\nu + \delta_1^2 / h^{11}]^{-n/2} \int [1 + (\delta_2 + \delta_1 H_{22}^{-1} H_{12}) C (\delta_2 + \delta_1 H_{22}^{-1} H_{12})]^{-\frac{n}{2}} d\delta_2$$

Now, MVT is of the general form:

$$p(x|\theta, V, v, m) \propto |V|^{\frac{1}{2}} [v + (x-\theta)' V (x-\theta)]^{-(m+v)/2}.$$

$$\therefore \int = |C|^{-\frac{1}{2}} \int |C|^{\frac{1}{2}} [1 + (\delta_2 + \delta_1 H_{22}^{-1} H_{12})' C (\delta_2 + \delta_1 H_{22}^{-1} H_{12})]^{-\frac{n}{2}} d\delta_2$$

$$\propto |C|^{-\frac{1}{2}}$$

$$\therefore p(\delta_1 | y, X) \propto [v + \delta_1^2 / h'']^{-n/2} |C|^{-\frac{1}{2}}$$

Now, $C = H_{22} / (v + \delta_1^2 / h'')$

$$\therefore |C| = (v + \delta_1^2/h'')^{-(k-1)} |H_{22}|$$

$$\propto (v + \delta_1^2/h'')^{-(k-1)}$$

$$\therefore |C|^{-\frac{1}{2}} \propto (v + \delta_1^2/h'')^{\frac{(k-1)}{2}}$$

$$\therefore p(\delta_1 | y, X) \propto [v + \delta_1^2/h'']^{-\frac{(n-k+1)}{2}} \propto [v + \delta_1^2/h'']^{-\frac{(v+1)}{2}}.$$

& this is of the form of a t-distribution.

So, $\frac{\delta_1}{(h'')^{\frac{1}{2}}} \sim t_v$.

$$\therefore \frac{(\beta_1 - \hat{\beta}_1)}{s(m'')^{\frac{1}{2}}} \sim t_v.$$

where m'' is the $(1,1)$th. element of $(X'X)^{-1}$

(because $h'' = \text{fctn}(h_{11})^{-1}$)

& $h_{11} = \text{fctn}((X'X)/s^2)$

$$\therefore (h'')^{\frac{1}{2}} \sim [s/(X'X)^{\frac{1}{2}}].$$

Now, let $F = \left(\frac{\delta_i^2}{h_{ii}}\right) = \frac{(\beta_i - \hat{\beta}_i)^2}{h_{ii}} = \frac{(\beta_i - \hat{\beta}_i)^2}{s^2 m^{ii}}$

Then $dF = \frac{2}{h_{ii}} \delta_i \, d\delta_i \propto 2F^{\frac{1}{2}} d\delta_i \propto F^{\frac{1}{2}} d\delta_i$

$$\therefore d\delta_i \propto F^{-\frac{1}{2}} dF$$

So, $p(\delta_i | y, X) \propto \left[v + \frac{\delta_i^2}{h_{ii}} \right]^{-(n-k+1)/2}$

$\Rightarrow P(F | y, X) \propto F^{-\frac{1}{2}} [v + F]^{-(n-k+1)/2}$

which is $F_{1,v}$.

So, $\frac{(\beta_i - \hat{\beta}_i)^2}{s^2 (m_{ii})} \sim F_{1,v}$.

Now consider inferences on a subset of β – i.e. we want a <u>joint</u> test on some partition of β.

Let $\delta' = (\beta - \hat{\beta})' = (\delta_1' : \delta_2')$

$$\therefore \delta' H \delta = (\delta_1' : \delta_2') \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

$$= (\delta_1' H_{11} + \delta_2' H_{21}, \; \delta_1' H_{12} + \delta_2' H_{22}) \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

$$= (\delta_1' H_{11} \delta_1 + \delta_2' H_{21} \delta_1 + \delta_1' H_{12} \delta_2 + \delta_2' H_{22} \delta_2)$$

$$= \delta_1' H_{11} \delta_1 + 2\delta_1' H_{12} \delta_2 + \delta_2' H_{22} \delta_2.$$

We want to find $p(\delta_1 | y, X)$ i.e. we want to get rid of δ_2, so <u>complete the square</u> on δ_2.

$$\therefore \delta' H \delta = [\delta_2 + \delta_1' H_{12} H_{22}^{-1}]' H_{22} [\delta_2 + \delta_1' H_{12} H_{22}^{-1}]$$
$$+ [\delta_1' H_{11} \delta_1 - \delta_1' H_{12} H_{22}^{-1} H_{21} \delta_1]$$

So, $p(\delta_1, \delta_2 | y, X) \propto \Big[v + [\delta_2 + \delta_1' H_{12} H_{22}^{-1}]' H_{22} [\delta_2 + \delta_1' H_{12} H_{22}^{-1}]$
$+ [\delta_1' H_{11} \delta_1 - \delta_1' H_{12} H_{22}^{-1} H_{21} \delta_1] \Big]^{-n/2}$

Now, ~~for given δ_1~~, we observe that

$p(\delta_2 | \delta_1, y, X)$ is MVS-t with condition mean

$$-H_{22}^{-1} H_{21} \delta_1$$

And $\delta_2 = \beta_2 - \hat{\beta}_2$

~~$\therefore$ conditional mean of β_2 is $\hat{\beta}_2 + \delta_2$~~

$$\therefore \; E(\beta_2 | \beta_1) = E[\delta_2 + \hat{\beta}_2 | \beta_1]$$
$$= E[\delta_2 | \beta_1] + \hat{\beta}_2$$
$$= \hat{\beta}_2 - H_{22}^{-1} H_{21} \delta_1$$
$$\therefore \; E(\beta_2 | \beta_1) = \hat{\beta}_2 - H_{22}^{-1} H_{21} (\beta_1 - \hat{\beta}_1).$$

Now, more important, continue to find $p(\delta_1 | X, y)$:

$$p(\delta_1 | y, X) = \int \cdots \int p(\delta_1, \delta_2 | y, x) \, d\delta_2 \qquad ; \text{ a } \exists \; k_2 \text{ integrals.}$$

$$\propto \int \cdots \int [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1 + (\delta_2 + H_{22}^{-1} H_{21} \delta_1)' H_{22} (\delta_2 + H_{22}^{-1} H_{21} \delta_1)]^{-\frac{n}{2}} d\delta_2$$

Let $C \equiv H_{22} / [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]$

Then $p(\delta_1, \delta_2 | y, x) \propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-n/2} \times$

$$\left[1 + \frac{(\delta_2 + H_{22}^{-1} H_{21} \delta_1)' H_{22} (\delta_2 + H_{22}^{-1} H_{21} \delta_1)}{[v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]} \right]^{-n/2}$$

$$\propto \left\{ [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-n/2} |C|^{-\frac{1}{2}} \right\} \times$$
$$\left\{ |C|^{\frac{1}{2}} [1 + (\delta_2 + H_{22}^{-1} H_{21} \delta_1)' C (\delta_2 + H_{22}^{-1} H_{21} \delta_1)]^{-n/2} \right\}$$

$$\therefore \; p(\delta_1 | y, x) \propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-n/2} |C|^{-\frac{1}{2}} \times$$
$$\int \cdots \int |C|^{\frac{1}{2}} [1 + (\delta_2 + H_{22}^{-1} H_{21} \delta_1)' C (\delta_2 + H_{22}^{-1} H_{21} \delta_1)]^{-n/2} d\delta_2$$

But the integrand is just M.V.S.-t.

$$\therefore \; p(\delta_1 | y, X) \propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-n/2} |C|^{-\frac{1}{2}}$$

And $|C|^{-\frac{1}{2}} = |H_{22} / [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]|^{-\frac{1}{2}}$

$$\propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{k_2/2}$$

$$\therefore \; p(\delta_1 | y, x) \propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-(n-k_2)/2}.$$

But, $(n - k_2) = (n - k) + (k - k_2) = (n - k) + k_1$
$= v + k_1$

$$\therefore \; p(\delta_1 | y, x) \propto [v + \delta_1'(H_{11} - H_{12} H_{22}^{-1} H_{21})\delta_1]^{-(v+k_1)/2}.$$

And this is just a M.V.S-t distribution.

Note: $E(\delta_1) = 0$

$\therefore E(\beta_1 - \hat{\beta}_1) = 0$

$\therefore E(\beta) = \hat{\beta}_1$

And, $E(\delta_1 \delta_1') = E[(\beta_1 - \hat{\beta}_1)(\beta_1 - \hat{\beta}_1)']$

$= \left(\frac{\nu}{\nu-2}\right)(H_{11} - H_{12} H_{22}^{-1} H_{21})^{-1}$ (see p. 386.)

Now, still using the vague prior, $p(\sigma, \beta) \propto 1/\sigma$, can we derive a posterior distribution for a <u>linear combination</u> of the β_i?

Let $\alpha = 1'\beta$; where $1'$ is a vector of <u>constant</u> (<u>not</u> necessarily unit) elements.

Note that $p(\beta, \sigma | y, X) = p(\beta | \sigma, y, X)\, p(\sigma | y, X)$

Now, we saw that $p(\beta | \sigma, y, X)$ was MVN, while $p(\sigma | y, X)$ is I.G.

Now $\Rightarrow p(\alpha | \sigma, y, X)$ is also MVN

& $E(\alpha) = \hat{\alpha} = 1' E(\beta) = 1'\hat{\beta}$.

& $Var(\alpha | \sigma, y, X) = E[(\alpha - \hat{\alpha})^2 | \sigma, y, X]$

$= 1' E[(\beta - \hat{\beta})'(\beta - \hat{\beta})] 1$

$= 1'(X'X)1\sigma^2$.

Now, $p(\alpha, \sigma | y, X) = p(\alpha | \sigma, y, X) \cdot p(\sigma | y, X)$

So, $p(\alpha | y, X) = \int p(\alpha, \sigma | y, X)\, d\sigma$

$= \int p(\alpha | \sigma, y, X)\, p(\sigma | y, X)\, d\sigma$

Now, $p(\alpha | \sigma, y, X) \propto \left(\frac{1}{\sigma}\right) \exp\left[-\frac{1}{2\sigma^2} \frac{(\alpha - \hat{\alpha})^2}{c}\right]$

where $c = 1'(X'X)1$

& $p(\sigma | y, X) \propto \left(\frac{1}{\sigma^{\nu+1}}\right) \exp\left[-\frac{\nu s^2}{2\sigma^2}\right]$

So, $p(\alpha, \sigma | y, X) \propto \left(\frac{1}{\sigma^{\nu+2}}\right) \exp\left[-\frac{1}{2\sigma^2}\left[\nu s^2 + \frac{(\alpha - \hat{\alpha})^2}{c}\right]\right]$.

$$\therefore\ p(\alpha \mid x, y) \propto \int \left(\frac{1}{\sigma^{\nu+2}}\right) \exp\left[-\frac{1}{2\sigma^2}\left[\nu s^2 + \frac{(\alpha-\hat{\alpha})^2}{c}\right]\right] d\sigma$$

$$\propto \left[\nu s^2 + \frac{(\alpha-\hat{\alpha})^2}{c}\right]^{-(\nu+1)/2}$$

$$\propto \left[\nu + \frac{(\alpha-\hat{\alpha})^2}{c s^2}\right]^{-(\nu+1)/2}$$

which is Student's t.

So $\left(\frac{\alpha-\hat{\alpha}}{s c^{\frac{1}{2}}}\right)$ is t_ν, as expected.

Now, suppose we base our posterior on an <u>informative</u> prior. In particular, use the posterior from a first sample (as derived already) as the prior for use with a second sample of data.

We have:

$$p(\beta, \sigma \mid y_1, X_1) \propto \left(\frac{1}{\sigma^{n_1+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_1 s_1^2 + (\beta-\hat{\beta}_1)' X_1' X_1 (\beta-\hat{\beta}_1)\right]\right\}$$

$$\propto \left\{\left(\frac{1}{\sigma^{\nu_1+1}}\right) \exp\left[-\frac{\nu_1 s_1^2}{2\sigma^2}\right]\right\} \left\{\left(\frac{1}{\sigma^{k}}\right) \exp\left[-\frac{1}{2\sigma^2}(\beta-\hat{\beta}_1)' X_1' X_1 (\beta-\hat{\beta}_1)\right]\right\}$$

And $p(\beta, \sigma \mid y_1, X_1) = p(\sigma \mid y_1, X_1) \cdot p(\beta_1 \mid \sigma, y_1, X_1)$

So we have the 2 marginal pdf's.

Now, take a 2nd sample. The L/F is given as

$$\ell(\beta, \sigma \mid y_2, X_2) \propto \left(\frac{1}{\sigma^{n_2}}\right) \exp\left[-\frac{1}{2\sigma^2}(y_2 - X_2\beta)'(y_2 - X_2\beta)\right]$$

$$\therefore\ p(\beta, \sigma \mid y_1, y_2, X_1, X_2) \propto \left(\frac{1}{\sigma^{n_1+n_2+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\left\{(y_1 - X_1\beta)'(y_1 - X_1\beta) + (y_2 - X_2\beta)'(y_2 - X_2\beta)\right\}\right]$$

Now, $\left[(y_1 - X_1\beta)'(y_1 - X_1\beta) + (y_2 - X_2\beta)'(y_2 - X_2\beta)\right]$

$$= \left[(y_1 - X_1\tilde{\beta}) - (X_1\beta - X_1\tilde{\beta})\right]'\left[(y_1 - X_1\tilde{\beta}) - (X_1\beta - X_1\tilde{\beta})\right]$$

$$+ \left[(y_2 - X_2\tilde{\beta}) - (X_2\beta - X_2\tilde{\beta})\right]'\left[(y_2 - X_2\tilde{\beta}) - (X_2\beta - X_2\tilde{\beta})\right]$$

$$= \left[(y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta}) + (y_2 - X_2\tilde{\beta})'(y_2 - X_2\tilde{\beta})\right] + \left[(X_1\beta - X_1\tilde{\beta})'(X_1\beta - X_1\tilde{\beta})\right.$$

$$\left. + (X_2\beta - X_2\tilde{\beta})'(X_2\beta - X_2\tilde{\beta})\right]$$

$$= \nu s^2 + \left[(\beta-\tilde{\beta})' X_1' X_1 (\beta-\tilde{\beta}) + (\beta-\tilde{\beta})' X_2' X_2 (\beta-\tilde{\beta})\right]$$

$$= \nu s^2 + (\beta-\tilde{\beta})' M (\beta-\tilde{\beta}) \quad ; \text{ say.}$$

So, $p(\beta, \sigma \mid y_1 y_2 X_1 X_2) \propto \left(\frac{1}{\sigma^{n_1+n_2+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta-\tilde{\beta})' M (\beta-\tilde{\beta})\right]\right\}$

And this is of the same form as was the case with Sample 1. So analyze it in the same way, if we wish.

Now, if we had pooled the 2 samples, got a joint L/F, & used a diffuse prior, we would have got the same end result.

Predictive p.d.f. —

q new observations ~ $\tilde{y}' = (y_{n+1}, ----, y_{n+q})$

& $\tilde{y} = \tilde{X}\beta + \tilde{u}$.

We want to obtain $p(\tilde{y} \mid y, X, \tilde{X})$.
We could take $p(\tilde{y}, \sigma, \beta \mid y, X, \tilde{X})$ and integrate w.r.t. σ, β.

Now, $p(\tilde{y}, \sigma, \beta \mid y, X, \tilde{X}) = p(\tilde{y} \mid \beta, \sigma, \tilde{X})\, p(\beta, \sigma \mid y, X)$

And, $p(\beta, \sigma \mid y, X) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\{\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\}\right]$

$$p(\tilde{y} \mid \beta, \sigma, \tilde{X}) \propto \left(\frac{1}{\sigma^q}\right) \exp\left[-\frac{1}{2\sigma^2}(\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)\right]$$

$$\therefore\ p(\tilde{y}, \beta, \sigma \mid y, X, \tilde{X}) \propto \left(\frac{1}{\sigma^{n+q+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[(y - X\beta)'(y - X\beta) + (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)\right]\right\}$$

Integrate w.r.t. σ :

$$p(\tilde{y}, \beta \mid X, y, \tilde{X}) \propto \int \left(\frac{1}{\sigma^{n+q+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[(y - X\beta)'(y - X\beta) + (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)\right]\right\} d\sigma$$

$$\propto \left[(y - X\beta)'(y - X\beta) + (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)\right]^{-(n+q)/2}.$$

$$\therefore\ p(\tilde{y} \mid X, y, \tilde{X}) \propto \int \left[(y - X\beta)'(y - X\beta) + (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)\right]^{-(n+q)/2} d\beta$$

Now, $(y - X\beta)'(y - X\beta) + (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta)$

$$= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta + \tilde{y}'\tilde{y} - \tilde{y}'\tilde{X}\beta - \beta'\tilde{X}'\tilde{y} + \beta'\tilde{X}'\tilde{X}\beta$$

$$= (y'y + \tilde{y}'\tilde{y}) + \beta'\left[(X'X) + (\tilde{X}'\tilde{X})\right]\beta - 2\beta'\left[X'y + \tilde{X}'\tilde{y}\right]$$

$$= (y'y) + (\tilde{y}'\tilde{y}) + \beta' M \beta - 2\beta'(X'y + \tilde{X}'\tilde{y})$$

Complete the square on β —

$$\Rightarrow\ y'y + \tilde{y}'\tilde{y} - (y'X + \tilde{y}'\tilde{X}) M^{-1} (X'y + \tilde{X}'\tilde{y})$$
$$+ \left[\beta - M^{-1}(X'y + \tilde{X}'\tilde{y})\right]' M \left[\beta - M^{-1}(X'y + \tilde{X}'\tilde{y})\right]$$

$$\therefore\ p(\tilde{\gamma},\beta\,|\,y,x,\tilde{x}) \propto \Big\{ (y'y+\tilde{\gamma}'\tilde{\gamma}) - (y'x+\tilde{\gamma}'\tilde{x})M^{-1}(x'y+\tilde{x}'\tilde{\gamma}) + [\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]'M[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]\Big\}^{-(n+q)/2}$$

$$\propto \Big[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (y'x+\tilde{\gamma}'\tilde{x})M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{-\frac{(n+q)}{2}} *$$

$$\left\{1 + \frac{[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]'M[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]}{[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (y'x+\tilde{\gamma}'\tilde{x})M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]}\right\}^{-\frac{(n+q)}{2}}$$

$$\therefore\ p(\tilde{\gamma}\,|\,y,x,\tilde{x}) \propto \Big[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{-(n+q)/2} *$$

$$\int \left\{1 + \frac{[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]'M[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]}{[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]}\right\}^{-(n+q)/2} d\beta.$$

Let $C = M \Big/ \left\{ y'y+\tilde{\gamma}'\tilde{\gamma} - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\right\}$

So integrand $= \left\{1 + [\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]'C[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]\right\}^{-\frac{(n+q)}{2}}$

$$\therefore\ p(\tilde{\gamma}\,|\,y,x,\tilde{x}) \propto \Big[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{-(n+q)/2} * |C|^{-\frac{1}{2}} *$$

$$\int \text{-----} \int |C|^{\frac{1}{2}} \left\{1 + [\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]'C[\beta - M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]\right\}^{-\frac{(n+q)}{2}} d\beta$$

But integrand is MVS-t.

So,

$$p(\tilde{\gamma}\,|\,y,x,\tilde{x}) \propto \Big[(y'y+\tilde{\gamma}'\tilde{\gamma}) - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{-(n+q)/2} |C|^{-\frac{1}{2}}$$

Now, $|C|^{-\frac{1}{2}} = \left\{\det.\left[M \Big/ [y'y+\tilde{\gamma}'\tilde{\gamma} - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})]\right]\right\}^{-\frac{1}{2}}$

$$\propto \Big[y'y+\tilde{\gamma}'\tilde{\gamma} - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{\frac{k}{2}}$$

So, $p(\tilde{\gamma}\,|\,y,x,\tilde{x}) \propto \Big[y'y+\tilde{\gamma}'\tilde{\gamma} - (x'y+\tilde{x}'\tilde{\gamma})'M^{-1}(x'y+\tilde{x}'\tilde{\gamma})\Big]^{-\frac{(\nu+q)}{2}}$

where $\nu = n-k$.

Now, simplify the expression in brackets on the RHS to get it into a more recognisable form –

$$[y'y + \tilde{y}'\tilde{y} - (X'y + \tilde{X}'\tilde{y})'M^{-1}(X'y + \tilde{X}'\tilde{y})]$$

$$= y'[I - X'M^{-1}X']y + \tilde{y}'[I - \tilde{X}'M^{-1}\tilde{X}']\tilde{y} - 2\tilde{y}'\tilde{X}M^{-1}X'y$$

$$= y'[I - X'M^{-1}X']y - y'XM^{-1}\tilde{X}'[I - \tilde{X}M^{-1}\tilde{X}']^{-1}\tilde{X}M^{-1}X'y$$

$$+ [\tilde{y} - (I - \tilde{X}M^{-1}\tilde{X}')^{-1}\tilde{X}M^{-1}X'y]'(I - \tilde{X}M^{-1}\tilde{X}')[\tilde{y} - (I - \tilde{X}M^{-1}\tilde{X}')^{-1}\tilde{X}M^{-1}X'y]$$

But,

$$(I - \tilde{X}M^{-1}\tilde{X}')^{-1} = (I + \tilde{X}(X'X)^{-1}\tilde{X}')$$

* see over →

So, $\underline{(I - \tilde{X}M^{-1}\tilde{X}')^{-1}\tilde{X}M^{-1}} = [I + \tilde{X}(X'X)^{-1}\tilde{X}']\tilde{X}M^{-1}$

$$= (I + \tilde{X}(X'X)^{-1}\tilde{X}')\tilde{X}(X'X + \tilde{X}'\tilde{X})^{-1}$$

$$= (I + \tilde{X}(X'X)^{-1}\tilde{X}')\tilde{X}[(X'X)\{I + (X'X)^{-1}(\tilde{X}'\tilde{X})\}]^{-1}$$

$$= \tilde{X}[I + (X'X)^{-1}\tilde{X}'\tilde{X}][I + (X'X)^{-1}\tilde{X}'\tilde{X}]^{-1}(X'X)^{-1}$$

$$= \tilde{X}(X'X)^{-1}$$

So the bracketed expression becomes –

$$y'[I - XM^{-1}X']y - y'XM^{-1}\tilde{X}'\tilde{X}(X'X)^{-1}X'y$$

$$+ (\tilde{y} - \tilde{X}(X'X)^{-1}X'y)'(I - \tilde{X}M^{-1}\tilde{X}')(\tilde{y} - \tilde{X}(X'X)^{-1}X'y)$$

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

$$\Rightarrow y'[I - XM^{-1}X']y - y'XM^{-1}\tilde{X}'\tilde{X}\hat{\beta} + (\tilde{y} - \tilde{X}\hat{\beta})'(I - \tilde{X}M^{-1}\tilde{X}')(\tilde{y} - \tilde{X}\hat{\beta})$$

$$= y'\{I - X[M^{-1} + M^{-1}\tilde{X}'\tilde{X}(X'X)^{-1}]X'\}y + (\tilde{y} - \tilde{X}\hat{\beta})'(I - \tilde{X}M^{-1}\tilde{X}')(\tilde{y} - \tilde{X}\hat{\beta})$$

$$= y'[I - X(X'X)^{-1}X']y + (\tilde{y} - \tilde{X}\hat{\beta})'(I - \tilde{X}M^{-1}\tilde{X}')(\tilde{y} - \tilde{X}\hat{\beta})$$

Now, $y'(I - X(X'X)^{-1}X')y = y'y - y'X\hat{\beta} = (y - X\hat{\beta})'(y - X\hat{\beta}) = \nu s^2$

$$\therefore p(\tilde{y}|y, X, \tilde{X}) \propto [\nu s^2 + (\tilde{y} - \tilde{X}\hat{\beta})'(I - \tilde{X}M^{-1}\tilde{X}')(\tilde{y} - \tilde{X}\beta)]^{-(\nu+q)/2}.$$

Let $H = \left(\frac{1}{s^2}\right)(I - \tilde{X}M^{-1}\tilde{X}')$

$$\therefore p(\tilde{y}|y, X, \tilde{X}) \propto [\nu + (\tilde{y} - \tilde{X}\hat{\beta})'H(\tilde{y} - \tilde{X}\hat{\beta})]^{-(\nu+q)/2}.$$

which is MVS – t_ν.

So, $E(\tilde{y}) = \tilde{X}\hat{\beta}$; $\mathrm{var}(\tilde{y}) = \left(\frac{\nu}{\nu-2}\right)H^{-1}$

$$= \left(\frac{\nu s^2}{\nu-2}\right)[I + \tilde{X}(X'X)^{-1}\tilde{X}'].$$

$$= [I - \tilde{X} M^{-1} \tilde{X}'][I + \tilde{X}(X'X)^{-1}\tilde{X}']$$

$$= [I - \tilde{X} M^{-1} \tilde{X}'] + \tilde{X}(X'X)^{-1}\tilde{X}' - \tilde{X} M^{-1} (\tilde{X}'\tilde{X})(X'X)^{-1}\tilde{X}'$$

Now, $(\tilde{X}'\tilde{X})(X'X)^{-1} = [M - (X'X)](X'X)^{-1}$

$$= M(X'X)^{-1} - I$$

$$\therefore \gamma = [I - \tilde{X} M^{-1} \tilde{X}'] + \tilde{X}(X'X)^{-1}\tilde{X}' - \tilde{X} M^{-1}[M(X'X)^{-1} - I]\tilde{X}'$$

$$= I - \tilde{X} M^{-1} \tilde{X}' + \tilde{X}(X'X)^{-1}\tilde{X}' - \tilde{X}(X'X)^{-1}\tilde{X}' + \tilde{X} M^{-1} \tilde{X}'$$

$$= I,$$ <u>as required</u>

So, for a single element, $\tilde{y}_i$, of $\tilde{y}$, we get:

$$\left(\frac{\tilde{y}_i - \tilde{X}_{(i)}\hat{\beta}}{\sqrt{h_{ii}}} \right) \sim t_v.$$

where h_{ii} is the (i,i)th. element of H^{-1}

<u>What if $(X'X)$ is singular</u> ? :

Then in order to be able to estimate, we must feed into the system sufficient <u>a priori</u> information to destroy the multicollinearity.

i.e. use an <u>informative</u> prior pdf. So choose a N.C. prior pdf for the parameters.

Now, $p(\beta, \sigma) = p(\beta|\sigma)\, p(\sigma)$

pick : $\begin{cases} p(\beta|\sigma) \propto \dfrac{|A|^{\frac{1}{2}}}{\sigma^K} \exp.\left[-\frac{1}{2\sigma^2}(\beta-\bar{\beta})'A(\beta-\bar{\beta})\right] \\ p(\sigma) \propto \left(\dfrac{1}{\sigma^{v_0+1}}\right) \exp\left[-\dfrac{v_0 c_0^2}{2\sigma^2}\right]; \quad v_0 > 1. \end{cases}$

$$\therefore p(\beta,\sigma) \propto \left(\frac{1}{\sigma^{K+v_0+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\left\{v_0 c_0^2 + (\beta-\bar{\beta})'A(\beta-\bar{\beta})\right\}\right]$$

Take the L.F. :

$$\ell(\beta,\sigma|y,X) \propto \left(\frac{1}{\sigma^n}\right) \exp\left[\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

$$\therefore p(\beta,\sigma|y) \propto \left(\frac{1}{\sigma^{n+K+v_0+1}}\right) \exp\left[-\frac{1}{2\sigma^2}\left[v_0 c_0^2 + (\beta-\bar{\beta})'A(\beta-\bar{\beta}) + (y-X\beta)'(y-X\beta)\right]\right]$$

Now, look at $\gamma = (\beta-\bar{\beta})'A(\beta-\bar{\beta}) + (y-X\beta)'(y-X\beta)$

$= \beta'A\beta - \beta'A\bar{\beta} - \bar{\beta}'A\beta + \bar{\beta}'A\bar{\beta} + y'y - y'X\beta - \beta'X'y + \beta'X'X\beta$

$= \beta'[A + X'X]\beta - \beta'[A\bar{\beta} + X'y + X'y + A\bar{\beta}] + y'y + \bar{\beta}'A\bar{\beta}$

$= \beta'[A + X'X]\beta - 2\beta'[A\bar{\beta} + X'y] + y'y + \bar{\beta}'A\bar{\beta}.$

Complete the square on β —

$\gamma = y'y + \bar{\beta}'A\bar{\beta} - (A\bar{\beta} + X'y)'(A + X'X)^{-1}(A\bar{\beta} + X'y)$

$+ [\beta - (A+X'X)^{-1}(A\bar{\beta}+X'y)]'(A+X'X)[\beta - (A+X'X)^{-1}(A\bar{\beta}+X'y)]$

Let $\check{\beta} = (A+X'X)^{-1}(A\bar{\beta} + X'y)$

$$\therefore \gamma = y'y + \bar{\beta}'A\bar{\beta} - (A\bar{\beta} + X'y)'\check{\beta} + (\beta-\check{\beta})'(A+X'X)(\beta-\check{\beta})$$

$$= y'y + \bar{\beta}'A\bar{\beta} + (\beta-\check{\beta})'(A+X'X)(\beta-\check{\beta}) - \check{\beta}'(A+X'X)\check{\beta}$$

$$\therefore p(\beta,\sigma|y) \propto \left(\frac{1}{\sigma^{n'+k+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[v_0c_0^2 + y'y + \bar{\beta}'A\bar{\beta} + (\beta-\check{\beta})'(A+X'X)(\beta-\check{\beta}) - \check{\beta}'(A+X'X)^{-1}\check{\beta}\right]\right\}$$

where $n' = v_0 + n$

Let $n'c^2 = [v_0c_0^2 + y'y + \bar{\beta}'A\bar{\beta} - \check{\beta}'(A+X'X)^{-1}\check{\beta}]$

$$\therefore p(\beta,\sigma|y) \propto \left(\frac{1}{\sigma^{n'+k+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[n'c^2 + (\beta-\check{\beta})'(A+X'X)(\beta-\check{\beta})\right]\right\}$$

$$\therefore p(\beta|y) = \int p(\beta,\sigma|y)\,d\sigma$$

$$\propto \left[n'c^2 + (\beta-\check{\beta})'(A+X'X)(\beta-\check{\beta})\right]^{-(n'+k)/2}.$$

Which is MVS-t_{v}. Post-Mean is $\check{\beta}$.

So we can obtain a Bayes Estimator ($\check{\beta}$) even though $(X'X)$ is singular. All you need are <u>informative</u> priors that are both tractable and also lead to a posterior which does <u>not</u> involve $(X'X)^{-1}$.

* $\boxed{\text{NB}}$ Why not look for the least informative prior that ill enable you to still handle the estimation problem? & Why not use an information-theoretic approach to getting such a "minimal-information" prior?

$n' = T + v_0$

$n'+K = T + v_0 + K$

Sampling-theory Approach to Singular $(X'X)$ —

$$y = X\beta + u$$
$$= XP\gamma + u$$

where P is orthogonal, so $P'P = I$.

$\therefore \gamma = P'\beta$; & P is $(k \times k)$.

Let X^{Σ} have rank q; then $\exists$ q non-zero C.R. of $(X'X)$.

Let $P'X'XP = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$, where D is diagonal, of order $(q \times q)$ & with the non-zero C.R. of $(X'X)$ as its diagonal elements.

Set up the "normal" equations —

$$P'X'y = P'X'XP\gamma$$
$$= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\gamma$$

$$\therefore \begin{pmatrix} P_1'X'y \\ P_2'X'y \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$$

$$\therefore \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} P_1'X'y \\ 0 \end{pmatrix} \qquad *.$$

i.e. $P_2'X'y = 0$.

Now, solve the system $*$:- (use g-inverse).

We have $(P'X'XP)\gamma = P'X'y$

i.e. $Ax = b$.

Then g-inverse $\Rightarrow$

$$\tilde{x} = A^g b + (I - A^g A)z \quad ; \text{ arbitrary } z.$$

So, $\tilde{\gamma} = (P'X'XP)^* P'X'y + [I - (P'X'XP)^*(P'X'XP)]z$

Now, there are a variety of ways of choosing the g-inverse and choosing z.

We have that:

$$\tilde{\gamma} = (P'X'XP)^* P'X'y + [I - (P'X'XP)^* P'X'XP] z$$

Let $(P'X'XP)^* = \begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix}$

Then this is a GI for $(P'X'XP)$

viz: $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I & DE \\ 0 & 0 \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$

$= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$ ✓

$$\therefore \tilde{\gamma} = \begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix} P'X'y + \left[I - \begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \right] z$$

$$\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix}\begin{pmatrix} P_1'X'y \\ 0 \end{pmatrix} + \left[I - \begin{pmatrix} I & 0 \\ CD & 0 \end{pmatrix} \right]\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

$$\therefore \begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \begin{pmatrix} D^{-1}P_1'X'y \\ CP_1'X'y \end{pmatrix} + \begin{pmatrix} 0 \\ z_2 - CDz_1 \end{pmatrix}$$

$$\therefore \tilde{\gamma}_1 = D^{-1}P_1'X'y$$

$$\tilde{\gamma}_2 = CP_1'X'y + z_2 - CDz_1$$

$$= z_2 + C(P_1'X'y - Dz_1)$$

$$= z_2 + CD(D^{-1}P_1'X'y - z_1)$$

$$\therefore \tilde{\gamma}_2 = z_2 + CD(\tilde{\gamma}_1 - z_1)$$

So $\tilde{\gamma}_1$ is independent of the choice of GI or z, but $\tilde{\gamma}_2$ depends on the choice of <u>both</u> GI & z.

Primarily we are interested in the $\tilde{\gamma}_1$ part, because that is the part of the coefficient vector not affected directly by the multicollinearity.

<u>Example</u>: Pick the Moore-Penrose G.I. —

$$(P'X'XP)^* = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Then for this GI or for any $\ni C = 0$, we have,

$$\left.\begin{aligned} \tilde{\gamma}_1 &= D^{-1}P_1'X'y \\ \tilde{\gamma}_2 &= z_2 . \end{aligned}\right\}$$

So $\tilde{\gamma}_2$ still depends on the choice of z, but no longer on the choice of the GI.

Now, $\beta = P\gamma$

So $\tilde{\beta} = P\tilde{\gamma}$

$$\therefore \tilde{\beta} = P\left[(P'X'XP)^* P'X'y + (I - (P'X'XP)^* P'X'XP)z\right]$$
$$= \left[P(P'X'XP)^*P'\right]X'y + \left\{I - \left[P(P'X'XP)^*P'\right]X'X\right\}Pz$$

Now, $(P'X'XP)^*$ is a GI for $(P'X'XP)$.

Hence, $(P'X'XP)(P'X'XP)^*(P'X'XP) = (P'X'XP)$

$$\therefore P(P'X'XP)(P'X'XP)^*(P'X'XP)P' = P(P'X'XP)P'$$

But $PP' = P'P = I$, since P is <u>orthogonal</u>.

Hence, $X'XP(P'X'XP)^*P'X'X = X'X$

viz. $(X'X)\left[P(P'X'XP)^*P'\right](X'X) = (X'X)$

So, $P(P'X'XP)^*P'$ is a GI for $(X'X) = (X'X)^*$

$$\therefore \tilde{\beta} = (X'X)^* X'y + (I - (X'X)^*(X'X))Pz$$

$$\tilde{\beta} = \begin{pmatrix}\tilde{\beta}_1\\ \tilde{\beta}_2\end{pmatrix} = \begin{pmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{pmatrix}\begin{pmatrix}\tilde{\gamma}_1\\ \tilde{\gamma}_2\end{pmatrix}$$
$$= \begin{pmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{pmatrix}\begin{pmatrix}D^{-1}P_1'X'y\\ z_2 + CD(\tilde{\gamma}_1 - z_1)\end{pmatrix}$$
$$= \begin{bmatrix}P_{11}D^{-1}P_1'X'y + P_{12}(z_2 + CD(\tilde{\gamma}_1 - z_1))\\ P_{21}D^{-1}P_1'X'y + P_{22}(z_2 + CD(\tilde{\gamma}_1 - z_1))\end{bmatrix}$$

So, although $\tilde{\gamma}_1$ is independent of the choices of the GI & of z, it turns out that $\tilde{\beta}_1$ (~~or~~ $\tilde{\beta}_2$) depends on both the GI & the choice of z.

Even if we pick the Moore-Penrose GI, or any other $\ni$ $C = 0$,

then <u>both</u> still depend on <u>z only</u>.

viz. $\tilde{\beta} = \begin{pmatrix}\tilde{\beta}_1\\ \tilde{\beta}_2\end{pmatrix} = \begin{bmatrix}P_{11}D^{-1}P_1'X'y + P_{12}z_2\\ P_{21}D^{-1}P_1'X'y + P_{22}z_2\end{bmatrix}$.

Bayesian Interpretation –

Now, give a Bayesian interpretation to this sampling theory approach to handling the singularity of $(X'X)$.

We have the L/F –

$$\mathcal{L}(\beta,\sigma|y) \propto \left(\frac{1}{\sigma^n}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[(y-X\beta)'(y-X\beta)\right]\right\}$$

But $\beta = P\gamma$

$\therefore \; |P| \propto$ const.

$$\text{So, } \mathcal{L}(\gamma_1,\gamma_2,\sigma|y) \propto \left(\frac{1}{\sigma^n}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[(y-XP\gamma)'(y-XP\gamma)\right]\right\}$$

Let $\tilde{\gamma}_1 = D^{-1}P_1'X'y$

& Let $a = (y - XP_1'\tilde{\gamma}_1)'(y - XP_1'\tilde{\gamma}_1)$

Now, $y = X\beta + u = XP\gamma + u$

$$\therefore (y - X\beta) = u = (y - XP\gamma)$$

$$= y - X(P_1, P_2)\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix}$$

$$= y - XP_1\gamma_1 \quad ; \text{ since } XP_2 = 0.$$

$$\therefore (y - XP\gamma) = (y - XP_1\gamma_1).$$

$$\therefore \mathcal{L}(\gamma_1,\gamma_2,\sigma|y) \propto \left(\frac{1}{\sigma^n}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[(y-XP_1\gamma_1)'(y-XP_1\gamma_1)\right]\right\}$$

But, $(y-XP_1\gamma_1)'(y-XP_1\gamma_1)$

$$= \left[(y-XP_1\tilde{\gamma}_1) - (XP_1\gamma_1 - XP_1\tilde{\gamma}_1)\right]'\left[(y-XP_1\tilde{\gamma}_1) - (XP_1\gamma_1 - XP_1\tilde{\gamma}_1)\right]$$

$$= (y-XP_1\tilde{\gamma}_1)'(y-XP_1\tilde{\gamma}_1) + (XP_1\gamma_1 - XP_1\tilde{\gamma}_1)'(XP_1\gamma_1 - XP_1\tilde{\gamma}_1)$$

$$= a + (\gamma_1 - \tilde{\gamma}_1)P_1'X'XP_1(\gamma_1 - \tilde{\gamma}_1)$$

But $P'X'XP = \begin{pmatrix} D & 0\\ 0 & 0\end{pmatrix}$

So $(P_1'X'XP_1) = D$

$$\therefore \mathcal{L}(\gamma_1,\gamma_2,\sigma|y) \propto \left(\frac{1}{\sigma^n}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[a + (\gamma_1 - \tilde{\gamma}_1)'D(\gamma_1 - \tilde{\gamma}_1)\right]\right\}$$

NB the L/F is quite independent of γ_2.

That is, it is now in a form where some of the (true values of the) coefficients in the equation are omitted from the function — we are concentrating only on those not affected directly by the singularity of $(X'X)$.

Now, take the general case, where

$$(P'X'XP)^* = \begin{pmatrix} D^{-1} & E \\ C & F \end{pmatrix}. \qquad ; C \neq 0.$$

Let $p(\gamma_1, \sigma) \propto \frac{1}{\sigma}$; $0 < \sigma < \infty$; $-\infty < \gamma_{1i} < \infty$; $i = 1, \ldots, q$.

& we have:

$$\gamma_2 = z_2 + CD(\gamma_1 - z_1).$$

Then by assigning values to C and to z_1, z_2, we are able to introduce our <u>a priori</u> knowledge.

$$\therefore p(\gamma_1, \sigma | y) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp.\left\{-\frac{1}{2\sigma^2}\left[a + (\gamma_1 - \tilde{\gamma}_1)' D (\gamma_1 - \tilde{\gamma}_1)\right]\right\}$$

And this pdf is proper for $n > q$ (i.e. +ve d.f.).

And $p(\gamma_1 | y) \propto \left[a + (\gamma_1 - \tilde{\gamma}_1)' D (\gamma_1 - \tilde{\gamma}_1)\right]^{-n/2}$

So, $E(\gamma_1) = \tilde{\gamma}_1$

& $\gamma_2 = z_2 + CD(\gamma_1 - z_1)$

$$\Rightarrow E(\gamma_2) = z_2 + CD(\tilde{\gamma}_1 - z_1)$$
$$= \tilde{\gamma}_2 \text{ , as before.}$$

So the use of the diffuse prior pdf leads to the same results as we obtained when using the GI approach.

Take the special case where $C = 0$ (eg the Moore-Penrose situation) —

$$\therefore \gamma_2 = z_2.$$

So if we assign a value to z_2, we assign a specific value to γ_2 — we are being "dogmatic". The prior pdf for γ_2 is then degenerate — all of its mass is concentrated at the point z_2. Further, $C = 0 \Rightarrow \gamma_2$ is independent of γ_1.

Clearly, to assume that $C = 0$, so that $\gamma_2 = z_2$, is very restrictive in practice. Can we loosen this assumption a little?

One way of doing this is to put

$$p(\gamma_2 \mid \sigma) \propto \left(\frac{1}{\sigma^{k-q}}\right) \exp\left\{-\frac{1}{2\sigma^2}(\gamma_2 - z_2)' Q (\gamma_2 - z_2)\right\}$$

<u>Note</u>: $\exists$ $(k-q)$ obs. for γ_2.

Then, since $p(\gamma_1, \sigma) \propto 1/\sigma$

$$\Rightarrow p(\gamma_1, \gamma_2, \sigma) = p(\gamma_1, \sigma)\, p(\gamma_2 \mid \sigma)$$

$$\propto \left(\frac{1}{\sigma^{k+1-q}}\right) \exp\left\{-\frac{1}{2\sigma^2}(\gamma_2 - z_2)' Q (\gamma_2 - z_2)\right\}$$

So, $$p(\gamma_1, \gamma_2, \sigma \mid y) \propto \left(\frac{1}{\sigma^{k+n+1-q}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[(\gamma_2 - z_2)' Q (\gamma_2 - z_2) + a + (\gamma_1 - \tilde{\gamma}_1)' D (\gamma_1 - \tilde{\gamma}_1)\right]\right\}$$

Let $\tilde{\beta} = P\tilde{\gamma}$; $\tilde{\gamma}' = (\tilde{\gamma}_1' : z_2')$; $F = \begin{pmatrix} D & 0 \\ 0 & Q \end{pmatrix}$

Then, $$p(\beta, \sigma \mid y) \propto \left(\frac{1}{\sigma^{n+k+1-q}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[a + (\beta - \tilde{\beta})' P F P' (\beta - \tilde{\beta})\right]\right\}$$

Note that $PFP' = (P_1 D P_1' + P_2 Q P_2')$ is non-singular, even though both $P_1 D P_1'$ and $P_2 Q P_2'$ are singular.

So the posterior is proper for $n > q$, and can be used for posterior inferences on β & σ.

Finally, consider the situation if a diffuse prior is used for β when $(X'X)$ is singular —

Then $p(\beta) \propto$ const. ; $-\infty < \beta_i < \infty$; $\forall i$.

So, for <u>given</u> σ we get:

$$p(\beta \mid \sigma, y) \propto \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta - \tilde{\beta})' X'X (\beta - \tilde{\beta})\right\}$$

For any $\tilde{\beta}$, sol'n. to the normal equ'ns. $X'X\tilde{\beta} = X'y$. And since $(X'X)$ is singular, $\tilde{\beta}$ is <u>not</u> unique.

Now, $\beta = P\gamma$

So, $$p(\gamma_1, \gamma_2 \mid \sigma, y) \propto \exp\left\{-\frac{1}{2\sigma^2}(\gamma_1 - \tilde{\gamma}_1)' D (\gamma_1 - \tilde{\gamma}_1)\right\}$$

which is independent of γ_2. So the posterior for γ_2 is diffuse, as is its prior. Even if we introduce no prior information, it is possible to make inferences concerning q linearly independent combinations of the elements of β.

Zellner: Chapter 4.

"Special Problems in Regression Analysis".

A) Autocorrelated Errors:

$$\left.\begin{aligned} y_t &= \beta x_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}\right\} \quad t = 1, 2, \ldots, T.$$

$$\epsilon_t \sim IN(0, \sigma^2).$$

So, $y_t = \rho y_{t-1} + \beta(x_t - \rho x_{t-1}) + \epsilon_t$; $\forall t = 1, \ldots, T.$

Now, consider y_0 —

If the same process operates for $t = 0, -1, -2, \ldots, -T_0$, where T_0 is <u>unknown</u> then,

$$y_0 = \rho y_{-1} + \beta(x_0 - \rho x_{-1}) + \epsilon_0$$

$$= \rho(y_{-1} - \beta x_{-1}) + \beta x_0 + \epsilon_0$$

$$\therefore (y_0 - \beta x_0) = M + \epsilon_0, \text{ say.}$$

i.e. M is the <u>initial level</u> of the process.

Now, $\epsilon_0 \sim N(0, \sigma^2)$

$$\therefore E(y_0) = \beta x_0 + M$$

$$\text{var}(y_0) = \sigma^2.$$

On the other hand, y_0 may be fixed & known. Whatever we assume for y_0, we end up with the same posterior pdf.

Now,

$$p(y_0, y \mid \beta, \sigma, \rho, M) = p(y_0 \mid \beta, \sigma, \rho, M)\, p(y \mid y_0, \beta, \sigma, \rho, M)$$

$$\propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{ -\frac{1}{2\sigma^2}(y_0 - \beta x_0 - M)^2 + \cancel{\ldots} - \frac{1}{2\sigma^2} \sum_1^T \left[y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1})\right]^2 \right\}$$

$0 < \sigma < \infty$; $-\infty < \beta < \infty$; $-\infty < \rho < \infty$; $-\infty < M < \infty$.

And let $p(\beta, \sigma, \rho, \mu) \propto (1/\sigma)$.

$$\therefore p(\beta, \rho, \sigma, \mu | y) \propto \left(\frac{1}{\sigma^{T+2}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[(y_0 - \beta x_0 - \mu)^2 + \sum_1^T (y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1}))^2\right]\right\}.$$

And, usually, μ is of little interest to us. So:

$$p(\rho, \sigma, \beta | y) = \cancel{\int_{-\infty}^{\infty} p(\rho, \sigma, \beta) d\mu} = \int_{-\infty}^{\infty} p(\beta, \sigma, \rho, \mu | y) d\mu$$

$$\propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2} \sum_1^T (y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1}))^2\right\} \cdot \int_{-\infty}^{\infty} \left(\frac{1}{\sigma}\right) \exp\left\{-\frac{1}{2\sigma^2}(y_0 - \beta x_0 - \mu)^2\right\}$$

$$\propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2} \sum_1^T (y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1}))^2\right\}.$$

Which is the same posterior we would have got by ignoring complications regarding y_0 etc.

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

$$\propto \int_0^{\infty} \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2} \sum_1^T (y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1}))^2\right\} d\sigma$$

$$\propto \left\{\sum_1^T (y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1}))^2\right\}^{-(T/2)}$$

Complete the square on β & use the properties of the Univariate t-- distn. to get:

$$p(\rho | y) = \int_{-\infty}^{\infty} p(\beta, \rho | y) d\beta$$

$$p(\rho | y) \propto \left[\sum (x_t - \rho x_{t-1})^2\right]^{-\frac{1}{2}} * \left\{\sum (y_t - \rho y_{t-1})^2 - \frac{\left[\sum (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})\right]^2}{\sum (x_t - \rho x_{t-1})^2}\right\}^{-\left(\frac{T-1}{2}\right)}.$$

& Complete square on ρ & integrate to get:

$$p(\beta | y) \propto \left[\sum (y_{t-1} - \beta x_{t-1})^2\right]^{-\frac{1}{2}} * \left\{\sum (y_t - \beta x_t)^2 - \frac{\left[\sum (y_{t-1} - \beta x_{t-1})(y_t - \beta x_t)\right]^2}{\sum (y_{t-1} - \beta x_{t-1})^2}\right\}^{-\left(\frac{T-1}{2}\right)}.$$

Note that for $p(\rho|y)$ to be a proper pdf, we require that $\Sigma(x_t - \rho x_{t-1})^2 > 0$

i.e. for a given ρ, $\exists$ some $t \ni$

$x_t \neq \rho x_{t-1}$. (This isn't too restrictive.)

The posterior pdf for β enables us to make inferences about this parameter which allow for departures from serial independence. Recall, that if we assume that $\rho = 0$, then $\dfrac{(\beta - \hat{\beta})}{s/(\Sigma x_t^2)^{\frac{1}{2}}} \sim t_v$.

And this will give a completely different posterior distribution.

Recall that
$$p(\beta|y) = \int_{-\infty}^{\infty} p(\beta, \rho|y)\,d\rho$$
$$= \int_{-\infty}^{\infty} p(\beta|\rho, y)\, p(\rho|y)\,d\rho$$

i.e. a weighted average of $p(\beta|\rho, y)$, with $p(\rho|y)$ as the weighting fctn. Note that $p(\beta|\rho, y)$ provides inferences about β for a given ρ, while $p(\rho|y)$ reflects the plausibility of assertions about ρ, in the light of our prior notions & the data evidence.

Note that $p(\beta|\rho, y) \propto [s^2(\rho)]^{-\frac{1}{2}} \left\{ 1 + \dfrac{[\beta - \hat{\beta}(\rho)]^2}{v s^2(\rho)} \right\}^{-(v+1)/2}$

where $v = T - 1$

$$\hat{\beta}(\rho) = \frac{\Sigma(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\Sigma(x_t - \rho x_{t-1})^2}$$

$$v s^2(\rho) = \frac{\Sigma[y_t - \rho y_{t-1} - \hat{\beta}(\rho)(x_t - \rho x_{t-1})]^2}{\Sigma(x_t - \rho x_{t-1})^2}$$

So, $\left(\dfrac{\beta - \hat{\beta}(\rho)}{s(\rho)} \right) \sim t_v$.

Let $y = X\beta + u$

$u = \rho u_{-1} + \epsilon$

$\therefore\ y - \rho y_{-1} = (X - \rho X_{-1})\beta + \epsilon$

Let $y_0 = \rho(y_{-1} - \beta X_{-1}) + X_0\beta + \epsilon$

$= M + X_0\beta + \epsilon$

& $M = \rho(y_{-1} - \beta X_{-1})$

So $p(y_0 | \beta, \sigma, \rho, M) \propto \left(\frac{1}{\sigma}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[y_0 - x_0'\beta - M\right]^2\right\}$

& $p(y | y_0, \beta, \sigma, \rho, M) \propto \left(\frac{1}{\sigma^T}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]\right.$

$\therefore\ p(y, y_0 | \beta, \sigma, \rho, M) = p(y | y_0, \beta, \sigma, \rho, M) \cdot p(y_0 | \beta, \sigma, \rho, M)$

$\propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[y_0 - x_0'\beta - M\right]^2 - \frac{1}{2\sigma^2}\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]\right.$

& $p(\beta, \sigma, \rho, M) \propto 1/\sigma$

$\therefore\ p(\beta, \sigma, \rho, M | y) \propto \left(\frac{1}{\sigma^{T+2}}\right) \exp\left\{-\frac{1}{2\sigma^2}(y_0 - x_0'\beta - M)^2 - \frac{1}{2\sigma^2}\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\right.$

$\left. \times \left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]\right\}$

$\therefore\ p(\beta, \sigma, \rho | y) \propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]\right\}$

$\therefore\ p(\beta, \rho | y) \propto \left\{\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]\right\}^{-(T/2)}$.

Now, the OLS estimator for

$(y - \rho y_{-1}) = (X - \rho X_{-1})\beta + \varepsilon$

is the Cochrane – Orcutt estimator –

$\hat{\beta}(\rho) = H^{-1}(X - \rho X_{-1})'(y - \rho y_{-1})$

where $H = (X - \rho X_{-1})'(X - \rho X_{-1})$

& $\nu s^2 = \left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]'\left[y - \rho y_{-1} - (X - \rho X_{-1})\beta\right]$

And, for the O.L.S. squares case we had —

$$p(\beta|y) \propto \left\{ \nu s^2 + [\beta-\hat\beta]'(X'X)^{-1}[\beta-\hat\beta] \right\}^{-T/2}$$
$$\propto \left\{ 1 + \frac{(\beta-\hat\beta)'(X'X)^{-1}(\beta-\hat\beta)}{\nu s^2} \right\}^{-T/2}.$$

So here —

$$p(\beta|\rho,y) \propto \left\{ 1 + \frac{[\beta-\hat\beta(\rho)]' H [\beta-\hat\beta(\rho)]}{\nu s^2} \right\}^{-T/2}.$$

And this is a MVS-t. distribution, with mean $\hat\beta(\rho)$.

We have assumed that H is +ve definite — if $k=1$, this requires that $x_t \neq \rho x_{t-1}$ for some t. If $k>1$, this requires that any linear combination of the columns of $(X-\rho X_{-1})$ must not be exactly 1st-order autoregressive.

Then,

$$p(\beta|y) = \int p(\beta,\rho|y)\,d\rho$$
$$= \int \left\{ [y-\rho y_{-1}-(X-\rho X_{-1})\beta]'[y-\rho y_{-1}-(X-\rho X_{-1})\beta] \right\}^{-T/2} d\rho$$

and $p(\rho|y) = \int p(\beta,\rho|y)\,d\beta$

Consider the 2nd one —

$$p(\rho|y) \propto \int \left\{ \nu s^2(\rho) + [\beta-\hat\beta(\rho)]' H [\beta-\hat\beta(\rho)] \right\}^{-T/2} d\beta$$

$$\propto [s^2(\rho)]^{-T/2} \int \left\{ \nu + \frac{[\beta-\hat\beta(\rho)]' H [\beta-\hat\beta(\rho)]}{s^2(\rho)} \right\}^{-T/2} d\beta$$

$$\propto [s^2(\rho)]^{-T/2} \left|\frac{H}{s^2(\rho)}\right|^{-\frac{1}{2}} \int \left|\frac{H}{s^2(\rho)}\right|^{\frac{1}{2}} \left\{ \nu + [\beta-\hat\beta(\rho)]' \left[\frac{H}{s^2(\rho)}\right] [\beta-\hat\beta(\rho)] \right\}^{-\frac{T}{2}}$$

$$\propto [s^2(\rho)]^{-T/2} [s^2(\rho)]^{K/2} |H|^{-\frac{1}{2}}$$

$$\propto [s^2(\rho)]^{-\nu/2} |H|^{-\frac{1}{2}}.$$

$$\therefore\ p(\rho|y) \propto |H|^{-\frac{1}{2}} \left\{ [y-\rho y_{-1} - (X-\rho X_{-1})\hat{\beta}(\rho)]' [y-\rho y_{-1} - (X-\rho X_{-1})\hat{\beta}(\rho)] \right\}^{-\frac{\nu}{2}}$$

$$\propto |H|^{-\frac{1}{2}} \left\{ (y-\rho y_{-1})' [I - (X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'] (y-\rho y_{-1}) \right\}^{-\nu/2}.$$

<u>Because:</u>

$$(y-\rho y_{-1})' [I - (X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'] (y-\rho y_{-1})$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

<u>And</u>:

$$[y-\rho y_{-1} - (X-\rho X_{-1})\hat{\beta}(\rho)]' [y-\rho y_{-1} - (X-\rho X_{-1})\hat{\beta}(\rho)]$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - \hat{\beta}(\rho)'(X-\rho X_{-1})'(y-\rho y_{-1})$$

$$- (y-\rho y_{-1})'(X-\rho X_{-1})\hat{\beta}(\rho) + \hat{\beta}(\rho)'(X-\rho X_{-1})'(X-\rho X_{-1})\hat{\beta}(\rho)$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - [H^{-1}(X-\rho X_{-1})'(y-\rho y_{-1})]'(X-\rho X_{-1})'(y-\rho y_{-1})$$

$$- (y-\rho y_{-1})'(X-\rho X_{-1})[H^{-1}(X-\rho X_{-1})'(y-\rho y_{-1})]$$

$$+ [H^{-1}(X-\rho X_{-1})'(y-\rho y_{-1})]'(X-\rho X_{-1})'(X-\rho X_{-1})[H^{-1}(X-\rho X_{-1})'(y-\rho y_{-1})]$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

$$- (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

$$+ (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} H \cdot H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - 2(y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

$$+ (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1})$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1}) - (y-\rho y_{-1})'(X-\rho X_{-1}) H^{-1} (X-\rho X_{-1})'(y-\rho y_{-1}).$$

<u>So the 2 are equivalent</u>.

Now, consider

$$p(\beta|y) = \int p(\beta, \rho|y)\, d\rho -$$

$$\int \left\{ [y-X\beta - \rho(y_{-1} - X_{-1}\beta)]' [y-X\beta - \rho(y_{-1} - X_{-1}\beta)] \right\}^{-T/2} d\rho$$

$$\propto \int \left\{ (y-X\beta)'(y-X\beta) + \rho^2 (y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) - 2\rho (y_{-1}-X_{-1}\beta)(y-X\beta) \right\}^{-T/2} d\rho$$

$$\propto \int \Big\{ (y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \left[\rho - \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-1} (y_{-1}-X_{-1}\beta)(y-X\beta) \right]^2$$

$$+ (y-X\beta)'(y-X\beta)$$

$$- \left[(y_{-1}-X_{-1}\beta)'(y-X\beta) \right]' \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-1} \left[(y_{-1}-X_{-1}\beta)'(y-X\beta) \right] \Big\}^{-T/2} d\rho$$

$$\propto \int \Big\{ H \left[\rho - \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-1} (y_{-1}-X_{-1}\beta)(y-X\beta) \right]^2$$

$$+ (y-X\beta)'(y-X\beta)$$

$$- \left[(y_{-1}-X_{-1}\beta)'(y-X\beta) \right]' H^{-1} \left[(y_{-1}-X_{-1}\beta)'(y-X\beta) \right] \Big\}^{-T/2} d\rho$$

$$\propto \{\gamma\}^{-T/2} \int \Big\{ 1 + \left[\rho - \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-1} (y_{-1}-X_{-1}\beta)(y-X\beta) \right]' \left(\frac{H}{\gamma} \right)$$

$$\left[\rho - \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-1} (y_{-1}-X_{-1}\beta)(y-X\beta) \right] \Big\}^{-T/2} d\rho$$

$$\propto (\gamma)^{-T/2} \left| \frac{H}{\gamma} \right|^{-\frac{1}{2}} \int \left| \frac{H}{\gamma} \right|^{\frac{1}{2}} \Big\{ \qquad \Big\}^{-T/2} d\rho$$

$$\propto \gamma^{-\left(\frac{T-1}{2}\right)} |H|^{-\frac{1}{2}}$$

$$\propto \left[(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta) \right]^{-\frac{1}{2}} * \left\{ (y-X\beta)'(y-X\beta) - \frac{\left[(y-X\beta)'(y_{-1}-X_{-1}\beta) \right]^2}{(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta)} \right\}^{-\left(\frac{T-1}{2}\right)}$$

What if we are interested in a single element of β? Integrate out all of the unwanted elements? Messy — especially if k is large. So, alternatively —

$$p(\beta_1, \rho | y) = p(\rho | y) \cdot p(\beta_1 | \rho, y)$$

And we have just derived $p(\rho | y)$, and we can obtain $p(\beta_1 | \rho, y)$ by integrating $p(\beta | \rho, y)$ w.r.t. the remaining $(k-1)$ elements of β.

Then, $$\frac{\beta_1 - \hat{\beta}_1(\rho)}{s(\rho)\sqrt{h''}} \sim t_v. \qquad *$$

where h'' is the $(1,1)$th. element of H^{-1}.

Then use <u>bivariate</u> integration to get:

$$p(\beta_1 | y) = \int p(\beta_1, \rho | y)\, d\rho.$$

We can do this because result * tells us the form of the pdf $p(\beta_1 | \rho, y)$. We already know the form of $p(\rho | y)$, so now we know the form of $p(\beta_1, \rho | y)$ then integrate out ρ, as above.

<u>Alternatively</u>: partition $\beta \ni$

$$\beta' = (\beta_1 : \bar{\beta}')$$

$$\text{Let } X = (x_1 : \bar{X})$$

net $\omega = y - \rho y_{-1} - (x_1 - \rho x_{1-1})\beta_1$

Now,

$$p(\beta_1, \bar{\beta}, \rho | y) \propto \left\{ [y - \rho y_{-1} - (X - \rho X_{-1})\beta]' [y - \rho y_{-1} - (X - \rho X_{-1})\beta] \right\}^{-T/2}$$

$$\propto \left\{ \left[y - \rho y_{-1} - \left[\begin{pmatrix} x_1 \\ \bar{X} \end{pmatrix} - \rho \begin{pmatrix} x_{1-1} \\ \bar{X}_{-1} \end{pmatrix}\right](\beta_1 : \bar{\beta})\right]' [\qquad\qquad] \right\}^{-T/2}$$

$$\propto \left\{ [y - \rho y_{-1} - (x_1 - \rho x_{1-1})\beta_1 - (\bar{X} - \rho \bar{X}_{-1})\bar{\beta}]' [\qquad\qquad] \right\}^{-T/2}$$

$$\propto \left\{ [\omega - (\bar{X} - \rho \bar{X}_{-1})\bar{\beta}]' [\qquad\qquad] \right\}^{-T/2}$$

Then, $p(\beta_1, \rho | y) = \int p(\beta_1, \bar{\beta}, \rho | y)\, d\bar{\beta}$

$$\propto \int \left\{ \omega'\omega + \bar{\beta}'(\bar{X} - \rho\bar{X}_{-1})'(\bar{X} - \rho\bar{X}_{-1})\bar{\beta} - 2\bar{\beta}'(\bar{X} - \rho\bar{X}_{-1})'\omega \right\}^{-T/2} d\beta$$

Let $\bar{H} = (\bar{X} - \rho\bar{X}_{-1})'(\bar{X} - \rho\bar{X}_{-1})$

$$\propto \int \left\{ \bar{H}\left[\bar{\beta} - \bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})'\omega\right]^2 + (\omega'\omega) - \bar{H}\left[(\bar{X} - \rho\bar{X}_{-1})'\omega\right]^2 \right\}^{-T/2} d\bar{\beta}$$

$$\propto \left\{ \omega'\omega - (\omega')(\bar{X} - \rho\bar{X}_{-1})\bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})'\omega \right\}^{-T/2}$$

$$* \int \left\{ 1 + [\bar{\beta} - \bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})'\omega] \frac{\bar{H}}{\{\qquad\}} [\bar{\beta} - \qquad] \right\}^{-T/2} d\bar{\beta}$$

$$\propto \left\{ \omega'\omega - \omega'(\bar{X} - \rho\bar{X}_{-1})\bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})'\omega \right\}^{-T/2} * \left\{ \qquad\qquad \right\}^{\frac{(k-1)}{2}}$$

$$* \, |\bar{H}|^{-\frac{1}{2}}$$

$$\propto \left\{ (\omega'\omega) - \omega'(\bar{X} - \rho\bar{X}_{-1})\,\bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})'\omega \right\}^{-(T-k+1)/2} * \left|\bar{H}\right|^{-\frac{1}{2}}$$

Then integrate w.r.t. ρ to get $p(\beta_1 | y)$. The advantage is that $\bar{H}^{-1}$ is inversion of a $(k-1) * (k-1)$ matrix, c.f. H^{-1} is inversion of a $(k * k)$ matrix.

B) Regressions with Unequal Variances:

Consider 2 regressions based on the same model, but 2 different samples –

$$y_1 = X_1\beta + u_1 \quad ; \quad u_1 \sim IN(0, \sigma_1^2)$$

$$y_2 = X_2\beta + u_2 \quad ; \quad u_2 \sim IN(0, \sigma_2^2)$$

If $\sigma_1^2 = \sigma_2^2$, or if $\sigma_1^2 = c\sigma_2^2$ (for known c) then pool the data for y & X, & compute:

$$y = X\beta + u.$$

But what if $\sigma_1^2 \neq \sigma_2^2$? Then consider 2 cases –

(a) σ_1^2 known, but σ_2^2 unknown.

(b) Both σ_1^2 & σ_2^2 unknown.

a) σ_1^2 Known, σ_2^2 Unknown:

$$\ell(\beta, \sigma_2 | y, \sigma_1) \propto \left(\frac{1}{\sigma_2^{n_2}}\right) \exp\left\{ -\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) - \frac{1}{2\sigma_2^2}(y_2 - X_2\beta)'(y_2 - X_2\beta) \right\}$$

$$p(\beta, \sigma_2) \propto \left(\frac{1}{\sigma_2}\right) \quad ; \quad \begin{cases} -\infty < \beta_i < \infty \quad ; \quad i = 1, 2, \ldots, k \\ 0 < \sigma_2 < \infty. \end{cases}$$

$$\therefore p(\beta, \sigma_2 | \sigma_1, y) \propto \left(\frac{1}{\sigma_2^{n_2+1}}\right) \exp\left\{ -\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) - \frac{1}{2\sigma_2^2}(y_2 - X_2\beta)'(y_2 - X_2\beta) \right\}.$$

$$\therefore p(\beta | \sigma_1, y) \propto \exp\left\{ -\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) \right\} * \left[(y_2 - X_2\beta)'(y_2 - X_2\beta) \right]^{-\frac{n_2}{2}}.$$

Let $Z_1 = (X_1'X_1)$; $Z_2 = (X_2'X_2)$; $\hat{\beta}_1 = Z_1^{-1}(X_1'y_1)$;
$\hat{\beta}_2 = Z_2^{-1}(X_2'y_2)$; $v_1 = n_1 - k$; $v_2 = n_2 - k$;
$v_2 s_2^2 = (y_2 - X_2\hat{\beta}_2)'(y_2 - X_2\hat{\beta}_2)$.

So, $\exp\{-\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta)\}$

$= \exp\{-\frac{1}{2\sigma_1^2}(y_1'y_1 + \beta'X_1'X_1\beta - 2y_1'X_1\beta)\}$

$= \exp\{-\frac{1}{2\sigma_1^2}(y_1'y_1 + \beta'Z_1\beta - 2y_1'X_1\beta)\}$

And $\hat{\beta}_1 = Z_1^{-1}(X_1'y_1)$

$\therefore Z_1\hat{\beta}_1 = X_1'y_1$

$\therefore -2\beta'X_1'y_1 = -2\beta Z_1\hat{\beta}_1$

$\therefore$ get : $\exp\{-\frac{1}{2\sigma_1^2}(y_1'y_1 + \beta'Z_1\beta - 2\beta'Z_1\hat{\beta}_1)\}$

And $\hat{\beta}_1'Z_1\hat{\beta}_1 = (y_1'X_1)Z_1^{-1}Z_1Z_1^{-1}X_1'y_1$

$= y_1'y_1$

So, $\exp\{-\frac{1}{2\sigma_1^2}(y_1'y_1 + \beta'X_1'X_1\beta - 2y_1'X_1\beta)\}$

$= \exp\{-\frac{1}{2\sigma_1^2}(\beta - \hat{\beta}_1)'Z_1(\beta - \hat{\beta}_1)\}$.

Further, $[(y_2 - X_2\beta)'(y_2 - X_2\beta)]^{-\frac{1}{2}n_2}$

$= \{y_2'y_2 + \beta'X_2'X_2\beta - 2y_2'X_2\beta\}^{-\frac{1}{2}n_2}$

$= \{[(y_2 - X_2\hat{\beta}_2) - (X_2\beta - X_2\hat{\beta}_2)]'[(y_2 - X_2\hat{\beta}_2) - (X_2\beta - X_2\hat{\beta}_2)]\}^{-\frac{1}{2}n_2}$

$= \{(y_2 - X_2\hat{\beta}_2)'(y_2 - X_2\hat{\beta}_2) + (X_2\beta - X_2\hat{\beta}_2)'(X_2\beta - X_2\hat{\beta}_2)\}^{-\frac{1}{2}n_2}$

$= \{v_2 s_2^2 + (\beta - \hat{\beta}_2)'X_2'X_2(\beta - \hat{\beta}_2)\}^{-\frac{1}{2}n_2}$

$= \{v_2 s_2^2 + (\beta - \hat{\beta}_2)'Z_2(\beta - \hat{\beta}_2)\}^{-\frac{1}{2}n_2}$

$\propto \left\{1 + \frac{(\beta - \hat{\beta}_2)'Z_2(\beta - \hat{\beta}_2)}{v_2 s_2^2}\right\}^{-\frac{1}{2}n_2}$.

But $n_2 = v_2 + k$.

$\therefore p(\beta|\sigma_1, y) \propto \exp\{-\frac{1}{2\sigma_1^2}(\beta - \hat{\beta}_1)'Z_1'(\beta - \hat{\beta}_1)\} * \left\{1 + \frac{(\beta - \hat{\beta}_2)'Z_2(\beta - \hat{\beta}_2)}{v_2 s_2^2}\right\}^{-\frac{(v_2+k)}{2}}$.

So the posterior density for β is the product of a MVN density & a MVT density. We call this product a "normal-t" density.

Now, expand the second factor (the MVT factor) of this density asymptotically, and approximate by taking only the first term —

Let $Q_2 = \frac{1}{S_2^2}(\beta-\hat{\beta}_2)' Z_2 (\beta-\hat{\beta}_2)$

Then the MVT term is of the form $\left(1+\frac{Q_2}{v_2}\right)^{-\frac{1}{2}(v_2+k)}$

$$\therefore \left(1+\frac{Q_2}{v_2}\right)^{-\frac{1}{2}(v_2+k)} = \exp\left[\log\left(1+\frac{Q_2}{v_2}\right)^{-\frac{1}{2}(v_2+k)}\right]$$

$$= \exp\left\{-\left(\frac{v_2+k}{2}\right)\log\left(1+\frac{Q_2}{v_2}\right)\right\}$$

Now, $\log\left(1+\frac{Q_2}{v_2}\right) = \left(\frac{Q_2}{v_2}\right) - \frac{1}{2}\left(\frac{Q_2}{v_2}\right)^2 + \frac{1}{3}\left(\frac{Q_2}{v_2}\right)^3 - \cdots\cdots$

$$= \left(\frac{Q_2}{v_2}\right) + R.$$

$$\therefore \left(1+\frac{Q_2}{v_2}\right)^{-\frac{1}{2}(v_2+k)} = \exp\left\{-\frac{(v_2+k)}{2}\left[R+\frac{Q_2}{v_2}\right]\right\}.$$

$$= \exp\left\{-\frac{1}{2}\left[(v_2+k)R + k\left(\frac{Q_2}{v_2}\right) + Q_2\right]\right\}$$

$$= \exp\left\{-\frac{Q_2}{2}\right\}\exp\left\{-\frac{1}{2}\left[k\left(\frac{Q_2}{v_2}\right) + (v_2+k)R\right]\right\}$$

But $e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots\cdots$

So, $\exp\left\{-\frac{1}{2}\left[k\frac{Q_2}{v_2} + (v_2+k)R\right]\right\}$

$$= 1 - \frac{1}{2}\left[k\frac{Q_2}{v_2} + (v_2+k)R\right] + \frac{1}{4}\left[k\,Q_2/v_2 + (v_2+k)R\right]^2 + \cdots\cdots$$

$$\therefore \left(1+\frac{Q_2}{v_2}\right)^{-\frac{1}{2}(v_2+k)} = \exp\left\{-\frac{Q_2}{2}\right\}\left\{1 - \frac{1}{2}\left[k\frac{Q_2}{v_2} + (v_2+k)R\right] + \frac{1}{4}\left[k\frac{Q_2}{v_2} + (v_2+k)R\right]^2 - \cdots\cdots\right\}$$

$$= \exp\left\{-\frac{Q_2}{2}\right\}\sum_0^\infty q_i v_2^{-i}.$$

$$\therefore p(\beta|\sigma_1, y) \propto \exp\left\{-\frac{1}{2\sigma_1^2}(\beta-\hat{\beta}_1)' Z_1 (\beta-\hat{\beta}_1)\right\} *$$

$$\exp\left\{-\frac{1}{2S_2^2}(\beta-\hat{\beta}_2)' Z_2 (\beta-\hat{\beta}_2)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_1^2}(\beta-\hat{\beta}_1)' Z_1 (\beta-\hat{\beta}_1) - \frac{1}{2S_2^2}(\beta-\hat{\beta}_2)' Z_2 (\beta-\hat{\beta}_2)\right\}.$$

Now, complete the square on β :

Exponent becomes:

$$-\tfrac{1}{2}\left\{ \frac{1}{\sigma_1^2}\left[\beta' Z_1 \beta + \hat\beta_1' Z_1 \hat\beta_1 - 2\beta' Z_1 \hat\beta_1\right] + \frac{1}{S_2^2}\left[\beta' Z_2 \beta + \hat\beta_2' Z_2 \hat\beta_2 - 2\beta' Z_2 \hat\beta_2\right] \right\}$$

$$= -\tfrac{1}{2}\left\{ \beta'\left[\frac{Z_1}{\sigma_1^2} + \frac{Z_2}{S_2^2}\right]\beta - 2\beta'\left[\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right] + \left[\frac{1}{\sigma_1^2}\hat\beta_1' Z_1 \hat\beta_1 + \frac{1}{S_2^2}\hat\beta_2' Z_2 \hat\beta_2\right] \right\}$$

$$= -\tfrac{1}{2}\left\{ \left[\beta - \frac{\left(\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right)}{A}\right] A \left[\beta - A^{-1}\left(\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right)\right] \right\}$$

$$= -\tfrac{1}{2}\left\{ \beta' A \beta + \left(\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right)' A^{-1}\left(\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right) - 2\beta'\left[\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right] + \left[\frac{1}{\sigma_1^2}\hat\beta_1' Z_1 \hat\beta_1 + \frac{1}{S_2^2}\hat\beta_2' Z_2 \hat\beta_2\right] - \left[\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right]' A^{-1}\left[\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right] \right\}$$

$$= -\tfrac{1}{2}\left\{ \beta' A \beta - 2\beta'\left[\frac{Z_1\hat\beta_1}{\sigma_1^2} + \frac{Z_2\hat\beta_2}{S_2^2}\right] + \frac{1}{\sigma_1^2}\hat\beta_1' Z_1 \hat\beta_1 + \frac{1}{S_2^2}\hat\beta_2' Z_2 \hat\beta_2 \right\}$$

$$= -\tfrac{1}{2}\left\{ \beta' A \beta - 2\beta'\left[\frac{1}{\sigma_1^2}X_1' y_1 + \frac{1}{S_2^2}X_2' y_2\right] + \frac{1}{\sigma_1^2} y_1' X_1 Z^{-1} X_1' y_1 + \frac{1}{S_2^2} y_2' X_2 Z^{-1} X_2' y_2 \right\}$$

Let $\tilde\beta = A^{-1}\left(\frac{1}{\sigma_1^2}X_1' y_1 + \frac{1}{S_2^2}X_2' y_2\right)$

$\therefore\ A\tilde\beta = \left(\frac{1}{\sigma_1^2}X_1' y_1 + \frac{1}{S_2^2}X_2' y_2\right)$

So the exponent becomes:

$$-\tfrac{1}{2}\left\{ \beta' A \beta - 2\beta' A \tilde\beta + \frac{1}{\sigma_1^2} y_1' X_1 Z_1^{-1} X_1' y_1 + \frac{1}{S_2^2} y_2' X_2 Z_2^{-1} X_2' y_2 \right\}$$

$$= -\tfrac{1}{2}\left\{ \beta' A \beta - 2\beta' A \tilde\beta + \gamma \right\}, \text{ say.}$$

$\Rightarrow \gamma = \tilde\beta' A \tilde\beta.$

So,

$$p(\beta \mid \sigma_1, y) \propto \exp\left\{ -\tfrac{1}{2}\left(\beta' A \beta - 2\beta' A \tilde\beta + \tilde\beta' A \tilde\beta\right)\right\}$$

$$\propto \exp\left\{ -\tfrac{1}{2}(\beta - \tilde\beta)' A (\beta - \tilde\beta)\right\}.$$

Now, $\tilde{\beta} = A^{-1}\left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{s_2^2}X_2'y_2\right)$

$\therefore \tilde{\beta} = \left(\frac{1}{\sigma_1^2}Z_1 + \frac{1}{s_2^2}Z_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{s_2^2}X_2'y_2\right)$

And this happens to coincide with the estimator (sampling theory) recommended by Theil. He gives a large-sample justification for this estimator & this is in keeping with the fact that we got our final situation by taking an asymptotic expansion.

(b) Both σ_1^2 and σ_2^2 Unknown:

Clearly, this is the more realistic case.

Then, $\ell(\beta,\sigma_1,\sigma_2|y) \propto \left(\frac{1}{\sigma_1^{n_1}\sigma_2^{n_2}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}(y_1-X_1\beta)'(y_1-X_1\beta) - \frac{1}{2\sigma_2^2}(y_2-X_2\beta)'(y_2-X_2\beta)\right\}$

Let $p(\beta,\sigma_1,\sigma_2) \propto \frac{1}{\sigma_1\sigma_2}$.

$\therefore p(\beta,\sigma_1,\sigma_2|y) \propto \left(\frac{1}{\sigma_1^{n_1+1}\cdot\sigma_2^{n_2+1}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}(y_1-X_1\beta)'(y_1-X_1\beta) - \frac{1}{2\sigma_2^2}(y_2-X_2\beta)'(y_2-X_2\beta)\right\}$

$\therefore p(\beta,\sigma_1|y) \propto \left(\frac{1}{\sigma_1^{n_1+1}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}(y_1-X_1\beta)'(y_1-X_1\beta)\right\} *$

$\int \frac{1}{\sigma_2^{n_2+1}}\cdot\exp\left[-\frac{1}{2\sigma_2^2}(y_2-X_2\beta)'(y_2-X_2\beta)\right]d\sigma_2$

$\propto \left(\frac{1}{\sigma_1^{n_1+1}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}(y_1-X_1\beta)'(y_1-X_1\beta)\right\} * \left[(y_2-X_2\beta)'(y_2-X_2\beta)\right]^{-\frac{n_2}{2}}$

$\therefore p(\beta|y) \propto \left[(y_1-X_1\beta)'(y_1-X_1\beta)\right]^{-\frac{n_1}{2}}\left[(y_2-X_2\beta)'(y_2-X_2\beta)\right]^{-\frac{n_2}{2}}$

Consider $\left[(y_i-X_i\beta)'(y_i-X_i\beta)\right]^{-n_i/2}$:

$= \left\{\left[(y_i-X_i\hat{\beta}_i)-(X_i\beta-X_i\hat{\beta}_i)\right]'\left[(y_i-X_i\hat{\beta}_i)-(X_i\beta-X_i\hat{\beta}_i)\right]\right\}^{-\frac{n_i}{2}}$

$= \left\{(y_i-X_i\hat{\beta}_i)'(y_i-X_i\hat{\beta}_i) + (X_i\beta-X_i\hat{\beta}_i)'(X_i\beta-X_i\hat{\beta}_i)\right\}^{-\frac{n_i}{2}}$

$= \left\{v_i s_i^2 + (\beta-\hat{\beta}_i)'X_i'X_i(\beta-\hat{\beta}_i)\right\}^{-\frac{(v_i+k)}{2}}$

$= \left\{v_i s_i^2 + (\beta-\hat{\beta}_i)'Z_i(\beta-\hat{\beta}_i)\right\}^{-(v_i+k)/2}.$

$\therefore p(\beta|y) \propto \left[v_1 s_1^2 + (\beta-\hat{\beta}_1)'Z_1(\beta-\hat{\beta}_1)\right]^{-\frac{(v_1+k)}{2}} *$

$\left[v_2 s_2^2 + (\beta-\hat{\beta}_2)'Z_2(\beta-\hat{\beta}_2)\right]^{-\frac{(v_2+k)}{2}}.$

$$\therefore p(\beta|y) \propto \left\{1 + \frac{(\beta-\hat{\beta}_1)' Z_1 (\beta-\hat{\beta}_1)}{v_1 s_1^2}\right\}^{-\frac{(v_1+k)}{2}} * \left\{1 + \frac{(\beta-\hat{\beta}_2)' Z_2 (\beta-\hat{\beta}_2)}{v_2 s_2^2}\right\}^{-\frac{(v_2+k)}{2}}$$

And this pdf is the product of 2 MVT-distns, so we call it a "multivariate double-t" pdf.

Now, just as we were able to take an asymptotic expansion & get analogy to Theil's estimator in the case where <u>one</u> variance was unknown, here with <u>both</u> variances unknown we can expand <u>each</u> term in $p(\beta|y)$ asymptotically, and it turns out that we get something very similar to a generalized least squares estimator; which is not very surprising really.

Using the earlier result, we have:

$$\left(1 + \frac{Q_i}{v_i}\right)^{-\frac{(v_i+k)}{2}} \doteq \exp\left\{-\frac{Q_i}{2}\right\} \sum_0^\infty p_j v_i^{-j}$$

So, $p(\beta|y) \dot{\propto} \exp\left[-\frac{1}{2}(Q_1+Q_2)\right] \sum_i \sum_j q_i p_j v_1^{-j} v_2^{-i}$

$$\dot{\propto} \exp\left\{-\frac{1}{2}\left[\frac{1}{s_1^2}(\beta-\hat{\beta}_1)' Z_1 (\beta-\hat{\beta}_1) + \frac{1}{s_2^2}(\beta-\hat{\beta}_2)' Z_2 (\beta-\hat{\beta}_2)\right]\right\}$$

Let $M_i = Z_i / s_i^2$

$D = M_1 + M_2$

$\bar{\beta} = D^{-1}(M_1\hat{\beta}_1 + M_2\hat{\beta}_2)$

$$\therefore p(\beta|y) \dot{\propto} \exp\left\{-\frac{1}{2}\left[(\beta-\hat{\beta}_1)' M_1 (\beta-\hat{\beta}_1) + (\beta-\hat{\beta}_2)' M_2 (\beta-\hat{\beta}_2)\right]\right\}$$

$$\dot{\propto} \exp\left\{-\frac{1}{2}(\beta-\bar{\beta})' D (\beta-\bar{\beta})\right\}.$$

Which gives the leading normal term

Now,

$$\bar{\beta} = D^{-1}(M_1\hat{\beta}_1 + M_2\hat{\beta}_2)$$

$$= \left(\frac{1}{s_1^2}X_1'X_1 + \frac{1}{s_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{s_1^2}(X_1'X_1)(X_1'X_1)^{-1}X_1'y_1 + \frac{1}{s_2^2}(X_2'X_2)(X_2'X_2)^{-1}X_2'y_2\right)$$

$$= \left(\frac{1}{s_1^2}X_1'X_1 + \frac{1}{s_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{s_1^2}X_1'y_1 + \frac{1}{s_2^2}X_2'y_2\right)$$

And $\bar{\beta}$ is the <u>mean</u> of the leading normal term in the asymptotic expansion of $p(\beta|y)$.

And $\bar{\beta}$ is just the GLS estimator with s_i^2 replacing σ_i^2; $i = 1, 2$.

viz: the GLS. estimator is —

$$(X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y)$$

$$= \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right)$$

Since $X' = (X_1' : X_2')$; $y' = (y_1' : y_2')$; $\Sigma = \begin{pmatrix} \sigma_1^2 I_{n_1} & 0 \\ 0 & \sigma_2^2 I_{n_2} \end{pmatrix}$.

If we set $\sigma_i^2 = s_i^2$, we get an approx. to the GLS. estimator which is generally given <u>asymptotic justification</u>.

Now, we can also get close to the GLS estimator from the Bayesian approach, by looking at <u>conditional</u> posterior pdf for β — $p(\beta \mid \sigma_1, \sigma_2, y)$

Now, $p(\beta \mid \sigma_1, \sigma_2, y) \propto (\sigma_1^{n_1+1} \cdot \sigma_2^{n_2+1})^{-1} \exp\left\{-\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) - \frac{1}{2\sigma_2^2}(y_2 - X_2\beta)'(y_2 - X_2\beta)\right\}$

$$\propto \exp\left\{-\frac{1}{2\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) - \frac{1}{2\sigma_2^2}(y_2 - X_2\beta)'(y_2 - X_2\beta)\right\}.$$

Let $\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y)$

$$= \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right)$$

So, $(\beta - \tilde{\beta})'(X'\Sigma^{-1}X)(\beta - \tilde{\beta})$

$$= (\beta - (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y))'(X'\Sigma^{-1}X)(\beta - (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y))$$

$$= \beta'(X'\Sigma^{-1}X)\beta + (X'\Sigma^{-1}y)'(X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y) - 2\beta'(X'\Sigma^{-1}X)(X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y) *$$

$$= \beta'\left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)\beta + \left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right)' * \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right)$$

$$- 2\beta'\left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)\left(\text{———}\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right)$$

$$= \left(\frac{1}{\sigma_1^2}\beta'X'X\beta\right) + \left(\frac{1}{\sigma_2^2}\beta'X_2'X_2\beta\right) + \left(\frac{1}{\sigma_1^2}y_1'y_1\right) + \left(\frac{1}{\sigma_2^2}(y_2'y_2)\right)$$

$$- 2\beta'\left[\frac{1}{\sigma_1^2}X_1'y_1 + \frac{1}{\sigma_2^2}X_2'y_2\right]$$

$$= \frac{1}{\sigma_1^2}(y_1 - X_1\beta)'(y_1 - X_1\beta) + \frac{1}{\sigma_2^2}(y_2 - X_2\beta)'(y_2 - X_2\beta)$$

$\therefore p(\beta \mid \sigma_1, \sigma_2, y) \propto \exp\left\{-\frac{1}{2}(\beta - \tilde{\beta})'(X'\Sigma^{-1}X)(\beta - \tilde{\beta})\right\}$

Which is normal with mean $\tilde{\beta}$ & c.v. matrix $(X'\Sigma^{-1}X)^{-1}$

$\therefore$ c.m. $= \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}$.

So, the <u>conditional</u> posterior mean for β is $\tilde{\beta}$, and

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y)$$

which is just the G.L.S. estimator.

As long as we have a large sample, there will be little loss involved in replacing σ_i^2 by s_i^2 in our estimator. So, the use of the conditional pdf may be satisfactory enough for large samples. However, in general we do <u>not</u> work with large samples. Then it is best to integrate out σ_1 & σ_2 to leave a <u>marginal</u> posterior pdf for β. Use <u>this</u> for making inferences. That is, use the <u>double-t</u> distribution in general, rather than its normal approx. or the conditional posterior (which is normal as well.)

Now, suppose that rather than being interested in the coefficients, we are instead interested in their <u>variances</u> i.e. in σ_1^2 & in σ_2^2.

Then let $\lambda = \left(\sigma_1^2/\sigma_2^2\right)$

$$\therefore \frac{d\lambda}{d\sigma_2} = -2\,\sigma_1^2/\sigma_2^3 \propto \sigma_1^2/\sigma_2^3 \propto \sigma_1\lambda^{-3/2}.$$

$$\therefore J \propto \sigma_1\lambda^{-3/2}$$

Now, $p(\lambda)d\lambda \propto p(\sigma_2)\sigma_1\lambda^{-3/2}d\sigma_2$

$$\therefore p(\lambda,\sigma_1,\beta|y) \propto p(\sigma_1,\sigma_2,\beta|y)\cdot\sigma_1\lambda^{-3/2}.$$

And so,

$$p(\lambda,\sigma_1,\beta|y) \propto \left(\frac{\lambda^{(n_2-2)/2}}{\sigma_1^{n_1+n_2+1}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}\left[(y_1-X_1\beta)'(y_1-X_1\beta) + \lambda(y_2-X_2\beta)'(y_2-X_2\beta)\right]\right\}.$$

Now, before integrating this pdf to get $p(\sigma_1,\lambda|y)$, complete the square on the exponent:

$$(y_1-X_1\beta)'(y_1-X_1\beta) + \lambda(y_2-X_2\beta)'(y_2-X_2\beta)$$

$$= y_1'y_1 + \beta'X_1'X_1\beta - 2\beta'X_1'y_1 + \lambda y_2'y_2 + \lambda\beta'X_2'X_2\beta - 2\lambda\beta'X_2'y_2$$

$$= \beta'(X_1'X_1 + \lambda X_2'X_2)\beta - 2\beta'(X_1'y_1 + \lambda X_2'y_2) + (y_1'y_1 + \lambda y_2'y_2)$$

<u>Let</u>: $C_1 = (X_1'X_1 + \lambda X_2'X_2)$; $C_2 = (X_1'y_1 + \lambda X_2'y_2)$.

Then: the exponent becomes —

$$\beta' C_1 \beta - 2\beta' C_2 + y_1' y_1 + \lambda y_2' y_2$$

$$= (\beta - C_1^{-1} C_2)' C_1 (\beta - C_1^{-1} C_2) + y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2$$

$$\therefore p(\beta, \sigma_1, \lambda | y) \propto \left(\frac{\lambda^{(n_2-2)/2}}{\sigma_1^{n_1+n_2+1}}\right) * \exp\left\{-\frac{1}{2\sigma_1^2}\left[(\beta - C_1^{-1}C_2)' C_1 (\beta - C_1^{-1}C_2) + y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2\right]\right\}$$

Now, the exponent is:

$$\exp\left\{-\frac{1}{2\sigma_1^2}(y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2)\right\} * \exp\left\{-\frac{1}{2\sigma_1^2}(\beta - C_1^{-1}C_2)' C_1 (\beta - C_1^{-1}C_2)\right\}.$$

$$\therefore p(\sigma_1, \lambda | y) \propto \left(\frac{\lambda^{(n_2-2)/2}}{\sigma_1^{n_1+n_2+1}}\right) \exp\left\{-\frac{1}{2\sigma_1^2}(y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2)\right\} * \int \exp\left\{-\frac{1}{2\sigma_1^2}(\beta - C_1^{-1}C_2)' C_1 (\beta - C_1^{-1}C_2)\right\} d\beta$$

And the integral is

$$\Sigma|^{\frac{1}{2}} \int |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_1^2}(\beta - C_1^{-1}C_2)' C_1 (\beta - C_1^{-1}C_2)\right\}$$

where $\Sigma^{-1} = \left(\frac{C_1}{\sigma_1^2}\right) \quad \therefore \Sigma = \left(\frac{\sigma_1^2}{C_1}\right)$

$$\therefore I \propto \frac{|C_1|^{-\frac{1}{2}}}{(\sigma_1^{-k})} \int \text{MVN}.$$

$$\propto \frac{|C_1|^{-\frac{1}{2}}}{\sigma_1^{-k}}$$

$$\therefore p(\sigma_1, \lambda | y) \propto |C_1|^{-\frac{1}{2}} \left(\frac{\lambda^{(n_2-2)/2}}{\sigma_1^{n_1+n_2-k+1}}\right) \exp\left\{-\frac{1}{2\sigma_1^2}(y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2)\right\}$$

And, $p(\lambda | y) = \int_0^\infty p(\sigma_1, \lambda | y)\, d\sigma_1$

$$\propto |C_1|^{-\frac{1}{2}} \lambda^{(n_2-2)/2} \int \left(\frac{1}{\sigma_1^{n_1+n_2-k+1}}\right) \exp\left\{-\frac{1}{2\sigma_1^2}(y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2)\right\}$$

$$\propto |C_1|^{-\frac{1}{2}} \lambda^{(n_2-2)/2} \left\{y_1' y_1 + \lambda y_2' y_2 - C_2' C_1^{-1} C_2\right\}^{-(n_1+n_2-k)/2}.$$

$$\propto \left(\lambda^{(n_2-2)/2} |X_1' X_1 + \lambda X_2' X_2|^{-\frac{1}{2}}\right) * \left\{y_1' y_1 + \lambda y_2' y_2 - (X_1' y_1 + \lambda X_2' y_2)' * (X_1' X_1 + \lambda X_2' X_2)^{-1} (X_1' y_1 + \lambda X_2' y_2)\right\}^{-\frac{(n_1+n_2-k}{2}}$$

And this joint pdf on λ (i.e. on σ_1^2/σ_2^2) is a little different from an F-dist'n, the difference arising because we assumed in our original model that the same β appears in each model. i.e. we have imposed a constraint & got away from the general situation where a different β appears in each model. Then an F-dist'n. does in fact arise.

(c) Two Regressions With Some Common Coefficients:

Suppose that the coefficients appearing in the Equations are not entirely the same — only some subvector is common to each.

eg
$$y_1 = W_1\beta_1 + W_2\beta_2 + u_1$$
$$y_2 = Z_1\beta_1 + Z_2\gamma_2 + u_2$$

NB $u_1, u_2 \sim I.N(0,\sigma^2)$: i.e. same variance.

y_1 is $(n_1 \times 1)$; y_2 is $(n_2 \times 1)$
$(W_1 : W_2)$ is $(n_1 * k_1)$; $(Z_1 : Z_2)$ is $(n_2 * k_2)$
β_1 is $(m*1)$; β_2 is $(m_1 * 1)$; γ_2 is $(m_2 * 1)$

$$\therefore \begin{cases} k_1 = m + m_1 \\ k_2 = m + m_2 \end{cases}$$

W_1 is $(n_1 * m)$, so W_2 is $(n_1 * m_1)$
Z_1 is $(n_2 * m)$, so Z_2 is $(n_2 * m_2)$

$$\therefore \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} W_1 & W_2 & 0 \\ Z_1 & 0 & Z_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

or, $y = X\beta + u$. $\sim$ $(n_1 + n_2)$ obs. in all.

assume a diffuse prior, then the basic Bayesian result is that

$$p(\beta|y) \propto \left\{ \nu s^2 + (\beta - \hat{\beta})' X'X (\beta - \hat{\beta}) \right\}^{-(n_1+n_2)/2}$$

Where $\nu = n_1 + n_2 - m - m_1 - m_2$

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\nu s^2 = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

What is the <u>marginal</u> posterior for β_1?

<u>Partition:</u> $\beta' = (\beta_1' : \gamma')$

where $\gamma' = (\beta_2' : \gamma_2')$

Let $(X'X) = \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right)$

Now – $(X'X) = \begin{pmatrix} \omega_1 & \omega_2 & 0 \\ z_1 & 0 & z_2 \end{pmatrix}' \begin{pmatrix} \omega_1 & \omega_2 & 0 \\ z_1 & 0 & z_2 \end{pmatrix}$

$$= \begin{pmatrix} \omega_1' & z_1' \\ \omega_2' & 0 \\ 0 & z_2' \end{pmatrix} \begin{pmatrix} \omega_1 & \omega_2 & 0 \\ z_1 & 0 & z_2 \end{pmatrix}$$

$$= \left(\begin{array}{c|cc} \omega_1'\omega_1 + z_1'z_1 & \omega_1'\omega_2 & z_1'z_2 \\ \hline \omega_2'\omega_1 & \omega_2'\omega_2 & 0 \\ z_2'z_1 & 0 & z_2'z_2 \end{array}\right)$$

So: $M_{11} = (\omega_1'\omega_1 + z_1'z_1)$

$M_{12} = (\omega_1'\omega_2 \,,\; z_1'z_2)$

$M_{21} = \begin{pmatrix} \omega_2'\omega_1 \\ z_2'z_1 \end{pmatrix}$

$M_{22} = \begin{pmatrix} \omega_2'\omega_2 & 0 \\ 0 & z_2'z_2 \end{pmatrix}$

Then, making use of the usual results derived for Ch 3 –

$$p(\beta_1 | y) \propto \left\{ \nu s^2 + (\beta_1 - \hat{\beta}_1)' H (\beta_1 - \hat{\beta}_1) \right\}^{-(n_1+n_2-m_1-m_2)/2}$$

where $H = M_{11} - M_{12} M_{22}^{-1} M_{21}$

$\therefore H = \omega_1'\omega_1 + z_1'z_1 - (\omega_1'\omega_2 \,,\; z_1'z_2) \begin{pmatrix} (\omega_2'\omega_2)^{-1} & 0 \\ 0 & (z_2'z_2)^{-1} \end{pmatrix} \begin{pmatrix} \omega_2'\omega_1 \\ z_2'z_1 \end{pmatrix}$

$= \omega_1'\omega_1 + z_1'z_1 - \left(\omega_1'\omega_2(\omega_2'\omega_2)^{-1} \,,\; z_1'z_2(z_2'z_2)^{-1}\right) \begin{pmatrix} \omega_2'\omega_1 \\ z_2'z_1 \end{pmatrix}$

$= \omega_1'\omega_1 + z_1'z_1 - \omega_1'\omega_2(\omega_2'\omega_2)^{-1}\omega_2'\omega_1 + z_1'z_2(z_2'z_2)^{-1}z_2'z_1$

& $\nu = n_1 + n_2 - m - m_1 - m_2$

So, $(\nu + m) = (n_1 + n_2 - m_1 - m_2)$.

$$\therefore\ p(\beta_1 | y) \propto \left\{ \nu s^2 + (\beta_1 - \hat{\beta}_1)' H (\beta_1 - \hat{\beta}_1) \right\}^{-(\nu+m)/2}$$

Note that $\hat{\beta}_1$ is a sub-vector of $\hat{\beta}$, where

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

& $\hat{\beta}_1 = H^{-1}(V_1 \tilde{\beta}_1 + V_2 \check{\beta}_2)$

where $V_1 = \omega_1'\omega_1 - \omega_1'\omega_2 (\omega_2'\omega_2)^{-1} \omega_2'\omega_1$

$V_2 = z_1' z_1 - z_1' z_2 (z_2' z_2)^{-1} z_2' z_1$

$\therefore H = V_1 + V_2.$

& $\tilde{\beta}$ is O.L.S. estimator of β_1 from 1st eqn —

So $\tilde{\beta}_1 = f\left\{ \left[(\omega_1 : \omega_2)'(\omega_1 : \omega_2) \right]^{-1} \left\{ (\omega_1 : \omega_2)' y_1 \right] \right\}$ (fcn of)

& $\check{\beta}_1$ is OLS est of β_1 for 2nd eqn —

$\check{\beta}_1 = f\left\{ \left[(z_1 : z_2)'(z_1 : z_2) \right]^{-1} \left[(z_1 : z_2)' y_2 \right] \right\}$ (fcn. of)

[So $\tilde{\beta}_1$ is part of the full OLS est. & same for $\check{\beta}_1$]

So $\hat{\beta}_1$ is a <u>weighted average</u> of the $\tilde{\beta}_1$ & $\check{\beta}_1$ that would be obtained if the 2 equations were handled separately.

Zellner: Chapter 6.

"Non-Linear Models."

(A) The Box-Cox Analysis:-

Let $y_\alpha > 0$; $\alpha = 1, 2, -----, n$.

Then consider the following transformation of the dependent variable —

$$y_\alpha \longrightarrow \left(\frac{y_\alpha^\lambda - 1}{\lambda}\right) \quad ; \quad \lambda \geq 1$$

$$\longrightarrow \log y_\alpha \quad ; \quad \lambda = 0.$$

Then the regression is:

$$\left(\frac{y_\alpha^\lambda - 1}{\lambda}\right) = \beta_1 + \beta_2 x_{2\alpha} + \beta_3 x_{3\alpha} + ---- + \beta_K x_{K\alpha} + u_\alpha.$$

Then, assume that for some <u>unknown</u> λ, the <u>transformed</u> dependent variables satisfy the conditions:

(a) Normally distributed.
(b) Same variance, say σ_y^2

So the 3 basic properties of the transformed dependent variable that interest us are:

(a) Normality.
(b) Stable Covariance.
(c) Simplicity of Structure. — in the sense that

$$E\left[\frac{y_\alpha^\lambda - 1}{\lambda}\right] = \beta_1 + \beta_2 x_{2\alpha} + --- + \beta_K x_{K\alpha}$$

is a simple function of the β's & of the x's.

Now, λ is <u>unknown</u> and so it will have to be estimated together with β & σ. So we are going to use the data to tell us what transformation we should use — i.e. to tell us what functional form the regression model should take. Now compare the MLE & Bayesian results.

Write: $(y_\alpha^\lambda - 1)/\lambda = \beta_1 + \beta_2 x_{2\alpha} + --- + \beta_K x_{K\alpha} + u_\alpha$

as: $y^{(\lambda)} = X\beta + u$.

Now, the u's are distributed $N(0,\sigma^2)$. But now the dependent variable $(Y^{(\lambda)})$ has a <u>different</u> distribution from that of the errors ~

$$p(y_\alpha) = J \cdot p(u_\alpha)$$

$$\& \quad J = \prod_{\alpha=1}^{n} \left|\frac{\partial u_\alpha}{\partial y_\alpha}\right| = \prod_{\alpha=1}^{n} y_\alpha^{\lambda-1} = \left(\prod_{\alpha=1}^{n} y_\alpha\right)^{\lambda-1} = \dot{y}^{n(\lambda-1)}$$

where $\dot{y}$ = geometric mean $= \left(\prod_{\alpha=1}^{n} y_\alpha\right)^{1/n}$.

So, $p(y|\lambda,\beta,\sigma^2) \propto J \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta)\right\}$

and $\ell(\lambda,\beta,\sigma^2|y) \propto J \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta)\right\}$

So

$$\mathcal{L} = \log \ell = \text{const.} + \log J - n/2 \log\sigma^2 - \frac{1}{2\sigma^2}(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta).$$

$$\therefore \left(\frac{\partial \mathcal{L}}{\partial \beta}\right) = -\frac{1}{2\sigma^2}\left[\frac{\partial}{\partial\beta}\left(y^{(\lambda)\prime}y^{(\lambda)} - 2\beta'X'y^{(\lambda)} + \beta'X'X\beta\right)\right]$$

$$= -\frac{1}{2\sigma^2}\left[-2X'y^{(\lambda)} + 2X'X\beta\right]$$

$$= 0 \text{ for MLE}$$

$$\therefore \hat{\beta}_{(\lambda)} = (X'X)^{-1}X'y^{(\lambda)}.$$

And, $\left(\frac{\partial \mathcal{L}}{\partial \sigma^2}\right) = \frac{1}{2\sigma^4}\left[(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta)\right] - \frac{n}{2\sigma^2}$

$$= 0 \text{ for MLE}$$

$$\therefore \hat{\sigma}^2_{(\lambda)} = 1/n\,(y^{(\lambda)} - X\hat{\beta}_{(\lambda)})'(y^{(\lambda)} - X\hat{\beta}_{(\lambda)}).$$

Now, <u>given λ</u>, we can compute $\hat{\beta}_{(\lambda)}$ and $\hat{\sigma}^2_{(\lambda)}$. But at present λ is unknown, & we will have to <u>estimate</u> its value —

$$\mathcal{L} = \text{const.} + \log J - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta)$$

$$\therefore \hat{\mathcal{L}}_{\underline{max}}(\lambda) = \text{const.} + \log J - n/2 \log(\hat{\sigma}^2_{(\lambda)}) - \frac{1}{2\hat{\sigma}^2_{(\lambda)}}\left[y^{(\lambda)}-X\hat{\beta}_{(\lambda)}\right]'\left[y^{(\lambda)}-X\hat{\beta}_{(\lambda)}\right]$$

$$= \text{const.} + \log J - n/2 \log(\hat{\sigma}^2_{(\lambda)}) - n/2$$

$$= \text{const.}' + \log J - n/2 \log(\hat{\sigma}^2_{(\lambda)})$$

$$= \text{const.}' + (\lambda-1)\sum_{\alpha=1}^{n} \log y_\alpha - n/2 \log(\hat{\sigma}^2_{(\lambda)}).$$

So, plot $\hat{L}_{\underline{max}}(\lambda)$ for different values of λ, and pick $\lambda = \hat{\lambda}$ to maximize this function. Then evaluate $\hat{\sigma}^2(\hat{\lambda})$ and $\hat{\beta}(\hat{\lambda})$.

Zarembka looked at a model where the regressors where also subjected to the Basic Box-Cox transform.

$$\left(\frac{M_\alpha^\lambda - 1}{\lambda}\right) = \beta_1 + \beta_2\left(\frac{y_\alpha^\lambda - 1}{\lambda}\right) + \beta_3\left(\frac{r_\alpha^\lambda - 1}{\lambda}\right) + u_\alpha$$

So, $\begin{cases} \text{if } \lambda = 1, \text{ then we have a } \underline{\text{linear}} \text{ relationship.} \\ \text{if } \lambda = 0, \text{ then we have a } \underline{\log - \log} \text{ relationship.} \end{cases}$

$\underline{\underline{e+c}}$

Let;
$$\left.\begin{aligned} (M_\alpha^\lambda - 1)/\lambda &= y_\alpha^{(\lambda)} \\ (y_\alpha^\lambda - 1)/\lambda &= x_{2\alpha}^{(\lambda)} \\ (r_\alpha^\lambda - 1)/\lambda &= x_{3\alpha}^{(\lambda)} \end{aligned}\right\}$$

$$\therefore \ell(\beta, \sigma^2, \lambda \mid y) \propto \left(\frac{J}{\sigma^n}\right) \exp\left[-\frac{1}{2\sigma^2}(y^{(\lambda)} - X^{(\lambda)}\beta)'(y^{(\lambda)} - X^{(\lambda)}\beta)\right]$$

where $\beta' = (\beta_1, \beta_2, \beta_3)$; $X^{(\lambda)} = (\iota, x_2^{(\lambda)}, x_3^{(\lambda)})$.

$$\& \ J = \prod_{\alpha=1}^{n} y_\alpha^{\lambda-1} = \prod_{\alpha=1}^{n} M_\alpha^{\lambda-1}.$$

$$L = \log \ell = k + \log J - n/2 \log \sigma^2 - \frac{1}{2\sigma^2}(y^{(\lambda)} - X^{(\lambda)}\beta)'(y^{(\lambda)} - X^{(\lambda)}\beta).$$

$$\left(\frac{\partial}{\partial \beta}\right) = -\frac{1}{2\sigma^2}\left[\frac{\partial}{\partial \beta}(y^{(\lambda)\prime} y^{(\lambda)} - 2\beta' X^{(\lambda)\prime} y^{(\lambda)} + \beta' X^{(\lambda)\prime} X^{(\lambda)} \beta)\right]$$

$$= -\frac{1}{2\sigma^2}\left[-2X^{(\lambda)\prime} y^{(\lambda)} + 2X^{(\lambda)\prime} X^{(\lambda)} \beta\right]$$

$= 0$ for MLE.

So, $X^{(\lambda)\prime} y^{(\lambda)} = X^{(\lambda)\prime} X^{(\lambda)} \hat{\beta}_{(\lambda)}$

$$\therefore \hat{\beta}_{(\lambda)} = (X^{(\lambda)\prime} X^{(\lambda)})^{-1} X^{(\lambda)\prime} y^{(\lambda)}.$$

$$\& \left(\frac{\partial L}{\partial \sigma^2}\right) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y^{(\lambda)} - X^{(\lambda)}\beta)'(y^{(\lambda)} - X^{(\lambda)}\beta)$$

$= 0$, for MLE.

$$\therefore \hat{\sigma}^2_{(\lambda)} = \frac{1}{n}(y^{(\lambda)} - X^{(\lambda)}\hat{\beta}_{(\lambda)})'(y^{(\lambda)} - X^{(\lambda)}\hat{\beta}_{(\lambda)}).$$

$$\text{Log } L_{\underline{max}}(\lambda) = -n/2 \log(\hat{\sigma}^2_{(\lambda)}) + (\lambda - 1)\sum_{\alpha=1}^{n} \log y_\alpha$$

Plot this for values of λ. And Zarembka finds a

maximal value of the function when $\hat{\lambda} = 0.19$. So pick $\lambda = 0$, or a log-log relationship.

Note that confidence intervals may be obtained for $\lambda \sim$ because $2[L_{max}(\hat{\lambda}) - L_{max}(\lambda)] \sim \chi^2_{(1)}$

Now, for Zarembka's model,

the approximate C.I. around $\hat{\lambda} = 0.19$ is given by

$$L_{max}(\hat{\lambda}) - L_{max}(\lambda) < \tfrac{1}{2}\chi^2_{(1)}(0.05) = 1.92$$

So, take $\hat{\lambda} = 0.19$, evaluate $L_{max}(\hat{\lambda})$. Take

$L_{max}(\hat{\lambda}) \pm 1.92$. Evaluate the values of $L_{max}(\lambda)$ solving for λ_1^*, λ_2^* Then take the interval around $\hat{\lambda}$, with $(\lambda^* - \hat{\lambda})$ on each side. viz form ~~$(\hat{\lambda} - \lambda_1^*, \hat{\lambda} - \lambda_2^*)$ as the C.I.~~

$$(\lambda_1^*, \lambda_2^*) \text{ as the C.I.}$$

Now introduce the possibility of autocorrelated disturbances.

Let $u_\alpha = \rho u_{\alpha-1} + \varepsilon_\alpha \quad ; \alpha = 1, 2, ----, n.$

So, $y_\alpha^{(\lambda)} = \beta_1 + \beta_2 x_{2\alpha}^{(\lambda)} + x_{3\alpha}^{(\lambda)} + \rho y_{\alpha-1}^{(\lambda)} - \beta_1\rho - \rho\beta_2 x_{2\alpha-1}^{(\lambda)}$

$$- \rho\beta_3 x_{3\alpha-1}^{(\lambda)} + \varepsilon_\alpha.$$

$$\therefore (y_\alpha^{(\lambda)} - y_{\alpha-1}^{(\lambda)}) = \beta_1(1-\rho) + \beta_2(x_{2\alpha}^{(\lambda)} - \rho x_{2\alpha-1}^{(\lambda)}) + \beta_3(x_{3\alpha}^{(\lambda)} - \rho x_{3\alpha-1}^{(\lambda)}) + \varepsilon_\alpha.$$

Then, if we are <u>given λ</u>, we can evaluate <u>conditional</u> MLE for β & for σ^2.

Having considered the MLE of the Box-Cox model, now turn to the Bayesian estimation of this model. Recall that the MLE is just an approximation to large-sample Bayesian analysis, under the normality assumption.

Let $p(\beta, \sigma, \lambda) = p_1(\beta, \sigma | \lambda)\, p_2(\lambda)$

Take $p_2(\lambda) \propto$ const.
We have to take a <u>conditional</u> prior for β & σ, i.e. take $p_1(\beta, \sigma | \lambda)$, since the general size & range of the transformed observations, $y^{(\lambda)}$, may depend strongly on λ.

So, put $p_1(\beta, \log\sigma \mid \lambda)\, d\beta_\lambda\, d(\log\sigma_\lambda) \propto g(\lambda)\, d\beta_\lambda\, d(\log\sigma_\lambda)$

i.e. β & $\log\sigma$ are uniform, but depend on a function of λ.

$$\therefore\ p_1(\beta, \sigma \mid \lambda) \propto \frac{g(\lambda)}{\sigma}.$$

Now, Box & Cox take $g(\alpha) = \dot{y}^{-k/n}$, where $\dot{y} = \left(\prod_{\alpha=1}^{n} y_\alpha\right)^{\lambda-1}$.

So, $p_1(\beta, \sigma \mid \lambda) \propto \left(\frac{1}{\sigma\, \dot{y}^{k/n}}\right)$

& $p(\beta, \sigma, \lambda) \propto \left(\frac{1}{\sigma\, \dot{y}^{k/n}}\right)$ too.

Now, $p(y \mid \lambda, \sigma^2, \beta) \propto \left(\frac{\dot{y}}{\sigma^n}\right) \exp\left\{-\frac{1}{2\sigma^2}(y^{(\lambda)} - X\beta)'(y^{(\lambda)} - X\beta)\right\}$

$$\therefore\ p(\lambda, \sigma, \beta \mid y) \propto \left(\frac{\dot{y}^{(n-k)/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}(y^{(\lambda)} - X\beta)'(y^{(\lambda)} - X\beta)\right\}$$

And, $(y^{(\lambda)} - X\beta)'(y^{(\lambda)} - X\beta) = \nu s^2(\lambda) + (\beta - \hat{\beta}(\lambda))' X'X (\beta - \hat{\beta}(\lambda))$

where $\hat{\beta}(\lambda) = (X'X)^{-1} X' y^{(\lambda)}$.

$\nu = (n-k)$

$\nu s^2(\lambda) = (y^{(\lambda)} - X\hat{\beta}(\lambda))'(y^{(\lambda)} - X\hat{\beta}(\lambda))$.

$$\therefore\ p(\lambda, \beta, \sigma \mid y) \propto \left(\frac{\dot{y}^{(n-k)/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2(\lambda) + (\beta - \hat{\beta}(\lambda))' X'X (\beta - \hat{\beta}(\lambda))\right]\right\}$$

So, $p(\lambda, \beta \mid y) \propto (\dot{y}^{\nu/n})\left[\nu s^2(\lambda) + (\beta - \hat{\beta}(\lambda))' X'X (\beta - \hat{\beta}(\lambda))\right]^{-n/2}$.

$$\propto (\dot{y}^{\nu/n})\left[s^2(\lambda)\right]^{-n/2}\left[\nu + (\beta - \hat{\beta}(\lambda))'\left(\frac{X'X}{s^2(\lambda)}\right)(\beta - \hat{\beta}(\lambda))\right]^{-\frac{n}{2}}$$

$$\propto \dot{y}^{\nu/n}\left(s^2(\lambda)\right)^{-n/2}\left|\frac{X'X}{s^2(\lambda)}\right|^{-\frac{1}{2}} \ *$$

$$\left\{\left|\frac{X'X}{s^2(\lambda)}\right|^{1/2}\left[\nu + (\beta - \hat{\beta}(\lambda))'\left(\frac{X'X}{s^2(\lambda)}\right)(\beta - \hat{\beta}(\lambda))\right]^{-n/2}\right\}$$

$$\therefore\ p(\lambda \mid y) \propto \dot{y}^{\nu/n} \cdot \left(s^2(\lambda)\right)^{-n/2}\left|\frac{X'X}{s^2(\lambda)}\right|^{-1/2}$$

$$\propto \dot{y}^{\nu/n} \cdot \left(s^2(\lambda)\right)^{-n/2} \cdot s^2(\lambda)^{k/2}$$

$$\therefore\ p(\lambda \mid y) \propto \dot{y}^{\nu/n} \cdot \left(s^2(\lambda)\right)^{-\nu/2}.$$

This pdf can be analysed numerically to permit inferences about λ. Also, the modal value of the posterior pdf coincides with the M.L.E.

Note that the <u>conditional</u> pdf of β <u>given λ</u> is a MVT dist'n, so we can make use of this fact to test the sensitivity of inferences about β to the assumptions regarding the value of λ.

(B) <u>The C.E.S. Production Function:</u>

Let the elasticity of substitution be $\varepsilon \ni 0 \le \varepsilon < \infty$.

Let the function be of the general form —

$$y_\alpha = \gamma \left[\delta K_\alpha^{-\rho} + (1-\delta) L_\alpha^{-\rho} \right]^{-v/\rho} e^{u_\alpha} \quad ; \alpha = 1, 2, ---, n.$$

For the 2-input case. And assume that $u_\alpha \sim NID(0, \sigma^2)$.

$$\therefore \text{Log } y_\alpha = \log \gamma + v \log \left\{ \left[\delta K_\alpha^{-\rho} + (1-\delta) L_\alpha^{-\rho} \right]^{-1/\rho} \right\} + u_\alpha.$$

Then,

$$\ell(\gamma, \delta, \rho, v, \sigma \mid y, K, L) \propto \left(\frac{1}{\sigma^n}\right) \exp. \left\{ -\frac{1}{2\sigma^2} \left[\hat{u}'\hat{u} + (\theta - \hat{\theta})' X' X (\theta - \hat{\theta}) \right] \right\}$$

Where: $\theta = \begin{pmatrix} \log \gamma \\ v \end{pmatrix}$; $\hat{\theta} = (X'X)^{-1} X' y$

$$y = \begin{pmatrix} \log y_1 \\ \vdots \\ \log y_n \end{pmatrix} \quad ; \quad X = \begin{pmatrix} 1 & \log \left[\delta k_1^{-\rho} + (1-\delta) L_1^{-\rho} \right]^{-1/\rho} \\ \vdots & \vdots \\ 1 & \log \left[\delta k_n^{-\rho} + (1-\delta) L_n^{-\rho} \right]^{-1/\rho} \end{pmatrix}$$

& $\hat{u}'\hat{u} = (y - X\hat{\theta})'(y - X\hat{\theta})$.

$$\text{Log } \ell = L = \text{const.} - n/2 \log \sigma^2 - \frac{1}{2\sigma^2} \left[\hat{u}'\hat{u} + (\theta - \hat{\theta})' X' X (\theta - \hat{\theta}) \right]$$

So, $\bar{\sigma}^2 = \frac{1}{n} \hat{u}'\hat{u}$; $\bar{\theta} = \hat{\theta}$; where a bar denotes a MLE, for given ρ and δ.

$$L_{\underline{max}} = \text{const} - n/2 \log (\hat{u}'\hat{u})$$

So $L_{\underline{max}}$ is a function of ρ & δ; since $\hat{u} = \hat{u}(\rho, \delta)$.
So search over the (ρ, δ) grid & find $(\hat{\rho}, \hat{\delta})$ such that $L_{\underline{max}}$ is maximal. Then compute $\bar{\sigma}^2(\hat{\rho}, \hat{\delta})$ and $\bar{\beta}(\hat{\rho}, \hat{\delta})$ for MLE.

Now, $0 < \delta < 1$, & $-1 < \rho < \infty$, since $\rho = -1 + 1/\varepsilon$.

So to ease the grid search, we might transform ρ —

Let $\lambda = 1/(1+\varepsilon)$
Then since $\rho = 1 + 1/\varepsilon$,

$$\Rightarrow \varepsilon = 1/(1+\rho)$$
$$\therefore \lambda = 1 / \left(1 + \frac{1}{1+\rho}\right) = (1+\rho)/(2+\rho).$$

So when $\rho = -1$, $\lambda = 0$.
But also, $(2+\rho)\lambda = 1+\rho$
$$\therefore \rho = (2\lambda - 1)/(1-\lambda)$$

So, if $\lambda = 1$, $\rho = \infty$.

So $0 \leq \lambda \leq 1$ & we just search for $(\hat{\lambda}, \hat{\delta})$ in the <u>unit square</u>.

Having obtained $\hat{\lambda}$, we get $\hat{\rho}$ & hence $\hat{\varepsilon}$.
Now, the mean & variance of $\hat{\varepsilon}$ do not exist in finite samples, but they do exist for the asymptotic normal distribution of $\hat{\varepsilon}$.

Thornber investigates by means of a Monte Carlo study, the risk functions associated with alternative estimators for ε.

The loss function that he employs is —

$$\mathcal{L}(\varepsilon, \hat{\varepsilon}) = \frac{(\varepsilon - \hat{\varepsilon})^2}{(1+\varepsilon)^2(1+\hat{\varepsilon})^2}.$$

Now, this loss function implies relatively greater loss for an ~~over~~estimate than for an underestimate.
To show this —

consider an underestimate of Δ ~

<u>viz</u>: $(\varepsilon - \hat{\varepsilon}) = \Delta$

$$\text{Then } \mathcal{L}_1(\varepsilon, \hat{\varepsilon}) = \frac{\Delta^2}{(1+\hat{\varepsilon}+\Delta)^2(1+\hat{\varepsilon})^2}$$

Consider an overestimate of Δ ~

<u>viz</u> $\hat{\varepsilon} - \varepsilon = \Delta$

$$\therefore \mathcal{L}_2(\varepsilon, \hat{\varepsilon}) = \frac{\Delta^2}{(1+\hat{\varepsilon}-\Delta)^2(1+\hat{\varepsilon})^2}$$

$$\therefore \left(\frac{\mathcal{L}_1}{\mathcal{L}_2}\right) = \frac{(1+\hat{\varepsilon}-\Delta)^2}{(1+\hat{\varepsilon}+\Delta)^2} < 1.$$

So $\mathcal{L}_1 < \mathcal{L}_2$, or $\mathcal{L}_2 > \mathcal{L}_1$.

Now, consider the prior pdf's used by Thornber:

$$p_1(\gamma, \upsilon, \sigma, \delta, \varepsilon) \propto \left[\frac{(1+\varepsilon)^2 e^{-\varepsilon}}{\gamma\sigma}\right] \quad ; \varepsilon \geq 0.$$

$$p_2(\gamma, \upsilon, \sigma, \delta, \varepsilon) \propto \left[\frac{\varepsilon(1+\varepsilon)^2 e^{-\varepsilon}}{\gamma\sigma}\right]$$

That is, the priors for $\upsilon, \delta, \log\gamma$ and $\log\sigma$ are taken as uniform, & $\begin{cases} p_1(\varepsilon) \propto (1+\varepsilon)^2 e^{-\varepsilon} \\ p_2(\varepsilon) \propto \varepsilon(1+\varepsilon)^2 e^{-\varepsilon}. \end{cases}$

Consider the normalizing factors for p_1 & p_2, the <u>marginal</u> prior pdf's for ε.

(a) $p_1(\varepsilon) \propto (1+\varepsilon)^2 e^{-\varepsilon}$.

So, the normalizing constant is the inverse of $I_1 = \int_0^\infty (1+\varepsilon)^2 e^{-\varepsilon} d\varepsilon$.

Let $u = (1+\varepsilon)^2$; $dv = e^{-\varepsilon} d\varepsilon$
$\therefore du = 2(1+\varepsilon) d\varepsilon$; $v = -e^{-\varepsilon}$

$$\therefore I_1 = uv - \int_0^\infty v\,du$$
$$= \left[-(1+\varepsilon)^2 e^{-\varepsilon}\right]_0^\infty + 2\int_0^\infty e^{-\varepsilon}(\varepsilon+1)\,d\varepsilon$$
$$= 1 + 2\int_0^\infty e^{-\varepsilon} d\varepsilon + 2\int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon$$
$$= 1 + 2\left[-e^{-\varepsilon}\right]_0^\infty + 2\int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon$$
$$= 3 + 2\int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon$$

Let $I_2 = \int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon$

$\therefore u = \varepsilon$; $dv = e^{-\varepsilon} d\varepsilon$
$\therefore du = d\varepsilon$; $v = -e^{-\varepsilon}$

$$\therefore I_2 = uv - \int_0^\infty v\,du$$
$$= \left[-\varepsilon e^{-\varepsilon}\right]_0^\infty + \int_0^\infty \varepsilon^{-\varepsilon} d\varepsilon$$
$$= 0 - \left[e^{-\varepsilon}\right]_0^\infty$$
$$= 1$$
$\therefore 2I_2 = 2$.

$\therefore I = 3 + 2 = 5$

So, $p_1(\varepsilon) = \frac{1}{5}(1+\varepsilon)^2 e^{-\varepsilon}$.

(b) $p_2(\varepsilon) \propto \varepsilon(1+\varepsilon)^2 e^{-\varepsilon}$.

So, the normalizing constant is the inverse of

$$I_1 = \int_0^\infty \varepsilon(1+\varepsilon)^2 e^{-\varepsilon} d\varepsilon$$

$$= \int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon + 2\int_0^\infty \varepsilon^2 e^{-\varepsilon} d\varepsilon + \int_0^\infty \varepsilon^3 e^{-\varepsilon} d\varepsilon$$

$$= 1 + 2I_2 + I_3$$

& $I_2 = \int_0^\infty \varepsilon^2 e^{-\varepsilon} d\varepsilon$.

Let $u = \varepsilon^2$; $dv = e^{-\varepsilon} d\varepsilon$

$\therefore du = 2\varepsilon\, d\varepsilon$; $v = -e^{-\varepsilon}$

$$\therefore I_2 = uv - \int_0^\infty v\, du$$

$$= \left[-\varepsilon^2 e^{-\varepsilon}\right]_0^\infty + 2\int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon$$

$$= 0 + 2$$

$$= 2$$

$$\therefore 2I_2 = 4$$

$$I_3 = \int_0^\infty \varepsilon^3 e^{-\varepsilon} d\varepsilon$$

Let $u = \varepsilon^3$; $dv = e^{-\varepsilon} d\varepsilon$

$\therefore du = 3\varepsilon^2 d\varepsilon$; $v = -e^{-\varepsilon}$

$$I_3 = uv - \int_0^\infty v\, du$$

$$= \left[-\varepsilon^3 e^{-\varepsilon}\right]_0^\infty + 3\int_0^\infty \varepsilon^2 e^{-3} d\varepsilon$$

$$= 0 + 3I_2$$

$$= 6$$

$$\therefore I_1 = 1 + 4 + 6$$

$$= 11$$

So, $p_2(\varepsilon) = \frac{1}{11}\varepsilon(1+\varepsilon)^2 e^{-\varepsilon}$.

Now, consider an alternative approach to the analysis of the C.E.S. production function.

Consider the deterministic form of the function, with 2 inputs & with CRTS —

$$V_\alpha = \gamma\left[\delta_1 x_{1\alpha}^g + (1-\delta_1) x_{2\alpha}^g\right]^{1/g}$$

$\ni \quad g = -\rho = (\varepsilon - 1)/\varepsilon.$

$$\therefore V_\alpha^g = \gamma^g\left[\delta_1 x_{1\alpha}^g + (1-\delta_1) x_{2\alpha}^g\right]$$

$$\therefore \left(\frac{V_\alpha^g - 1}{g}\right) = \frac{\left[\gamma^g \delta_1 x_{1\alpha}^g + \gamma^g (1-\delta_1) x_{2\alpha}^g\right] - 1}{g}$$

$$\therefore V_\alpha^{(g)} = \left[\gamma^g \delta_1 x_{1\alpha}^g - \gamma^g \delta_1 x_{2\alpha}^g + \gamma^g x_{2\alpha}^g - 1\right]/g$$

$$= \left[\gamma^g \delta_1 (x_{1\alpha}^g - 1) + \gamma^g \delta_1 - \gamma^g \delta_1 (x_{2\alpha}^g - 1) - \gamma^g \delta_1 + \gamma^g (x_{2\alpha}^g - 1) + \gamma^g - 1\right]/g$$

$$= \gamma^g \delta_1 x_{1\alpha}^{(g)} - \gamma^g \delta_1 x_{2\alpha}^{(g)} + \gamma^{(g)} + \gamma^g (x_{2\alpha}^{(g)})$$

$$= \gamma^g \delta_1 x_{1\alpha}^{(g)} - \gamma^g \delta_1 x_{2\alpha}^{(g)} + \gamma^{(g)} + g\gamma^{(g)} x_{2\alpha}^{(g)} + x_{2\alpha}^{(g)}$$

$$\therefore V_\alpha^{(g)} = x_{2\alpha}^{(g)} + \gamma^g \delta_1 \left[x_{1\alpha}^{(g)} - x_{2\alpha}^{(g)}\right] + \gamma^{(g)}\left[1 + g x_{2\alpha}^{(g)}\right]$$

And assume that the observed output, y_α, satisfies

$$y_\alpha^{(\lambda)} = V_\alpha^{(g)} + u_\alpha$$

where $y_\alpha^{(\lambda)} = (y_\alpha^\lambda - 1)/\lambda$; $\lambda = \phi g$, for free ϕ.

Note that this is more general than if we just took $y_\alpha^{(g)}$ instead of $y_\alpha^{(\lambda)}$.

$$\therefore y_\alpha^{(\lambda)} = x_{2\alpha}^{(g)} + \beta_1 (x_{1\alpha}^{(g)} - x_{2\alpha}^{(g)}) + \beta_2 (1 + g x_{2\alpha}^{(g)}) + u_\alpha$$

where $\beta_1 = \gamma^g \delta_1$; $\beta_2 = \gamma^{(g)}$.

Then, if $g = 0$, the fcn. becomes C-D.
if $\lambda = 1$ & $g = 1$, it is linear in the variables.
if $\lambda = 0$ & $g = 1$, it is ~~log-linear~~ semi-log.

So, by using the more general parameter, λ, instead of g for the dependent variable, we are able to cover a far wider range of possible functional forms.

Now, let $w = y^{(\lambda)} - x_2^{(g)}$; $X = \left(x_1^{(g)} - x_2^{(g)} \,\middle|\, \iota + x_2^{(g)}\right)$
& $\beta' = (\beta_1, \beta_2)$.

Then, if we assume that $u_\alpha \sim NID(0, \sigma^2)$, $\forall \alpha$;

$$\Rightarrow \ell(\beta, \sigma, \lambda, g \mid y) \propto \left(\frac{J}{\sigma^n}\right) \exp\left\{-\frac{1}{2\sigma^2}(w - X\beta)'(w - X\beta)\right\}$$

$$\propto \left(J/\sigma^n\right) \exp\left\{-\frac{1}{2\sigma^2}\left[n\hat{\sigma}^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\right]\right\}$$

where:
$$\hat{\beta} = (X'X)^{-1} X' w$$
$$\hat{\sigma}^2 = \frac{1}{n}(w - X\hat{\beta})'(w - X\hat{\beta})$$
$$J = \prod_{\alpha=1}^{n} y_\alpha^{\lambda - 1}.$$

Now, $\hat{\beta}$ & $\hat{\sigma}^2$ are functions of λ & g. For <u>given</u> λ & g, $\hat{\beta}$ and $\hat{\sigma}^2$ are M.L.E. Now, to get global maximum estimates, look at:

$$\underline{L_{max}}(g, \lambda) = \text{const.} + (\lambda - 1)\sum_{\alpha=1}^{n} \log y_\alpha - n/2 \log \hat{\sigma}^2.$$

Search over the (g, λ) grid to find a maximal value for $\underline{L_{max}}$. Then this yields $(\hat{g}, \hat{\lambda})$, which yields $\hat{\beta}(\hat{\lambda}, \hat{g})$ and $\hat{\sigma}^2(\hat{g}, \hat{\lambda})$. These are the desired MLE. Given $\hat{\beta}$ we then have $\hat{\beta}_1$ & $\hat{\beta}_2$. We already have $\hat{g}$ and $\hat{\lambda}$, so we can obtain $\hat{\gamma}, \hat{\delta}, \hat{\phi}$.

Now, the advantage of using this type of approach to handling the CES fctn is that is easily generalized to the case of more than 2 inputs. This point is noted by Chetty & Sankar, who attribute their use of this approach to a suggestion made to them by Zellner.

$$V = \gamma\left[\delta_1 x_1^g + \delta_2 x_2^g + \cdots + \delta_{k-1} x_{k-1}^g + (1 - \delta_1 - \cdots - \delta_{k-1}) x_k^g\right]^{v/g}$$

& $g = -\rho = (\varepsilon - 1)/\varepsilon$

$$\therefore V^{g/v} = \gamma^{g/v}\left[\delta_1(x_1^g - x_k^g) + \delta_2(x_2^g - x_k^g) + \cdots\cdots + \delta_{k-1}(x_{k-1}^g - x_k^g) + x_k^g\right]$$

& so,
$$V^{(g/v)} = v x_k^{(g)} + \beta_1(x_1^{(g)} - x_k^{(g)}) + \beta_2(x_2^{(g)} - x_k^{(g)}) + \cdots + \beta_{k-1}(x_{k-1}^{(g)} - x_k^{(g)}) + \beta_k(1 + g x_k^{(g)}$$

where $\beta_i = v\delta_i \gamma^{g/v}$; $i = 1, \ldots, k-1$.
$\beta_k = \gamma^{g/v}$.

Assume, for generality, that

$$y^{(\lambda)} = V^{(g/v)} + u.$$

where $\lambda = \phi(g/v)$, with free parameter ϕ.

Then $W = X\beta + u$

where,

$$W = y^{(\lambda)} \; ; \; \beta' = (\beta_1, \ldots, \beta_K)$$

$$u' = (u_1, \ldots, u_n)$$

$$X = \left(x_K^{(g)}, (x_1^{(g)} - x_K^{(g)}), \ldots, (x_{K-1}^{(g)} - x_K^{(g)}), (\iota + g x_K^{(g)}) \right)$$

$$\therefore \ell(\beta, \lambda, \sigma, g, v \mid y) \propto \left(\frac{J}{\sigma^n}\right) \exp.\left\{-\frac{1}{2\sigma^2}(W - X\beta)'(W - X\beta)\right\}$$

Then, for <u>given</u> λ and g, the <u>conditional</u> maximum likelihood values are:

$$\hat{\beta} = (X'X)^{-1}X'W = (X'X)^{-1}X'y^{(\lambda)}.$$

$$\& \quad \hat{\sigma}^2 = \tfrac{1}{n}(W - X\hat{\beta})'(W - X\hat{\beta})$$

$$= \tfrac{1}{n}(y^{(\lambda)} - X\hat{\beta})'(y^{(\lambda)} - X\hat{\beta}).$$

$$\therefore \; \underline{L_{max}} = \text{const.} + (\lambda - 1)\sum_{\alpha=1}^{n} \log y_\alpha - n/2 \log \hat{\sigma}^2$$

Search over the (λ, g) grid & obtain $(\hat{\lambda}, \hat{g})$ so that $\underline{L_{max}}(\lambda, g)$ is max.

Then we get MLE ~ $\hat{\beta}(\hat{\lambda}, \hat{g})$; $\hat{\sigma}^2(\hat{\lambda}, \hat{g})$.

(c) <u>Generalized production functions:</u>

GPF's are just another broad class of functions which are usually non-linear in both parameters & variables. They permit generalization in two directions —

(a) We want prod'n. fctns. with a preassigned but unknown elasticity of substitution.

(b) We want the returns to scale to vary with output according to some pre-assigned function.

We follow Zellner & Revankar & set up a function as follows:

$$\text{Let} \quad \frac{dV}{df} = \frac{V\alpha(V)}{f\alpha_f}$$

with solution $V = g(f)$

where $\alpha(V)$ = r.t.s. as a fctn. of output, V.
$f = f(K, L)$ is a neo-classical prod'n. fctn.
α_f = r.t.s. associated with f.
and $\alpha(V)$ is chosen so that $(dV/df) > 0 \quad \forall \; 0 < f < \infty$.

Then this leads to a production function of the form

$$Ve^{\theta V} = \gamma K^{\alpha(1-\delta)} L^{\alpha\delta}$$

$$\therefore \log V_i + \theta V_i = \beta_1 + \beta_2 \log K_i + \beta_3 \log L_i + u_i$$

where $\beta_1 = \log\gamma$; $\beta_2 = \alpha(1-\delta)$; $\beta_3 = \alpha\delta$.

Let the $u_i \sim NID(0, \sigma^2)$; $\forall i$.

$$\therefore \ell(\beta, \theta, \sigma \mid \text{data}) \propto (J/\sigma^n) \exp\left[-\tfrac{1}{2\sigma^2}(Z_\theta - X\beta)'(Z_\theta - X\beta)\right]$$

where $Z_\theta = (\log V_i + \theta V_i)$
$\beta' = (\beta_1, \beta_2, \beta_3)$
$X = (\iota \mid \log K \mid \log L)$
$J = \prod_{i=1}^{n} \left(\frac{1+\theta V_i}{V_i}\right)$.

Zellner then works through the MLE estimation of this set-up.

Now look at a Bayesian analysis —

Now, we require a prior pdf for the parameters:

Now, $\beta_2 + \beta_3 = \alpha$.

So, initialize in terms of $\beta_1, \beta_2, \alpha, \sigma$.

& $\log U_i + \theta U_i = \beta_1 + \beta_2 \operatorname{Log} K_i + \beta_3 \operatorname{Log} L_i + u_i$

$\Rightarrow \operatorname{Log} U_i + \theta U_i = \beta_1 + \beta_2 \operatorname{Log}(K_i/L_i) + \alpha \operatorname{Log} L_i + u_i$

Take, for <u>given θ</u> :

$$p(\beta_1, \beta_2, \sigma, \alpha \mid \theta) \propto g(\theta)\, p_1(\beta_2 \mid \alpha)\, p_2(\alpha)\, p_3(\sigma)$$

where : $0 < \theta, \sigma < \infty$; $0 < \beta_2 < \alpha$; $0 < \alpha < \infty$.

& $g(\theta) \propto J^{-3/n}$

$$p_1(\beta_2 \mid \alpha) \propto \left(\frac{\beta_2}{\alpha}\right)^{q_1 - 1} \left(1 - \frac{\beta_2}{\alpha}\right)^{q_2 - 1}$$; i.e. <u>β-dist'n</u>.

$$p_2(\alpha) \propto \text{const.}$$

$$p_3(\sigma) \propto 1/\sigma.$$

Note that $g(\theta)$ is just a proportionality factor — it is <u>not</u> a pdf at all.

So we need some marginal prior for θ, say $p_4(\theta)$.

Then if $p_4(\theta) \propto$ const. & if $q_1 = q_2$,

$$\Rightarrow p(\beta_1, \beta_2, \sigma, \alpha, \theta) \propto \left(\frac{1}{J^{3/n} \cdot \sigma}\right)$$

Now, transform this pdf to be in terms of β_3 —

& $\alpha = \beta_2 + \beta_3$

So $\dfrac{\partial \alpha}{\partial \beta_3} = 1 \quad \therefore J = 1$

$$\therefore p(\beta_1, \beta_2, \beta_3, \sigma, \theta) \propto \left(\frac{1}{J^{3/n} \cdot \sigma}\right)$$

And we had:

$$\ell(\beta, \theta, \sigma \mid \text{data}) \propto \left(\frac{J}{\sigma^n}\right) \exp\left\{-\frac{1}{2\sigma^2}(z_\theta - X\beta)'(z_\theta - X\beta)\right\}$$

$$\therefore p(\beta, \sigma, \theta \mid \text{data}) \propto J^{(n-3)/n} \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}(z_\theta - X\beta)'(z_\theta - X\beta)\right\}$$

$$\propto \left(\frac{J^{U/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\nu S_\theta^2 + (\beta - \hat{\beta}_\theta)' X' X (\beta - \hat{\beta}_\theta)\right]\right\}$$

where $\nu = n-3$; $\nu S_\theta^2 = (z_\theta - X\hat{\beta}_\theta)'(z_\theta - X\hat{\beta}_\theta)$

& $\hat{\beta}_\theta = (X'X)^{-1}X'z_\theta$.

Then $p(\beta \mid \theta, \sigma, y)$ is M.V.N., with conditional mean $\hat{\beta}_\theta$ & conditional C.M. $(X'X)^{-1}\sigma^2$.

Now, how do we obtain the <u>marginal</u> posterior pdf's ?

$$p(\sigma, \theta \mid data) = \int p(\beta, \sigma, \theta \mid data)\, d\beta$$

$$\propto \int \left(\frac{J^{\nu/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{\nu S_\theta^2}{2\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left[(\beta-\hat{\beta}_\theta)'X'X(\beta-\hat{\beta}_\theta)\right]\right\} d\beta$$

$$\propto \left(\frac{J^{\nu/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{\nu S_\theta^2}{2\sigma^2}\right\} \int \exp\left\{-\frac{1}{2\sigma^2}(\beta-\hat{\beta}_\theta)'X'X(\beta-\hat{\beta}_\theta)\right\} d\beta$$

$$\propto \left(\frac{J^{\nu/n}}{\sigma^{n+1}}\right) \exp\left\{-\frac{\nu S_\theta^2}{2\sigma^2}\right\} \cdot \left(\frac{1}{\sigma^3}\right) \quad ; \ \underline{NB} \ k=3.$$

$$\propto \left(\frac{J^{\nu/n}}{\sigma^{\nu+1}}\right) \exp\left\{-\frac{\nu S_\theta^2}{2\sigma^2}\right\}.$$

$$\therefore \ p(\theta \mid data) = \int p(\theta, \sigma \mid data)\, d\sigma$$

$$\propto (J^{\nu/n}) \int \frac{1}{\sigma^{\nu+1}} \exp\left\{-\frac{1}{2\sigma^2}(\nu S_\theta^2)\right\} d\sigma$$

$$\propto (J^{\nu/n})(\nu S_\theta^2)^{-\nu/2}$$

$$\propto \frac{J^{\nu/n}}{(S_\theta^2)^{\nu/2}}.$$

And the mode of this last marginal posterior pdf coincides with the MLE.

Zellner:

Chapter 7: "Time Series Models: Some Selected Examples".

A) 1st-Order Autoregressive Model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t .$$

$$u_t \sim IN(0, \sigma^2) \; ; \; \forall t.$$

Suppose we are given y_0 at time $t=0$. See Ch. 4. for other alternatives.

Then, $\ell(\beta_1, \beta_2, \sigma | y, y_0) \propto \frac{1}{\sigma^T} . \exp\left\{ \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2 \right\}$

Assume $p(\beta_1, \beta_2, \sigma) \propto \left(\frac{1}{\sigma}\right)$.

i.e. we are <u>not</u> constraining the model to be stable. We are allowing $-\infty < \beta_2 < \infty$, so it <u>may</u> be that $|\beta_2| > 1$, & so the system will explode.

$$\therefore p(\beta, \sigma | y, y_0) \propto \left(\frac{1}{\sigma^{T+1}}\right) \exp. \left[-\frac{1}{2\sigma^2} \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2 \right]$$

$$\therefore p(\beta | y, y_0) \propto \left\{ \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2 \right\}^{-T/2} .$$

Let $\nu = T-2$; $\nu s^2 = \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2$.

Now, $X = \begin{pmatrix} 1 & y_{t-1} \\ \vdots & \vdots \\ 1 & y_{t-T} \end{pmatrix}$

So, $(X'X) = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix} = H$; say.

& $(X'y) = \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix}$

$$\therefore \hat{\beta} = (X'X)^{-1}(X'y) = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix} \begin{pmatrix} \sum y_t \\ \sum y_t y_{t-1} \end{pmatrix}$$

So, $\sum (y_t - \beta_1 - \beta_2 y_{t-1})^2 = \sum \left[(y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1}) - \left\{ (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2) y_{t-1} \right\} \right]^2$

$$= \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 + \sum \left[(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2) y_{t-1} \right]^2$$

$$-2 \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1}) \left\{ (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2) y_{t-1} \right\}$$

$$= \nu s^2 + T(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)^2 \sum y_{t-1}^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \sum y_{t-1}$$

$$-2(\beta_1 - \hat{\beta}_1) \sum y_t + 2T\hat{\beta}_1 (\beta_1 - \hat{\beta}_1) + 2\hat{\beta}_2 (\beta_1 - \hat{\beta}_1) \sum y_{t-1}$$

$$-2(\beta_2 - \hat{\beta}_2) \sum y_t + 2T\hat{\beta}_2 (\beta_2 - \hat{\beta}_2) + 2\hat{\beta}_2 (\beta_2 - \hat{\beta}_2) \sum y_{t-1}$$

$$\nu s^2 + \left[(\beta_1 - \hat{\beta}_1), \; (\beta_2 - \hat{\beta}_2) \right] \begin{bmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix} \begin{bmatrix} \beta_1 - \hat{\beta}_1 \\ \beta_2 - \hat{\beta}_2 \end{bmatrix} + \gamma$$

$$= \nu s^2 + (\beta - \hat{\beta})' H (\beta - \hat{\beta}) + \gamma.$$

$$\gamma = 2(\beta_1 + \beta_2 - \hat{\beta}_1 - \hat{\beta}_2)\left[T\hat{\beta}_1 + \hat{\beta}_2 \Sigma y_{t-1} - \Sigma y_t\right]$$

But, $\left[T\hat{\beta}_1 + \hat{\beta}_2 \Sigma y_{t-1} - \Sigma y_t\right] = -\Sigma(y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})$

$$= -\Sigma(r_t)$$
$$= 0$$

So,
$$p(\beta \mid y, y_0) \propto \left\{\nu s^2 + (\beta - \hat{\beta})' H (\beta - \hat{\beta})\right\}^{-(\nu+2)/2}.$$

And this is a bivariate t-distribution, with mean $\hat{\beta}$.

so the <u>marginal</u> posteriors for β_1 and β_2 will be univariate t —

$$\frac{(\beta_1 - \hat{\beta}_1)}{s(h^{11})^{1/2}} \sim t_\nu. \quad ; \quad \frac{(\beta_2 - \hat{\beta}_2)}{s(h^{22})^{\frac{1}{2}}} \sim t_\nu.$$

<u>NB</u> : $h^{11} = \Sigma y_{t-1}^2$; $h^{22} = T$.

Now, suppose that we are interested in $p(\sigma \mid y, y_0)$.

Then, $p(\sigma \mid y, y_0) = \int p(\beta, \sigma \mid y, y_0)\, d\beta$

Use the Ch. 2 result —
if $p(\mu, \sigma \mid y) \propto \left(\frac{1}{\sigma^{n+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left(\nu s^2 + n(\mu - \hat{\mu})^2\right)\right\}$

Then $p(\sigma \mid y) \propto \sigma^{-(\nu+1)} \exp\left\{\frac{-\nu s^2}{2\sigma^2}\right\}$

So here:
$$p(\sigma \mid y, y_0) \propto \sigma^{-(\nu+1)} \exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}$$

So, the results for the marginal posterior distributions of β or of σ are the same as in the basic theoretical model.

But this just because we have paid no attention to limitations on β_2 — such limitations are now important if the model is not to explode.

in particular, impose the condition:
$$|\beta_2| < 1.$$

Now, if $|\beta_2| < 1$, this does in fact imply stability in a time series sense —

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t$$
$$= \beta_1 + \beta_2 (\beta_1 + \beta_2 y_{t-2} + u_{t-1}) + u_t$$
$$= \beta_1 + \beta_1 \beta_2 + \beta_2^2 y_{t-2} + \beta_2 u_{t-1} + u_t$$

etc

So, $y_t = \beta_1 \sum_0^\infty \beta_2^i + \sum_0^\infty \beta_2^i u_{t-i}$.

So, $y_0 = \beta_1 \sum_0^\infty \beta_2^i + \sum_0^\infty \beta_2^i u_{-i}$.

$$= \left(\frac{\beta_1}{1-\beta_2}\right) + \sum_0^\infty \beta_2^i u_{-i}.$$

$$\therefore E(y_0) = \left(\frac{\beta_1}{1-\beta_2}\right) + \sum_0^\infty \beta_2^i E(u_{-i})$$

$$\therefore E(y_0) = \left(\frac{\beta_1}{1-\beta_2}\right)$$

And, ~~var~~ if $x = f(x_1, x_2, ---- x_n)$

$$\Rightarrow var(x) = \sum_1^n \left(\frac{\partial x}{\partial x_i}\right)_0^2 Var(x_i)$$

where a zero subscript $\Rightarrow$ a mean value.

So, $var(y_0) = 1 \cdot \sigma^2 + \beta_2^2 \sigma^2 + \beta_2^4 \sigma^2 + ---$

$$= \sigma^2 \left[1 + \beta_2^2 + (\beta_2^2)^2 + ---- \right]$$

$$\therefore var(y_0) = \left(\frac{\sigma^2}{1-\beta_2^2}\right)$$

And normality is preserved.

So $l(\beta_1, \beta_2, \sigma | y_0) \propto \frac{(1-\beta_2^2)^{\frac{1}{2}}}{\sigma} \exp\left\{-\frac{(1-\beta_2^2)}{2\sigma^2}\left(y_0 - \frac{\beta_1}{1-\beta_2}\right)^2\right\}$.

$$l(\beta_1, \beta_2, \sigma | y, y_0) \propto \left(\frac{1}{\sigma^{T+1}}\right)(1-\beta_2^2)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(1-\beta_2^2)\left(y_0 - \frac{\beta_1}{1-\beta_2}\right)^2 + \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2\right]\right\}$$

Let σ, β_1 & β_2 be independently distributed. Let β_1 & $\log \sigma$ be uniformly dist., & let prior of β_2 be $p(\beta_2)$

$$\therefore p(\beta_1, \beta_2, \sigma) \propto \left(\frac{p(\beta_2)}{\sigma}\right). \quad ; \begin{cases} -\infty < \beta_1 < \infty \\ 0 < \sigma < \infty \\ |\beta_2| < 1 \end{cases}$$

$$\therefore\ p(\beta_1, \beta_2, \sigma \mid y, y_0) \propto \left[\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}}\right] *$$

$$\exp\left\{-\frac{1}{2\sigma^2}\left[\sum(y_t - \beta_1 - \beta_2 y_{t-1})^2 + (1-\beta_2^2)\left(y_0 - \frac{\beta_1}{1-\beta_2}\right)^2\right]\right\}$$

Look at the exponent –

$$(1-\beta_2^2)\left(y_0 - \frac{\beta_1}{1-\beta_2}\right)^2 + \sum(y_t - \beta_1 - \beta_2 y_{t-1})^2$$

$$= (1-\beta_2^2)y_0^2 + (1-\beta_2^2)\left(\frac{\beta_1}{1-\beta_2}\right)^2 - 2y_0\beta_1\left(\frac{1-\beta_2^2}{1-\beta_2}\right)$$

$$+ \sum(y_t - \beta_1)^2 + \sum(\beta_2 y_{t-1})^2 - 2\sum(y_t - \beta_1)\beta_2 y_{t-1}$$

$$= y_0^2(1-\beta_2^2) + \frac{\beta_1^2(1+\beta_2)}{(1-\beta_2)} - 2y_0\beta_1(1+\beta_2) + \sum y_t^2 + T\beta_1^2 - 2\beta_1\sum y_t$$

$$+ \beta_2^2\sum y_{t-1}^2 - 2\beta_2\sum y_t y_{t-1} + 2\beta_1\beta_2\sum y_{t-1}$$

$$= \beta_1^2\left[T + \frac{(1+\beta_2)}{(1-\beta_2)}\right] - 2\beta_1\left[\sum y_t + y_0(1+\beta_2) - \beta_2\sum y_{t-1}\right]$$

$$+ \left[y_0^2(1-\beta_2^2) + \sum y_t^2 + \beta_2^2\sum y_{t-1}^2 - 2\beta_2\sum y_t y_{t-1}\right]$$

$$= c\beta_1^2 - 2\beta_1\left[y_0(1+\beta_2) + \sum(y_t - \beta_2 y_{t-1})\right] + \left[y_0^2(1-\beta_2^2) + \sum(y_t - \beta_2 y_{t-1})^2\right]$$

$$= c\left[\beta_1 - \left\{\sum(y_t - \beta_2 y_{t-1}) + y_0(1+\beta_2)\right\}\left(\frac{1}{c}\right)\right]^2 + \left[y_0^2(1-\beta_2^2) + \sum(y_t - \beta_2 y_{t-1})^2\right]$$

$$- \left(\frac{1}{c}\right)\left[\sum(y_t - \beta_2 y_{t-1}) + (1+\beta_2)y_0\right]^2$$

$$= c(\beta_1 - \hat{\beta}_1)^2 + \gamma \qquad ;\ \text{w/ } c = T + \frac{(1+\beta_2)}{(1-\beta_2)}$$

$$\&\ \gamma = \frac{1}{c}\left(T\Sigma_1 + \left(\frac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right)$$

where $\Sigma_1 = \sum(y_t - \bar{y} - \beta_2(y_{t-1} - \bar{y}_{-1}))^2$

$\Sigma_2 = \sum\left[y_t - \beta_2 y_{t-1} - (1-\beta_2)y_0\right]^2$

$\bar{y} = \frac{1}{T}\sum y_t \quad ; \quad \bar{y}_{-1} = \frac{1}{T}\sum y_{t-1}$

So, $$p(\beta, \sigma \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}}\right) *$$

$$\exp\left\{-\frac{1}{2\sigma^2}\left[c(\beta_1 - \hat{\beta}_1)^2 + \frac{1}{c}\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right]\right\}.$$

Now, we have:

$$\hat{\beta}_1 = [(1+\beta_2)y_0 + \Sigma(y_t - \beta_2 y_{t-1})]/c$$

But, let $h_1 = (T/\sigma^2)$; $h_2 = \dfrac{(1+\beta_2)}{\sigma^2(1-\beta_2)}$

So $(h_1 + h_2) = T/\sigma^2 + \frac{1}{\sigma^2}\left(\frac{1+\beta_2}{1-\beta_2}\right)$

$$= \frac{1}{\sigma^2}\left(T + \frac{1+\beta_2}{1-\beta_2}\right)$$

$$= (c/\sigma^2)$$

And, $h_1 \Sigma(y_t - \beta_2 y_{t-1})/T$

$$= \frac{1}{\sigma^2}\Sigma(y_t - \beta_2 y_{t-1})$$

And, $h_2(1-\beta_2)y_0 = \dfrac{(1+\beta_2)(1-\beta_2)y_0}{(1-\beta_2)\sigma^2}$

$$= \frac{1}{\sigma^2}(1+\beta_2)y_0$$

$$\therefore \hat{\beta}_1 = \frac{h_1 \Sigma(y_t - \beta_2 y_{t-1})/T + h_2(1-\beta_2)y_0}{(h_1 + h_2)}$$

Now, suppose that we are <u>given β_2 & σ</u>. Then the <u>conditional</u> posterior for β_1 is

$$p(\beta_1 | \beta_2, \sigma, y, y_0) \propto \frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}} \exp\left\{-\frac{c}{2\sigma^2}(\beta_1 - \hat{\beta}_1)^2\right\} *$$

$$\exp\left\{-\frac{1}{2c\sigma^2}\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right\}$$

$$\propto \exp\left\{-\frac{c}{2\sigma^2}(\beta_1 - \hat{\beta}_1)^2\right\}$$

So the <u>conditional posterior</u> for β_1 has mean $\hat{\beta}_1$, and variance (σ^2/c).

Now $c = \dfrac{(1+\beta_2)}{(1-\beta_2)} + T$

$$\therefore T \to \infty \Rightarrow (\sigma^2/c) \to 0.$$

& $T \to \infty \Rightarrow \hat{\beta}_1 \to \Sigma(y_t - \beta_2 y_{t-1})/T.$

To get the marginal posterior for β_1, integrate $p(\beta_1, \beta_2, \sigma | y, y_0)$ w.r.t. σ, & then use bivariate numerical integration w.r.t. β_2.

What about the marginal posterior pdf for β_2?

$$p(\beta_1, \beta_2, \sigma \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}}\right) \exp.\left\{-\frac{1}{2\sigma^2}\left[c(\beta_1-\hat{\beta}_1)^2 + \left(\frac{1}{c}\right)\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right\}\right.$$

$$\therefore p(\beta_2, \sigma \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}}\right) \exp\left\{-\frac{1}{2c\sigma^2}\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right\}$$

$$* \int \exp\left\{-\frac{1}{2\sigma^2}\left(c\,(\beta_1-\hat{\beta}_1)^2\right)\right\} d\beta_1$$

Now, $\int \left(\frac{c}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{c}{2\sigma^2}(\beta_1-\hat{\beta}_1)^2\right\} d\beta_1 = 1.$

$$\therefore \int \exp\left\{-\frac{c}{2\sigma^2}(\beta_1-\hat{\beta}_1)^2\right\} d\beta_1 = \left(\sigma/\sqrt{c}\right).$$

$$\therefore p(\beta_2 \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+1}\, c^{\frac{1}{2}}}\right) * \exp\left\{-\frac{1}{2c\sigma^2}\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right\}$$

$$\propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}\, c^{\frac{1}{2}}\, \sigma^{-1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\cdot\frac{1}{c}\left(T\Sigma_1 + \left(\frac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right)\right\}.$$

Now, $(h_1 + h_2) = (c/\sigma^2)$

So, $$\frac{(h_1\Sigma_1 + h_2\Sigma_2)}{(h_1+h_2)} = \frac{(T/\sigma^2)\Sigma_1 + \frac{(1+\beta_2)}{(1-\beta_2)}\frac{1}{\sigma^2}\Sigma_2}{(c/\sigma^2)}$$

$$= \left(\frac{1}{c}\right)\left(T\Sigma_1 + \frac{(1+\beta_2)}{(1-\beta_2)}\Sigma_2\right)$$

$$\therefore p(\beta_2, \sigma \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+2}(h_1+h_2)^{\frac{1}{2}}}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{h_1\Sigma_1 + h_2\Sigma_2}{h_1+h_2}\right]\right\}.$$

Now, Consider, $$p(\beta_2, \sigma \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{\sigma^{T+1}\, c^{\frac{1}{2}}}\right) * \exp\left\{-\frac{1}{2c\sigma^2}\left[T\Sigma_1 + \left(\frac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right]\right\}$$

$$p(\beta_2 \mid y, y_0) \propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{c^{\frac{1}{2}}}\right) \int_0^\infty \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}\cdot\frac{1}{c}\left[T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right]\right\} d\sigma$$

$$\propto \left(\frac{p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}}{c^{\frac{1}{2}}}\right)\left\{\frac{1}{c}\left(T\Sigma_1 + \frac{1+\beta_2}{1-\beta_2}\Sigma_2\right)\right\}^{-T/2}$$

$$\propto \frac{p(\beta)(1-\beta_2^2)^{\frac{1}{2}}}{1} \cdot c^{(T-1)/2} \cdot \left[T\Sigma_1 + \left(\frac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right]^{-T/2}; \quad |\beta_2| < 1.$$

Now, consider the following expression —

$$\frac{p(\beta_2)}{(\Sigma_1)^{T/2}} \cdot \left[\frac{1-\beta_2^2}{1+(1/T)(1+\beta_2)/(1-\beta_2)}\right]^{\frac{1}{2}} * \left[\frac{1+(\frac{1}{T})(1+\beta_2)/(1-\beta_2)}{1+(\frac{1}{T})(\frac{1+\beta_2}{1-\beta_2})(\frac{\Sigma_2}{\Sigma_1})}\right]^{T/2}$$

$$= p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}\left[\tfrac{1}{T}\cdot c\right]^{\frac{1}{2}}(\Sigma_1)^{-T/2}\left[\frac{\frac{1}{T}(c)}{\frac{1}{T}\left[T+\left(\frac{1+\beta_2}{1-\beta_2}\right)\frac{\Sigma_2}{\Sigma_1}\right]}\right]^{T/2}$$

$$= p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}\, c^{(T-1)/2}\left(\tfrac{1}{T}\right)^{\frac{1}{2}}\left[T\Sigma_1+\left(\tfrac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right]^{-T/2}$$

$$\times\; p(\beta_2)(1-\beta_2^2)^{\frac{1}{2}}\, c^{(T-1)/2}\left(T\Sigma_1+\left(\tfrac{1+\beta_2}{1-\beta_2}\right)\Sigma_2\right)^{-T/2}.$$

So;

$$p(\beta_2|y,y_0) \propto \frac{p(\beta_2)}{(\Sigma_1)^{T/2}}\left[\frac{1-\beta_2^2}{1+(1/T)(1+\beta_2)/(1-\beta_2)}\right]^{\frac{1}{2}} * \left[\frac{1+(\frac{1}{T})(1+\beta_2)/(1-\beta_2)}{1+(\frac{1}{T})(\frac{1+\beta_2}{1-\beta_2})(\frac{\Sigma_2}{\Sigma_1})}\right]^{T/2}.$$

~~Now, note that the second big term~~

If we are to use this then we need information concerning $p(\beta_2)$ & also concerning the pdf of y_0.

But, consider the asymptotic properties of this marginal posterior pdf —

$$\begin{aligned}\text{Log } p(\beta_2|y,y_0) \propto\; & \log p(\beta_2) - T/2 \log \Sigma_1 + \tfrac{1}{2}\log(1-\beta_2^2) \\ & - \tfrac{1}{2}\log\left(1+1/T\,(1+\beta_2)/(1-\beta_2)\right) + T/2\log\left(1+\tfrac{1}{T}\left(\tfrac{1+\beta_2}{1-\beta_2}\right)\right) \\ & - T/2\log\left(1+\left(\tfrac{1}{T}\right)\left(\tfrac{1+\beta_2}{1-\beta_2}\right)\tfrac{\Sigma_2}{\Sigma_1}\right)\end{aligned}$$

And as $T \to \infty$, the dominant term is $-T/2 \log \Sigma_1$

So, as $T \to \infty$,

$$p(\beta_2|y,y_0) \propto (\Sigma_1)^{-T/2}.$$

i.e. $p(\beta_2|y,y_0) \propto \left\{\sum(y_t-\bar{y}-\beta_2(y_{t-1}-\bar{y}_{-1}))^2\right\}^{-T/2}$

And this is <u>independent</u> of $p(\beta_2)$ & of assumptions concerning y_0. So, in large samples we have a neat & simple univariate t approximation to the marginal posterior pdf for β_2.

However, if the sample is small, then we are going to require information concerning both y_0 & $p(\beta_2)$. Consider the latter —

we want to restrict $|\beta_2| < 1$.

Now, consider the <u>Beta pdf</u> —

$$p(z) \propto (z)^{a-1}(1-z)^{b-1} \quad ; \quad 0 < z < \infty.$$

Now, let
$$a = k_2$$
$$b = k_1$$
$$z = \tfrac{1}{2}(1+\beta_2)$$
$$\therefore \; 1-z = \tfrac{1}{2}(1-\beta_2)$$

$$\& \; dz \propto d\beta_2$$

So $p(\beta_2) \propto (1+\beta_2)^{k_2-1}(1-\beta_2)^{k_1-1}$

And $0 < z < 1 \Rightarrow 0 < \tfrac{1}{2}(1+\beta_2) < 1$

$\Rightarrow -1 < \beta_2 < 1$, as required.

So put $p(\beta_2) \propto (1-\beta_2)^{k_1-1}(1+\beta_2)^{k_2+1}$.

We still have to assign the weights k_1 & k_2 <u>a priori</u>.

If we put $k_1 = k_2 = \tfrac{1}{2}$,
then $p(\beta_2) \propto (1-\beta_2)^{-\frac{1}{2}}(1+\beta_2)^{-\frac{1}{2}}$

$$\therefore \; p(\beta_2) \propto (1-\beta_2^2)^{-\frac{1}{2}} \quad ; \quad |\beta_2| < 1.$$

And it can be shown that this is <u>approximately</u> the same as Jeffrey's invariant prior
<u>viz</u> $p(\beta_2) \propto |\mathrm{Inf}_{\beta_2}|^{\frac{1}{2}}$.

So put $p(\beta_1, \beta_2, \sigma) \propto \dfrac{1}{\sigma(1-\beta_2^2)^{\frac{1}{2}}}$.

(B) — Omitted.

(C) - <u>Stability of a 2nd-Order Autoregressive Structure</u> -

Bayesian techniques may be used to make inferences about the dynamic properties of solutions to difference equations. The sort of question in which we may be interested is —

"On the basis of our data, what is the posterior probability that the model's solution will be explosive or oscillatory?"

Take a 2nd-order autoregressive model —

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t$$

$$\ni \quad u_t \sim NID(0, \sigma^2) \quad ; \forall t.$$

Assume that the initial values y_0 & y_{-1} are <u>given</u>.

Then,

$$l(\alpha_1, \alpha_2, \sigma | y) \propto \left(\frac{1}{\sigma^T}\right) exp\left\{ -\frac{1}{2\sigma^2}\left(\sum_1^T (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2\right)\right\}.$$

Assume that $p(\alpha_1, \alpha_2, \sigma) \propto 1/\sigma$; $-\infty < \alpha_i < \infty$; $0 < \sigma < \infty$.

Then $p(\alpha_1, \alpha_2 | y) \propto \left[\sum_1^T (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2\right]^{-T/2}$

$$\propto \left[\nu s^2 + (\alpha - \hat{\alpha})' H (\alpha - \hat{\alpha})\right]^{-T/2}$$

where $\nu = T - 2$; $\nu s^2 = \sum_1^T (y_t - \hat{\alpha}_1 y_{t-1} - \hat{\alpha}_2 y_{t-2})^2$

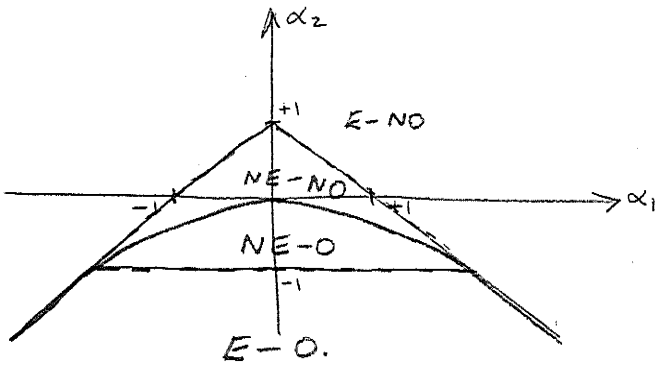
$$\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = H^{-1} \begin{pmatrix} \Sigma y_t y_{t-1} \\ \Sigma y_t y_{t-2} \end{pmatrix}$$

$$\ni H = \begin{pmatrix} \Sigma y_{t-1}^2 & \Sigma y_{t-1} y_{t-2} \\ \Sigma y_{t-1} y_{t-2} & \Sigma y_{t-2}^2 \end{pmatrix}.$$

Then the joint posterior pdf for (α_1, α_2) is <u>bivariate t</u>.

If we are given values for y, then we can make inferences about α_1 & α_2, & hence about the dynamic properties of the model.

Since we have $p(\alpha_1, \alpha_2 | y)$, we can normalize & then integrate over various ranges for α_1 & α_2, & this gives us the posterior probabilities relating to the properties of the solution to the model.



If the volume of the posterior pdf over the "O-NE" region is 0.9, then we say that the probability that the model is oscillatory & non-explosive is 0.9. etc.
Also, $P(NE) = P(NE-O) + P(NE-NO)$. etc.

The characteristic equation for the model is

$$x^2 - \alpha_1 x - \alpha_2 = 0$$

$$\therefore x_1 = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \quad ; \quad x_2 = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}.$$

And the system will <u>oscillate</u> if $\alpha_1^2 + 4\alpha_2 < 0$.

So, it may be of interest to find Pr. (oscillates).

i.e. $Pr(\alpha_1^2 + 4\alpha_2 < 0)$.

Let $v_1 = \alpha_1$
$v_2 = \alpha_1^2 + 4\alpha_2$

$$\therefore J^{-1} = \left| \frac{\partial v_i}{\partial \alpha_j} \right| = \begin{vmatrix} \frac{\partial v_1}{\partial \alpha_1} & \frac{\partial v_1}{\partial \alpha_2} \\ \frac{\partial v_2}{\partial \alpha_1} & \frac{\partial v_2}{\partial \alpha_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 2\alpha_1 & 4 \end{vmatrix} = 4$$

$\therefore J = (1/4)$

$$\therefore p(v_1, v_2 | y) = \frac{1}{4} p(\alpha_1, \alpha_2 | y)$$

$$\propto [vs^2 + (\alpha - \hat{\alpha})' H (\alpha - \hat{\alpha})]^{-T/2}.$$

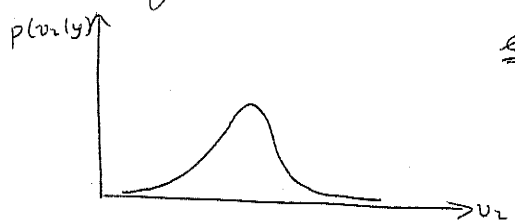
$$\therefore\ p(v_1, v_2 | y) \propto \Big[\nu s^2 + (v_1 - \hat{\alpha}_1)^2 h_{11} + \Big(\frac{v_2 - v_1^2}{4} - \hat{\alpha}_2 \Big)^2 h_{22} + 2 (v_1 - \hat{\alpha}_1)\Big(\frac{v_2 - v_1^2}{4} - \hat{\alpha}_2\Big) h_{21} \Big]^{-T/2}$$

Now, we are interested in $p(v_2 | y)$.

So, $p(v_2 | y) = \int_{-\infty}^{\infty} p(v_1, v_2 | y)\, dv_1$

So obtain this posterior <u>marginal</u> pdf, & normalize. Then see when $v_2 < 0$.

For the given set of data, compute $p(v_2 | y)$ for different values of v_2. Plot –



<u>eg</u>

Is the majority of the area to the right of $v_2 = 0$? If so, it <u>doesn't</u> oscillate.

Why not generalize this type of analysis to the model –

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + X(t)\beta + u_t$$

where $X(t)$ is a $(1 \times k)$ row vector of independent variables.

D) "Distributed Lag" Models:—

Consider the usual "geometric" model—

$$y_t = \alpha \sum_0^\infty \lambda^i x_{t-i} + u_t \quad ; \quad -\infty < \alpha < \infty, \quad 0 \le \lambda < 1.$$

Then the Koyck Transformation yields

$$y_t = \lambda y_{t-1} + \alpha x_t + (u_t - \lambda u_{t-1})$$

Assume that $u_t \sim NID(0, \sigma^2)$; $\forall t$.

Then $\text{var.}[u_t - \lambda u_{t-1}] = \text{var}(u_t) + \lambda^2 \text{var}(u_{t-1})$
$= \sigma^2(1+\lambda^2)$.

Further, $E[(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2})]$

$= E(u_t u_{t-1}) + \lambda^2 E(u_{t-1} u_{t-2}) - \lambda E(u_{t-1}^2) - \lambda E(u_t u_{t-2})$

$= 0 + 0 - \lambda\sigma^2 - 0$ (By independence assumption).
$= -\lambda\sigma^2$.

Further, $E[(u_t - \lambda u_{t-1})(u_{t-k} - \lambda u_{t-k-1})]$; $k \geq 1$.

$= E[u_t u_{t-k}] + \lambda^2 E[u_{t-1} u_{t-k-1}] - \lambda E[u_{t-1} u_{t-k}]$
$- \lambda E[u_t u_{t-k-1}]$
$= 0$.

So, $E[(u - \lambda u_{-1})(u - \lambda u_{-1})'] = \sigma^2 G$

$$\ni G = \begin{bmatrix} (1+\lambda^2) & -\lambda & 0 & \cdots & 0 \\ -\lambda & (1+\lambda^2) & -\lambda & 0 \cdots & 0 \\ 0 & -\lambda & (1+\lambda^2) & -\lambda & 0 \cdots 0 \\ \vdots & & \ddots & \ddots & \ddots & -\lambda \\ 0 & \cdots & \cdots & 0 & -\lambda & (1+\lambda^2) \end{bmatrix}$$

Let $y' = (y_1, \ldots, y_T)$ & assume y_0 given.

Then, $p(y \mid \lambda, \alpha, \sigma, y_0) \propto \frac{|G|^{-\frac{1}{2}}}{\sigma^T} \exp\left\{-\frac{1}{2\sigma^2}(y - \lambda y_{-1} - \alpha x)' G^{-1} (y - \lambda y_{-1} - \alpha x)\right\}$

Assume that $p(\alpha, \lambda, \sigma) \propto \left(\frac{1}{\sigma}\right)$

$\therefore p(\alpha, \lambda, \sigma \mid y, y_0) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma^{T+1}}\right) \exp.\left\{-\frac{1}{2\sigma^2}(y - \lambda y_{-1} - \alpha x)' G^{-1} (y - \lambda y_{-1} - \alpha x)\right\}$

So, $p(\lambda, \alpha | y, y_0) \propto \{(y-\lambda y_{-1}-\alpha x)' G^{-1} (y-\lambda y_{-1}-\alpha x)\}^{-T/2} \cdot |G|^{-\frac{1}{2}}$.

The normalizing constant can be obtained from the ~~properties of the Bivariate norm~~ bivariate numerical routine. Then <u>joint</u> inferences may be made about α and λ. The marginal posterior for λ can be obtained by completing the square on α and integrating w.r.t. α, using the properties of the univariate t dist'n. This is feasible <u>analytically.</u>

To get $p(\alpha | y, y_0)$, however, we have to use <u>numerical</u> techniques (univariate) because we have restricted $0 \le \lambda < 1$, so the t-dist'n. is <u>not</u> applicable.

Now consider a more elaborate model —

$$y_t = \alpha \sum_0^\infty \lambda^i x_{t-i} + \sum_0^\infty \lambda^i u_{t-i} \quad ; \quad t = 1, 2, \ldots, T.$$

Then, $\lambda y_{t-1} = \alpha \sum_0^\infty \lambda^{i+1} x_{t-i-1} + \sum_0^\infty \lambda^{i+1} u_{t-i-1}$

$$= \alpha \sum_1^\infty \lambda^i x_{t-i} + \sum_1^\infty \lambda^i u_{t-i}$$

$\therefore (y_t - \lambda y_{t-1}) = \alpha x_t + u_t.$

$\therefore \; y_t - \lambda y_t + \alpha x_t + u_t$

And $u_t \sim NID(0, \sigma^2)$; $\forall t$.

Then, for <u>given</u> y_0 —

$$l'(x, \lambda, \sigma | y, y_0) \propto \left(\frac{1}{\sigma^T}\right) \exp\left\{-\frac{1}{2\sigma^2}(y-\lambda y_{-1}-\alpha x)'(y-\lambda y_{-1}-\alpha x)\right\}$$

And $p(\alpha, \lambda, \sigma) \propto (1/\sigma)$

So, $p(\alpha, \lambda, \sigma | y, y_0) \propto \left(\frac{1}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}(y-\lambda y_{-1}-\alpha x)'(y-\lambda y_{-1}-\alpha x)\right\}$

$\therefore p(\alpha, \lambda | y, y_0) \propto \{(y-\lambda y_{-1}-\alpha x)'(y-\lambda y_{-1}-\alpha x)\}^{-T/2}$

Now, if it were not for the fact that we have restricted $0 \le \lambda < 1$, the joint posterior for α and λ would be a bivariate t-dist'n.

So, to make <u>joint</u> inferences on λ and α, obtain the constant term for $p(\alpha, \lambda | y, y_0)$ by bivariate numerical integration.

Further, to get the marginal posterior pdf's —

(a) $p(\lambda | y, y_0)$ can be obtained <u>analytically</u> by completing the square on α, and using the properties of the univariate t-dist'n.

(b) The restrictions on λ necessitate the use of <u>numerical</u> (univariate) integration to obtain $p(\alpha | y, y_0)$.

Suppose that we have our model in the form

$$y_t = \lambda y_{t-1} + \alpha x_t + u_t$$

And we now wish to assume autocorrelated disturbances —

$$u_t = \rho u_{t-1} + \epsilon_t \;;$$

$$\ni \epsilon_t \sim NID(0, \tau^2).$$

Then $\rho y_{t-1} = \rho\lambda y_{t-2} + \rho\alpha x_{t-1} + \rho u_{t-1}$

$$\therefore (y_t - \rho y_{t-1}) = \lambda(y_{t-1} - \rho y_{t-2}) + \alpha(x_t - \rho x_{t-1}) + \epsilon_t$$

Now, assume that $p(\alpha, \lambda, \rho, \tau) \propto (1/\tau)$.

where $-\infty < \rho, \alpha < \infty$; $0 < \lambda < 1$; $0 < \tau < \infty$.
Now, assuming that we are <u>given</u> y_0 & y_{-1}, let
$y' = (y_{-1}, y_0, y_1, ---, y_T)$.

So, $p(\lambda, \rho, \alpha, \tau | y) \propto \left(\frac{1}{\tau^{T+1}}\right) \exp\left\{-\frac{1}{2\tau^2}\sum_1^T [y_t - \lambda y_{t-1} - \alpha x_t - \rho(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})]^2\right\}$.

Then,

$$p(\lambda, \rho, \alpha | y) \propto \left\{\sum_1^T [y_t - \lambda y_{t-1} - \alpha x_t - \rho(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})]^2\right\}^{-T/2}.$$

$$\propto \left\{\sum_1^T [(y_t - \rho y_{t-1}) - \lambda(y_{t-1} - \rho y_{t-2}) - \alpha(x_t - \rho x_{t-1})]^2\right\}^{-T/2}.$$

Look at the <u>2nd</u> line. If it hadn't been for the fact that we assumed that $0 < \lambda < 1$, then $p(\alpha, \lambda | \rho, y)$ would be bivariate t. Now, we can analyze this joint conditional posterior pdf numerically & make inferences for α and λ, <u>given</u> ρ. But what if we choose the wrong value of ρ when we estimate it (since ρ is always unknown)? Then Zellner shows that the results obtained for $p(\alpha | \rho, y)$ and $p(\lambda | \rho, y)$ are very sensitive to a mis-specification of the value of ρ.

So, if we suspect that $\rho \neq 0$, then it is better to make use of $p(\alpha, \lambda | y)$ & hence $p(\alpha | y)$ & $p(\lambda | y)$ when making inferences, than to rely upon $p(\alpha, \lambda | \rho, y)$ and $p(\alpha | \rho, y)$, $p(\lambda | \rho, y)$.

We now show how to obtain $p(\alpha, \lambda | y)$, $p(\alpha | y)$, $p(\lambda | y)$ for such situations.

The derivation of $p(\alpha, \lambda | y)$:

$$p(\alpha, \lambda | y) = \int_{-\infty}^{\infty} p(\alpha, \lambda, \rho | y)\, d\rho$$

$$\propto \int_{-\infty}^{\infty} \left\{ \sum_1^T \left[y_t - \lambda y_{t-1} - \alpha x_t - \rho(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1}) \right]^2 \right\}^{-T/2} d\rho.$$

We have a quadratic in ρ ~
$ax^2 + bx + c$

Complete the square ~ $a(x + b/2a)^2 + (c - \frac{b^2}{4a})$.

$$\Rightarrow \sum_1^T (y_t - \lambda y_{t-1} - \alpha x_t)^2 + \rho^2 \sum_1^T (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2$$
$$- 2\rho \sum_1^T (y_t - \lambda y_{t-1} - \alpha x_t)(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})$$

$$\Rightarrow \left[\sum_1^T (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2 \right] \left[\rho - \frac{\sum_1^T (y_t - \lambda y_{t-1} - \alpha x_t)(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})}{\sum (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2} \right]$$
$$+ \left[\sum (y_t - \lambda y_{t-1} - \alpha x_t)^2 - \frac{\left\{ \sum (y_t - \lambda y_{t-1} - \alpha x_t)(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1}) \right\}^2}{\sum (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2} \right]$$

Now, $\left\{ a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a}) \right\}^{-T/2}$

$$= (c - \tfrac{b^2}{4a})^{-T/2} \left\{ 1 + \frac{a(x + b/2a)^2}{(c - b^2/4a)} \right\}^{-T/2}$$

$$\therefore p \propto (c - \tfrac{b^2}{4a})^{-T/2} \int_{-\infty}^{\infty} \left[1 + \frac{a(x + b/2a)^2}{(c - b^2/4a)} \right]^{-T/2} dx$$

$$\propto (c - \tfrac{b^2}{4a})^{-T/2} \left(\frac{a}{c - b^2/4a} \right)^{-\frac{1}{2}} \int \left(\frac{a}{c - b^2/4a} \right)^{1/2} \left[1 + \frac{a(x + \frac{b}{2a})^2}{(c - b^2/4a)} \right]^{-T/2} dx$$

$$\propto a^{-\frac{1}{2}} (c - \tfrac{b^2}{4a})^{-\frac{(T-1)}{2}}$$

So;

$$p(\alpha, \lambda | y) \propto \left\{ \sum (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2 \right\}^{-\frac{1}{2}} *$$
$$\left\{ \sum (y_t - \lambda y_{t-1} - \alpha x_t)^2 - \frac{\left[\sum (y_t - \lambda y_{t-1} - \alpha x_t)(y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1}) \right]^2}{\sum (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})^2} \right\}^{-\gamma}$$

where $\gamma = \frac{(T-1)}{2}$

We can now make joint inferences about α & λ. If we apply <u>numerical</u> integration we can derive $p(\alpha | y)$ and $p(\lambda | y)$.

Also of interest may be functions of α and λ.
eg. the "long-run" quantity:
$$\eta = \left(\frac{\alpha}{1-\lambda}\right).$$

To make posterior inferences about η, change our variables from (α, λ) to (η, λ).

Then $J = \begin{vmatrix} \frac{\partial \alpha}{\partial \eta} & \frac{\partial \alpha}{\partial \lambda} \\ \frac{\partial \lambda}{\partial \eta} & \frac{\partial \lambda}{\partial \lambda} \end{vmatrix}$

$$\therefore J = \begin{vmatrix} (1-\lambda) & 0 \\ \left(\frac{\alpha}{\eta^2}\right) & 1 \end{vmatrix} = (1-\lambda).$$

So, $p(\eta, \lambda | y) \propto (1-\lambda) p(\alpha, \lambda | y) \propto (1-\lambda) p[\eta(1-\lambda), \lambda]$.

$$\therefore p(\eta | y) \propto \int_0^1 (1-\lambda) p(\eta(1-\lambda), \lambda) \, d\lambda.$$

And this can be evaluated exactly numerically.

E) Applications to Consumption Functions:-

Suppose that we have the model

$$C_t = k Y_t^* + u_t$$

where Y_t^* is "expected" income. (i.e. P.I.H.)

And $(Y_t^* - Y_{t-1}^*) = (1-\lambda)(Y_t - Y_{t-1}^*)$

$$\therefore Y_t^* = Y_{t-1}^* + Y_t - \lambda Y_t - Y_{t-1}^* + \lambda Y_{t-1}^*$$
$$= (1-\lambda) Y_t + \lambda (Y_{t-1}^*)$$

$$\therefore Y_t^* = (1-\lambda) Y_t + \lambda [\lambda Y_{t-2}^* + (1-\lambda) Y_{t-1}]$$
$$= (1-\lambda)[Y_t + \lambda Y_{t-1}] + \lambda^2 Y_{t-2}^* \quad \text{etc.}$$

So, $Y_t^* = (1-\lambda)[Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \text{-------}]$.

And we assume that $0 < \lambda < 1$.

$$\therefore C_t = k(1-\lambda)[Y_t + \lambda Y_{t-1} + \cdots + \cdots] + u_t$$

$$\& \ \lambda C_{t-1} = k(1-\lambda)[\lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots] + \lambda u_{t-1}$$

$$\therefore (C_t - \lambda C_{t-1}) = k(1-\lambda) Y_t + (u_t - \lambda u_{t-1})$$

$$\therefore C_t = \lambda C_{t-1} + k(1-\lambda) Y_t + (u_t - \lambda u_{t-1}).$$

Now, consider the various assumptions which could be made concerning the structure of the disturbances.

I) $(u_t - \lambda u_{t-1}) = \varepsilon_{1t} \sim NID(0, \sigma_1^2)$.

II) $u_t \sim NID(0, \sigma_2^2)$

III) u_t are 1st order auto.
viz $u_t = \rho u_{t-1} + \varepsilon_{3t}$; $\varepsilon_{3t} \sim NID(0, \sigma_3^2)$

IV) $(u_t - \lambda u_{t-1})$ is 1st. order auto
viz $(u_t - \lambda u_{t-1}) = \gamma(u_{t-1} - \lambda u_{t-2}) + \varepsilon_{4t}$; $\varepsilon_{4t} \sim NID(0, \sigma_4^2)$

Note: (a) $\rho = \lambda \Rightarrow (III) \equiv (I)$.

(b) $\rho = 0 \Rightarrow (III) \equiv (II)$

(c) $\gamma = 0 \Rightarrow (IV) \equiv (I)$

Assume that the initial condition, C_0, is given.

Case (I) :

$$p(C|k,\lambda,\sigma_1) \propto \left(\frac{1}{\sigma_1^T}\right) \exp\left\{-\frac{1}{2\sigma_1^2}\sum_1^T [C_t - \lambda C_{t-1} - k(1-\lambda)Y_t]^2\right\}$$

$$p(k,\lambda,\sigma_1) \propto \left(\frac{1}{\sigma_1}\right) \quad ; \quad \begin{cases} 0 < \lambda, k < 1 \\ 0 < \sigma < \infty \end{cases}$$

$$\therefore p(k,\lambda,\sigma_1|C) \propto \left(\frac{1}{\sigma_1^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma_1^2}\sum_1^T [C_t - \lambda C_{t-1} - k(1-\lambda)Y_t]^2\right\}$$

$$p(k,\lambda|C) \propto \left\{\sum_1^T [C_t - \lambda C_{t-1} - k(1-\lambda)Y_t]^2\right\}^{-T/2} ; \quad 0 < \lambda, k < 1.$$

Now, to get $p(\lambda|C)$ we obtain $\int_0^1 p(\lambda,k|C)\,dk$.

But, this is a truncated student pdf that we are integrating, since the range of k is $(0,1)$. So handle this numerically. Similarly when we obtain $p(k|C)$.

<u>Case (II)</u>: $u_t \sim NID(0, \sigma_2^2)$

So, $E\{(u-\lambda u_{-1})'(u-\lambda u_{-1})\} = \sigma_2^2 G$, where G is as defined in the previous section of these notes.

So,
$$p(c|k,\lambda,\sigma_2) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma_2^T}\right) \exp\left\{-\frac{1}{2\sigma_2^2}[c-\lambda c_{-1}-k(1-\lambda)Y]'G^{-1}[c-\lambda c_{-1}-k(1-\lambda)Y]\right\}$$

& $p(k,\lambda,\sigma_2) \propto \frac{1}{\sigma_2}$.

So,
$$p(k,\lambda,\sigma_2|c) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma_2^{T+1}}\right) \exp.\left\{-\frac{1}{2\sigma_2^2}[c-\lambda c_{-1}-k(1-\lambda)Y]'G^{-1}[c-\lambda c_{-1}-k(1-\lambda)Y]\right\}$$

$$p(k,\lambda|c) \propto \left\{[c-\lambda c_{-1}-k(1-\lambda)Y]'G^{-1}[c-\lambda c_{-1}-k(1-\lambda)Y]\right\}^{-T/2} . |G|^{-\frac{1}{2}}.$$

for $0<\lambda, k, <1$.

By applying bivariate numerical integration, we can normalize this pdf & make <u>joint</u> posterior inferences for k & λ. The integrate (bivariate numerical) again to get the two marginal posterior pdf's.

<u>Case (III)</u>:- $u_t = \rho u_{t-1} + \epsilon_{3t}$; $\epsilon_{3t} \sim NID(0, \sigma_3^2)$

And we have: $C_t = \lambda C_{t-1} + k(1-\lambda)y_t + (u_t - \lambda u_{t-1})$

Let $\eta_t = C_t - u_t$

$\therefore \lambda\eta_{t-1} = \lambda C_{t-1} - \lambda u_{t-1}$

So, $\eta_t = \lambda\eta_{t-1} + k(1-\lambda)y_t$

and $\rho\eta_{t-1} = \lambda\rho\eta_{t-2} + \rho k(1-\lambda)y_{t-1}$

$\therefore (\eta_t - \rho\eta_{t-1}) = \lambda(\eta_{t-1} - \rho\eta_{t-2}) + k(1-\lambda)(y_t - \rho y_{t-1})$

$$\begin{aligned}
\therefore \eta_t(\rho) &= \lambda\eta_{t-1}(\rho) + k(1-\lambda)y_t(\rho) \\
&= \lambda^2\eta_{t-2}(\rho) + \lambda k(1-\lambda)y_{t-1}(\rho) + k(1-\lambda)y_t(\rho) \\
&= \lambda^3\eta_{t-3}(\rho) + \lambda^2 k(1-\lambda)y_{t-2}(\rho) + \lambda k(1-\lambda)y_{t-1}(\rho) + k(1-\lambda)y_t(\rho)
\end{aligned}$$

$$\begin{aligned}
\therefore \eta_t(\rho) &= \lambda^t\eta_0(\rho) + k(1-\lambda)[y_t(\rho) + \lambda y_{t-1}(\rho) + \cdots + \lambda^{t-1}y_1(\rho)] \\
&= \lambda^t\eta_0' + k(1-\lambda)[y_t(\rho) + \lambda y_{t-1}(\rho) + \cdots + \lambda^{t-1}y_1(\rho)].
\end{aligned}$$

Let $C_t(\rho) = C_t - \rho C_{t-1}$

& $\eta_t = C_t - u_t$

$\therefore \eta_t(\rho) = \eta_t - \rho\eta_{t-1}$

$= C_t - \rho C_{t-1} - u_t + \rho u_{t-1}$
$= C_t(\rho) - u_t + \rho u_{t-1}$

$\therefore C_t(\rho) = \eta_t(\rho) + u_t - \rho u_{t-1}$
$= \eta_t(\rho) + \varepsilon_{3t}.$

$\therefore C_t(\rho) = \lambda^t \eta_0' + k(1-\lambda)\left[y_t(\rho) + \lambda y_{t-1}(\rho) + \cdots + \lambda^{t-1} y_1(\rho)\right] + \varepsilon_{3t}.$

$\therefore C_t(\rho) = \lambda^t \eta_0' + k(1-\lambda) Z_t(\rho,\lambda) + \varepsilon_{3t}.$

$\therefore \ell(C \mid \lambda, k, \rho, \sigma_3, \eta_0') \propto \left(\frac{1}{\sigma_3^T}\right) \exp.\left\{-\frac{1}{2\sigma_3^2}\sum_1^T\left[C_t(\rho) - \lambda^t\eta_0' - (1-\lambda)k Z_t(\rho,\lambda)\right]^2\right\}$

& $p(\lambda, k, \rho, \sigma_3, \eta_0') \propto \left(\frac{1}{\sigma_3}\right)$; $0 < \lambda, k < 1$
$0 < \sigma_3 < \infty$
$-\infty < \eta_0', \rho < \infty.$

$\therefore p(\lambda, k, \rho, \sigma_3, \eta_0 \mid C) \propto \left(\frac{1}{\sigma_3^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma_3^2}\sum_1^T\left[C_t(\rho) - \lambda^t\eta_0' - (1-\lambda)k Z_t(\rho,\lambda)\right]^2\right\}$

$\therefore p(\lambda, k, \rho, \eta_0' \mid C) \propto \left\{\sum_1^T\left[C_t(\rho) - \lambda^t\eta_0' - (1-\lambda)k Z_t(\rho,\lambda)\right]^2\right\}^{-T/2}$

This joint distribution can be integrated numerically w.r.t. ρ and η_0' to yield a bivariate posterior distribution in λ and k. The latter can be normalized numerically, and if desired the marginal posterior pdf's for λ and k can be obtained.

<u>Case IV</u>:

$$(u_t - \lambda u_{t-1}) = \gamma(u_{t-1} - \lambda u_{t-2}) + \varepsilon_{4t}.$$

$\therefore p(C \mid \lambda, k, \gamma, \sigma_4) \propto \left(\frac{1}{\sigma_4^T}\right) \exp\left\{-\frac{1}{2\sigma_4^2}\varepsilon_4'\varepsilon_4\right\}$

& $\varepsilon_4 = \left[(u - \lambda u_{-1}) - \gamma(u_{-1} - \lambda u_{-2})\right].$

$= \left\{C - \lambda C_{-1} - k(1-\lambda)y - \gamma\left[C_{-1} - \lambda C_{-2} - k(1-\lambda)y_{-1}\right].\right\}$

& $p(\lambda, k, \gamma, \sigma_4) \propto \left(\frac{1}{\sigma_4}\right)$

$\therefore p(\lambda, k, \gamma, \sigma_4 \mid C) \propto \left(\frac{1}{\sigma_4^{T+1}}\right) \exp.\left\{-\frac{1}{2\sigma_4^2}(\varepsilon_4'\varepsilon_4)\right\}$

$\therefore p(\lambda, k, \gamma) \propto \left\{\varepsilon_4'\varepsilon_4\right\}^{-T/2}$; for $\begin{cases} 0 < \lambda, k < 1 \\ 0 < \sigma_4 < \infty \\ -\infty < \gamma < \infty \end{cases}$

$$\therefore\ p(k,\lambda,\gamma|c) \propto \left\{[c-\lambda c_{-1}-k(1-\lambda)y-\gamma(c_{-1}-\lambda c_{-2}-k(1-\lambda)y_{-1}]' * [c-\lambda c_{-1}-k(1-\lambda)y-\gamma(c_{-1}-\lambda c_{-2}-k(1-\lambda)y_{-1}]\right\}^{-T/2}$$

$$\propto \left\{\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t-\gamma(c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1})]^2\right\}^{-T/2}$$

And to obtain the joint posterior pdf for k and λ, complete the square on γ and use the properties of the univariate t-dist'n. to integrate out the γ term.

viz:

$$\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_{t-1}-\gamma(c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1})]^2$$

$$= \sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_{t-1}]^2 + \gamma^2\sum_1^T [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2$$

$$-2\gamma\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_{t-1}][c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}].$$

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

$\Longrightarrow$

$$a = \sum_1^T [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2$$

$$b = -2\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t][c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]$$

$$c = \sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t]^2$$

Now, referring back to the notes for the last section in this chapter, we find that for this functional form we finally arrive at the result —

$$p \propto a^{-\frac{1}{2}}\left(c-\frac{b^2}{4a}\right)^{-(T-1)/2}.$$

$$\therefore\ p(\lambda,\alpha|c) \propto \left[\sum_1^T [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2\right]^{-\frac{1}{2}} * \left\{\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t]^2 - \frac{\left[\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t][c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]\right]^2}{\sum [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2}\right\}^{-\frac{(T-1)}{2}}$$

$$\text{let } \tilde{\gamma} = \left\{\frac{\sum [c_t-\lambda c_{t-1}-k(1-\lambda)y_t][c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]}{\sum [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2}\right\}$$

$$\therefore\ p(\lambda,\alpha|c) \propto \left[\sum_1^T [c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]^2\right]^{-\frac{1}{2}} * \left\{\sum_1^T [c_t-\lambda c_{t-1}-k(1-\lambda)y_t]^2 - \tilde{\gamma}\sum [c_t-\lambda c_{t-1}-k(1-\lambda)y_t][c_{t-1}-\lambda c_{t-2}-k(1-\lambda)y_{t-1}]\right\}^{-\frac{(T-1)}{2}}$$

(as is quoted in Zellner.)

Now this pdf is bivariate in k & λ but follows no particular distributional form. Further, it is a <u>truncated</u> distribution, since the ranges of λ & k are each in $(0,1)$.

So bivariate <u>numerical</u> integration may be used to normalize $p(\lambda, k \mid c)$, and to enable joint posterior inferences to be made; and also, the same techniques may be used to obtain $p(\lambda \mid c)$ and $p(k \mid c)$, if desired.

E). Generalizations of Distributed Lags:

1): Let $y_t = \beta_1 \sum_0^\infty \lambda^i x_{t-i,1} + \beta_2 \sum_0^\infty \lambda^i x_{t-i,2} + \cdots\cdots$

$$\cdots + \beta_k \sum_0^\infty \lambda^i x_{t-i,k} + u_t$$

$$\therefore y_t = \sum_0^\infty \lambda^i X(t-i)\beta + u_t$$

$$\& \ \lambda y_{t-1} = \sum_0^\infty \lambda^{i+1} X(t-i-1)\beta + \lambda u_{t-1}$$

$$= \sum_1^\infty \lambda^i X(t-i)\beta + \lambda u_{t-1}$$

$$\therefore y_t = \lambda y_{t-1} + X(t)\beta + (u_t - \lambda u_{t-1})$$

Now, suppose that $u_t - \lambda u_{t-1} = \rho(u_{t-1} - \lambda u_{t-2}) + \varepsilon_t$

where $\varepsilon_t \sim N(0, \tau^2)$.

Then, $(y_t - \lambda y_{t-1}) = X(t)\beta + \rho[y_{t-1} - \lambda y_{t-2} - X(t-1)\beta] + \varepsilon_t$

$$\therefore (y_t - \lambda y_{t-1}) - \rho(y_{t-1} - \lambda y_{t-2}) = [X(t) - \rho X(t-1)]\beta + \varepsilon_t.$$

Suppose that our initial conditions, y_0 and y_{-1}, are <u>given</u>.

Then $p(\lambda, \rho, \tau, \beta) \propto \left(\frac{1}{\tau}\right)$; $\begin{cases} 0 < \lambda < 1 & ; -\infty < \beta_i < \infty \\ 0 < \tau < \infty & ; -\infty < \rho < \infty. \end{cases}$

$$\therefore p(\lambda, \rho, \tau, \beta \mid y, y_0) \propto \left(\frac{1}{\tau^{T+1}}\right) \exp\Big\{ -\frac{1}{2\tau^2} [y - \lambda y_{-1} - \rho(y_{-1} - \lambda y_{-2}) - (X - \rho X_{-1})\beta]' * [y - \lambda y_{-1} - \rho(y_{-1} - \lambda y_{-2}) - (X - \rho X_{-1})\beta] \Big\}.$$

Let $W = [y - \lambda y_{-1} - \rho(y_{-1} - \lambda y_{-2})]$

$$\therefore p(\lambda, \rho, \tau, \beta \mid y, y_0) \propto \left(\frac{1}{\tau^{T+1}}\right) \exp\Big\{ -\frac{1}{2\tau^2} [W - (X - \rho X_{-1})\beta]' [W - (X - \rho X_{-1})\beta] \Big\}$$

$$\therefore\ p(\lambda,\rho,\beta|y,y_0) \propto \left\{[\omega-(X-\rho X_{-1})\beta]'[\omega-(X-\rho X_{-1})\beta]\right\}^{-T/2}.$$

Then use the properties of the MVT dist'n. to integrate out β, having first expanded & completed the square on β.

$$p(\lambda,\rho,\beta|y,y_0) \propto \left\{\omega'\omega-\omega'(X-\rho X_{-1})\beta-\beta'(X-\rho X_{-1})'\omega \right.$$
$$\left. +\beta'(X-\rho X_{-1})'(X-\rho X_{-1})\beta\right\}^{-T/2}$$

Let $H = (X-\rho X_{-1})'(X-\rho X_{-1})$

$$\therefore\ p(\lambda,\rho,\beta|y,y_0) \propto \left\{\omega'\omega-2\beta'(X-\rho X_{-1})\omega+\beta'H\beta\right\}^{-T/2}.$$

Complete the square on β —

$$?^{-T/2} = \left\{[\beta-(X-\rho X_{-1})'\omega]'H[\beta-(X-\rho X_{-1})'\omega]\right.$$
$$\left. +\omega'\omega-[(X-\rho X_{-1})'\omega]'H^{-1}[(X-\rho X_{-1})'\omega]\right\}^{-T/2}$$

Now, see the earlier notes concerning the use of MVT when we have a function of this general form.

$$p(\lambda,\rho|y,y_0) \propto |a|^{-\frac{1}{2}}\left(c-\frac{b^2}{4a}\right)^{-(T-k)/2}.$$

Note the differences because we are now in a MVT dist'n. and not just a bivariate t-dist'n.

$$\therefore\ p(\lambda,\rho|y,y_0) \propto |H|^{-\frac{1}{2}}\left\{\omega'\omega-\omega'(X-\rho X_{-1})H^{-1}(X-\rho X_{-1})'\omega\right\}^{-\frac{(T-k)}{2}}.$$
$$\propto |H|^{-\frac{1}{2}}\left\{\omega'[I-(X-\rho X_{-1})H^{-1}(X-\rho X_{-1})']\omega\right\}^{-(T-k)/2}.$$

Then this pdf may be normalized by means of bivariate <u>numerical</u> integration, & the same technique may be adopted to obtain the 2 marginal posterior pdfs — $p(\lambda|y,y_0)$ and $p(\rho|y,y_0)$.

2): Now consider a second generalization of the basic distributed lag model:

$$y_t = \alpha \sum_0^\infty \lambda^i x_{t-i} + Z(t)\gamma + u_t$$

$$\therefore \lambda y_{t-1} = \alpha \sum_0^\infty \lambda^{i+1} x_{t-i-1} + \lambda Z(t-1)\gamma + \lambda u_{t-1}$$

$$\therefore y_t = \lambda y_{t-1} + \alpha x_t + [Z(t) - \lambda Z(t-1)]\gamma + (u_t - \lambda u_{t-1})$$

$$\therefore (y - \lambda y_{-1}) = [x \,\vdots\, (Z - \lambda Z_{-1})] \begin{bmatrix} \alpha \\ \hline \gamma \end{bmatrix} + (u - \lambda u_{-1})$$

Now, assume that $u_t \sim NID(0, \sigma^2)$

$$\therefore \ell(\beta, \lambda, \sigma \mid y, y_0) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma^T}\right) \exp\left\{-\frac{1}{2\sigma^2}(y - \lambda y_{-1} - W\beta)' G^{-1} * (y - \lambda y_{-1} - W\beta)\right\}$$

where $W = [x \,\vdots\, (Z - \lambda Z_{-1})]$; $\beta = \begin{bmatrix} \alpha \\ \hline \gamma \end{bmatrix}$.

Let $p(\beta, \lambda, \sigma) \propto \left(\frac{1}{\sigma}\right)$; $0 < \sigma < \infty$; $-\infty < \beta_i < \infty$; $0 < \lambda < 1$.

$$\therefore p(\beta, \lambda, \sigma \mid y, y_0) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}(y - \lambda y_{-1} - W\beta)' G^{-1} (y - \lambda y_{-1} - W\beta)\right\}$$

Let $\omega = y - \lambda y_{-1}$

$$\therefore p(\beta, \lambda, \sigma \mid y, y_0) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma^{T+1}}\right) \exp\left\{-\frac{1}{2\sigma^2}(\omega - W\beta)' G^{-1} (\omega - W\beta)\right\}$$

$$\cancel{\therefore p(\beta, \lambda \mid y, y_0) \propto \left(\frac{|G|^{-\frac{1}{2}}}{\sigma^{T+1}}\right) \exp\{}$$

$$\therefore p(\beta, \lambda \mid y, y_0) \propto |G|^{-\frac{1}{2}} \left\{(\omega - W\beta)' G^{-1} (\omega - W\beta)\right\}^{-T/2}.$$

So, $p(\beta \mid \lambda, y, y_0)$ is M.V.T. So we could see the sensitivity of this conditional posterior pdf for β as we vary assumptions on λ if we wish.

Suppose that we want $p(\lambda \mid y, y_0)$:

Complete the square on β in $p(\beta, \lambda \mid y, y_0)$ and use the properties of the MVT dist'n. to integrate out β.

$$\{ (\omega - W\beta)' G^{-1} (\omega - W\beta) \}^{-T/2}$$

$$= (\omega' G^{-1} \omega - \omega' G^{-1} W \beta - \beta' W' G^{-1} \omega + \beta' W' G^{-1} W \beta)^{-T/2}.$$

$$\therefore \quad a = W' G^{-1} W$$
$$b = -2 W' G^{-1} \omega$$
$$c = \omega' G^{-1} \omega.$$

$$\therefore \quad p(\lambda | y, y_0) \propto |a|^{-\frac{1}{2}} |G|^{-\frac{1}{2}} \left(c - \frac{b^2}{4a} \right)^{-(T-K)/2}.$$

$$\propto \frac{|G|^{-\frac{1}{2}} \cdot |W' G^{-1} W|^{-\frac{1}{2}}}{\{ \omega' G^{-1} \omega - (W' G^{-1} \omega)' (W' G^{-1} W)^{-1} (W' G^{-1} \omega) \}^{(T-K)/2}}.$$

Let $\hat{\beta} = (W' G^{-1} W)^{-1} (W' G^{-1} \omega)$

$$\therefore \quad \hat{\beta}' (W' G^{-1} W) \hat{\beta}$$

$$= (W' G^{-1} \omega)' (W' G^{-1} W)^{-1} (W' G^{-1} W)(W' G^{-1} W)^{-1} (W' G^{-1} \omega)$$

$$= (W' G^{-1} \omega)' (W' G^{-1} W)^{-1} (W' G^{-1} \omega)$$

$$\therefore \quad p(\lambda | y, y_0) \propto \frac{|G|^{-\frac{1}{2}} |W' G^{-1} W|^{-\frac{1}{2}}}{\{ \omega' G^{-1} \omega - \hat{\beta}' (W' G^{-1} W) (W' G^{-1} \omega) \}^{(T-K)/2}}.$$

<u>as is required</u>.

And this <u>is</u> indeed a fcn. of λ, since $\omega = y - \lambda y_{-1}$.

And we could also get $p(\beta | y, y_0)$ or $p(\beta_i | y, y_0)$ if we so desire.

Chapter 8: Multivariate Regression Models.

(A) The Multivariate Regression Model :-

$$Y = (y_1, ----, y_m) = \begin{pmatrix} y_{11} & --- & y_{1m} \\ \vdots & & \vdots \\ y_{n1} & --- & y_{nm} \end{pmatrix}$$

i.e. n observations on m equations.

$$\therefore \quad Y = XB + U$$

where X is $(n \times k)$ and of rank k,

$$B = (\beta_1, ----, \beta_m)$$

$$= \begin{pmatrix} \beta_{11} & --- & \beta_{1m} \\ \vdots & & \vdots \\ \beta_{k1} & --- & \beta_{km} \end{pmatrix}$$

i.e. k exogenous variables per equation — & these are the <u>same</u> variables in each equation.

$$U = (u_1, ---, u_m) = \begin{pmatrix} u_{11} & ---- & u_{1m} \\ \vdots & & \vdots \\ u_{n1} & --- & u_{nm} \end{pmatrix}$$

i.e. n observations on each of m disturbances.

Let the rows of U be independently distributed. Then there is no time dependence and so no serial correlation.

Let each row be distributed $N(0, \Sigma)$, where Σ is an $(m \times m)$ covariance matrix.

Then,

$$p(Y \mid X, B, \Sigma) \propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2} tr\left[(Y-XB)'(Y-XB)\Sigma^{-1}\right]\right\}$$

But, $(Y-XB)'(Y-XB)$

$$= (Y-X\hat{B})'(Y-X\hat{B}) + (B-\hat{B})'X'X(B-\hat{B})$$

$$= S + (B-\hat{B})'X'X(B-\hat{B})$$

where $S = (Y-X\hat{B})'(Y-X\hat{B})$

and $\hat{B} = (X'X)^{-1}X'Y$.

$$\therefore\ \ell(B,\Sigma \mid Y,X) \propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\left[S + (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]\right.$$

$$\propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\, S\Sigma^{-1} - \tfrac{1}{2}\,\mathrm{tr}\,(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$$

Now, Σ has $m(m+1)/2$ distinct elements —

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix} \qquad \& \ \sigma_{ij} = \sigma_{ji} \quad \forall\, i \neq j.$$

~~And $\exists$ m elements on the diagonal. Ignoring these $\exists$ $\frac{1}{2}m-1$~~

& no. of distinct elements $= m + (m-1) + (m-2) + \cdots + 2 + 1$
$= $ sum of A.P.
$= \frac{1}{2}m(m+1)$.

Now, take a diffuse prior pdf — then
$p(B,\Sigma) = p(B)p(\Sigma)$

& $p(B) \propto$ const.

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$$

$$\therefore\ p(B,\Sigma) \propto |\Sigma|^{-(m+1)/2}.$$

Note that if $m=1$, then $p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$ reduces to

$p(\sigma_{11}) \propto \left(\frac{1}{\sigma_{11}}\right)$, as is usual for the Jeffreys' diffuse prior.

Let $\sigma^{\mu\mu'}$ be the (μ,μ')th. element of Σ^{-1}.

Then transform $\Sigma \to \Sigma^{-1}$.

$$\therefore\ (\sigma_{11}, \sigma_{12}, \ldots, \sigma_{mm}) \to (\sigma^{11}, \sigma^{12}, \ldots, \sigma^{mm}).$$

Which is a transformation involving $\frac{1}{2}m(m+1)$ variables.

Then $J = \left|\dfrac{\partial(\sigma_{11}, \sigma_{12}, \ldots \sigma_{mm})}{\partial(\sigma^{11}, \sigma^{12}, \ldots \sigma^{mm})}\right| = |\Sigma|^{m+1}$

* For this see the separate notes from Anderson's book.

Now, $p(B, \Sigma) \propto |\Sigma|^{-(m+1)/2}$

& $J = |\Sigma|^{m+1}$

$$\therefore p(B, \Sigma^{-1}) \propto |\Sigma|^{m+1} \cdot |\Sigma|^{-(m+1)/2}$$
$$\propto |\Sigma|^{(m+1)/2}$$
$$\propto |\Sigma^{-1}|^{-(m+1)/2}.$$

So the prior is invariant to the transformation of Σ to Σ^{-1}.

There is an alternative way of arriving at the prior probability dist'n. $p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2}$.

Take an (informative) Wishart prior pdf & let the d.f. tend to zero.

i.e. let $p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{v/2} \exp\left\{-\frac{1}{2} tr. \left[(m-v+1)\Lambda \Sigma^{-1}\right]\right\}$

& let $(m-v+1) \to 0$

$\therefore v \to m+1$

$$\therefore p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp\left\{-\frac{1}{2} tr\, 0\right\}$$
$$\propto |\Sigma^{-1}|^{-(m+1)/2}.$$

Now, the likelihood function with which we are working is,

$$l(B, \Sigma | y, X) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} tr.\, S\Sigma^{-1} - \frac{1}{2} tr\left[(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]\right\}$$

& $p(B, \Sigma) \propto |\Sigma|^{-(m+1)/2}$

$$\therefore p(B, \Sigma | y, X) \propto |\Sigma|^{-(n+m+1)/2} * \exp\left\{-\frac{1}{2} tr\, S\Sigma^{-1} - \frac{1}{2} tr (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$$

Now, write $p(B, \Sigma | y, X) = p(B | \Sigma, y, X)\, p(\Sigma | y, X)$

Let $p(B | \Sigma, y, X) \propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} tr (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$

& $p(\Sigma | y, X) \propto |\Sigma|^{-(n-k+m+1)} \exp\left\{-\frac{1}{2} tr\, \Sigma^{-1} S\right\}$

Then the <u>conditional</u> posterior for B is normal in form, while that for Σ (i.e. the <u>marginal</u> posterior for Σ) is in the form of an <u>Inverted Wishart</u> pdf.

Now, $p(B|\Sigma, y, X) \propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} \text{tr.} (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$

$$\propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2}(B-\hat{B})'\Sigma^{-1} \otimes (X'X)(B-\hat{B})\right\}$$

$\therefore$ this normal conditional posterior pdf has mean $\hat{B}$, and c.m. of $\Sigma \otimes (X'X)^{-1}$.

[N.B. $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$.]

Now, what if we are interested in just <u>one</u> of the equations from the system — say the first. Then we just look at β_1.

Then, $p(\beta_1 | \Sigma, y, X) \propto \left(\frac{1}{\sigma_{11}^{k/2}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}(\beta_1-\hat{\beta}_1)'X'X(\beta_1-\hat{\beta}_1)\right\}$

& this is $N(\hat{\beta}_1, (X'X)^{-1}\sigma_{11})$.

Now look at $p(\Sigma|y, X)$ again:

$p(\Sigma|y, X) \propto |\Sigma|^{-(n-k+m+1)} \exp\left\{-\frac{1}{2} \text{tr.} \Sigma^{-1} S\right\}$

Let $\Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right)$

where Σ_{11} is $(p \times p)$ $\ni$ $p < m$.

Then, $p(\Sigma_{11}|y, X) \propto |\Sigma_{11}|^{-[\nu-2(m-p)]/2}$ *

$$\exp\left\{-\frac{1}{2} \text{tr.} \Sigma_{11}^{-1} S_{11}\right\}$$

where $\nu = (n-k+m+1)$.

If $p = 1$,

then $p(\sigma_{11}|y, X) \propto \left[\frac{1}{\sigma_{11}^{[\nu-2(m-1)]/2}}\right] \exp\left\{-\frac{S_{11}}{2\sigma_{11}}\right\}$

And this is just of the "<u>Inverted Gamma</u>" form.

Now, suppose that we are interested in the <u>marginal</u> posterior pdf for β_1, not the conditional posterior pdf.

$p(\beta_1 | \Sigma, y, X)$ depends on σ_{11}, but <u>not</u> on any of the other elements of Σ.

$$\therefore p(\beta_1 | y, X) \propto \int p(\beta_1 | \Sigma, y, X)\, p(\sigma_{11} | y, X)\, d\sigma_{11}$$

$$\propto \int \left(\frac{1}{\sigma_{11}^{k/2}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}(\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right\} *$$

$$\left(\frac{1}{\sigma_{11}^{(\nu-2m+2)/2}}\right) \exp\left\{-\frac{s_{11}}{2\sigma_{11}}\right\} \cdot d\sigma_{11}$$

& $\nu = n-k+m+1$

$$\therefore p(\beta_1 | y, X) \propto \int \left(\frac{1}{\sigma_{11}^{\frac{1}{2}(n-m+3)}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}\left[s_{11} + (\beta_1-\hat{\beta}_1)' * X'X(\beta_1-\hat{\beta}_1)\right]\right\} d\sigma_{11}$$

$$\propto \left[s_{11} + (\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right]^{-1/2(n-m-1)}$$

by virtue of the properties of the t-dist'n.

We now have a M.V. Student t-dist'n.

If $m=1$, then we have just the one eqn. in the system,

so, $p(\beta_1 | y, X, m=1) \propto \left[s_{11} + (\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right]^{-n/2}$

This corresponds to the Chap. 3 result.

Now, what about the posterior pdf for the joint B? (i.e. the joint marginal posterior for B).

Now, $p(B | y, X) = \int p(B, \Sigma | y, X)\, d\Sigma$

$$= \int p(B, \Sigma^{-1} | y, X)\, d\Sigma^{-1}.$$

And, $p(B, \Sigma^{-1} | y, X) \propto |\Sigma^{-1}|^{[n-(m+1)]/2} *$

$$\exp\left\{-\tfrac{1}{2} tr\left[S + (B-\hat{B})' X'X (B-\hat{B})\right]\Sigma^{-1}\right\}$$

$$\therefore p(B \mid y, X) \propto \int |\Sigma^{-1}|^{[n-(m+1)]/2} * \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[S + (B-\hat{B})'X'X * (B-\hat{B})\right]\Sigma^{-1}\right\} d\Sigma^{-1}$$

$$\propto \frac{1}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{n/2}} *$$

$$\int \frac{\exp.\left\{-\tfrac{1}{2}\operatorname{tr}\left[S + (B-\hat{B})'X'X(B-\hat{B})\right]\Sigma^{-1}\right\} * |\Sigma^{-1}|^{(n-m-1)/2}\, d\Sigma^{-1}}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{-n/2}}.$$

$$\propto \frac{1}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{n/2}},$$

since the integral is just the normalizing constant for a Wishart pdf, this constant being indep. of B.

To demonstrate this, the usual form of the Wishart pdf is:

$$p(A \mid \Sigma, \nu, m) = k \frac{|A|^{(\nu-m-1)/2}}{|\Sigma|^{\nu/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$\therefore k = 1 \Big/ \int \frac{|A|^{(\nu-m-1)/2}}{|\Sigma|^{\nu/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$= \int \frac{|\Sigma|^{\nu/2}}{|A|^{(\nu-m-1)/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$= \int \frac{|\Sigma^{-1}|^{-\nu/2}}{|A|^{(\nu-m-1)/2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

Then let $A = \left(S + (B-\hat{B})'X'X(B-\hat{B})\right)$

& let $\nu = -(n-m-1)$

$$\therefore \nu - m - 1 = -n + m + 1 - m - 1 = -n$$

$$\therefore \Rightarrow \int \frac{|\Sigma^{-1}|^{(n-m-1)/2}}{|A|^{-n/2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}.$$

as required.

Further the form of $p(B \mid y, X)$ is a <u>generalized</u> m.v. student pdf. This will be the product of a whole lot of m.v. student pdf's.

(B) <u>Predictive p.d.f. for the Traditional Mult. Regression Model:</u>

Take the M.V.R. model:

$$Y = XB + U.$$

Now we want a <u>predictive</u> p.d.f. on further observations on Y. Let these observations constitute W, a $(p \times m)$ matrix of p <u>additional</u> observations on the m endogenous variables. The same generating model is assumed as previously.

Thus, $W = ZB + V$

where Z is a $(p \times k)$ matrix of p future observations on the k exogenous variables common to each equation, and V is a $(p \times m)$ matrix of p future observations on the m disturbances.

Each row of V is independent, with mean zero and C.M. Σ, the same as that for U.

Then the predictive p.d.f. for W is given by:

$$p(W|Y, X, Z) = \int \cdots \int p(B, \Sigma^{-1}, W|Y, X, Z)\, dB\, d\Sigma^{-1}$$

$$= \int \cdots \int p(W|B, \Sigma^{-1}, Z)\, p(B, \Sigma^{-1}|Y, X)\, dB\, d\Sigma^{-1}$$

Now, $p(W|B, Z, \Sigma^{-1}) \propto |\Sigma^{-1}|^{p/2} \exp\{-\frac{1}{2} \text{tr.} [(W-ZB)'(W-ZB)\Sigma^{-1}]\}$

And, $p(B, \Sigma^{-1}|Y, X) \propto |\Sigma^{-1}|^{(n-m-1)/2} *$

$$\exp\{-\tfrac{1}{2} \text{tr.} [S + (B-\hat{B})'X'X(B-\hat{B})]\Sigma^{-1}\}.$$

$$\therefore p(W|X, Y, Z) \propto \int \cdots \int |\Sigma^{-1}|^{p/2} \cdot |\Sigma^{-1}|^{(n-m-1)/2} *$$

$$\exp\{-\tfrac{1}{2} \text{tr.} [(W-ZB)'(W-ZB)\Sigma^{-1}] - \tfrac{1}{2} \text{tr} [S +$$

$$(B-\hat{B})'X'X(B-\hat{B})]\Sigma^{-1}\}\, dB\, d\Sigma^{-1}$$

$$\propto \int \cdots \int |\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \text{tr}\, A\Sigma^{-1}\}\, dB\, d\Sigma^{-1}$$

where $A = [(w - zB)'(w - zB) + (y - xB)'(y - xB)]$

since $(y - xB)'(y - xB) = s + (B - \hat{B})' X'X (B - \hat{B})$.

So, the integrand in $p(w \mid y, x, z)$ is proportional to:

$$|\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\}$$

And this is in the form of a Wishart pdf. with $\nu = (n+p-m-1)$.

$$\therefore \int \cdots \int |\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\} \, d\Sigma^{-1}$$

$$\propto |A|^{-(n+p)/2}$$

Now we have to integrate $|A|^{-(n+p)/2}$ w.r.t. the elements of B:-

$$A = (w - zB)'(w - zB) + (y - xB)'(y - xB)$$

$$\therefore |A|^{-(n+p)/2} = |(w - zB)'(w - zB) + (y - xB)'(y - xB)|^{-(n+p)/2}.$$

Let $M = (X'X + Z'Z)$

Let $\tilde{B} = M^{-1}(X'y + Z'w)$

Now, $(y - XB)'(y - XB) + (w - ZB)'(w - ZB)$

$= y'y + B'X'XB - 2B'X'y + w'w + B'Z'ZB - 2B'Z'w$

$= y'y + w'w + B'[X'X + Z'Z]B - 2B'[X'y + Z'w]$

Completing the square on B:

$$= [B - (X'X + Z'Z)^{-1}(X'y + Z'w)]'[X'X + Z'Z][B - (X'X + Z'Z)^{-1} * (X'y + Z'w)]$$

$$+ y'y + w'w - [X'y + Z'w]'[X'X + Z'Z]^{-1}[X'y + Z'w]$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - (X'y + Z'w)'M^{-1}(X'y + Z'w) + y'y + w'w$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - \tilde{B}'MM^{-1}M\tilde{B} + y'y + w'w$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - \tilde{B}'M\tilde{B}. + y'y + w'w.$$

$$\therefore |A|^{-(n+p)/2} = \left| (B-\tilde{B})'M(B-\tilde{B}) - \tilde{B}'M\tilde{B} + y'y + w'w \right|^{-(n+p)/2}.$$

$$\therefore p(w|x,y,z) \propto \int \cdots \int \left| y'y + w'w + (B-\tilde{B})'M(B-\tilde{B}) - \tilde{B}'M\tilde{B} \right|^{-(n+p)/2} dB$$

$$\propto \left| y'y + w'w - \tilde{B}'M\tilde{B} \right|^{-(n+p-k)/2}.$$

Now, simplify this expression –

$$(y'y + w'w - \tilde{B}'M\tilde{B})$$

$$= (y'y + w'w) - (y'x + w'z)M^{-1}(x'y + z'w)$$

$$= y'[I - xM^{-1}x']y + w'[I - zM^{-1}z']w - y'xM^{-1}z'w$$

$$- w'zM^{-1}x'y$$

Now complete the square on w:

$$= [w - (I - zM^{-1}z')^{-1} zM^{-1}x'y]'(I - zM^{-1}z')[w - (I - zM^{-1}z')^{-1}$$

$$* zM^{-1}x'y] + y'(I - xM^{-1}x')y \cancel{- y'xM^{-1}}$$

$$- y'xM^{-1}z'[I - zM^{-1}z']^{-1} zM^{-1}x'y$$

Let $C = (I - zM^{-1}z')$

$$\Rightarrow [w - C^{-1}zM^{-1}x'y]'C[w - C^{-1}zM^{-1}x'y]$$

$$+ y'[I - xM^{-1}x' - xM^{-1}z'C\, zM^{-1}x']y$$

Now, $C^{-1} = [I - zM^{-1}z']^{-1} = I + z(x'x)^{-1}z'$.

To show this:

$$[I - zM^{-1}z'][I + z(x'x)^{-1}z']$$

$$= I - z[M^{-1} - (x'x)^{-1}]z' - zM^{-1}z'z(x'x)^{-1}z'$$

$$= I - z[M^{-1} - (x'x)^{-1} + M^{-1}z'z(x'x)^{-1}]z'$$

$$= I - zM^{-1}[I - M(x'x)^{-1} + (z'z)(x'x)^{-1}]z'$$

$$= I - zM^{-1}[(x'x) - M + (z'z)](x'x)^{-1}z'$$

But, $M = (x'x + z'z)$

$\therefore (x'x) - M + (z'z) = 0$

$\therefore [\quad][\quad] = I.$

So, returning to the main theme:

$$C^{-1} Z M^{-1} = [I + Z(X'X)^{-1} Z'] Z M^{-1}$$

$$= Z [I + (X'X)^{-1} Z'Z] M^{-1}$$

$$= Z (X'X)^{-1} [(X'X) + (Z'Z)] M^{-1}$$

$$= Z (X'X)^{-1} M M^{-1}$$

$$= Z (X'X)^{-1}$$

And, further:

$$X M^{-1} X' + X M^{-1} Z' C^{-1} Z M^{-1} X'$$

$$= X [M^{-1} + M^{-1} Z' C^{-1} Z M^{-1}] X'$$

$$= X M^{-1} [I + Z' C^{-1} Z M^{-1}] X'$$

$$= X M^{-1} [I + Z'Z (X'X)^{-1}] X'$$

$$= X M^{-1} [X'X + Z'Z] (X'X)^{-1} X'$$

$$= X M^{-1} M (X'X)^{-1} X'$$

$$= X (X'X)^{-1} X'$$

So:

$$y'y + w'w - \tilde{B}' M \tilde{B}$$

$$= y' [I - X(X'X)^{-1} X'] y + (w - Z(X'X)^{-1} X'y)' C * (w - Z(X'X)^{-1} X'y).$$

Let $\hat{B} = (X'X)^{-1} X'y$

$$\Rightarrow y' [I - X(X'X)^{-1} X'] y + (w - Z\hat{B})' C (w - Z\hat{B})$$

But, $y' [I - X(X'X)^{-1} X'] y = (y - X(X'X)^{-1} X'y)' (y - X(X'X)^{-1} X'y)$

$$= (y - X\hat{B})' (y - X\hat{B})$$

$$= S.$$

So, $y'y + w'w - \tilde{B}' M \tilde{B}$

$$= S + (w - Z\hat{B})' C (w - Z\hat{B}).$$

$$\therefore p(W|X,Y,Z) \propto |S + (W - Z\hat{B})'(I - ZM^{-1}Z')(W - Z\hat{B})|^{-\frac{(n+p-k)}{2}}$$

And this is in the same form as the joint posterior pdf for B in the M.V.R. model:

Recall:

$$p(B|X,Y) \propto |S + (B-\hat{B})'(X'X)(B-\hat{B})|^{-n/2}$$

And this is the "generalized" MVS-t dist'n. So we know that the <u>marginal</u> predictive pdf for any row or column of W will be MVS-t.

Further, if we partition $W = (W_1 \vdots W_2)$, then the marginal predictive pdf for W_1 will also be in the "generalized" MVS-t. form.

(C). The Trad. M.V.R. Model With Exact Restrictions:

Sometimes we will know that some elements of B are zero (eg. some variables do not affect one of the dependent variables) or that $\exists$ an exact linear restriction on B. Then the joint kth pdf for B may be "conditionalized" to take this fact into account. Then the conditionalized posterior pdf may be used to make inferences about the remaining non-zero coefficients in B.

If exact zero restrictions pertain only to one β_i vector in B, say β_1, then this means that <u>some</u> of the variables in X are not relevant to the 1st equation. If all of the restrictions are confined to just β_1 then the MVS posterior pdf for β_1 can be analyzed to obtain a <u>conditional</u> posterior pdf which incorporates the restrictive information. There is no need to modify the posterior pdf's related to any other of the β_i's.

Partition $\beta_1 \ni \beta_1' = (\beta_a' \vdots \beta_0')$, where $\beta_0 = 0$.

Then the posterior pdf —

$p(\beta_a|Y, \beta_0 = 0)$ may be obtained. We can make use of the properties of the MVS distribution to impose other linear constraints on β_1. However, if the constraints relate to more than one column of B,

Then the situation is more complex. In this case we have:

$$y_x = X_{x1}\beta_{x1} + X_{x0}\beta_{x0} + u_x \quad ; x = 1, 2, \ldots, m.$$

where $\beta_{x0} = 0$. Then in this case the partitioning of X is not the same for all equations, & this is the complicating factor.

To tackle this situation we look at the joint posterior pdf for β, expand it, and then conditionalize the leading term in this expansion by setting $\beta_{x0} = 0$; $\forall x$.

Now,

$$p(\beta|y, X) \propto |S + (B-\hat{B})'X'X(B-\hat{B})|^{-n/2}$$

let $\begin{cases} \bar{S} = \frac{1}{n} S \\ M = \frac{1}{n}(X'X) \end{cases}$

$$\therefore p(B|y, X) \propto |\bar{S} + (B-\hat{B})'M(B-\hat{B})|^{-n/2}$$

let H be $\ni$ $H\bar{S}H' = I$

and, $H(B-\hat{B})'M(B-\hat{B})H' = D$,

where $D = D(\lambda_i)$ is a diagonal matrix with positive elements are the c.r. (the λ_i's.) of $(B-\hat{B})'M(B-\hat{B})$.

Now, $H\bar{S}H' = I$

$$\Rightarrow \quad H = (\bar{S}H')^{-1}$$
$$H' = (H\bar{S})^{-1}$$

$$\therefore H'H = (H\bar{S})^{-1}(\bar{S}H')^{-1}$$
$$= [(H\bar{S})^{-1}H'^{-1}]\bar{S}^{-1}$$
$$= \bar{S}^{-1}$$

$$\therefore |S + (B-\hat{B})'M(B-\hat{B})|^{-n/2}$$
$$= (|H'H|^{-1})^{-n/2}\left[|I + H'H(B-\hat{B})'M(B-\hat{B})|\right]^{-n/2}$$
$$= |H'H|^{n/2}|I + H'H(B-\hat{B})'M(B-\hat{B})|^{-n/2}$$

$$= |\bar{S}|^{-n/2} \; |I + H(B-\hat{B})' M (B-\hat{B}) H'|^{-n/2}$$

$$= |\bar{S}|^{-n/2} \exp\{\log |I + D|^{-n/2}\}$$

(where $D = H(B-\hat{B})' M (B-\hat{B}) H'$)

$$= |\bar{S}|^{-n/2} \exp\{-n/2 \log |I + D|\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 \sum_i \log(1+\lambda_i)\}$$

(since $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \therefore (I+D) = \begin{bmatrix} 1+\lambda_1 & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}$

& $\log \prod (1+\lambda_i) = \sum \log(1+\lambda_i)$).

$$\Rightarrow |\bar{S}|^{-n/2} \exp\{-n/2 [\sum_i (\lambda_i - \tfrac{1}{2}\lambda_i^2 + \tfrac{1}{3}\lambda_i^3 - \cdots)]\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 [\sum_i \lambda_i - \tfrac{1}{2}\sum_i \lambda_i^2 + \tfrac{1}{3}\sum_i \lambda_i^3 - \cdots]\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 [\text{tr.}D - \tfrac{1}{2}\text{tr}D^2 + \tfrac{1}{3}\text{tr}D^3 - \cdots]\}.$$

And, $(\text{tr.}D) = \text{tr.}\, H(B-\hat{B})' M (B-\hat{B}) H'$

$$= \text{tr.}\, H'H(B-\hat{B})' M (B-\hat{B})$$

Let $\bar{Q} = (B-\hat{B})' M (B-\hat{B})$

$\therefore \bar{S}^{-1}\bar{Q} = H'H(B-\hat{B})' M (B-\hat{B})$

$\therefore \text{tr.}D = \text{tr.}\, \bar{S}^{-1}\bar{Q}$

$\text{tr.}D^2 = \text{tr.}\, \bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}$

$\text{tr}D^3 = \text{tr.}\, \bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}$ <u>etc.</u>

$\therefore p(B|y) \propto |\bar{S} + (B-\hat{B})' M (B-\hat{B})|^{-n/2}$

$\propto |\bar{S}|^{-n/2} \exp\{-n/2 [\text{tr.}D - \tfrac{1}{2}\text{tr.}D^2 + \tfrac{1}{3}\text{tr.}D^3 - \cdots]\}$

But $M = \frac{1}{n} X'X$

$\bar{Q} = (B-\hat{B})' M (B-\hat{B})$

$= \frac{1}{n}(B-\hat{B})' X'X (B-\hat{B})$

$= \frac{1}{n} Q$, say.

$$\therefore\ p(B|y) \propto \exp\left\{-\tfrac{1}{2}\left[k\ \mathrm{tr.}\ \bar{S}^{-1}Q - \tfrac{1}{2n^2}\ \mathrm{tr.}\bar{S}^{-1}Q\bar{S}^{-1}Q - \cdots\right]\right\}$$

$$\propto \exp\left(-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right)\exp\left\{\tfrac{1}{4n}\ \mathrm{tr.}\ \bar{S}^{-1}Q\bar{S}^{-1}Q - \tfrac{1}{6n^2}\ \mathrm{tr}\,\bar{S}^{-1}Q\bar{S}^{-1}Q\bar{S}^{-1}Q - \cdots\right\}.$$

And $e^x = 1 + x + x^2/2! + \cdots$

$$\text{So, } p(B|y) \propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right] * \left\{1 + \tfrac{1}{4n}\ \mathrm{tr.}\ \bar{S}^{-1}Q\bar{S}^{-1}Q + \cdots\right\}$$

$$\propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right]$$

$$\propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}(B-\hat{B})'X'X(B-\hat{B})\right]$$

$$\propto \exp.\left[-\tfrac{1}{2}(B-\hat{B})'\bar{S}^{-1}\otimes(X'X)\ (B-\hat{B})\right]$$

Let $\beta_{\cdot 1}$ denote the non-zero coefficients
" $\beta_{\cdot 0}$ " " zero "
" $\bar{S}^{\alpha\ell}$ be the (α,ℓ)th. element of $\bar{S}^{-1}$
" $d' = \left[(\beta_{11}-\hat{\beta}_{11})', -\hat{\beta}_{10}', \cdots, (\beta_{m1}-\hat{\beta}_{m1})', -\hat{\beta}_{m0}'\right]$

i.e. this is the $(\beta-\hat{\beta})$, conditionalized $\ni$ the zero restrictions are incorporated.

Then, $X = (X_1 : X_0)$; $(X'X) = (X_1 : X_0)'(X_1 : X_0)$

$\therefore\ (X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})$ is the (α,ℓ)th. el. of $(X'X)$

So, $\left\{\bar{S}^{\alpha\ell}(X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})\right\}$ is a typical element of the partitioned matrix.

$$\therefore\ p(\beta_{\cdot 1}|y, \beta_{\cdot 0}=0) \propto \exp\left\{-\tfrac{1}{2}d'\left[\bar{S}^{\alpha\ell}(X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})\right] * d\right\}$$

$$\cancel{\propto \exp\left\{-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})'(X_{\alpha 1} : X_{\ell 1})^{\prime}\right.}$$

$$\propto \exp\left\{-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})'(X_{\alpha 1}X_{\ell 1}\bar{S}^{\alpha\ell})(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})\right.$$
$$\left. + \hat{\beta}_{\cdot 0}'(X_{\alpha 0}'X_{\ell 0}\bar{S}^{\alpha\ell})\hat{\beta}_{\cdot 0} - 2\hat{\beta}_{\cdot 0}'(X_{\alpha 0}'X_{\ell 1}\bar{S}^{\alpha\ell})(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})\right\}$$

Let $V = (X_{\alpha 1}'X_{\ell 1}\bar{S}^{\alpha\ell})$; $R = (X_{\alpha 0}'X_{\ell 1}\bar{S}^{\alpha\ell})$

$$\therefore\ p(\beta_{\cdot 1}|y, \beta_{\cdot 0}) \propto \exp\left[-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1}-V^{-1}R'\hat{\beta}_{\cdot 0})' * V(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1}-V^{-1}R'\hat{\beta}_{\cdot 0})\right]$$

And this is approximately a <u>normal</u> pdf. i.e. we have a <u>conditionalized</u> posterior pdf with mean $(\hat{\beta}_{\cdot 1} + V^{-1}R'\hat{\beta}_{\cdot 0})$ and c.m. of

$$V^{-1} = (X_{\alpha 1}' X_{\alpha 1} \bar{\sigma}^{\alpha\alpha})^{-1}.$$

(D) <u>Traditional Model with an Informative Prior p.d.f.:</u>

Now we wish to introduce <u>informative</u> prior information about B. If this information is reasonably accurate then we should be able to improve the precision of our inferences by using it in favour of a diffuse prior. Also, if we compare the prior and posterior pdf's, we shall be able to discern to what extent the sample information has modified our original beliefs about the model.

Now, if we were to use a simple natural conjugate prior pdf on the traditional M.V. regression model, then we would have to place restrictions on the variances and covariances appearing in the equations of the system. This is because $(X'X)^{-1}$ enters the covariance structure via $\Sigma \otimes (X'X)^{-1}$. Then if we used a simple natural conjugate prior, we would find for example that the ratios of variances of corresponding coefficients in equations 1 and 2 would be equal. This problem is avoided if we use a general MVN pdf as our prior for all of the coeff. in the model. But the price that we pay then is a loss of easy analysis — the prior pdf will no longer combine neatly with the LIF.

So, our prior pdf is —

$$p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp\left\{-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right\}$$

where $\bar{\beta}$ is an $(mk \times 1)$ vector, the mean of the prior pdf, assigned by the analyst.

C is an $(mk \times mk)$ prior C.M, whose value is also assigned by the user. (N.B. C is <u>constant</u>, so the leading term in $p(\beta, \Sigma^{-1})$ involving C is absorbed into the proportionality sign.).

The usual diffuse prior is used for Σ^{-1} in the above expression.

Now,

$$l(B, \Sigma \mid y, X) \propto |\Sigma|^{-n/2} \exp\left[-\tfrac{1}{2}\operatorname{tr} S\Sigma^{-1} - \tfrac{1}{2}\operatorname{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]$$

$$\propto |\Sigma^{-1}|^{N/2} \exp\left[-\tfrac{1}{2}\operatorname{tr} S\Sigma^{-1} - \tfrac{1}{2}\operatorname{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]$$

So,

$$p(B, \Sigma^{-1} | y, X) \propto |\Sigma^{-1}|^{(n-m-1)/2} * \exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right] *$$
$$\exp\left[-\tfrac{1}{2} tr\, \Sigma^{-1}\left[S + (B-\hat{B})' X' X (B-\hat{B})\right]\right]$$

where $\hat{B} = (X'X)^{-1} X' y$.

Now, integrate w.r.t. Σ^{-1}, & obtain —

$$p(B | y, X) \propto \exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right] * \left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-\frac{n}{2}}$$

And this is the product of a MVN pdf. and a <u>generalized</u> MVS pdf. This is rather messy to handle as it stands. One way of getting around the analytical problem is to expand the determinant as we did before, & normalize on the first term so obtained —

$$\left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-n/2}$$
$$= |\bar{S}|^{-n/2} \exp\left[-\tfrac{n}{2}\left[tr.D - \tfrac{1}{2} tr.D^2 + \tfrac{1}{3} tr.D^3 - \cdots\cdots\right]\right]$$

where $D = \frac{1}{n} H (B-\hat{B})' X' X (B-\hat{B}) H'$

& $H \bar{S} H' = I$, & $\bar{S} = \frac{1}{n} S$.

$$\therefore \left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-n/2}$$
$$\doteq \exp\left[-\tfrac{1}{2}(\beta-\hat{\beta})' \bar{S}^{-1} \otimes (X'X) (\beta-\hat{\beta})\right]$$

$$\therefore p(\beta | y, X) \propto \exp\left[-\tfrac{1}{2}(\beta-\hat{\beta})' \bar{S}^{-1} \otimes (X'X)(\beta-\hat{\beta})\right] *$$
$$\exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right]$$
$$\propto \exp\left[-\tfrac{1}{2}(\beta-b)' F (\beta-b)\right]$$

where $F = \left(C^{-1} + \bar{S}^{-1} \otimes (X'X)\right)$

$$b = \left(C^{-1} + \bar{S}^{-1} \otimes (X'X)\right)^{-1} \left(C^{-1}\bar{\beta} + \bar{S}^{-1} \otimes X'X\hat{\beta}\right)$$

Then b is the mean of the leading normal term of the expansion, & F^{-1} is the CM of the leading term approximating the posterior pdf of B.

(E) The "Seemingly Unrelated" Regression Model:

In a sense, this is a generalization of the M.V. regression model, in that the matrix X appearing in each equation of the traditional model is now allowed to be different from equation to equation. viz:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}$$

y_α is an $(n \times 1)$ vector of obs. on the αth. dep. var.

X_α " " $(n \times k_\alpha)$ matrix " " " " k_α indep. var.

β_α " " $(k_\alpha \times 1)$ vector " " " coeff.

u_α " " $(n \times 1)$ " " " " " disturbances.

$$\therefore \quad y = Z\beta + u \quad , \text{ say.}$$

where Z is the <u>block-diagonal</u> matrix shown above.

$$\& \quad E(u) = 0$$
$$E(u,u') = \Sigma \otimes I_n.$$

And Σ is $(m \times m)$ and pos. def. symm.

Then,

$$\ell(\beta, \Sigma | y) \propto |\Sigma^{-1}|^{n/2} \exp\left[-\tfrac{1}{2}(y - Z\beta)' \Sigma^{-1} \otimes I_n (y - Z\beta)\right]$$
$$\propto |\Sigma^{-1}|^{n/2} \exp\left[-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\right]$$

where, $$A = \begin{bmatrix} (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & \cdots & (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) \\ \vdots & & \vdots \\ (y_m - X_m\beta_m)'(y_1 - X_1\beta_1) & \cdots & (y_m - X_m\beta_m)'(y_m - X_m\beta_m) \end{bmatrix}$$

And assume that $p(\beta, \Sigma^{-1}) = p(\beta)\, p(\Sigma^{-1})$.

$\therefore\ p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2}$

$\therefore\ p(\beta, \Sigma^{-1}|y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\left[-\frac{1}{2}\operatorname{tr} A\Sigma^{-1}\right]$

Then $p(\beta|\Sigma^{-1}, y)$ is MVN —

$p(\beta|\Sigma^{-1}, y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\left[-\frac{1}{2}(y-Z\beta)'\Sigma^{-1}\otimes I_n(y-Z\beta)\right]$

$\therefore\ E[\beta|\Sigma^{-1}, y] = (Z'(\Sigma^{-1}\otimes I_n)Z)^{-1} Z'(\Sigma^{-1}\otimes I_n)y = \hat{\beta}$

& $\text{Cov.}(\beta|\Sigma^{-1}, y) = (Z'(\Sigma^{-1}\otimes I_n)Z)^{-1}$

And the <u>conditional</u> posterior mean of β, <u>given</u> Σ^{-1} is just the sampling theory GLS estimator. And the same applies to the CM above. Also, under the assumption of normality, as is expressed by the above L.F., $E(\beta|\Sigma^{-1}, y) = \hat{\beta}$ is a MLE of β.

Now: (i) If all of the X's are the same, or if they are proportional to one another, then $\hat{\beta}$ reduces to a vector of O.L.S. estimators.

and/or: (ii) If Σ is diagonal, then again $\hat{\beta}$ reduces to a vector of O.L.S. estimators.

Then $\hat{\beta}_\alpha = (X_\alpha' X_\alpha)^{-1} X_\alpha' y_\alpha$; $\alpha = 1, 2, ---, m$.

Now, the sampling theory estimator of β is equivalent to the mean of the <u>conditional</u> posterior pdf for β, given Σ^{-1}. Now, in the sampling theory approach, β^* is independent of Σ^{-1}, which is usually unknown & is replaced by $\hat{\Sigma}^{-1}$, formed from the residuals of the equations as estimated by O.L.S.

So then, in the sampling theory case we get:

$$\beta^* = [Z'(\hat{\Sigma}^{-1}\otimes I_n)Z]^{-1}[Z'(\hat{\Sigma}^{-1}\otimes I_n)y]$$

And note that β^* has the same <u>large sample</u> properties as has $\hat{\beta}$ in the Bayes case.

So β^* is what we would obtain in the Bayes case if we proceeded with $\Sigma = \hat{\Sigma}$. In <u>large samples</u>, if $\hat{\Sigma}$ is consistent, then Σ & $\hat{\Sigma}$ are little different, so the assumption that $\Sigma = \hat{\Sigma}$ will produce quite satisfactory results. However, in <u>small</u> samples,

things will not be so satisfactory. Then it is better to use the <u>marginal</u> posterior pdf for β and take <u>its</u> mean as the estimator of β. Note that by adopting a Bayes technique, we have an advantage over the sampling theorist when it comes to <u>small</u> samples, in this particular case. We can get an <u>exact</u> small-sample result, whereas he is unable to do so.

Now,

$$p(\beta, \Sigma^{-1} | y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\}$$

$$\therefore\ p(\beta | y) \propto \int \cdots \int |\Sigma^{-1}|^{(n-m-1)/2} \exp[-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}]\, d\Sigma^{-1}$$

$$\propto |A|^{-n/2}$$

$$\therefore\ p(\beta | y) \propto \begin{vmatrix} (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & \cdots & (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) \\ \vdots & & \vdots \\ (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) & \cdots & (y_m - X_m\beta_m)'(y_m - X_m\beta_m) \end{vmatrix}^{-\frac{n}{2}}$$

Now, the <u>marginal</u> posterior pdf for β vaguely resembles a <u>generalized</u> MVS pdf, but in fact it <u>cannot</u> be brought exactly into that form, because not all of the X's are the same. For $p(\beta|y)$ in the form showing we don't have any practical way of analyzing this pdf.

An alternative way of viewing the SUR ~~model~~ model is to write it as a "restricted" traditional M.V.R. model. <u>i.e.</u>

$$(y_1, y_2, \cdots, y_m) = (X_1\ X_2 \cdots X_m) \begin{pmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_m \end{pmatrix} + (u_1 \cdots u_m)$$

Then the (zero) restrictions on the coefficient matrix appear quite explicitly. If $(X_1 \cdots X_m)$ has full rank, then we can use the methods of section (C) above to handle the zero restrictions, as for the traditional model.

Chapter 8: Multivariate Regression Models.

(A) The Multivariate Regression Model :-

$$Y = (y_1, ----, y_m) = \begin{pmatrix} y_{11} & --- & y_{1m} \\ \vdots & & \vdots \\ y_{n1} & --- & y_{nm} \end{pmatrix}$$

i.e. n observations on m equations.

$$\therefore \quad Y = XB + U$$

where X is $(n \times k)$ and of rank k,

$$B = (\beta_1, ----, \beta_m)$$

$$= \begin{pmatrix} \beta_{11} & --- & \beta_{1m} \\ \vdots & & \vdots \\ \beta_{k1} & --- & \beta_{km} \end{pmatrix}$$

i.e. k exogenous variables per equation — & these are the <u>same</u> variables in each equation.

$$U = (u_1, ---, u_m) = \begin{pmatrix} u_{11} & ---- & u_{1m} \\ \vdots & & \vdots \\ u_{n1} & --- & u_{nm} \end{pmatrix}$$

i.e. n observations on each of m disturbances.

Let the rows of U be independently distributed. Then there is no time dependence and so no serial correlation.

Let each row be distributed $N(0, \Sigma)$, where Σ is an $(m \times m)$ covariance matrix.

Then,

$$p(Y \mid X, B, \Sigma) \propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2} tr\left[(Y-XB)'(Y-XB)\Sigma^{-1}\right]\right\}$$

But, $(Y-XB)'(Y-XB)$

$$= (Y-X\hat{B})'(Y-X\hat{B}) + (B-\hat{B})'X'X(B-\hat{B})$$

$$= S + (B-\hat{B})'X'X(B-\hat{B})$$

where $S = (Y-X\hat{B})'(Y-X\hat{B})$

and $\hat{B} = (X'X)^{-1}X'Y$.

$$\therefore\ \ell(B,\Sigma \mid Y,X) \propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\left[S + (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]\right.$$

$$\propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\, S\Sigma^{-1} - \tfrac{1}{2}\,\mathrm{tr}\,(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$$

Now, Σ has $m(m+1)/2$ distinct elements —

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix} \qquad \& \ \sigma_{ij} = \sigma_{ji} \quad \forall\, i \neq j.$$

~~And $\exists$ m elements on the diagonal. Ignoring these $\exists$ $\frac{1}{2}m-1$~~

& no. of distinct elements $= m + (m-1) + (m-2) + \cdots + 2 + 1$
$= $ sum of A.P.
$= \frac{1}{2}m(m+1)$.

Now, take a diffuse prior pdf — then
$p(B,\Sigma) = p(B)p(\Sigma)$

& $p(B) \propto$ const.

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$$

$$\therefore\ p(B,\Sigma) \propto |\Sigma|^{-(m+1)/2}.$$

Note that if $m=1$, then $p(\Sigma) \propto |\Sigma|^{-(m+1)/2}$ reduces to

$p(\sigma_{11}) \propto \left(\frac{1}{\sigma_{11}}\right)$, as is usual for the Jeffreys' diffuse prior.

Let $\sigma^{\mu\mu'}$ be the (μ,μ')th. element of Σ^{-1}.

Then transform $\Sigma \to \Sigma^{-1}$.

$$\therefore\ (\sigma_{11}, \sigma_{12}, \ldots, \sigma_{mm}) \to (\sigma^{11}, \sigma^{12}, \ldots, \sigma^{mm}).$$

Which is a transformation involving $\frac{1}{2}m(m+1)$ variables.

Then $J = \left|\dfrac{\partial(\sigma_{11}, \sigma_{12}, \ldots \sigma_{mm})}{\partial(\sigma^{11}, \sigma^{12}, \ldots \sigma^{mm})}\right| = |\Sigma|^{m+1}$

* For this see the separate notes from Anderson's book.

Now, $p(B, \Sigma) \propto |\Sigma|^{-(m+1)/2}$

& $J = |\Sigma|^{m+1}$

$$\therefore p(B, \Sigma^{-1}) \propto |\Sigma|^{m+1} \cdot |\Sigma|^{-(m+1)/2}$$
$$\propto |\Sigma|^{(m+1)/2}$$
$$\propto |\Sigma^{-1}|^{-(m+1)/2}.$$

So the prior is invariant to the transformation of Σ to Σ^{-1}.

There is an alternative way of arriving at the prior probability dist'n. $p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2}$.

Take an (informative) Wishart prior pdf & let the d.f. tend to zero.

i.e. let $p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{v/2} \exp\left\{-\frac{1}{2} tr. \left[(m-v+1)\Lambda \Sigma^{-1}\right]\right\}$

& let $(m-v+1) \to 0$

$\therefore v \to m+1$

$$\therefore p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp\left\{-\frac{1}{2} tr\, 0\right\}$$
$$\propto |\Sigma^{-1}|^{-(m+1)/2}.$$

Now, the likelihood function with which we are working is,

$$l(B, \Sigma | y, X) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} tr.\, S\Sigma^{-1} - \frac{1}{2} tr\left[(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]\right\}$$

& $p(B, \Sigma) \propto |\Sigma|^{-(m+1)/2}$

$$\therefore p(B, \Sigma | y, X) \propto |\Sigma|^{-(n+m+1)/2} * \exp\left\{-\frac{1}{2} tr\, S\Sigma^{-1} - \frac{1}{2} tr (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$$

Now, write $p(B, \Sigma | y, X) = p(B | \Sigma, y, X)\, p(\Sigma | y, X)$

Let $p(B | \Sigma, y, X) \propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} tr (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$

& $p(\Sigma | y, X) \propto |\Sigma|^{-(n-k+m+1)} \exp\left\{-\frac{1}{2} tr\, \Sigma^{-1} S\right\}$

Then the <u>conditional</u> posterior for B is normal in form, while that for Σ (i.e. the <u>marginal</u> posterior for Σ) is in the form of an <u>Inverted Wishart</u> pdf.

Now, $p(B|\Sigma, y, X) \propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} \text{tr.} (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right\}$

$$\propto |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2}(B-\hat{B})'\Sigma^{-1} \otimes (X'X)(B-\hat{B})\right\}$$

$\therefore$ this normal conditional posterior pdf has mean $\hat{B}$, and c.m. of $\Sigma \otimes (X'X)^{-1}$.

[N.B. $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$.]

Now, what if we are interested in just <u>one</u> of the equations from the system — say the first. Then we just look at β_1.

Then, $p(\beta_1 | \Sigma, y, X) \propto \left(\frac{1}{\sigma_{11}^{k/2}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}(\beta_1-\hat{\beta}_1)'X'X(\beta_1-\hat{\beta}_1)\right\}$

& this is $N(\hat{\beta}_1, (X'X)^{-1}\sigma_{11})$.

Now look at $p(\Sigma|y, X)$ again:

$p(\Sigma|y, X) \propto |\Sigma|^{-(n-k+m+1)} \exp\left\{-\frac{1}{2} \text{tr.} \Sigma^{-1} S\right\}$

Let $\Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right)$

where Σ_{11} is $(p \times p)$ $\ni$ $p < m$.

Then, $p(\Sigma_{11}|y, X) \propto |\Sigma_{11}|^{-[\nu-2(m-p)]/2}$ *

$$\exp\left\{-\frac{1}{2} \text{tr.} \Sigma_{11}^{-1} S_{11}\right\}$$

where $\nu = (n-k+m+1)$.

If $p = 1$,

then $p(\sigma_{11}|y, X) \propto \left[\frac{1}{\sigma_{11}^{[\nu-2(m-1)]/2}}\right] \exp\left\{-\frac{S_{11}}{2\sigma_{11}}\right\}$

And this is just of the "<u>Inverted Gamma</u>" form.

Now, suppose that we are interested in the <u>marginal</u> posterior pdf for β_1, not the conditional posterior pdf.

$p(\beta_1 | \Sigma, y, X)$ depends on σ_{11}, but <u>not</u> on any of the other elements of Σ.

$$\therefore p(\beta_1 | y, X) \propto \int p(\beta_1 | \Sigma, y, X)\, p(\sigma_{11} | y, X)\, d\sigma_{11}$$

$$\propto \int \left(\frac{1}{\sigma_{11}^{k/2}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}(\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right\} *$$

$$\left(\frac{1}{\sigma_{11}^{(\nu-2m+2)/2}}\right) \exp\left\{-\frac{s_{11}}{2\sigma_{11}}\right\} \cdot d\sigma_{11}$$

& $\nu = n-k+m+1$

$$\therefore p(\beta_1 | y, X) \propto \int \left(\frac{1}{\sigma_{11}^{\frac{1}{2}(n-m+3)}}\right) \exp\left\{-\frac{1}{2\sigma_{11}}\left[s_{11} + (\beta_1-\hat{\beta}_1)' * X'X(\beta_1-\hat{\beta}_1)\right]\right\} d\sigma_{11}$$

$$\propto \left[s_{11} + (\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right]^{-1/2(n-m-1)}$$

by virtue of the properties of the t-dist'n.

We now have a M.V. Student t-dist'n.

If $m=1$, then we have just the one eqn. in the system,

so, $p(\beta_1 | y, X, m=1) \propto \left[s_{11} + (\beta_1-\hat{\beta}_1)' X'X (\beta_1-\hat{\beta}_1)\right]^{-n/2}$

This corresponds to the Chap. 3 result.

Now, what about the posterior pdf for the joint B? (i.e. the joint marginal posterior for B).

Now, $p(B | y, X) = \int p(B, \Sigma | y, X)\, d\Sigma$

$$= \int p(B, \Sigma^{-1} | y, X)\, d\Sigma^{-1}.$$

And, $p(B, \Sigma^{-1} | y, X) \propto |\Sigma^{-1}|^{[n-(m+1)]/2} *$

$$\exp\left\{-\tfrac{1}{2} tr\left[S + (B-\hat{B})' X'X (B-\hat{B})\right]\Sigma^{-1}\right\}$$

$$\therefore p(B \mid y, X) \propto \int |\Sigma^{-1}|^{[n-(m+1)]/2} * \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[S + (B-\hat{B})'X'X * (B-\hat{B})\right]\Sigma^{-1}\right\} d\Sigma^{-1}$$

$$\propto \frac{1}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{n/2}} *$$

$$\int \frac{\exp.\left\{-\tfrac{1}{2}\operatorname{tr}\left[S + (B-\hat{B})'X'X(B-\hat{B})\right]\Sigma^{-1}\right\} * |\Sigma^{-1}|^{(n-m-1)/2}\, d\Sigma^{-1}}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{-n/2}}.$$

$$\propto \frac{1}{|S + (B-\hat{B})'X'X(B-\hat{B})|^{n/2}},$$

since the integral is just the normalizing constant for a Wishart pdf, this constant being indep. of B.

To demonstrate this, the usual form of the Wishart pdf is:

$$p(A \mid \Sigma, \nu, m) = k \frac{|A|^{(\nu-m-1)/2}}{|\Sigma|^{\nu/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$\therefore k = 1 \Big/ \int \frac{|A|^{(\nu-m-1)/2}}{|\Sigma|^{\nu/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$= \int \frac{|\Sigma|^{\nu/2}}{|A|^{(\nu-m-1)/2}} \cdot \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

$$= \int \frac{|\Sigma^{-1}|^{-\nu/2}}{|A|^{(\nu-m-1)/2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}$$

Then let $A = \left(S + (B-\hat{B})'X'X(B-\hat{B})\right)$

& let $\nu = -(n-m-1)$

$$\therefore \nu - m - 1 = -n + m + 1 - m - 1 = -n$$

$$\therefore \Rightarrow \int \frac{|\Sigma^{-1}|^{(n-m-1)/2}}{|A|^{-n/2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}A\right\}.$$

as required.

Further the form of $p(B \mid y, X)$ is a <u>generalized</u> m.v. student pdf. This will be the product of a whole lot of m.v. student pdf's.

(B) <u>Predictive p.d.f. for the Traditional Mult. Regression Model:</u>

Take the M.V.R. model:

$$Y = XB + U.$$

Now we want a <u>predictive</u> p.d.f. on further observations on Y. Let these observations constitute W, a $(p \times m)$ matrix of p <u>additional</u> observations on the m endogenous variables. The same generating model is assumed as previously.

Thus, $W = ZB + V$

where Z is a $(p \times k)$ matrix of p future observations on the k exogenous variables common to each equation, and V is a $(p \times m)$ matrix of p future observations on the m disturbances.

Each row of V is independent, with mean zero and C.M. Σ, the same as that for U.

Then the predictive p.d.f. for W is given by:

$$p(W|Y, X, Z) = \int \cdots \int p(B, \Sigma^{-1}, W|Y, X, Z)\, dB\, d\Sigma^{-1}$$

$$= \int \cdots \int p(W|B, \Sigma^{-1}, Z)\, p(B, \Sigma^{-1}|Y, X)\, dB\, d\Sigma^{-1}$$

Now, $p(W|B, Z, \Sigma^{-1}) \propto |\Sigma^{-1}|^{p/2} \exp\{-\frac{1}{2} \text{tr.} [(W-ZB)'(W-ZB)\Sigma^{-1}]\}$

And, $p(B, \Sigma^{-1}|Y, X) \propto |\Sigma^{-1}|^{(n-m-1)/2} *$

$$\exp\{-\tfrac{1}{2} \text{tr.} [S + (B-\hat{B})'X'X(B-\hat{B})]\Sigma^{-1}\}.$$

$$\therefore p(W|X, Y, Z) \propto \int \cdots \int |\Sigma^{-1}|^{p/2} \cdot |\Sigma^{-1}|^{(n-m-1)/2} *$$

$$\exp\{-\tfrac{1}{2} \text{tr.} [(W-ZB)'(W-ZB)\Sigma^{-1}] - \tfrac{1}{2} \text{tr} [S +$$

$$(B-\hat{B})'X'X(B-\hat{B})]\Sigma^{-1}\}\, dB\, d\Sigma^{-1}$$

$$\propto \int \cdots \int |\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \text{tr}\, A\Sigma^{-1}\}\, dB\, d\Sigma^{-1}$$

where $A = [(w - zB)'(w - zB) + (y - xB)'(y - xB)]$

since $(y - xB)'(y - xB) = s + (B - \hat{B})' X'X (B - \hat{B})$.

So, the integrand in $p(w \mid y, x, z)$ is proportional to:

$$|\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\}$$

And this is in the form of a Wishart pdf. with $\nu = (n+p-m-1)$.

$$\therefore \int \cdots \int |\Sigma^{-1}|^{(n+p-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\} \, d\Sigma^{-1}$$

$$\propto |A|^{-(n+p)/2}$$

Now we have to integrate $|A|^{-(n+p)/2}$ w.r.t. the elements of B:-

$$A = (w - zB)'(w - zB) + (y - xB)'(y - xB)$$

$$\therefore |A|^{-(n+p)/2} = |(w - zB)'(w - zB) + (y - xB)'(y - xB)|^{-(n+p)/2}.$$

Let $M = (X'X + Z'Z)$

Let $\tilde{B} = M^{-1}(X'y + Z'w)$

Now, $(y - XB)'(y - XB) + (w - ZB)'(w - ZB)$

$= y'y + B'X'XB - 2B'X'y + w'w + B'Z'ZB - 2B'Z'w$

$= y'y + w'w + B'[X'X + Z'Z]B - 2B'[X'y + Z'w]$

Completing the square on B:

$$= [B - (X'X + Z'Z)^{-1}(X'y + Z'w)]'[X'X + Z'Z][B - (X'X + Z'Z)^{-1} * (X'y + Z'w)]$$

$$+ y'y + w'w - [X'y + Z'w]'[X'X + Z'Z]^{-1}[X'y + Z'w]$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - (X'y + Z'w)'M^{-1}(X'y + Z'w) + y'y + w'w$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - \tilde{B}'MM^{-1}M\tilde{B} + y'y + w'w$$

$$= (B - \tilde{B})'M(B - \tilde{B}) - \tilde{B}'M\tilde{B}. + y'y + w'w.$$

$$\therefore |A|^{-(n+p)/2} = \left| (B-\tilde{B})'M(B-\tilde{B}) - \tilde{B}'M\tilde{B} + y'y + w'w \right|^{-(n+p)/2}.$$

$$\therefore p(w|x,y,z) \propto \int \cdots \int \left| y'y + w'w + (B-\tilde{B})'M(B-\tilde{B}) - \tilde{B}'M\tilde{B} \right|^{-(n+p)/2} dB$$

$$\propto \left| y'y + w'w - \tilde{B}'M\tilde{B} \right|^{-(n+p-k)/2}.$$

Now, simplify this expression –

$$(y'y + w'w - \tilde{B}'M\tilde{B})$$

$$= (y'y + w'w) - (y'x + w'z)M^{-1}(x'y + z'w)$$

$$= y'[I - xM^{-1}x']y + w'[I - zM^{-1}z']w - y'xM^{-1}z'w$$

$$- w'zM^{-1}x'y$$

Now complete the square on w:

$$= [w - (I - zM^{-1}z')^{-1} zM^{-1}x'y]'(I - zM^{-1}z')[w - (I - zM^{-1}z')^{-1}$$

$$* zM^{-1}x'y] + y'(I - xM^{-1}x')y \cancel{- y'xM^{-1}}$$

$$- y'xM^{-1}z'[I - zM^{-1}z']^{-1} zM^{-1}x'y$$

Let $C = (I - zM^{-1}z')$

$$\Rightarrow [w - C^{-1}zM^{-1}x'y]'C[w - C^{-1}zM^{-1}x'y]$$

$$+ y'[I - xM^{-1}x' - xM^{-1}z'C\, zM^{-1}x']y$$

Now, $C^{-1} = [I - zM^{-1}z']^{-1} = I + z(x'x)^{-1}z'$.

To show this:

$$[I - zM^{-1}z'][I + z(x'x)^{-1}z']$$

$$= I - z[M^{-1} - (x'x)^{-1}]z' - zM^{-1}z'z(x'x)^{-1}z'$$

$$= I - z[M^{-1} - (x'x)^{-1} + M^{-1}z'z(x'x)^{-1}]z'$$

$$= I - zM^{-1}[I - M(x'x)^{-1} + (z'z)(x'x)^{-1}]z'$$

$$= I - zM^{-1}[(x'x) - M + (z'z)](x'x)^{-1}z'$$

But, $M = (x'x + z'z)$

$\therefore (x'x) - M + (z'z) = 0$

$\therefore [\quad][\quad] = I.$

So, returning to the main theme:

$$C^{-1} Z M^{-1} = [I + Z(X'X)^{-1} Z'] Z M^{-1}$$

$$= Z [I + (X'X)^{-1} Z'Z] M^{-1}$$

$$= Z (X'X)^{-1} [(X'X) + (Z'Z)] M^{-1}$$

$$= Z (X'X)^{-1} M M^{-1}$$

$$= Z (X'X)^{-1}$$

And, further:

$$X M^{-1} X' + X M^{-1} Z' C^{-1} Z M^{-1} X'$$

$$= X [M^{-1} + M^{-1} Z' C^{-1} Z M^{-1}] X'$$

$$= X M^{-1} [I + Z' C^{-1} Z M^{-1}] X'$$

$$= X M^{-1} [I + Z'Z (X'X)^{-1}] X'$$

$$= X M^{-1} [X'X + Z'Z] (X'X)^{-1} X'$$

$$= X M^{-1} M (X'X)^{-1} X'$$

$$= X (X'X)^{-1} X'$$

So:

$$y'y + w'w - \tilde{B}' M \tilde{B}$$

$$= y' [I - X(X'X)^{-1} X'] y + (w - Z(X'X)^{-1} X'y)' C * (w - Z(X'X)^{-1} X'y).$$

Let $\hat{B} = (X'X)^{-1} X'y$

$$\Rightarrow y' [I - X(X'X)^{-1} X'] y + (w - Z\hat{B})' C (w - Z\hat{B})$$

But, $y' [I - X(X'X)^{-1} X'] y = (y - X(X'X)^{-1} X'y)' (y - X(X'X)^{-1} X'y)$

$$= (y - X\hat{B})' (y - X\hat{B})$$

$$= S.$$

So, $y'y + w'w - \tilde{B}' M \tilde{B}$

$$= S + (w - Z\hat{B})' C (w - Z\hat{B}).$$

$$\therefore p(W|X,Y,Z) \propto |S + (W - Z\hat{B})'(I - ZM^{-1}Z')(W - Z\hat{B})|^{-\frac{(n+p-k)}{2}}$$

And this is in the same form as the joint posterior pdf for B in the M.V.R. model:

Recall:

$$p(B|X,Y) \propto |S + (B-\hat{B})'(X'X)(B-\hat{B})|^{-n/2}$$

And this is the "generalized" MVS-t dist'n. So we know that the <u>marginal</u> predictive pdf for any row or column of W will be MVS-t.

Further, if we partition $W = (W_1 \vdots W_2)$, then the marginal predictive pdf for W_1 will also be in the "generalized" MVS-t. form.

(C). The Trad. M.V.R. Model With Exact Restrictions:

Sometimes we will know that some elements of B are zero (eg. some variables do not affect one of the dependent variables) or that $\exists$ an exact linear restriction on B. Then the joint kth pdf for B may be "conditionalized" to take this fact into account. Then the conditionalized posterior pdf may be used to make inferences about the remaining non-zero coefficients in B.

If exact zero restrictions pertain only to one β_i vector in B, say β_1, then this means that <u>some</u> of the variables in X are not relevant to the 1st equation. If all of the restrictions are confined to just β_1 then the MVS posterior pdf for β_1 can be analyzed to obtain a <u>conditional</u> posterior pdf which incorporates the restrictive information. There is no need to modify the posterior pdf's related to any other of the β_i's.

Partition $\beta_1 \ni \beta_1' = (\beta_a' \vdots \beta_0')$, where $\beta_0 = 0$.

Then the posterior pdf —

$p(\beta_a|Y, \beta_0 = 0)$ may be obtained. We can make use of the properties of the MVS distribution to impose other linear constraints on β_1. However, if the constraints relate to more than one column of B,

Then the situation is more complex. In this case we have:

$$y_x = X_{x1}\beta_{x1} + X_{x0}\beta_{x0} + u_x \quad ; x = 1, 2, \ldots, m.$$

where $\beta_{x0} = 0$. Then in this case the partitioning of X is not the same for all equations, & this is the complicating factor.

To tackle this situation we look at the joint posterior pdf for β, expand it, and then conditionalize the leading term in this expansion by setting $\beta_{x0} = 0$; $\forall x$.

Now,

$$p(\beta|y, X) \propto |S + (B-\hat{B})'X'X(B-\hat{B})|^{-n/2}$$

let $\begin{cases} \bar{S} = \frac{1}{n} S \\ M = \frac{1}{n}(X'X) \end{cases}$

$$\therefore p(B|y, X) \propto |\bar{S} + (B-\hat{B})'M(B-\hat{B})|^{-n/2}$$

let H be $\ni$ $H\bar{S}H' = I$

and, $H(B-\hat{B})'M(B-\hat{B})H' = D$,

where $D = D(\lambda_i)$ is a diagonal matrix with positive elements are the c.r. (the λ_i's.) of $(B-\hat{B})'M(B-\hat{B})$.

Now, $H\bar{S}H' = I$

$$\Rightarrow \quad H = (\bar{S}H')^{-1}$$
$$H' = (H\bar{S})^{-1}$$

$$\therefore H'H = (H\bar{S})^{-1}(\bar{S}H')^{-1}$$
$$= [(H\bar{S})^{-1}H'^{-1}]\bar{S}^{-1}$$
$$= \bar{S}^{-1}$$

$$\therefore |S + (B-\hat{B})'M(B-\hat{B})|^{-n/2}$$
$$= (|H'H|^{-1})^{-n/2}\left[|I + H'H(B-\hat{B})'M(B-\hat{B})|\right]^{-n/2}$$
$$= |H'H|^{n/2}|I + H'H(B-\hat{B})'M(B-\hat{B})|^{-n/2}$$

$$= |\bar{S}|^{-n/2} \; |I + H(B-\hat{B})' M (B-\hat{B}) H'|^{-n/2}$$

$$= |\bar{S}|^{-n/2} \exp\{\log |I + D|^{-n/2}\}$$

(where $D = H(B-\hat{B})' M (B-\hat{B}) H'$)

$$= |\bar{S}|^{-n/2} \exp\{-n/2 \log |I + D|\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 \sum_i \log(1+\lambda_i)\}$$

(since $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \therefore (I+D) = \begin{bmatrix} 1+\lambda_1 & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}$

& $\log \prod (1+\lambda_i) = \sum \log(1+\lambda_i)$).

$$\Rightarrow |\bar{S}|^{-n/2} \exp\{-n/2 [\sum_i (\lambda_i - \tfrac{1}{2}\lambda_i^2 + \tfrac{1}{3}\lambda_i^3 - \cdots)]\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 [\sum_i \lambda_i - \tfrac{1}{2}\sum_i \lambda_i^2 + \tfrac{1}{3}\sum_i \lambda_i^3 - \cdots]\}$$

$$= |\bar{S}|^{-n/2} \exp\{-n/2 [\text{tr.}D - \tfrac{1}{2}\text{tr}D^2 + \tfrac{1}{3}\text{tr}D^3 - \cdots]\}.$$

And, $(\text{tr.}D) = \text{tr.}\, H(B-\hat{B})' M (B-\hat{B}) H'$

$$= \text{tr.}\, H'H(B-\hat{B})' M (B-\hat{B})$$

Let $\bar{Q} = (B-\hat{B})' M (B-\hat{B})$

$\therefore \bar{S}^{-1}\bar{Q} = H'H(B-\hat{B})' M (B-\hat{B})$

$\therefore \text{tr.}D = \text{tr.}\, \bar{S}^{-1}\bar{Q}$

$\text{tr.}D^2 = \text{tr.}\, \bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}$

$\text{tr}D^3 = \text{tr.}\, \bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}\,\bar{S}^{-1}\bar{Q}$ <u>etc.</u>

$\therefore p(B|y) \propto |\bar{S} + (B-\hat{B})' M (B-\hat{B})|^{-n/2}$

$\propto |\bar{S}|^{-n/2} \exp\{-n/2 [\text{tr.}D - \tfrac{1}{2}\text{tr.}D^2 + \tfrac{1}{3}\text{tr.}D^3 - \cdots]\}$

But $M = \frac{1}{n} X'X$

$\bar{Q} = (B-\hat{B})' M (B-\hat{B})$

$= \frac{1}{n}(B-\hat{B})' X'X (B-\hat{B})$

$= \frac{1}{n} Q$, say.

$$\therefore\ p(B|y) \propto \exp\left\{-\tfrac{1}{2}\left[k\ \mathrm{tr.}\ \bar{S}^{-1}Q - \tfrac{1}{2n^2}\ \mathrm{tr.}\bar{S}^{-1}Q\bar{S}^{-1}Q - \cdots\right]\right\}$$

$$\propto \exp\left(-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right)\exp\left\{\tfrac{1}{4n}\ \mathrm{tr.}\ \bar{S}^{-1}Q\bar{S}^{-1}Q - \tfrac{1}{6n^2}\ \mathrm{tr}\,\bar{S}^{-1}Q\bar{S}^{-1}Q\bar{S}^{-1}Q - \cdots\right\}.$$

And $e^x = 1 + x + x^2/2! + \cdots$

$$\text{So, } p(B|y) \propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right] * \left\{1 + \tfrac{1}{4n}\ \mathrm{tr.}\ \bar{S}^{-1}Q\bar{S}^{-1}Q + \cdots\right\}$$

$$\propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}Q\right]$$

$$\propto \exp\left[-\tfrac{1}{2}\ \mathrm{tr.}\ \bar{S}^{-1}(B-\hat{B})'X'X(B-\hat{B})\right]$$

$$\propto \exp.\left[-\tfrac{1}{2}(B-\hat{B})'\bar{S}^{-1}\otimes(X'X)\ (B-\hat{B})\right]$$

Let $\beta_{\cdot 1}$ denote the non-zero coefficients
" $\beta_{\cdot 0}$ " " zero "
" $\bar{S}^{\alpha\ell}$ be the (α,ℓ)th. element of $\bar{S}^{-1}$
" $d' = \left[(\beta_{11}-\hat{\beta}_{11})', -\hat{\beta}_{10}', \cdots, (\beta_{m1}-\hat{\beta}_{m1})', -\hat{\beta}_{m0}'\right]$

i.e. this is the $(\beta-\hat{\beta})$, conditionalized $\ni$ the zero restrictions are incorporated.

Then, $X = (X_1 : X_0)$; $(X'X) = (X_1 : X_0)'(X_1 : X_0)$

$\therefore\ (X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})$ is the (α,ℓ)th. el. of $(X'X)$

So, $\left\{\bar{S}^{\alpha\ell}(X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})\right\}$ is a typical element of the partitioned matrix.

$$\therefore\ p(\beta_{\cdot 1}|y, \beta_{\cdot 0}=0) \propto \exp\left\{-\tfrac{1}{2}d'\left[\bar{S}^{\alpha\ell}(X_{\alpha 1} : X_{\alpha 0})'(X_{\ell 1} : X_{\ell 0})\right] * d\right\}$$

$$\cancel{\propto \exp\left\{-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})'(X_{\alpha 1} : X_{\ell 1})^{\prime}\right.}$$

$$\propto \exp\left\{-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})'(X_{\alpha 1}X_{\ell 1}\bar{S}^{\alpha\ell})(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})\right.$$
$$\left. + \hat{\beta}_{\cdot 0}'(X_{\alpha 0}'X_{\ell 0}\bar{S}^{\alpha\ell})\hat{\beta}_{\cdot 0} - 2\hat{\beta}_{\cdot 0}'(X_{\alpha 0}'X_{\ell 1}\bar{S}^{\alpha\ell})(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1})\right\}$$

Let $V = (X_{\alpha 1}'X_{\ell 1}\bar{S}^{\alpha\ell})$; $R = (X_{\alpha 0}'X_{\ell 1}\bar{S}^{\alpha\ell})$

$$\therefore\ p(\beta_{\cdot 1}|y, \beta_{\cdot 0}) \propto \exp\left[-\tfrac{1}{2}(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1}-V^{-1}R'\hat{\beta}_{\cdot 0})' * V(\beta_{\cdot 1}-\hat{\beta}_{\cdot 1}-V^{-1}R'\hat{\beta}_{\cdot 0})\right]$$

And this is approximately a <u>normal</u> pdf. i.e. we have a <u>conditionalized</u> posterior pdf with mean $(\hat{\beta}_{\cdot 1} + V^{-1}R'\hat{\beta}_{\cdot 0})$ and c.m. of

$$V^{-1} = (X_{\alpha 1}' X_{\alpha 1} \bar{\sigma}^{\alpha\alpha})^{-1}.$$

(D) <u>Traditional Model with an Informative Prior p.d.f.:</u>

Now we wish to introduce <u>informative</u> prior information about B. If this information is reasonably accurate then we should be able to improve the precision of our inferences by using it in favour of a diffuse prior. Also, if we compare the prior and posterior pdf's, we shall be able to discern to what extent the sample information has modified our original beliefs about the model.

Now, if we were to use a simple natural conjugate prior pdf on the traditional M.V. regression model, then we would have to place restrictions on the variances and covariances appearing in the equations of the system. This is because $(X'X)^{-1}$ enters the covariance structure via $\Sigma \otimes (X'X)^{-1}$. Then if we used a simple natural conjugate prior, we would find for example that the ratios of variances of corresponding coefficients in equations 1 and 2 would be equal. This problem is avoided if we use a general MVN pdf as our prior for all of the coeff. in the model. But the price that we pay then is a loss of easy analysis — the prior pdf will no longer combine neatly with the LIF.

So, our prior pdf is —

$$p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp\left\{-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right\}$$

where $\bar{\beta}$ is an $(mk \times 1)$ vector, the mean of the prior pdf, assigned by the analyst.

C is an $(mk \times mk)$ prior C.M, whose value is also assigned by the user. (N.B. C is <u>constant</u>, so the leading term in $p(\beta, \Sigma^{-1})$ involving C is absorbed into the proportionality sign.).

The usual diffuse prior is used for Σ^{-1} in the above expression.

Now,

$$l(B, \Sigma \mid y, X) \propto |\Sigma|^{-n/2} \exp\left[-\tfrac{1}{2}\operatorname{tr} S\Sigma^{-1} - \tfrac{1}{2}\operatorname{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]$$

$$\propto |\Sigma^{-1}|^{N/2} \exp\left[-\tfrac{1}{2}\operatorname{tr} S\Sigma^{-1} - \tfrac{1}{2}\operatorname{tr}(B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1}\right]$$

So,

$$p(B, \Sigma^{-1} | y, X) \propto |\Sigma^{-1}|^{(n-m-1)/2} * \exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right] *$$
$$\exp\left[-\tfrac{1}{2} tr\, \Sigma^{-1}\left[S + (B-\hat{B})' X' X (B-\hat{B})\right]\right]$$

where $\hat{B} = (X'X)^{-1} X' y$.

Now, integrate w.r.t. Σ^{-1}, & obtain —

$$p(B | y, X) \propto \exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right] * \left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-\frac{n}{2}}$$

And this is the product of a MVN pdf. and a <u>generalized</u> MVS pdf. This is rather messy to handle as it stands. One way of getting around the analytical problem is to expand the determinant as we did before, & normalize on the first term so obtained —

$$\left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-n/2}$$
$$= |\bar{S}|^{-n/2} \exp\left[-\tfrac{n}{2}\left[tr.D - \tfrac{1}{2} tr.D^2 + \tfrac{1}{3} tr.D^3 - \cdots\cdots\right]\right]$$

where $D = \frac{1}{n} H (B-\hat{B})' X' X (B-\hat{B}) H'$

& $H \bar{S} H' = I$, & $\bar{S} = \frac{1}{n} S$.

$$\therefore \left| S + (B-\hat{B})' X' X (B-\hat{B}) \right|^{-n/2}$$
$$\doteq \exp\left[-\tfrac{1}{2}(\beta-\hat{\beta})' \bar{S}^{-1} \otimes (X'X) (\beta-\hat{\beta})\right]$$

$$\therefore p(\beta | y, X) \propto \exp\left[-\tfrac{1}{2}(\beta-\hat{\beta})' \bar{S}^{-1} \otimes (X'X)(\beta-\hat{\beta})\right] *$$
$$\exp\left[-\tfrac{1}{2}(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})\right]$$
$$\propto \exp\left[-\tfrac{1}{2}(\beta-b)' F (\beta-b)\right]$$

where $F = \left(C^{-1} + \bar{S}^{-1} \otimes (X'X)\right)$

$$b = \left(C^{-1} + \bar{S}^{-1} \otimes (X'X)\right)^{-1} \left(C^{-1}\bar{\beta} + \bar{S}^{-1} \otimes X'X\hat{\beta}\right)$$

Then b is the mean of the leading normal term of the expansion, & F^{-1} is the CM of the leading term approximating the posterior pdf of B.

(E) The "Seemingly Unrelated" Regression Model:

In a sense, this is a generalization of the M.V. regression model, in that the matrix X appearing in each equation of the traditional model is now allowed to be different from equation to equation. viz:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}$$

y_α is an $(n \times 1)$ vector of obs. on the αth. dep. var.

X_α " " $(n \times k_\alpha)$ matrix " " " " k_α indep. var.

β_α " " $(k_\alpha \times 1)$ vector " " " coeff.

u_α " " $(n \times 1)$ " " " " " disturbances.

$$\therefore \quad y = Z\beta + u \quad , \text{ say.}$$

where Z is the <u>block-diagonal</u> matrix shown above.

$$\& \quad E(u) = 0$$
$$E(u,u') = \Sigma \otimes I_n.$$

And Σ is $(m \times m)$ and pos. def. symm.

Then,

$$\ell(\beta, \Sigma | y) \propto |\Sigma^{-1}|^{n/2} \exp\left[-\tfrac{1}{2}(y - Z\beta)' \Sigma^{-1} \otimes I_n (y - Z\beta)\right]$$
$$\propto |\Sigma^{-1}|^{n/2} \exp\left[-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\right]$$

where, $$A = \begin{bmatrix} (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & \cdots & (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) \\ \vdots & & \vdots \\ (y_m - X_m\beta_m)'(y_1 - X_1\beta_1) & \cdots & (y_m - X_m\beta_m)'(y_m - X_m\beta_m) \end{bmatrix}$$

And assume that $p(\beta, \Sigma^{-1}) = p(\beta)\, p(\Sigma^{-1})$.

$\therefore\ p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2}$

$\therefore\ p(\beta, \Sigma^{-1}|y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\left[-\frac{1}{2}\operatorname{tr} A\Sigma^{-1}\right]$

Then $p(\beta|\Sigma^{-1}, y)$ is MVN —

$p(\beta|\Sigma^{-1}, y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\left[-\frac{1}{2}(y-Z\beta)'\Sigma^{-1}\otimes I_n(y-Z\beta)\right]$

$\therefore\ E[\beta|\Sigma^{-1}, y] = (Z'(\Sigma^{-1}\otimes I_n)Z)^{-1} Z'(\Sigma^{-1}\otimes I_n)y = \hat{\beta}$

& $\text{Cov.}(\beta|\Sigma^{-1}, y) = (Z'(\Sigma^{-1}\otimes I_n)Z)^{-1}$

And the <u>conditional</u> posterior mean of β, <u>given</u> Σ^{-1} is just the sampling theory GLS estimator. And the same applies to the CM above. Also, under the assumption of normality, as is expressed by the above L.F., $E(\beta|\Sigma^{-1}, y) = \hat{\beta}$ is a MLE of β.

Now: (i) If all of the X's are the same, or if they are proportional to one another, then $\hat{\beta}$ reduces to a vector of O.L.S. estimators.

and/or: (ii) If Σ is diagonal, then again $\hat{\beta}$ reduces to a vector of O.L.S. estimators.

Then $\hat{\beta}_\alpha = (X_\alpha' X_\alpha)^{-1} X_\alpha' y_\alpha$; $\alpha = 1, 2, ---, m$.

Now, the sampling theory estimator of β is equivalent to the mean of the <u>conditional</u> posterior pdf for β, given Σ^{-1}. Now, in the sampling theory approach, β^* is independent of Σ^{-1}, which is usually unknown & is replaced by $\hat{\Sigma}^{-1}$, formed from the residuals of the equations as estimated by O.L.S.

So then, in the sampling theory case we get:

$$\beta^* = [Z'(\hat{\Sigma}^{-1}\otimes I_n)Z]^{-1}[Z'(\hat{\Sigma}^{-1}\otimes I_n)y]$$

And note that β^* has the same <u>large sample</u> properties as has $\hat{\beta}$ in the Bayes case.

So β^* is what we would obtain in the Bayes case if we proceeded with $\Sigma = \hat{\Sigma}$. In <u>large samples</u>, if $\hat{\Sigma}$ is consistent, then Σ & $\hat{\Sigma}$ are little different, so the assumption that $\Sigma = \hat{\Sigma}$ will produce quite satisfactory results. However, in <u>small</u> samples,

things will not be so satisfactory. Then it is better to use the <u>marginal</u> posterior pdf for β and take <u>its</u> mean as the estimator of β. Note that by adopting a Bayes technique, we have an advantage over the sampling theorist when it comes to <u>small</u> samples, in this particular case. We can get an <u>exact</u> small-sample result, whereas he is unable to do so.

Now,

$$p(\beta, \Sigma^{-1} | y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}\}$$

$$\therefore\ p(\beta | y) \propto \int \cdots \int |\Sigma^{-1}|^{(n-m-1)/2} \exp[-\tfrac{1}{2} \operatorname{tr} A \Sigma^{-1}]\, d\Sigma^{-1}$$

$$\propto |A|^{-n/2}$$

$$\therefore\ p(\beta | y) \propto \begin{vmatrix} (y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & \cdots & (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) \\ \vdots & & \vdots \\ (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) & \cdots & (y_m - X_m\beta_m)'(y_m - X_m\beta_m) \end{vmatrix}^{-\frac{n}{2}}$$

Now, the <u>marginal</u> posterior pdf for β vaguely resembles a <u>generalized</u> MVS pdf, but in fact it <u>cannot</u> be brought exactly into that form, because not all of the X's are the same. For $p(\beta|y)$ in the form showing we don't have any practical way of analyzing this pdf.

An alternative way of viewing the SUR ~~model~~ model is to write it as a "restricted" traditional M.V.R. model. <u>i.e.</u>

$$(y_1, y_2, \cdots, y_m) = (X_1\ X_2 \cdots X_m) \begin{pmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_m \end{pmatrix} + (u_1 \cdots u_m)$$

Then the (zero) restrictions on the coefficient matrix appear quite explicitly. If $(X_1 \cdots X_m)$ has full rank, then we can use the methods of section (C) above to handle the zero restrictions, as for the traditional model.

Zellner: Chapter 10.

"Comparing & Testing Hypotheses."

In comparing different hypotheses we may wish to see how much the data alters our prior odds in favour of one or other hypothesis. That is, what are the posterior odds in favour of each hypothesis?

This information may be all that we are interested in, but in addition to this it may well be that we want to make decisions and take actions on the basis of these posterior odds. Then we will, naturally, be concerned with the loss structure involved when such actions are made.

(A) Posterior Probabilities for Hypotheses:-

Let H_0 and H_1 be two mutually exclusive and jointly exhaustive simple hypotheses.

Under H_0, y has a p.d.f. $p(y|\theta=\theta_0)$

Under H_1, y has a pdf. $p(y|\phi=\phi_1)$

Let ω be a dichotomous random variable ∋

$$\omega = \begin{cases} 0 \; ; & \text{if } H_0 \text{ is true.} \\ 1 \; ; & \text{if } H_1 \text{ is true.} \end{cases}$$

Then the prior probabilities are:

$$p(H_0) = p(\omega=0)$$
$$p(H_1) = p(\omega=1)$$

And, $p(H_0) + p(H_1) = p(\omega=0) + p(\omega=1) = 1.$

Now, consider the following —

$$p(y,\omega) = p(y|\omega)\,p(\omega)$$
$$= p(\omega|y)\,p(y)$$

$$\therefore \; p(\omega|y) = \frac{p(y|\omega)p(\omega)}{p(y)}$$

And $p(y) = \sum_i p(y|\omega_i)p(\omega_i) = p(y|\omega=0)p(\omega=0) + p(y|\omega=1)p(\omega=1)$
$\neq 0$; by assumption.

$$\therefore\ p(H_0|y) = p(\omega=0|y) = \frac{p(y|\omega=0)\,p(\omega=0)}{p(y)}$$

$$= \frac{p(y|\theta=\theta_0)\,p(H_0)}{p(y)}$$

And, $$p(H_1|y) = p(\omega=1|y) = \frac{p(y|\omega=1)\,p(\omega=1)}{p(y)}$$

$$= \frac{p(y|\phi=\phi_1)\,p(H_1)}{p(y)}.$$

Consider then the <u>posterior odds</u> in favour of H_0:

$$K_{01} = \frac{p(H_0|y)}{p(H_1|y)} = \left[\frac{p(y|\theta=\theta_0)\,p(H_0)}{p(y)}\right] \div \left[\frac{p(y|\phi=\phi_1)\,p(H_1)}{p(y)}\right]$$

$$= \left[\frac{p(H_0)}{p(H_1)}\right] * \left[\frac{p(y|\theta=\theta_0)}{p(y|\phi=\phi_1)}\right]$$

~~i.e. (posterior odds) = p~~

i.e. (posterior odds) = (prior odds) * (likelihood ratio)

So, if we are given the prior odds and the two L/F. then we can arrive at the posterior odds quite simply.

But, having been <u>given</u> the posterior odds, what say that we want to make use of them in decision problems? i.e. can we make use of the posterior odds in a neat, systematic fashion to decide on our actions with regard to the acceptance or otherwise of a given hypothesis?

In the situation so far we have 2 <u>states</u>: H_0 and H_1, and we have 2 <u>actions</u> accept or reject H_0 (or, alternatively — accept H_0; accept H_1). This set-up may be depicted —

		States	
		H_0 true.	H_1 true.
Actions	Acc H_0.	$L(H_0, \hat{H}_0) = 0$	$L(H_1, \hat{H}_0)$
	Acc H_1	$L(H_0, \hat{H}_1)$.	$L(H_1, \hat{H}_1) = 0$

Now, consider the <u>expected losses</u> involved in a posterior sense (after having seen the data).

Then,

$$E(\mathcal{L}\,|\,\hat{H}_0) = \mathcal{L}(H_0, \hat{H}_0)\,p(H_0|y) + \mathcal{L}(H_1, \hat{H}_0)\,p(H_1|y)$$
$$= 0 + \mathcal{L}(H_1, \hat{H}_0)\,p(H_1|y).$$

And,

$$E(\mathcal{L}\,|\,\hat{H}_1) = \mathcal{L}(H_1, \hat{H}_1)\,p(H_1|y) + \mathcal{L}(H_0, \hat{H}_1)\,p(H_0|y)$$
$$= 0 + \mathcal{L}(H_0, \hat{H}_1)\,p(H_0|y).$$

Now, we look at $E(\mathcal{L}\,|\,\hat{H}_0)$ and $E(\mathcal{L}\,|\,\hat{H}_1)$ and on the basis of these values we make a <u>decision</u> whether to accept H_0 or to accept H_1.

<u>viz</u>:

<u>(1)</u> If $E(\mathcal{L}\,|\,\hat{H}_0) < E(\mathcal{L}\,|\,\hat{H}_1)$, then accept H_0; i.e. $\hat{H}_0$.

<u>(2)</u> If $E(\mathcal{L}\,|\,\hat{H}_1) < E(\mathcal{L}\,|\,\hat{H}_0)$, then accept H_1; ie $\hat{H}_1$.

If the posterior expected losses are equal, then <u>either</u> action ($\hat{H}_0$ or $\hat{H}_1$) will give rise to the same degree of loss.

Now,
$$\begin{cases} E(\mathcal{L}\,|\,\hat{H}_0) = \mathcal{L}(H_1, \hat{H}_0).p(H_1|y) \\ E(\mathcal{L}\,|\,\hat{H}_1) = \mathcal{L}(H_0, \hat{H}_1).p(H_0|y) \end{cases}$$

So, the <u>decision rule</u> may be expressed as —

<u>(1)</u> Accept H_0 if:

$$E(\mathcal{L}\,|\,\hat{H}_0) < E(\mathcal{L}\,|\,\hat{H}_1)$$

i.e. if $\quad p(H_1|y)\,\mathcal{L}(H_1, \hat{H}_0) < p(H_0|y)\,\mathcal{L}(H_0, \hat{H}_1)$

$$\Rightarrow \left[\frac{\mathcal{L}(H_1, \hat{H}_0).\,p(y|\phi=\phi_1)\,p(H_1)}{p(y)}\right] < \left[\frac{\mathcal{L}(H_0, \hat{H}_1)\,p(y|\theta=\theta_0)\,p(H_0)}{p(y)}\right]$$

$$\Rightarrow \mathcal{L}(H_1, \hat{H}_0)\,p(H_1)\,p(y|\phi=\phi_1) < \mathcal{L}(H_0, \hat{H}_1)\,p(y|\theta=\theta_0)\,p(H_0)$$

$$\Rightarrow \left[\frac{p(H_1)}{p(H_0)}\right].\left[\frac{\mathcal{L}(H_1, \hat{H}_0)}{\mathcal{L}(H_0, \hat{H}_1)}\right] < \left[\frac{p(y|\theta=\theta_0)}{p(y|\phi=\phi_1)}\right].$$

$$\Rightarrow \left[\frac{p(y|H_0)}{p(y|H_1)}\right] > \left[\frac{p(H_1)}{p(H_0)} \cdot \frac{L(H_1, \hat{H}_0)}{L(H_0, \hat{H}_1)}\right]$$

So, we pick H_0 (i.e. select <u>action</u> $\hat{H}_0$) iff

Likelihood ratio > ratio of <u>prior</u> <u>expected</u> losses.

(2) Similarly, accept H_1 iff.

$$\left(\frac{p(y|H_1)}{p(y|H_0}\right) > \left[\frac{p(H_0) \cdot L(H_0, \hat{H}_1)}{p(H_1) \cdot L(H_1, \hat{H}_0)}\right]$$

Now, our usual decision theoretic approach to sampling theory gives us the well known <u>Likelihood Ratio Test</u> when testing between 2 hypotheses.

And in that particular case, the rule is:

<u>Accept H_0</u> if $\frac{p(y|H_0)}{p(y|H_1)} > \lambda$

<u>Accept H_1</u> if $\frac{p(y|H_0)}{p(y|H_1)} < \lambda$.

Now, the main difference between the sampling theory approach and the Bayesian approach is that in the case of the former, λ is determined by the <u>significance level</u> chosen for the test, & this is unavoidably arbitrary. The loss structure is introduced into the selection of λ only in a very <u>implicit</u> fashion. This seems rather unsatisfactory.

On the contrary, in the Bayesian approach the loss structure is used <u>explicitly</u> to find λ. Namely, we put λ equal to ratio of the <u>prior expected</u> losses. (<u>NB</u> recall we pick the lower of the 2 <u>posterior</u> expected losses, & this leads to a <u>ratio</u> test involving <u>prior</u> expected losses.)

That is, we put:

$$\lambda = \left(\frac{p(H_1)}{p(H_0)}\right)\left(\frac{L(H_1, \hat{H}_0)}{L(H_0, \hat{H}_1)}\right).$$

Now, clearly the value of λ, and hence the form of the ratio test, will depend very much on the functional form taken by the loss structure.

A natural class of loss functions are those that are <u>symmetric</u>:

<u>viz</u>
$$\begin{cases} L(H_0, \hat{H}_1) = L(H_1, \hat{H}_0) \\ L(H_0, \hat{H}_0) = L(H_1, \hat{H}_1) = 0. \end{cases}$$

In sampling theory terminology, this is the same as saying that a Type I and Type II error have the same losses associated with them.

Then, if $\frac{p(y|H_0)}{p(y|H_1)} > \left(\frac{p(H_1)\cdot L}{p(H_0)\cdot L}\right)$

$$> \frac{p(H_1)}{p(H_0)}$$, Then pick H_0.

So in the case of a symmetric loss structure, the <u>critical value</u>, λ, in the likelihood ratio test is just the ratio of the <u>prior</u> odds for H_1.

<u>i.e.</u> select H_0 as true if the <u>proportional</u> data evidence overwhelms the <u>proportional</u> prior evidence for H_1.

Now, can we represent the decision rule for the selection of H_0 in terms of the posterior odds?

Well, we had that:

$$E(L|\hat{H}_0) = L(H_1, \hat{H}_0)\, p(H_1|y)$$

$$E(L|\hat{H}_1) = L(H_0, \hat{H}_1)\, p(H_0|y)$$

And the <u>symmetry</u> assumption $\Longrightarrow L(H_1, \hat{H}_0) = L(H_0, \hat{H}_1) = L$.

So, $E(L|\hat{H}_0) = L \cdot p(H_1|y)$

$$E(L|\hat{H}_1) = L \cdot p(H_0|y)$$

And our natural decision rule is to select H_0 if the posterior expected loss for H_0 is less than that for H_1:

<u>i.e.</u> if $E(L|\hat{H}_0) < E(L|\hat{H}_1)$.

Now, $\Longrightarrow L \cdot p(H_1|y) < L \cdot P(H_0|y)$

$$\Longrightarrow \left(\frac{p(H_0|y)}{p(H_1|y)}\right) > 1.$$

Similarly, accept H_1 if:

$$\left(\frac{p(H_1|y)}{p(H_0|y)}\right) > 1.$$

So, in the <u>special case</u> of a symmetric loss function, we just pick an hypothesis if the <u>posterior odds</u> are in its favour.

Now, so far we have considered only <u>simple</u> hypotheses. This is generally quite unrealistic in practice, as we more usually encounter <u>composite</u> hypotheses. To extend the analysis developed so far to a more general framework.

Let H_0 and H_1 be <u>composite</u> mutually exclusive and totally exhaustive hypotheses, $\ni$

$$\begin{cases} p(H_0) = p(\omega = 0) \\ p(H_1) = p(\omega = 1) \end{cases}$$ are the prior pdf's.

Let Θ be the parameter vector associated with H_0, under which the pdf for y is $p(y|\Theta) \equiv p(y|\omega=0, \Theta)$; and let ϕ be the parameter vector associated with H_1, under which the pdf for y is $p(y|\phi) \equiv p(y|\omega=1, \Theta)$. Then the joint dist'n. is:

$$p(y, \Theta, \omega) = p(\omega, \Theta|y)p(y)$$

$$= p(y|\omega, \Theta)\, p(\omega, \Theta)$$

So, $$p(\omega, \Theta|y) = \left[\frac{p(y|\omega, \Theta)\, p(\omega, \Theta)}{p(y)}\right] = \left[\frac{p(y|\omega, \Theta)\, p(\Theta|\omega)\, p(\omega)}{p(y)}\right].$$

Now, $p(\Theta|\omega = 0) = p(\Theta)$

$p(\Theta|\omega = 1) = p(\phi)$

~~and $p(\omega=0) = p(\Theta)$; $p(\omega=1) = p(\phi)$.~~

So, $$p(H_0|y) = \int p(\omega=0, \Theta|y)\, d\Theta$$

$$= \int \frac{p(y|\omega=0, \Theta)\, p(\Theta, \omega)\, d\Theta}{p(y)} = \int \frac{p(y|\omega=0, \Theta)\, p(\Theta|\omega=0)\, p(\omega=0)}{p(y)}$$

$$= \int \frac{p(\omega=0)\, p(y|\omega=0, \Theta) \cdot p(\Theta)\, d\Theta}{p(y)}$$

$$= \left[\frac{p(H_0) \int p(y|\Theta)\, p(\Theta)\, d\Theta}{p(y)}\right]$$

and, $$p(H_1|y) = p(\omega=1|y) = \int p(\omega=1, \Theta|y)\, d\phi$$

$$= \frac{\int p(\omega=1)\, p(y|\omega=1, \Theta)\, p(\phi)\, d\phi}{p(y)}$$

$$= \left(p(H_1) \int p(y|\phi)\, p(\phi)\, d\phi\right) / (p(y)).$$

So, the posterior odds are given by –

$$\kappa_{01} = \frac{p(H_0|y)}{p(H_1|y)} = \left[\frac{p(H_0)}{p(H_1)}\right] \cdot \left[\frac{\int p(y|\theta)p(\theta)d\theta}{\int p(y|\phi)\,p(\phi)\,d\phi}\right].$$

That is, the posterior odds are the product of the prior odds and the ratio of the averaged likelihoods.

See how this contrasts with sampling theory version of this result in which the maximized likelihoods are used — the MLE are used as if they were the true values of the parameters.

B) Analysis of Hypotheses Using Diffuse Prior Information:-

Lindleys' procedure to be outlined here is valid only if the prior information is fairly diffuse in the locality of the likelihood function. For many, but not all problems, Lindleys' procedure leads to tests which are computationally equivalent to the standard sampling theory results. But the interpretation is fundamentally different.

Further, if the prior information is not in fact diffuse, then the form that it takes will have a considerable impact on the test results.

Lindleys procedure involves the derivation of a test procedure based on a posterior pdf, starting from a diffuse prior pdf. Let the posterior pdf be $p(\theta|y)$, & assume further that $p(\theta|y)$ is unimodal.

We want to construct a test of the hypothesis $H_0 : \theta = \theta_0$, at the α% significance level.

So construct an interval $(a, b) \ni \Pr\{a < \theta < b|y\} = 1-\alpha$.

That is, minimize $(b-a)$ s.t. $\Pr\{a < \theta < b|y\} = 1-\alpha$.

And $\Rightarrow$ $p(a) = p(b)$. Then, if θ_0 falls in (a, b) we accept H_0 at the α% significance.

Lindley chooses to use the posterior pdf for θ as a basis for making inferences about values of θ. If $\theta = \theta_0$ falls within a region where the posterior density is not high, then we have little belief in $\theta = \theta_0$, & so H_0 would be rejected.

Now, the posterior distribution is based on the L/F explicitly, & so it contains all of the sample information. Compare this situation with that frequently

arising with respect to the sampling theory approach. There the tests are often based on test statistics which are <u>not sufficient</u> statistics & hence do not contain all of the sample knowledge.

Also, the hypothesis $H_0 : \theta = \theta_0$ is being judged on the basis of our ~~our~~ posterior knowledge — that is, on <u>all</u> of the knowledge that we currently have. We are <u>not</u> basing our judgement on the fact that some test statistic takes an unusual value if $\theta = \theta_0$. So sampling theory tests should <u>not</u> be interpreted as measuring the degree of belief that the hypothesis $H_0 : \theta = \theta_0$ is true.

Thirdly, in the Bayesian approach the analyst looks at the <u>whole</u> of the posterior distribution, whereas the sampling theorist looks just at the significance level, & this is only a <u>partial</u> expression of posterior beliefs.

<u>Example:</u> Let $y' = (y_1, y_2, \ldots y_n) \sim NID(\mu, \sigma_0^2)$, where σ_0^2 is <u>known</u>.

When will we reject $H_0 : \mu = 5$, at the 5% level? Suppose that $p(\mu) \propto$ constant.

$$\therefore p(\mu \mid y, \sigma = \sigma_0) \propto \exp.\left\{-\frac{1}{2\sigma_0^2}\sum_1^n (y_i - \mu)^2\right\}$$

$$\propto \exp.\left\{-\frac{1}{2\sigma_0^2}\sum\left[(y_i - \hat{\mu}) - (\mu - \hat{\mu})\right]^2\right\}$$

$$\propto \exp\left\{-\frac{n}{2\sigma_0^2}(\mu - \hat{\mu})^2\right\}.$$

So the <u>posterior</u> pdf for μ is $N(\hat{\mu}, \sigma_0/\sqrt{n})$.

$$\therefore \quad Z = \left(\frac{\mu - \hat{\mu}}{\sigma/\sqrt{n}}\right) \text{ is } N(0, 1).$$

$$\therefore \Pr\{-1.96 < Z < 1.96\} = 0.95$$

$$\therefore \Pr\left\{\hat{\mu} - \frac{1.96\sigma_0}{\sqrt{n}} < \mu < \hat{\mu} + \frac{1.96\sigma_0}{\sqrt{n}}\right\} = 0.95.$$

$$\therefore \text{if } 5 \in \left(\hat{\mu} - \frac{1.96\sigma_0}{\sqrt{n}}, \hat{\mu} + \frac{1.96\sigma_0}{\sqrt{n}}\right); \text{ accept } H_0.$$

<u>cf.</u> the sampling theory interpretation of a computationally equivalent test — Here $Z' = \left(\frac{\hat{\mu} - 5}{\sigma_0/\sqrt{n}}\right)$ is considered <u>random</u> (<u>not</u> μ itself), and H_0 is accepted if $Z' \in (-1.96, 1.96)$.

<u>Example</u>: As an example of a case where there is no sampling theory equivalent test, consider the case of the autocorrelation coefficient, ρ —

$$p(\rho|y) \propto |H|^{-\frac{1}{2}}\left\{(y-\rho y_{-1})'[I-(X-\rho X_{-1})'H^{-1}(X-\rho X_{-1})](y-\rho y_{-1})\right\}^{-\frac{\nu}{2}}$$

Then employ numerical integration techniques to find the shortest interval, (a, b), $\ni$ $\Pr.\{a<\rho<b\} = 1-\alpha$.

To test $H_0: \rho=\rho_0$,
accept H_0 if $(a<\rho_0<b)$, otherwise reject.

And clearly, Lindley's general procedure may be extended to the case where we have a <u>joint</u> hypothesis about 2 or more parameters, say Θ. We just obtain the Bayesian highest posterior density confidence region with probability content $(1-\alpha)$. If H_0 is $\Theta=\Theta_0$, then accept H_0 if Θ_0 lies in this region.

The basic rationale underlying Lindley's procedure is to accept H_0 when the suggested value of the parameter under H_0 is in an interval of high posterior density, & to reject otherwise. This procedure has no obvious decision theoretic justification. It is based solely on what appears to be reasonable, <u>a posteriori</u>.

In some cases, Lindley's procedure is <u>computationally</u> equivalent to the sampling theory solution. In other cases, there <u>is no</u> sampling theory solution. In <u>large samples</u> Lindley's procedure corresponds to M.L. procedure under the assumption of an asymptotically normal L/F.

C) Hypotheses & Non-Diffuse Prior Information:

If prior information is <u>not</u> in fact diffuse, then it is most important that this fact be taken into a/c. when testing hypotheses.

So, for example, if we test one set of explanatory variables, find them to be insignificant, & so modify the set & test again, then we have some prior information when it comes to testing the second time, & we really should take account of this fact. Sampling theory techniques do <u>not</u> allow us take care of this particular problem very satisfactorily.

There are a no. of situations in which we will definitely wish to introduce non-diffuse prior information. For example, consider the simple hypothesis

$$H_0 : \theta = \theta_0.$$

Then if $\theta = \theta_0$ is the value suggested by the theory (as it is likely to be — otherwise we probably wouldn't bother testing H_0), then our prior info. is <u>not</u> diffuse at all — in fact it is tightly packed close to θ_0. That is, θ_0 is seen <u>a priori</u> as more likely than other values — is this position maintained <u>a posteriori</u>?

Can we introduce <u>non-diffuse</u> prior information successfully?

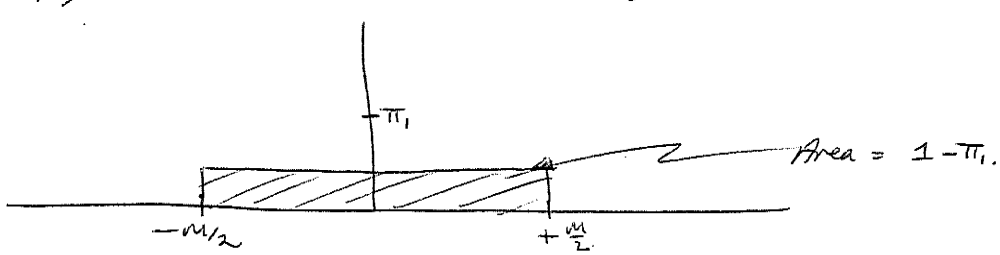
Consider the following —

$$y' = (y_1, \ldots, y_n) \sim NID(\mu, 1)$$

$$H_0 : \mu = 0.$$

& $p(\mu)$ is $\ni$ $\begin{cases} p(\mu = 0) = \pi_1 \\ p(\mu \neq 0) = 1 - \pi_1. \end{cases}$

& $p(\mu \neq 0)$ is spread uniformly on $(-\frac{M}{2}, +\frac{M}{2})$



$H_0 : \mu = 0$; $H_1 : \mu \neq 0.$

Now, the prior odds are $\left(\frac{\pi_1}{1-\pi_1}\right)$.

Now, under H_0, the posterior pdf is obtained as

$$P(y \mid \mu=0, \sigma=1) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\tfrac{1}{2}\sum_i^n y_i^2\right\}$$
$$= (2\pi)^{-\frac{n}{2}} \exp\left\{-\tfrac{1}{2}(\nu s^2 + n\bar{y}^2)\right\}$$

& under H_1:

$$P(y \mid \mu \neq 0, \sigma=1) = (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}\sum_i^n (y_i-\mu)^2\right\}.$$
$$= (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}(\nu s^2 + n(\mu-\bar{y})^2)\right\}$$

And, $$K_{01} = \frac{P(H_0 \mid y)}{P(H_1 \mid y)}$$
$$= \frac{P(H_0)}{P(H_1)} \cdot \frac{\int p(\theta)\, p(y \mid \theta)\, d\theta}{\int p(\phi)\, p(y \mid \phi)\, d\phi}$$

& $\frac{p(H_0)}{p(H_1)} = \left(\frac{\pi_1}{1-\pi_1}\right)$

& $$\int p(y \mid \theta)\, p(\theta)\, d\theta = \int (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}[\nu s^2 + n\bar{y}^2]\right\} \cdot 1\, d\mu$$
$$= (2\pi)^{-n/2} \exp\left\{-\tfrac{\nu s^2}{2}\right\} \cdot \exp\left\{-\tfrac{n}{2}\bar{y}^2\right\}$$

& $$\int p(y \mid \phi)\, p(\phi)\, d\phi = \int (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}[\nu s^2 + n(\mu-\bar{y})^2]\right\} \cdot \tfrac{1}{M}\, d\mu$$
$$= (2\pi)^{-n/2} \exp\left\{-\tfrac{\nu s^2}{2}\right\} \cdot \int \exp\left\{-\tfrac{n}{2}(\mu-\bar{y})^2\right\} \tfrac{1}{M}\, d\mu$$

$$\therefore K_{01} = \left(\frac{\pi_1}{1-\pi_1}\right) \frac{\exp\left\{-\tfrac{n}{2}\bar{y}^2\right\} \cdot M}{\int_{-M/2}^{M/2} \exp\left\{-\tfrac{n}{2}(\mu-\bar{y})^2\right\} d\mu}$$

Now note the following points —

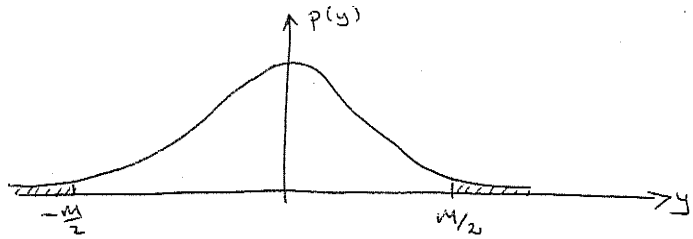
(1) We now have a basis for testing H_0 using non-diffuse prior information, as well as sample information.

(2) Given the loss function, L, we can make a decision between H_0 and H_1.

(3) The result of utilizing K_{01} may be different from the corresponding sampling theory result.

Now consider point no. (3) above —

Assume that $\bar{y}$ falls well within the interval $(-\frac{M}{2}, +\frac{M}{2})$.



Then if we look only at the curve between $-M/2$ & $+M/2$ instead of between $-\infty$ and $+\infty$, if $\bar{y}$ is well within the area, then the loss of information from this approximation (i.e. the 2 shaded areas above) is <u>relatively small</u>.

Now, because $\sigma^2 = 1$, $\Rightarrow$ $\mu \sim N(\bar{y}, 1/n)$

So, $$\int_{-\infty}^{\infty} \left(\frac{\sqrt{n}}{\sqrt{2\pi}}\right) \exp\left\{-\frac{n}{2}(\mu-\bar{y})^2\right\} d\mu = 1$$

$$\therefore \int_{-\infty}^{\infty} \exp\left\{-\tfrac{n}{2}(\mu-\bar{y})^2\right\} d\mu = \left(\tfrac{2\pi}{n}\right)^{1/2}$$

Now taking account of the admissibility of the range approximation just noted above —

$$\int_{-M/2}^{+M/2} \exp\left\{-\tfrac{n}{2}(\mu-\bar{y})^2\right\} d\mu \simeq \left(\tfrac{2\pi}{n}\right)^{1/2}$$

$$\therefore K_{01} \simeq \left(\frac{\pi_1}{1-\pi_1}\right) \frac{\exp\left\{-\frac{n}{2}\bar{y}^2\right\} . M}{(2\pi/n)^{1/2}}$$

$$= \left(\frac{\pi_1}{1-\pi_1}\right) M . \left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\tfrac{z_\alpha^2}{2}\right\}$$

where $z_\alpha = \bar{y}\sqrt{n}$.

Now, if $z_\alpha > 1.96$, a sampling theorist would reject at the 5% level the hypothesis that $\mu = 0$.

So, put $z_\alpha = 1.96$.

$$\therefore K_{01} = \left(\frac{\pi_1}{1-\pi_1}\right) M . \left(\frac{n}{2\pi}\right)^{1/2} \exp\{c\}$$

$$= f(\pi_1, M, n)$$

And it is in fact possible for K_{01} to be <u>high</u> (thus <u>favouring</u> H_0) even if $z_\alpha = 1.96$.

eg put $\pi_1 = \frac{1}{2}$, $M = 1$, $K_{01} > 1$ for $n \geq 300$.

This is <u>Lindley's Paradox</u>.

This just serves to underline the fact that you should not equate $(1-\alpha)$ with a degree of belief in an hypothesis at a posterior level.

Now, consider Lindley's paradox by considering an even more informative prior pdf —

$$H_0 : \mu = 0 \quad ; \quad \ni \quad p(H_0) = \pi_1$$

$$H_1 : \mu \neq 0 \quad ; \quad \ni \quad p(H_1) = 1-\pi_1 .$$

Now, if H_1 true, then $\mu \neq 0$, so let $p(\mu)$ be a cts. proper pdf for μ.

i.e. $\int p(\mu)d\mu = 1$ under H_1.

Let $Pr\{\omega = 0\}$ = prior $p(H_0)$
$Pr\{\omega = 1\}$ = prior $p(H_1)$

Now,
$$p(\omega, \mu | y) = \frac{p(y|\omega,\mu)\, p(\omega,\mu)}{p(y)}$$
$$= \frac{p(y|\omega,\mu)\, p(\mu|\omega)\, p(\omega)}{p(y)}$$

Then the posterior probability for true H_0 is

$$P(\omega=0|y) = \frac{p(y|\omega=0, \mu=0)\, P(\mu|\omega=0)\, p(\omega=0)}{p(y)}$$
$$= [\pi_1\, p(y|\omega=0, \mu=0)\, p(\mu|\omega=0)] / p(y)$$
$$= \frac{\pi_1 \cdot p(y|\omega=0, \mu=0) \cdot 1}{p(y)}$$

Since, $p(\mu|\omega=0) = \begin{cases} 1 & ; \mu = 0 \\ 0 & ; \mu \neq 0 \end{cases}$

$$\therefore p(\omega=0|y) = \frac{\pi_1\, p(y|\omega=0, \mu=0)}{p(y)}$$

& $$p(\omega=1|y) = [p(\omega=1) \int p(y|\mu, \omega=1)\, p(\mu|\omega=1)\, d\mu] / p(y)$$
$$= \frac{(1-\pi_1) \int p(y|\mu, \omega=1)\, p(\mu|\omega=1)\, d\mu}{p(y)}$$

Now, $p(\mu \mid \omega = 1) = p(\mu)$, since $p(\mu)$ is defined only under H_1 true.

So,

$$p(\omega = 1 \mid y) = \frac{(1-\pi_1) \int p(y \mid \mu, \omega = 1)\, p(\mu)\, d\mu}{p(y)}$$

$$\therefore K_{01} = \left(\frac{\pi_1}{1-\pi_1}\right) \left[\frac{p(y \mid \omega = 0, \mu = 0)}{\int p(y \mid \mu, \omega = 1)\, p(\mu)\, d\mu} \right].$$

Now, take the case where $y' = (y_1, ----, y_n) \ni y_i \sim NID(\mu, \sigma_0^2)$; σ_0^2 <u>known</u>.

Then,

$$K_{01} = \left(\frac{\pi_1}{1-\pi_1}\right) \frac{(\sigma_0\sqrt{2\pi})^{-n} \exp\left\{-\frac{\nu s^2}{2\sigma_0^2}\right\} \exp\left\{-\frac{n\bar{y}^2}{2\sigma_0^2}\right\}}{(\sigma_0\sqrt{2\pi})^{-n} \exp\left\{\frac{-\nu s^2}{2\sigma_0^2}\right\} \int p(\mu) \exp\left\{-\frac{n}{2\sigma_0^2}(\mu-\bar{y})^2\right\} d\mu}$$

$$= \left(\frac{\pi_1}{1-\pi_1}\right) \left[\frac{\exp\left\{\frac{-n\bar{y}^2}{2\sigma_0^2}\right\}}{\int p(\mu) \exp\left\{\frac{-n}{2\sigma_0^2}(\mu-\bar{y})^2\right\} d\mu} \right]$$

And if we are given $p(\mu)$ and π_1, then we can evaluate K_{01}.

Consider 2 extreme cases —

(a) Assume $\exists\ (-a, a) \ni \int_{-a}^{a} p(\mu)\, d\mu = 1$.

If $\bar{y}$ lies well within $(-a, a)$ & if $(\sigma_0/\sqrt{n})$ is so <u>large</u> that $\left(\frac{n(\mu-\bar{y})^2}{\sigma_0^2}\right)$ is small for $-a < \mu < a$

Then $K_{01} \simeq 1$; since $e^0 = 1$, & $\int_{-a}^{a} p(\mu)\, d\mu = 1$.

That is, when the standard deviation of $\bar{y}$ (viz $\sigma_0/\sqrt{n}$) is much larger than the range of μ, then we cannot distinguish between 2 alternative hypotheses — this sounds fairly plausible.

b). Suppose that $(\sigma_0/\sqrt{n})$ is fairly <u>small</u>. Then $var(\bar{y})$ is fairly small.

Then $\exp\left[\frac{-n}{2\sigma_0^2}(\mu-\bar{y})^2\right]$ can be very <u>large</u> in $(-a, a)$. Then K_{01} increases with n — <u>i.e.</u> the odds tend to favour H_0 as n gets larger.

Non-diffuse prior information affects the <u>posterior</u>

odds in both small and large samples, so care must be taken in specifying the prior pdf if this specification is non-diffuse.

Compare this situation to that of Bayesian estimation. There in a large sample the information in the data swamps any prior information, & so the likelihood fcn. becomes all important in large samples.

D) Comparing Regression Models:

What is the relationship between the sampling theory approach of choosing the model with the highest R^2, & the Bayesian approach of comparing & selecting among alternative models? Further, is there a general Bayesian procedure for model selection?

Consider the case of 2 alternative models —

$$M_1 : y = X_1 \beta_1 + u_1$$

$$M_2 : y = X_2 \beta_2 + u_2$$

Note: Both X_1 & X_2 and $(n \times k)$, so both β_1 & β_2 are $(k \times 1)$ i.e. the same no. of exogenous variables appear in both M_1 & M_2. Further, β_1 & β_2 have no common elements.

If M_1 true, then $u_{1t} \sim NID(0, \sigma_1^2)$; $\forall t$.

If M_2 true, then $u_{2t} \sim NID(0, \sigma_2^2)$; $\forall t$.

Choose natural conjugate prior —

$$p(\beta_i ; \sigma_i) = p(\beta_i | \sigma_i)\, p(\sigma_i) \quad ; \; i = 1, 2.$$

where $$p(\beta_i | \sigma_i) = \left(\frac{|C_i|^{1/2}}{(2\pi)^{k/2} \sigma_i^k} \right) \exp\left[-\frac{1}{2\sigma_i^2} (\beta_i - \bar{\beta}_i)' C_i (\beta_i - \bar{\beta}_i) \right]$$

& $$p(\sigma_i) = \left(\frac{K_i}{\sigma_i^{q_i+1}} \right) \exp\left[-\frac{q_i \bar{s}_i^2}{2\sigma_i^2} \right]$$

where $K_i = \left[2 (q_i \bar{s}_i^2 / 2)^{q_i/2} \right] / \Gamma(q_i/2)$.

That is, $p(\sigma_i)$ is an Inverted Gamma.

Note that in the normal pdf for $(\beta_i|\sigma_i)$, the mean of β_i is $\bar{\beta}_i$ & C.M. is $(\sigma^2 C^{-1})$. Both C & β_i must be assigned.

In the I.G. dist'n., both q_i & $\bar{s}_i^2$ have to be assigned.

If we have prior odds on M_1 & M_2 then we get posterior odds —

$$K_{01} = \frac{p(M_1)}{p(M_2)} \frac{\int p(y|\beta_1,\sigma_1,M_1)\, p(\beta_1|\sigma_1)\, p(\sigma_1)\, d\beta_1\, d\sigma_1}{\int p(y|\beta_2,\sigma_2,M_2)\, p(\beta_2|\sigma_2)\, p(\sigma_2)\, d\beta_2\, d\sigma_2}$$

& $p(y|\beta_i,\sigma_i,M_i)$ is the L/F, given M_i true.

So, $$p(y|\beta_i,\sigma_i,M_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma_i}\right)^n *$$
$$\exp\left\{-\frac{1}{2\sigma_i^2}\left[\nu s_i^2 + (\beta_i-\hat{\beta}_i)' X_i' X_i (\beta_i - \hat{\beta}_i)\right]\right\}$$

where $\nu = n-k$; $\nu s_i^2 = (y - X_i\hat{\beta}_i)'(y - X_i\hat{\beta}_i)$; $\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y$.

Now, we have to evaluate:

$$\int p(y|\beta_i,\sigma_i,M_i)\, p(\beta_i|\sigma_i)\, p(\sigma_i)\, d\beta_i\, d\sigma_i; \quad i=1,2.$$

Consider, for example, $i=1$:

$$p(\sigma_1) = \left(\frac{k_1}{\sigma_1^{q_1+1}}\right) \exp\left[-\frac{q_1\bar{s}_1^2}{2\sigma_1^2}\right]$$

$$(y|\beta_1,\sigma_1,M_1) = \left(\frac{1}{(2\pi)^{n/2}\sigma_1^n}\right) \exp.\left\{-\frac{1}{2\sigma_1^2}\left[\nu s_1^2 + (\beta_1-\hat{\beta}_1)' X_1' X_1 (\beta_1-\hat{\beta}_1)\right]\right\}$$

$$p(\beta_1|\sigma_1) = \frac{|C_1|^{1/2}}{(2\pi)^{k/2}\sigma_1^k} \cdot \exp\left\{-\frac{1}{2\sigma_1^2}(\beta_1-\bar{\beta}_1)' C_1 (\beta_1-\bar{\beta}_1)\right\}.$$

So, $$\int = \frac{k_1|C_1|^{1/2}}{(2\pi)^{(n+k)/2}} \cdot \int \frac{1}{\sigma_1^{(n+k+q_1+1)}} \cdot \exp\left\{-\frac{1}{2\sigma_1^2}\left[\frac{q_1\bar{s}_1^2}{1} \cdot \nu s_1^2 + (\beta_1-\bar{\beta}_1)' C_1 (\beta_1-\bar{\beta}_1) + (\beta_1-\hat{\beta}_1)' X_1' X_1 (\beta_1-\hat{\beta}_1)\right]\right\}$$

Integrate w.r.t. β_1 — first complete the square on β_1 in the exponent:

$$\beta_1-\bar{\beta}_1)' C_1 (\beta_1-\bar{\beta}_1) + (\beta_1-\hat{\beta}_1)' X_1' X_1 (\beta_1-\hat{\beta}_1)$$

$$= \beta_1' C_1 \beta_1 + \bar{\beta}_1' C_1 \bar{\beta}_1 - 2\beta_1' C_1 \bar{\beta}_1 + \beta_1' X_1' X_1 \beta_1 + \hat{\beta}_1' X_1' X_1 \hat{\beta}_1$$
$$- 2\beta_1' X_1' X_1 \hat{\beta}_1$$

$$\beta_1'[C_1 + (X_1'X_1)]\beta_1 - 2\beta_1'[C_1\bar{\beta}_1 + (X_1'X_1)\hat{\beta}_1] + \bar{\beta}_1' C_1 \bar{\beta}_1 + \hat{\beta}_1' X_1' X_1 \hat{\beta}_1.$$

And let $A_1 = C_1 + (X_1'X_1)$.

$\therefore$ we obtain:

$$[\beta_1 - A_1^{-1}(C_1\bar{\beta}_1 + X_1'X_1\hat{\beta}_1)]'A_1[\beta_1 - A_1^{-1}(C_1\bar{\beta}_1 + X_1'X_1\hat{\beta}_1)]$$
$$+\bar{\beta}_1'C_1\bar{\beta}_1 + \hat{\beta}_1'(X_1'X_1)\hat{\beta}_1 - (C_1\bar{\beta}_1 + X_1'X_1\hat{\beta}_1)'A_1^{-1}(C_1\bar{\beta}_1 + X_1'X_1\hat{\beta}_1)$$

Let $\tilde{\beta}_1 = A_1^{-1}(C_1\bar{\beta}_1 + X_1'X_1\hat{\beta}_1)$

$$\Rightarrow [\beta_1 - \tilde{\beta}_1]'A_1[\beta_1 - \tilde{\beta}_1] + \bar{\beta}_1'C_1\bar{\beta}_1 + \hat{\beta}_1'(X_1'X_1)\hat{\beta}_1 - \tilde{\beta}_1'A_1\tilde{\beta}_1.$$

$$= (\beta_1 - \tilde{\beta}_1)'A_1(\beta_1 - \tilde{\beta}_1) + (\bar{\beta}_1'C_1\bar{\beta}_1) + (\hat{\beta}_1'X_1'X_1\hat{\beta}_1) - \tilde{\beta}_1'C_1\tilde{\beta}_1 - \tilde{\beta}_1'X_1'X_1\tilde{\beta}_1.$$

$$= (\beta_1 - \tilde{\beta}_1)'A_1(\beta_1 - \tilde{\beta}_1) + (\bar{\beta}_1 - \tilde{\beta}_1)'C_1(\bar{\beta}_1 - \tilde{\beta}_1) + (\hat{\beta}_1 - \tilde{\beta}_1)'X_1'X_1(\hat{\beta}_1 - \tilde{\beta}_1)$$

$$\text{lo, } \int = \left[\frac{k_1|C_1|^{1/2}}{(2\pi)^{(n+k)/2}}\right] * \int \frac{1}{\sigma^{(n+k+q_1+1)}} \cdot \exp\Big\{-\frac{1}{2\sigma_1^2}\Big[q_1\bar{s}_1^2 + vs_1^2$$

$$+ (\beta_1 - \tilde{\beta}_1)'A_1(\beta_1 - \tilde{\beta}_1) + (\bar{\beta}_1 - \tilde{\beta}_1)'C_1(\bar{\beta}_1 - \tilde{\beta}_1) + (\hat{\beta}_1 - \tilde{\beta}_1)'X_1'X_1(\hat{\beta}_1 - \tilde{\beta}_1)\Big]\Big\} \, d\beta_1\, d\sigma_1$$

$$\left(\frac{k_1|C_1|^{1/2}}{(2\pi)^{n/2}}\right)\left(|A|^{-1/2}\right) \int \frac{1}{\sigma_1^{(n+q_1+1)}} \exp\Big\{-\frac{1}{2\sigma_1^2}\Big[q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{1b}\Big]\Big\}$$

$$* \left(\frac{1}{\sigma_1^k}\right)\left(\frac{1}{2\pi}\right)^{k/2} |A|^{1/2} \exp\Big\{-\frac{1}{2\sigma_1^2}\Big[(\beta_1 - \tilde{\beta}_1)'A_1(\beta_1 - \tilde{\beta}_1)\Big]\Big\} d\beta_1\, d\sigma_1$$

where $Q_{1a} = (\bar{\beta}_1 - \tilde{\beta}_1)'C_1(\bar{\beta}_1 - \tilde{\beta}_1)$; $Q_{2a} = (\hat{\beta}_1 - \tilde{\beta}_1)'X_1'X_1(\hat{\beta}_1 - \tilde{\beta}_1)]$

$$= \frac{k_1|C|^{1/2}|A|^{-1/2}}{(2\pi)^{n/2}} \int \left(\frac{1}{\sigma^{n+q_1+1}}\right) \exp\Big\{-\frac{1}{2\sigma_1^2}\Big[q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{2a}\Big]\Big\} d\sigma_1$$

Consider now the integration of this fctn. w.r.t. σ_1:

The integral itself is —

$$\int_0^\infty \left(\frac{1}{\sigma_1^{(n+q_1+1)}}\right) \exp\Big\{-\frac{1}{2\sigma_1^2}\Big[q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{2a}\Big]\Big\} d\sigma_1$$

Which is of the general form

$$\int_0^\infty \left(\frac{1}{\sigma^{a+1}}\right) \exp\Big\{-\frac{b}{2\sigma^2}\Big\} d\sigma.$$

Let $x = (b/2\sigma^2)$

$\therefore dx = -b\sigma^{-3}\,d\sigma$

$\therefore d\sigma = (-\sigma^3/b)\,dx$

& $\sigma = (b/2x)^{1/2}$

$\therefore (\sigma^{a+1})^{-1} = (b/2x)^{-\frac{1}{2}(a+1)}$

$$\therefore \int_0^\infty \left(\frac{1}{\sigma^{a+1}}\right) \exp\left(\frac{-b}{2\sigma^2}\right) d\sigma = \int_\infty^0 \left(\frac{b}{2x}\right)^{-\frac{1}{2}(a+1)} e^{-x} \left(-\sigma^3/b\right) dx$$

$$= \int_0^\infty \left(b/2x\right)^{-\frac{1}{2}(a+1)} e^{-x} (\sigma^3/b)\, dx$$

But, $(\sigma^3/b) = (b^{1/2})/(2x)^{3/2}$

$$\Rightarrow \int_0^\infty \left(\frac{b^{-1/2} \cdot b^{-a/2} \cdot b^{1/2}}{(2x)^{-a/2} \cdot (2x)^{-1/2} \cdot (2x)^{3/2}}\right) e^{-x}\, dx$$

$$= \int_0^\infty \frac{b^{-a/2}}{(2x)^{-a/2+1}} \cdot e^{-x}\, dx$$

$$= \int_0^\infty \left(\frac{b^{-a/2}}{2^{-a/2+1}}\right) \cdot \left(x^{a/2-1}\right) e^{-x}\, dx$$

$$= \left(\tfrac{1}{2}\right)\left(2/b\right)^{a/2} \int_0^\infty \left(x^{a/2-1}\right) e^{-x}\, dx$$

$$= \left(\tfrac{1}{2}\right)\left(2/b\right)^{a/2} \Gamma\left(a/2\right)$$

So, $a = n+q_1$; $b = \left[q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{1b}\right]$

$$\Rightarrow \left(\frac{k_1 |C_1|^{1/2} |A|^{-1/2}}{(2\pi)^{n/2}}\right)\left(\tfrac{1}{2}\right)\left[\frac{2}{q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{1b}}\right]^{(n+q_1)/2} * \Gamma\left(\frac{n+q_1}{2}\right)$$

$$= \left(\tfrac{1}{2}\right)\frac{k_1}{(2\pi)^{n/2}}\left(\frac{|C_1|}{|A_1|}\right)^{1/2} \frac{2^{(n+q_1)/2} \cdot \Gamma\left(\frac{n+q_1}{2}\right)}{\left[q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{1b}\right]^{(n+q_1)/2}}.$$

$$K_{12} = \frac{p(M_1)}{p(M_2)}\left(\frac{k_1}{k_2}\right)\left[\frac{|C_1|/|A_1|}{|C_2|/|A_2|}\right]^{1/2} \cdot \frac{2^{(n+q_1)/2}}{2^{(n+q_2)/2}} \cdot \frac{\Gamma\left(\frac{n+q_1}{2}\right)}{\Gamma\left(\frac{n+q_2}{2}\right)} *$$

$$\frac{\left(q_1\bar{s}_1^2 + vs_1^2 + Q_{1a} + Q_{1b}\right)^{-(n+q_1)/2}}{\left(q_2\bar{s}_2^2 + vs_2^2 + Q_{2a} + Q_{2b}\right)^{-(n+q_2)/2}}.$$

But,
$$K_i = 2\left(\frac{q_i\bar{s}_i^2}{2}\right)^{q_i/2} \Big/ \Gamma\left(q_i/2\right)$$

$$\therefore \left[K_i \cdot 2^{(n+q_i)/2}\right] = \left[\frac{1}{\Gamma\left(\frac{q_i}{2}\right)} \cdot \left(q_i\bar{s}_i^2\right)^{q_i/2} \cdot 2^{n/2+1}\right] \quad ; \quad i = 1, 2.$$

$$\therefore K_{12} = \frac{p(M_1)}{p(M_2)} \cdot \left[\frac{|C_1|/|A_1|}{|C_2|/|A_2|}\right]^{1/2} \frac{\Gamma\left(\frac{n+q_1}{2}\right)/\Gamma\left(\frac{q_1}{2}\right)}{\Gamma\left(\frac{n+q_2}{2}\right)/\Gamma\left(q_2/2\right)} * \frac{\left(q_1\bar{s}_1^2\right)^{q_1/2}}{\left(q_2\bar{s}_2^2\right)^{q_2/2}} .$$

$$* \frac{\left(q_1\bar{s}_1^2 + \nu s_1^2 + Q_{1a} + Q_{1b}\right)^{-(n+q_1)/2}}{\left(q_2\bar{s}_2^2 + \nu s_2^2 + Q_{2a} + Q_{2b}\right)^{-(n+q_2)/2}} .$$

Let $\delta_i = \left(q_i\bar{s}_i^2 + \nu s_i^2 + Q_{ia} + Q_{ib}\right)/n.$

$$\therefore \frac{\left(q_1\bar{s}_1^2 + \nu s_1^2 + Q_{1a} + Q_{1b}\right)^{-(n+q_1)/2}}{\left(q_2\bar{s}_2^2 + \nu s_2^2 + Q_{2a} + Q_{2b}\right)^{-(n+q_2)/2}} = \frac{\left(n\delta_1\right)^{-(n+q_1)/2}}{\left(n\delta_2\right)^{-(n+q_2)/2}} .$$

$$= \left(\frac{n^{-n/2} \cdot n^{-q_1/2}}{n^{-n/2} \cdot n^{-q_2/2}}\right)\left(\frac{\delta_1}{\delta_2}\right)^{-n/2}\left(\frac{\delta_1^{-q_1/2}}{\delta_2^{-q_2/2}}\right)$$

$$= \left(\frac{\delta_1}{\delta_2}\right)^{-n/2}\left(\frac{\left(n\delta_1\right)^{-q_1/2}}{\left(n\delta_2\right)^{-q_2/2}}\right).$$

So,
$$K_{12} = \frac{p(M_1)}{p(M_2)}\left[\frac{|C_1|/|A_1|}{|C_2|/|A_2|}\right]^{+1/2} * \frac{\Gamma\left(\frac{n+q_1}{2}\right)/\Gamma\left(\frac{q_1}{2}\right)}{\Gamma\left(\frac{n+q_2}{2}\right)/\Gamma\left(q_2/2\right)} * \left(\frac{\delta_1}{\delta_2}\right)^{-n/2}$$

$$* \frac{\left(n\delta_1\right)^{-q_1/2}}{\left(n\delta_2\right)^{-q_2/2}} * \frac{\left(q_1\bar{s}_1^2\right)^{q_1/2}}{\left(q_2\bar{s}_2^2\right)^{q_2/2}}$$

$$= \frac{p(M_1)}{p(M_2)}\left[\frac{|C_1|/|A_1|}{|C_2|/|A_2|}\right]^{1/2}\left(\frac{\delta_1}{\delta_2}\right)^{-n/2}\left\{\frac{\Gamma\left(\frac{n+q_1}{2}\right)/\Gamma\left(\frac{q_1}{2}\right)}{\Gamma\left(\frac{n+q_2}{2}\right)/\Gamma\left(\frac{q_2}{2}\right)} * \frac{\left(q_1\bar{s}_1^2/n\delta_1\right)^{q_1/2}}{\left(q_2\bar{s}_2^2/n\delta_2\right)^{q_2/2}}\right\}$$

$$\therefore K_{12} = \frac{p(M_1)}{p(M_2)}\left[\frac{|C_1|/|A_1|}{|C_2|/|A_2|}\right]^{1/2}\left(\frac{\delta_1}{\delta_2}\right)^{-n/2}\left(\frac{\bar{s}_1^2/\delta_1}{\bar{s}_2^2/\delta_2}\right)\frac{f_{q_1 n}\left(\bar{s}_1^2/\delta_1\right)}{f_{q_2 n}\left(\bar{s}_2^2/\delta_2\right)}$$

Now, look at the four terms making up the posterior odds, K_{12} :

(a) $p(M_1)/p(M_2)$ = prior odds.

(b) $|C_i|/|A_i|$ is the relative precision for the prior pdf w.r.t. the precision for the posterior pdf

Now, K_{12} increases if $|C_1|/|A_1|$ increases relative to $|C_2|/|A_2|$ This implies more prior information in

M_1 relative to that in M_2.

(c) Look at $(\delta_1/\delta_2)^{-n/2}$. Note that δ_i reflects what the sample has to say about the goodness of fit of the models —

$$\delta_i = (\nu s_i^2 + Q_{1a} + Q_{1b} + q_i \bar{s}_i^2)/n.$$

And νs_i^2 = residual sum of squares; so high $\nu s_i^2 \Rightarrow$ low goodness of fit.

& $Q_{1a} = (\bar{\beta}_1 - \tilde{\beta}_1)' C_1 (\bar{\beta}_1 - \tilde{\beta}_1)$; the closer together are $\bar{\beta}_1$ & $\tilde{\beta}_1$, the smaller is Q_{1a}. So, if Q_{1a} & Q_{1b} are high, then the sample & prior information are incompatible.

So low $\delta_i \Rightarrow$ good fit.

If δ_1 high relative to δ_2, then M_1 is poor rel. to M_2.

So $\left(\frac{\delta_1}{\delta_2}\right)^{-n/2}$ high $\Longrightarrow$ favour M_1. (NB negative index).

d) $\left(\frac{\bar{s}_1^2/\delta_1}{\bar{s}_2^2/\delta_2}\right) \frac{f_{q_1,n}(\bar{s}_1^2/\delta_1)}{f_{q_2,n}(\bar{s}_2^2/\delta_2)}$ reveals a dependence of K_{12} on the prior information regarding σ_1 & σ_2. If the prior & sample information regarding σ_i diverge, then K_{12} is affected.

Now, what happens to K_{12} if $n \to \infty$?

If $n \to \infty$ and $p(M_1) = p(M_2)$, then

$$K_{12} \longrightarrow \left(\frac{s_1^2}{s_2^2}\right)^{-(n+q)/2} \cdot \frac{\exp(-q\bar{s}^2/2s_1^2)}{\exp(-q\bar{s}^2/2s_2^2)}$$

where $\bar{s}_1^2 = \bar{s}_2^2 = \bar{s}^2$; $q_1 = q_2 = q$.

Then: if $s_1^2 = s_2^2$, $\Longrightarrow K_{12} = 1$

if $s_1^2 > s_2^2$, $\Longrightarrow K_{12} < 1$

if $s_1^2 < s_2^2$, $\Longrightarrow K_{12} > 1$

If the prior information becomes diffuse, then $|C_i| \to 0$, & $q_1 = q_2 = q \to 0$. Assume that $p(M_1) = p(M_2)$ & $\bar{s}_1^2 = \bar{s}_2^2$

Then : $k_{12} = (s_1^2/s_2^2)^{-n/2}$

If $s_1^2 = s_2^2$; $k_{12} = 1$

$s_1^2 < s_2^2$; $k_{12} > 1$

$s_1^2 > s_2^2$; $k_{12} < 1$

Now, we saw that if we act to choose the model with the lower expected loss & if the loss fcn. is symmetric, then we just pick the model with the high posterior probability.

So, if we see $k_{12} > 1$, choose M_1 over M_2. And we see above that this arises when $s_1^2 < s_2^2$; ie when M_1 has a higher R^2 than has M_2. This is true if either e have a symmetric loss function & a diffuse prior; or we have large n, & the other assumptions noted.

E) Comparing Distributed Lag Models:

$$C_t = \lambda C_{t-1} + (1-\lambda) Y_t + u_t - \lambda u_{t-1}$$

M1 : $u_t - \lambda u_{t-1} = \epsilon_{1t} \sim NID(0, \sigma_1^2)$

M2 : $u_t = \epsilon_{2t} \sim NID(0, \sigma_2^2)$

If $\lambda \neq 0$, then M1 & M2 are distinct.

Try 3 alternative prior pdf's —

(a) $p(\lambda, k, \sigma_i) \propto (1/\sigma_i)$; $i = 1, 2$; $0 < \lambda, k < 1$.

(b) $p(\lambda, k, \sigma_i) \propto [\lambda^{34}(1-\lambda)^{14} k^{89}(1-k)^9]/\sigma_i$

(c) $p(\lambda, k, \sigma_i) \propto [\lambda(1-\lambda)^7 k^{89}(1-k)^9]/\sigma_i$

i.e. independent priors, with $p(k)$, $p(\lambda)$, β-dist'n.

$$\therefore K_{12} = \frac{p(M_1)}{p(M_2)} \frac{\iiint p_1(C|\lambda,k,\sigma_1)\, p(\lambda,k,\sigma_1)\, d\lambda\, dk\, d\sigma_1}{\iiint p_2(C|\lambda,k,\sigma_2)\, p(\lambda,k,\sigma_2)\, d\lambda\, dk\, d\sigma_2}$$

Now, $p_1(C|\lambda,k,\sigma_1) \propto \left(\frac{1}{\sigma_1^T}\right) \exp\left\{-\frac{1}{2\sigma_1^2}[C-\lambda C_{-1}-k(1-\lambda)Y]' * [C-\lambda C_{-1}-k(1-\lambda)Y]\right\}$

And,

$$p_2(C|\lambda,k,\sigma_2) \propto \frac{|G|^{-1/2}}{\sigma_2^T} \exp\left\{-\frac{1}{2\sigma_2^2}[C-\lambda C_{-1}-k(1-\lambda)Y] * G^{-1}[C-\lambda C_{-1}-k(1-\lambda)Y]\right\}$$

Now, $\iiint p(\lambda,k,\sigma_1)\, p_1(C|\lambda,k,\sigma_1)\, d\lambda\, dk\, d\sigma_1$

$$\iiint p(\lambda,k)\left(\frac{1}{\sigma_1^{T+1}}\right)\exp\left\{-\frac{1}{2\sigma_1^2}[C-\lambda C_{-1}-k(1-\lambda)Y]'[C-\lambda C_{-1}-k(1-\lambda)Y]\right. d\lambda\, dk\, d\sigma$$

$$= \gamma \iint p(\lambda,k)\left\{[C-\lambda C_{-1}-k(1-\lambda)Y]'[C-\lambda C_{-1}-k(1-\lambda)Y]\right\}^{-T/2} d\lambda\, dk.$$

$$\iiint p(\lambda,k,\sigma_2)\, p_2(C|\lambda,k,\sigma_2)\, dk\, d\lambda\, d\sigma_2$$

$$\gamma \iint p(\lambda,k)\left\{[C-\lambda C_{-1}-k(1-\lambda)Y]'G^{-1}[C-\lambda C_{-1}-k(1-\lambda)Y]\right\}^{-T/2} dk\, d\lambda.$$

$$\therefore K_{12} = \frac{p(M_1)}{p(M_2)} \frac{\iint p(\lambda,k)\left\{[C-\lambda C_{-1}-k(1-\lambda)Y]'[C-\lambda C_{-1}-k(1-\lambda)Y]\right\}^{-T/2} d\lambda\, dk}{\iint p(\lambda,k)\left\{[C-\lambda C_{-1}-k(1-\lambda)Y]'G^{-1}[C-\lambda C_{-1}-k(1-\lambda)Y]\right\}^{-T/2} d\lambda\, dk}$$

And these integrals may be evaluated using bivariate numerical techniques, given $p(\lambda,k)$ as in (a), (b), (c) above

<u>Example</u>: Consider Koyck's distributed lag model –

$$y_t = \sum_0^\infty \alpha_i x_{t-i} + u_t$$

$$\alpha_i = k\binom{r+i-1}{i}(1-\lambda)^r \lambda^i \quad ; \; 0<\lambda<1 \; ; \; r>0.$$

$$= k w_i$$

Then w_i is just a probability from a <u>Pascal</u> pdf.

Mean $= r\lambda/(1-\lambda)$; var. $= r\lambda/(1-\lambda)^2$.

If $r=1$, we have a <u>geometric distribution.</u>

Let $L^i x_t = x_{t-i}$

$\therefore \sum_0^\infty \binom{r+i-1}{i} L^i (1-L)^r = 1$; since we have a pdf.

$\therefore \sum_0^\infty \binom{r+i-1}{i} L^i = (1-L)^{-r}$

$$\therefore y_t = \sum_0^\infty \alpha_i L^i x_t + u_t$$
$$= \sum_0^\infty k \binom{r+i-1}{i} (1-\lambda)^r \lambda^i L^i x_t + u_t$$
$$= \sum_0^\infty k \binom{r+i-1}{i} (1-\lambda)^r (\lambda L)^i x_t + u_t$$
$$= k(1-\lambda)^r x_t \cdot \sum_0^\infty \binom{r+i-1}{i} (\lambda L)^i x_t + u_t$$
$$= k(1-\lambda)^r x_t \cdot (1-\lambda L)^{-r} x_t + u_t.$$

$\therefore (1-\lambda L)^r y_t = k(1-\lambda)^r x_t + (1-\lambda L)^r u_t$

Now, assume that $(1-\lambda L)^r u_t = \varepsilon_t \sim NID(0,\sigma^2)$.

The problem then becomes — given our observations & prior assumptions about the parameters, what are the <u>posterior</u> probabilities of

$H_1 : r=1$; $H_2 : r=2$; $H_3 : r=3$; $H_4 : r=4$. ?

$M_1 : r=1 \Rightarrow y_t = \lambda y_{t-1} + k(1-\lambda)x_t + \varepsilon_{1t}$

$M_2 : r=2 \Rightarrow y_t = 2\lambda y_{t-1} - \lambda^2 y_{t-2} + k(1-\lambda)^2 x_t + \varepsilon_{2t}$

$M_3 : r=3 \Rightarrow y_t = 3\lambda y_{t-1} + 3\lambda^2 y_{t-2} + \lambda^3 y_{t-3} + k(1-\lambda)^3 x_t + \varepsilon_{3t}$

$M_4 : r=4 \Rightarrow y_t = 4\lambda y_{t-1} - 6\lambda^2 y_{t-2} + 4\lambda^3 y_{t-3} - \lambda^4 y_{t-4} + k(1-\lambda)^4 x_t + \varepsilon_{4t}$

For $r=i$, we have a prior pdf, $p(\lambda, k, \sigma_i | r=i.)$

Now,

$$K_{12} = \frac{Pr\{r=1\}}{Pr\{r=2\}} \cdot \frac{\iiint p(\lambda, k, \sigma_1 | r=1)\, \ell_1 \, d\lambda\, dk\, d\sigma_1}{\iiint p(\lambda, k, \sigma_2 | r=2)\, \ell_2 \, d\lambda\, dk\, d\sigma_2}$$

Then we do the same for each pair of models, get the posterior <u>odds</u>, & hence the posterior <u>probabilities</u>.

Let the latter be π_i ; $i = 1, \ldots 4$.
Then $\sum_1^4 \pi_i = 1 = (\pi_1 + \pi_2 + \pi_3 + \pi_4$

$$\therefore \quad 1 = \pi_1 \left(1 + \pi_2/\pi_1 + \pi_3/\pi_1 + \pi_4/\pi_1\right)$$

$$1 = \pi_1 \left(1 + k_{21} + k_{31} + k_{41}\right)$$

$$\therefore \quad \pi_1 = 1/\left(1 + k_{21} + k_{31} + k_{41}\right) \qquad \underline{\underline{\text{etc}}}$$

i.e. in general, $\pi_i = 1/\left(1 + \sum_{j \neq i} k_{ji}\right)$.

So, we use numerical integration to get the k_{ij}, & then get the π_i from these.

Zellner: Chapter 11.

"Control Problems"

In control problems there are three basic ingredients —

(a) A criterion function, to be maximized or minimized.

(b) A model containing some variables appearing in the criterion function.

(c) A subset of the model's variables which can be controlled.

The problem usually can be solved without too much difficulty for non-stochastic systems. If the system is stochastic, then the problem is far more complex.

The Bayesian approach is particularly useful in this area, since it treats stochastic elements and uncertainty about parameters in a systematic & unified fashion.

When the control problem involves optimization over several time periods with stochastic elements and uncertainty about the parameter values present, another basic complication arises. The way in which we set the control variables in any period will affect the information that we get in later periods regarding parameter values. Then a full solution to the optimization problem must take account of the flow of information about parameter values as we move through time. This is an "adaptive control" problem. The Bayesian solution gives a sequence of optimizing actions (i.e. settings of the control variables) which achieves the control goal while simultaneously showing how we learn from new data.

(A) One-Period Control Problems:

Let $y_t = \beta x_t + u_t$

Assume that the values of x_t can be controlled, & also assume that $u_t \sim NID(0, \sigma^2)$; $\forall t = 1, 2, \ldots, T$.

Now, consider period $T+1$.

Let $Z \equiv y_{T+1}$ & let $w \equiv x_{T+1}$

$\therefore Z = w\beta + u_{T+1}$ & $u_{T+1} \sim NID(0, \sigma^2)$

Suppose that we have not yet observed Z, nor determined a value for w. Let 'a' denote the target value for Z.

Let $L(Z, a) = (Z-a)^2$ ~ i.e. a quadratic loss fctn.

Now, we can show that minimizing expected loss when the loss fctn. is quadratic, is equivalent to maximizing expected utility —

Let $U = c_0 + 2c_1 z - c_2 z^2$; $c_1, c_2 > 0$

Then, completing the square on z:

$$U = -c_2\left(z - \frac{c_1}{c_2}\right)^2 + \left(c_0 + \frac{c_1^2}{c_2}\right)$$

$$= a_0 - c_2(z-a)^2$$

where $a = (c_1/c_2)$; $a_0 = c_0 + (c_1^2/c_2)$.

Now, U is a decreasing function of $(z-a)^2$. So if we minimize $E(z-a)^2$, we maximize $E(U)$.

Now, $z = y_{T+1}$, so z is a random variable. Hence $L(z,a)$ is random, & so we consider $E[L(z,a)]$.

So, $\min_{(w)} E[L(z,a)] = \min_{(w)} E[(z-a)^2]$; is our problem.

Now, $z = \beta w + u$; so z depends on the control variable, w.

Now, $$E[L(z,a)] = \int_{-\infty}^{\infty} L(z,a)\, p(z|y,w)\, dz$$

$$= \int_{-\infty}^{\infty} (z-a)^2\, p(z|y,w)\, dz.$$

Now, $p(z|y,w)$ is the predictive pdf for z, given the data & the value $w = x_{T+1}$

Now, $p(z|y,w)$ is in the form of a Student-t dist'n. If we have used the diffuse prior — $p(\beta,\sigma) \propto 1/\sigma$.

i.e. $$p(z|y,w) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\Gamma(\frac{1}{2})}\left(\frac{g}{\nu}\right)^{1/2}\left[1 + \frac{g}{\nu}\left(z - w\hat{\beta}\right)^2\right]^{-(\nu+1)/2}.$$

where $\nu = (T-1)$; $\hat{\beta} = \left(\sum_1^T x_t y_t\right) / \left(\sum_1^T x_t^2\right)$

$$g = \left[s^2\left(1 + w^2/\Sigma x_t^2\right)\right] \; ; \; \& \; \nu s^2 = \Sigma(y_t - \hat{\beta} x_t)^2.$$

Then the mean of $p(z|y,w)$ is $w\hat{\beta}$, which is clearly a function of w. And g is also a function of w, the control variable.

Now, $$E(z-a)^2 = E\left[z - E(z) - (a - E(z))\right]^2$$
$$= Var(z) + (a - E(z))^2$$

Now, $t = \sqrt{g}\,(z - w\hat{\beta})$ is of the t-dist'n. & has mean zero, and variance $v/(v-2)$.

$$\text{So, } z = \left(w\hat{\beta} + t/\sqrt{g}\right) = \left(w\hat{\beta} + \frac{1}{\sqrt{g}}\cdot t\right)$$

$$\therefore \operatorname{var}(z) = \left(\frac{1}{\sqrt{g}}\right)^2 \operatorname{Var}(t)$$

$$= v/g(v-2).$$

$$\therefore E(z-a)^2 = \left(\frac{v}{g(v-2)}\right) + (a - E(z))^2$$

$$= \frac{v}{g(v-2)} + (a - w\hat{\beta})^2$$

$$= \left(\frac{vs^2}{v-2}\right)\left[1 + \frac{w^2}{\Sigma x_t^2}\right] + (a - w\hat{\beta})^2$$

Now, to find the value of w to min. $E(z-a)^2$:

$$\frac{\partial E(z-a)^2}{\partial w} = \left(\frac{vs^2}{v-2}\right)\left(\frac{2w}{\Sigma x_t^2}\right) - 2(a - w\hat{\beta})\hat{\beta}$$

$$= 0$$

$$\therefore w^*\left[\left(\frac{2}{\Sigma x_t^2}\right)\left(\frac{vs^2}{(v-2)}\right) + 2\hat{\beta}^2\right] - 2a\hat{\beta} = 0$$

$$\therefore w^* = \left(\frac{2a\hat{\beta}}{2\hat{\beta}^2}\right)\left(1 + \frac{2vs^2}{(v-2)\hat{\beta}^2\Sigma x_t^2}\right)^{-1}$$

$$= \left(\frac{a}{\hat{\beta}}\right)\left[1 + \frac{vs^2}{\Sigma x_t^2 (v-2)\hat{\beta}^2}\right]^{-1}$$

Let $m_{xx} = \Sigma x_t^2$; $\bar{s}^2 = \left(\frac{vs^2}{v-2}\right)$; $t_0^2 = \hat{\beta}^2 m_{xx}/\bar{s}^2$.

$$\therefore w^* = \left(\frac{a}{\hat{\beta}}\right)\left(\frac{1}{1 + 1/t_0^2}\right).$$

So, w^* is the product of 2 factors — first, $(a/\hat{\beta})$ is the target value as a ratio of the mean of the posterior pdf. Second, we have a factor which is a function of t_0^2. Now $t_0^2 = \hat{\beta}^2 m_{xx}/\bar{s}^2$, and since $(\bar{s}^2/m_{xx})$ is the posterior variance of β, then t_0^2 is the square of the coefficient of variation associated with the posterior pdf for β.

Now, $(m_{xx}/\bar{s}^2)$ measures the precision of estimation. As precision increases, so too does t_0^2,

$$\text{So } \left(\frac{1}{1 + 1/t_0^2}\right) \longrightarrow 1.$$

So, if our estimate of β is <u>precise</u>, then $w^* \to (a/\hat{\beta})$. However, if we approximate w^* by $(a/\hat{\beta})$ we may have a sub-optimal setting of the control variable.

Now, if we were to adopt the "certainty equivelance" approach, then we would proceed conditional on the assumption that $\beta = \hat{\beta}$.

Then, $L(z,a) \simeq (w\hat{\beta} - a)^2$

Then, $\dfrac{\partial L(z,a)}{\partial w} = 2\hat{\beta}(w\hat{\beta} - a) = 0$

$$\therefore w_{ce} = (a/\hat{\beta})$$

So the certainty equivelance solution to our minimization problem is just the <u>first</u> factor of our <u>optimal</u> solution, w^*. The second factor, which reflects the precision with which β has been estimated, does not appear in w_{ce}, since there we have <u>equated</u> $\hat{\beta}$ with β. We are <u>certain</u> that the two are <u>equivelant</u>.

If we use the (sub-optimal) control value w_{ce}, then this gives rise to a <u>higher</u> expected loss than if w^* is employed. Compare the two results for $L(z,a) = (z-a)^2$.

$$E[L \mid w = w^*] = E[(z-a)^2 \mid w = w^*]$$

$$= \left(\frac{vs^2}{v-2}\right)\left[1 + \left(\frac{1}{\Sigma x_t^2}\right)\left[\frac{a^2}{\hat{\beta}^2}\left(\frac{t_0^2}{1+t_0^2}\right)^2\right]\right]$$

$$+ \left[a - \hat{\beta}\left(\frac{a}{\hat{\beta}}\right)\left(\frac{t_0^2}{1+t_0^2}\right)\right]^2$$

$$= \bar{s}^2 + \left(\frac{\bar{s}^2}{\Sigma x_t^2}\right)\left[\frac{a^2}{\hat{\beta}^2}\left(\frac{t_0^2}{1+t_0^2}\right)^2\right] + a^2\left[1 - \left(\frac{t_0^2}{1+t_0^2}\right)\right]^2$$

$$= \bar{s}^2 + \left(\frac{\hat{\beta}^2}{t_0^2}\right)\left[\frac{a^2}{\hat{\beta}^2}\left(\frac{t_0^2}{1+t_0^2}\right)^2\right] + a^2\left[1 - \left(\frac{t_0^2}{1+t_0^2}\right)\right]^2$$

$$= \bar{s}^2 + \left(\frac{a^2}{t_0^2}\right)\left(\frac{t_0^2}{1+t_0^2}\right)^2 + a^2\left[1 + \left(\frac{t_0^2}{1+t_0^2}\right)^2 - \frac{2t_0^2}{1+t_0^2}\right]$$

$$= \bar{s}^2 + \left(\frac{a^2}{t_0^2}\right)\left(\frac{t_0^2}{1+t_0^2}\right)^2 + a^2 + a^2\left(\frac{t_0^2}{1+t_0^2}\right)^2 - \frac{2a^2t_0^2}{1+t_0^2}.$$

$$= \bar{s}^2 + a^2\left(\frac{t_0^2}{1+t_0^2}\right)^2\left[1 + \frac{1}{t_0^2}\right] + a^2 - \frac{2a^2t_0^2}{1+t_0^2}$$

$$= \bar{s}^2 + a^2\left(\frac{t_0^2}{1+t_0^2}\right)^2\left[\frac{t_0^2+1}{t_0^2}\right] + a^2 - 2a^2\left(\frac{t_0^2}{1+t_0^2}\right)$$

$$= \bar{s}^2 + \frac{a^2t_0^2}{1+t_0^2} + a^2 - \frac{2a^2t_0^2}{1+t_0^2}$$

$$= \bar{s}^2 + a^2 - \left(\frac{a^2t_0^2}{1+t_0^2}\right)$$

$$= \bar{s}^2 + \frac{a^2 + a^2t_0^2 - a^2t_0^2}{1+t_0^2}$$

$$= \bar{s}^2 + \left(\frac{a^2}{1+t_0^2}\right).$$

And, $E[L \mid w = w_{ce}] = E[(z-a)^2 \mid w = w_{ce}]$

$$= \left(\frac{US^2}{U-2}\right)\left[1 + \left(\frac{1}{\Sigma x_t^2}\right)\left(\frac{a^2}{\hat{\beta}^2}\right)\right] + \left(a - \hat{\beta}\,\frac{a}{\hat{\beta}}\right)^2$$

$$= \bar{s}^2 + \frac{\hat{\beta}^2}{t_0^2}\left(\frac{a^2}{\hat{\beta}^2}\right)$$

$$= \left(\bar{s}^2 + \frac{a^2}{t_0^2}\right) \quad ////$$

$$\therefore E[L \mid w = w_{ce}] - E[L \mid w = w^*]$$

$$= \bar{s}^2 + \frac{a^2}{t_0^2} - \bar{s}^2 - \frac{a^2}{1+t_0^2}$$

$$= a^2\left[\frac{1 + t_0^2 - t_0^2}{t_0^2(1+t_0^2)}\right]$$

$$= \frac{a^2}{t_0^2(1+t_0^2)}.$$

So, the <u>increase</u> in expected loss when w_{ce} is used instead of w^* depends not only on <u>precision</u>, as measured by t_0, but also on the value of the target, 'a'. Only if $a=0$ is there ~~no~~ no increase in expected loss when w_{ce} is used in place of w^*. And this is because we assumed that the regression goes through the origin <u>i.e.</u> there is a <u>zero</u> intercept.

For target values far from the origin, then the <u>extra</u> loss when w_{ce} is used may well be substantial.

Write the <u>relative expected loss</u> as:

$$REL = \frac{E[L \mid w = w^*]}{E[L \mid w = w_{ce}]} = \left(\frac{\bar{s}^2 + a^2/(1+t_0^2)}{\bar{s}^2 + a^2/t_0^2}\right)$$

$$= \frac{\left[1 + \left\{\frac{a^2}{\bar{s}^2}\right\}/(1+t_0^2)\right]}{\left[1 + \left\{a^2/\bar{s}^2\right\}/(t_0^2)\right]} \leq 1.$$

$$= \cancel{\left(\frac{t_0^2}{1+t_0^2}\right)}$$

In particular, the reduction in expected loss obtained by using w^* (rather than w_{ce}) is <u>greater</u> the <u>smaller</u> is t_0^2 and the <u>larger</u> is $(a/\bar{s})^2$.

Consider the first statement — as t_0^2 gets small, it becomes relatively better to use w^*. This makes sense, since $t_0^2 = \hat{\beta}^2/(\bar{s}^2/m_{xx})$. And $\bar{s}^2/m_{xx}$ is a variance related to $\hat{\beta}$.

$\therefore$ as t_0^2 gets small, variance $(\hat{\beta})$ gets big.

So, the less <u>accurate</u> is your estimate of β, the more profitable it becomes to use w^* instead of w_{ce}. ✓

Example: Let $\begin{cases} y_t = \Delta GNP \\ x_t = \Delta M_t \end{cases}$

Data for 1921 – 1929 given, use a diffuse prior pdf, & find x_{T+1} for 1930 & y_{T+1} for 1930 = \$10m.

Take $L(z,a) = (z-10)^2$; $z = y_{T+1}$.

The regression estimates are $\hat{y}_t = \underset{(0.8813)}{2.0676}\, x_t$

$\therefore\ w^* = 2.843$; $E(L \mid w = w^*) = 55.29$

& $w_{ce} = 4.837$; $E(L \mid w = w_{ce}) = 60.03$.

So there is an 8% gain in using w^* cf. w_{ce}.

Now let us generalize the analysis a little. In particular, modify the loss fctn. so that it takes account of possible costs involved in changing the control variable.

Let $L = (z-a)^2 + c(w - x_T)^2$

where c is <u>known</u>, non-negative, & constant. And $w = x_{T+1}$.

$$\therefore\ E(L) = E(z-a)^2 + E\left[c(x - w_T)^2\right]$$

$$= var(z) + (a - E(z))^2 + c(w - x_T)^2$$

$$= \left(\frac{VS^2}{V-2}\right)\left(1 + \frac{w^2}{\Sigma x_t^2}\right) + (a - w\hat{\beta})^2 + (c - x_T)^2$$

$$= \bar{S}^2\left(1 + \frac{w^2}{M_{xx}}\right) + \hat{\beta}^2\left(a/\hat{\beta} - w\right)^2 + c(w - x_T)^2$$

And we wish to minimize $E(L)$.

$$\begin{cases} \text{We min. 1st term if } w = 0. \\ \text{" " 2nd " " } w = (a/\hat{\beta}) \\ \text{" " 3rd " " } w = x_T. \end{cases}$$

Now, $$\frac{\partial E(L)}{\partial w} = \bar{S}^2\left(\frac{2w}{M_{xx}}\right) + 2\hat{\beta}^2(-1)\left(a/\hat{\beta} - w\right) + 2c(1)(w - x_T) = 0$$

$$\therefore\ w^{**}\left[\frac{2\bar{S}^2}{M_{xx}} + 2\hat{\beta}^2 + 2c\right] - 2a\hat{\beta} - 2c\,x_T = 0$$

$$\therefore w^{**} = \left(\frac{a\hat{\beta} + cx_T}{\hat{\beta}^2 + c + \bar{s}^2/M_{xx}} \right)$$

and this just a weighted average of zero, $(a/\hat{\beta})$, and x_T.

Then: (a) If $c = 0$; $w^{**} = w^*$.

(b) If $c \to \infty$; $w^{**} \to x_T$ (as cost rises, we'll make little shift from x_T.)

(c) $\begin{cases} \text{If } x_T < w^*, \text{ then } x_T < w^{**} < w^* \\ \text{If } x_T > w^*, \text{ then } w^* < w^{**} < x_T. \end{cases}$

B) Single-Period Control of Multiple Regressions:

Let $y' = (y_1, y_2, ---, y_T)$ be generated according to

$$y = X\beta + u$$

Assume that all k explanatory variables can be controlled. (This assumption may be relaxed. Zellner looks at this possibility later in the chapter.)

Let $u_t \sim NID(0, \sigma^2)$; $t = 1, 2, ---, T$.

Assume $p(\beta, \sigma) \propto 1/\sigma$; $0 < \sigma < \infty$; $-\infty < \beta_i < \infty$; $i = 1, --, k$.

Let $z = y_{T+1}$, $w = (x_{1T+1}, x_{2T+1}, ----, x_{kT+1})$

$$\therefore z = w'\beta + u_{T+1}.$$

Where $u_{T+1} \sim NID(0, \sigma^2)$

Let $L(z, a) = (z-a)^2$

$$\therefore p(z|y, w) \propto \left[v + (z - w'\hat{\beta})^2 H \right]^{-(v+1)/2}$$

where $\hat{\beta} = (X'X)^{-1}X'y$

$$H = \frac{1}{s^2}\left(1 + w'(X'X)^{-1}w\right)^{-1}$$

$$v = (T-k)$$

$$vs^2 = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

& $E(z) = w'\hat{\beta}$,

$$var(z) = \left(\frac{H^{-1}v}{v-2}\right) = \bar{s}^2\left[1 + w'(X'X)^{-1}w\right]$$

where $\bar{s}^2 = \left(\frac{vs^2}{v-2}\right)$.

Now, $E(z-a)^2 = var(z) + (a - E(z))^2$

$$= \bar{s}^2\left[1 + w'(X'X)^{-1}w\right] + (a - w'\hat{\beta})^2$$

$$\therefore \frac{\partial E(L)}{\partial w} = 2\bar{s}^2(X'X)^{-1}w - 2a\hat{\beta} + 2\hat{\beta}\hat{\beta}'w$$

$$= 0$$

$$\therefore \left[\bar{s}^2(X'X)^{-1} + \hat{\beta}\hat{\beta}'\right]w^* = a\hat{\beta}$$

$$\therefore w^* = \frac{a\hat{\beta}}{\left[\bar{s}^2(X'X)^{-1} + \hat{\beta}\hat{\beta}'\right]}$$

$$\therefore w^* = \frac{X'X\hat{\beta}a}{(\bar{s}^2 + \hat{\beta}'X'X\hat{\beta})}$$

$$\therefore a = w^*\hat{\beta}'\left(1 + \frac{1}{(\hat{\beta}'X'X\hat{\beta})/\bar{s}^2}\right)$$

$$= w^{*\prime}\hat{\beta}^*, \text{ say.}$$

Then $\hat{\beta}^* = \hat{\beta}\left(1 + \frac{1}{(\hat{\beta}'X'X\hat{\beta})/\bar{s}^2}\right)$

And this is just the <u>implied estimator</u> for $\hat{\beta}$.

Then $E(L \mid w = w^*) = \bar{s}^2\left[1 + w^{*\prime}(X'X)^{-1}w^*\right] + (a - w^{*\prime}\hat{\beta})^2$

$$= \bar{s}^2\left[1 + \frac{a'\hat{\beta}'(X'X)(X'X)^{-1}(X'X)\hat{\beta}a}{\bar{s}^2 + \hat{\beta}X'X\hat{\beta}}\right.$$

$$\left. + \left(a - \frac{a'\hat{\beta}X'X\hat{\beta}}{\bar{s}^2 + \hat{\beta}'X'X\hat{\beta}}\right)^2\right]$$

$$= \bar{s}^2\left[1 + \frac{a^2}{\bar{s}^2 + \hat{\beta}'X'X\hat{\beta}}\right]$$

This is the expected loss when w is set at its <u>optimal</u> value, w^*. The first term, $\bar{s}^2$, does <u>not</u> disappear as $n \to \infty$, but the 2nd term $\to 0$ as $n \to \infty$

& Also, $\hat{\beta}^* = \hat{\beta}\left(1 + \frac{1}{(\hat{\beta}'X'X\hat{\beta})/\bar{s}^2}\right) \to \hat{\beta}$, as $n \to \infty$.

<u>Note</u>: Section (C) on Multivariate Regression is omitted.

(D) Sensitivity of Control to Form of the Loss Function:

Consider the whole problem of robustness under changes in the loss function. i.e. suppose that we make an incorrect assumption about the form of the loss function — to what extent will this affect our results?

Zellner performs a Monte Carlo study with 15 observations on the model:

$$y_t = 2.0 x_t + u_t.$$

where: $u_t \sim N(0,9)$
$x_t \sim N(0, 0.64)$

Let: $L(z,a) = |z-a|^{\alpha}$; $\alpha = 0.5, 1, 2, 4.$

For each of the 4 alternative loss fctns., compute the value

$$\int_{-\infty}^{\infty} L(z,a)\, p(z \mid y, \omega)\, dz$$

where $$p(z \mid y, \omega) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2).\Gamma(1/2)} \left(\frac{g}{v}\right)^{1/2} \left[1 + \left(\frac{g}{v}\right)(z - \omega\hat{\beta})^2\right]^{-\frac{(v+1)}{2}}$$

where $v = 14$.

Zellner finds that for this problem the optimal solution for the squared error loss function is remarkably robust under changes in the form of the loss function, as long as we restrict ourselves to <u>symmetric</u> loss functions.