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

Prepared by David Giles, 1973

Released, April 2012

dgiles@uvic.ca
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 proceed from premises to conclusion.

Not sufficient for all inference because:
1. Impossibility of complete certainty in knowledge. (Can’t be sure that the sun will rise tomorrow.)
2. A given set of data can be generated by an infinity of theories. Deductive logic can’t help us to choose between them.
3. May choose the simplest —
   a. Desirable that its implications be far-reaching though.
   b. Valuable because simple models make strong statements which are readily testable.

Although deductive inference is important, it is insufficient by itself.

Take Jeffrey’s view of inductive logic — it contains deductive logic within its bounds, but not mutually exclusive.

**Inductive Inference**

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

But induction is more than mere description. Here Jeffrey is in conflict with Mach.

Always an element of uncertainty with induction.
Reductive Inference

"Reduction" suggests that something may be.

It involves studying facts to derive theories to explain them.

The link of reduction with the unusual fact is emphasized.

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

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

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

Two possibilities—
(a) A goal being given, or how to reach it?
(b) Discovering a fact or 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 "incidental" ones as "useful guides."
One of the implications of Jeffreys' 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 conditional (upon our information). We revise the probability as our information changes. This is the essence of learning from experience.

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

\[
\begin{align*}
\text{Initial Info. (I))} & \rightarrow \text{Prior } p(H|I) \\
\text{New Data (x)} & \rightarrow \text{Likelihood } p(x|H) \\
\text{Bayes Theorem} & \rightarrow \text{Posterior } p(H|x, I)
\end{align*}
\]

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 and operational approach to a wide variety of inferential problems. The same principles are adopted in each case. c.f. other methods of inference.
Bayes' Theorem

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

\[
p(y, \theta) = p(y | \theta) p(\theta) = p(\theta | y) p(y)
\]

So,

\[
p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)}
\]

\[
= \frac{p(\theta)p(y | \theta)}{\int_\theta p(\theta)p(y | \theta) d\theta}
\]

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, \( y \) is not a p.d.f., it is to be viewed as a function of \( \theta \).

The latter constitutes the entire evidence of the experiment.

Zellner gives the example \( y_i \sim \text{IN}(\mu, \sigma^2) \) where \( \sigma^2 \) is known, \( \mu \sim \text{N}(\mu_0, \sigma_0^2) \).

Then let \( \hat{\mu} = \frac{1}{n} \sum y_i \); \( \sigma^2 = \frac{1}{n} \sum (y_i - \hat{\mu})^2 \)

so \( p(\mu | y, \sigma^2) \propto \exp \left\{ -\frac{1}{2} \left[ \frac{(\mu - \mu_0)^2}{\sigma_0^2} + \frac{n}{\sigma^2} (\mu - \hat{\mu})^2 \right] \right\} \)

Then, as in the text, note, complete the square in \( \mu \) to obtain:

\[
p(\mu | y, \sigma^2) \propto \exp \left[ -\left( \frac{\sigma_0^2 + \sigma^2/n}{\sigma_0^2 + \sigma^2/n} \right) \left( \mu - \frac{\hat{\mu} \sigma_0^2 + \mu_0 \sigma^2/n}{\sigma_0^2 + \sigma^2/n} \right)^2 \right]
\]

\( \mu \sim \text{N}(\mu(\mu), \text{var}(\mu)) \)

\[
\therefore \quad \mu(\mu) = \left( \frac{\sigma_0^2 + \mu_0 \sigma^2/n}{\sigma_0^2 + \sigma^2/n} \right) = \frac{\hat{\mu} \sigma_0^2 (\sigma^2/n)^{-1} + \mu_0 (\sigma^2/n)^{-1}}{(\sigma^2/n)^{-1} + (\sigma_0^2/n)^{-1}}
\]

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

\[
\text{var}(\mu) = \left( \frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n} \right) = \frac{1}{(\sigma^2/n)^{-1} + (\sigma_0^2/n)^{-1}}
\]
\[
\begin{align*}
\text{Let, } & p(x, y) \text{ be a predicate.} \\
\text{Then, } & p(x, y) \text{ is true if and only if } p(y, x) \\
\text{are equivalent.} \\
\text{We treat } & p(x, y) \text{ as a premise.} \\
\text{We have, } & (p(x, y) \land p(y, x)) \text{ is true.} \\
\text{Hence, } & (p(x, y) \land p(y, x)) \text{ is true.} \\
\text{Therefore, } & \text{there is a prior relation we compute.} \\
\end{align*}
\]
Prior p.d.f.'s

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

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

1. **Data-based (DB)** — information in samples
2. **Non-data-based (NDB)** — arising from introspection, casual observation, or theoretical considerations.

Information in the data can be used to make comparisons among alternative prior beliefs on hypotheses. Later in the book, Jeffreys analyzes this problem and develops a framework for making a choice among alternative conflicting beliefs on hypotheses.

If 2 investigators with the same model \( \theta \) with DB prior based on different past data, the posterior densities may differ. If their past samples are pooled, then their prior becomes 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

\[
\int p(\mu) \, d\mu = \int_0^\infty \, d\mu \to \infty.
\]

But Bayes’ Theorem still applies without modification.

Then if \( y_i \sim \text{IN} (\mu, \sigma^2) \), \( \sigma^2 \) known.

We have:

\[
p(\mu | y_i, \sigma^2) \propto k \left( \frac{1}{\sigma_0^2} \right)^n \exp \left\{ -\frac{1}{2\sigma_0^2} \sum (y_i - \mu)^2 \right\}
\]

\[
\propto \exp \left\{ -\frac{1}{2\sigma_0^2} \sum (y_i - \mu)^2 - (\mu - \bar{y})^2 \right\}
\]
\[ p(\mu | y, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ (\mu - \bar{y})^2 + n(\mu - \mu_0)^2 \right] \right\} \]
\[ \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mu - \bar{y})^2 \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\mu - \mu_0)^2 \right\} \]
\[ \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mu - \bar{y})^2 \right\} . \]

where \( \bar{y} = \frac{1}{n} \sum y_i \).

So, \( \mu (\mu) = \bar{y} \), \( \text{var}(\mu) = \frac{\sigma^2}{n} \) and Normal.

c.f. the earlier example with a normal prior. Here we let \( \sigma \to \infty \), \( \sigma \) get the same result as here.

When we have little 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 pdf's are useful in this respect.

Let \( p(y | \mu, n) \) be the pdf for \( y \).

If \( p(y | \mu, n) = p_1(t | \mu, n), p_2(y) \)
\[ \Rightarrow t = (t_1, \ldots, t_k) \]
where \( p_2 \) is independent of \( \mu \) then the \( t_i \)'s are termed sufficient statistics. \( \mathbf{NO} : t_i = t_i(y) \).

A Natural Conjugate prior pdf for \( \theta \) is

\[ f(\theta | t) \propto p_1(t | \theta, n) \]
\[ = k p_1(t | \theta, n) \]
where \( k = k(t, n) \).

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

\[ f(\theta | t_0, n_0) \propto p_1(t_0 | \theta, n_0) \]

or use this as the Informative prior pdf.

\( \mathbf{NB} : \) If \( p(y | \theta, n) = p_1(t | \theta, n), p_2(y) \)
Then \( p(\theta | y, n) \propto p(\theta) p(y | \theta, n) \)
\[ \propto p(\theta) p_1(t | \theta, n) p_2(y) \]
\[ \propto p(\theta) p_1(t | \theta, n). \]
And as \( n \to \infty \), \( p(\theta | y, n) \propto p_1(t | \theta, n) \) also.
Marginal & Conditional
Posterior Distributions

Let \( \theta' = (\theta'_1 : \theta'_n) \)

Then,
\[
p(\theta_1 | y) = \int p(\theta_1, \theta_2 | y) d\theta_2 \\
= \int p(\theta_1 | y) d\theta_2 \\
= \int p(\theta_1, \theta_2 | y) p(\theta_2 | y) d\theta_2.
\]

As the marginal posterior pdf for \( \theta_1 \) is a weighted average of the conditional posterior pdf, the weighting function being the marginal posterior for \( \theta_2 \).

Let \( y' = (y_1, \ldots, y_n) \); \( y_i \sim \text{IN}(\mu, \sigma^2) \); both parameters unknown.

a. \( p(\mu, \sigma | y) \propto p(\mu | \sigma) p(\sigma | y) \)

b. \( p(\mu | \sigma) \propto \frac{1}{\sigma} \); \( -\infty < \mu < \infty \); \( 0 < \sigma < \infty \).

So,
\[
p(\mu, \sigma | y) \propto \sigma^{-1\mu} \exp \left\{ -\frac{1}{2\sigma^2} \sum_i (y_i - \mu)^2 \right\}.
\]

\[
\alpha \, \sigma^{-1\mu} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n-1)\sigma^2 + n(\mu - \bar{y})^2 \right] \right\}.
\]

c. \( p(\mu | y) = \int p(\mu, \sigma | y) d\sigma \)

\[
\int_0^\infty \sigma^{-1\mu} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n-1)\sigma^2 + n(\mu - \bar{y})^2 \right] \right\} d\sigma.
\]

And, as noted in the notes from Reid's book, by putting
\[
a = \left[ (n-1)\sigma^2 + n(\mu - \bar{y})^2 \right] \]
\[
\alpha \, \sigma = \left( \frac{a}{2\sigma^2} \right).
\]

We get:
\[
p(\mu | y) \propto \left\{ (n-1)\sigma^2 + n(\mu - \bar{y})^2 \right\}^{-\frac{1}{2}} = \left\{ (n-1)\sigma^2 + n(\mu - \bar{y})^2 \right\}^{-\frac{1}{2}}.
So \( p(\mu | y) \sim \text{u.s.} (t) \)
\[
\theta \sim \mu.
\]
So the variable \( t = \frac{(\mu - \bar{y})}{s/\sqrt{n}} \) takes a Student \( t \)-distribution with \( v = (n-1) \) d.f.

Now consider the marginal pdf for \( \sigma \):
\[
p(\sigma | y) = \int p(\sigma, \mu | y) d\mu
\]
\[
= \int_0^\infty \sigma^{-(v+1)} \exp \left\{ -\frac{1}{2\sigma} \left( \frac{y^2}{\sigma^2} + \frac{(\mu - \bar{y})^2}{s^2} \right) \right\} d\mu
\]

Now, let \( z = \frac{(\mu - \bar{y})}{\sigma/\sqrt{n}} \)
\[
\Rightarrow \frac{dz}{d\mu} \propto \frac{1}{\sigma}
\]
\[
\Rightarrow \frac{d\mu}{d\sigma} \propto \sigma
\]

So,
\[
p(\sigma | y) \propto \int_0^\infty \sigma^{-n} \exp \left\{ -\frac{y^2}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2} z^2 \right\} dz
\]
\[
= \sigma^{-n} \exp \left\{ -\frac{y^2}{2\sigma^2} \right\} \int_0^\infty \exp \left\{ -\frac{1}{2} z^2 \right\} dz
\]
\[
\Rightarrow p(\sigma | y) \propto \sigma^{-n} \exp \left\{ -\frac{y^2}{2\sigma^2} \right\} \quad ; \quad 0 < \sigma < \infty
\]

Now, the \text{Inverted Gamma} distribution takes the form
\[
p(y | \nu, \alpha) = \left( \frac{2}{\Gamma(\nu)} \right)^{\frac{\nu}{2}} \frac{y^{\nu - 1}}{y^{\nu + 1}} e^{-y/2} \quad ; \quad 0 < y < \infty
\]
or, put \( \sigma = \nu, \alpha = \nu/2 \Rightarrow \gamma = 2/s^2 \)
\[
\Rightarrow p(\sigma | y, \nu) = \frac{2}{\Gamma(\nu)(\nu/2)} \left( \frac{1}{\sigma^{\nu+1}} \right) e^{-y/2\sigma^2} \quad ; \quad 0 < \sigma < \infty
\]
So \( p(\sigma | y, \nu) \propto \sigma^{-(\nu+1)} \exp \left\{ -\frac{y^2}{2\sigma^2} \right\} \).

So over posterior marginal is \( \sigma \) \text{Inverted Gamma form.}
Point Estimates for Parameters.

Consider an estimate $\hat{\theta} = \hat{g}(y)$, of a large $\theta$.

given as $L(\hat{\theta}; \theta)$.

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

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

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

Ex. $L = (\theta - \hat{\theta})' C (\theta - \hat{\theta})$

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

So pick $\hat{\theta} = \bar{\theta}$, a min. loss.

i.e. For the definite quadratic loss, the 

optimal bayes estimator is the mean of the posterior density.

A relationship between classical sampling theory

& Bayesian estimation -

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

$c(\hat{\theta}) = \int L(\hat{\theta}; \theta) p(y | \theta) dy$

q. No best $\hat{\theta}$, $\theta$, so take average -

$$\min_{\theta} \int \int L(\theta; \hat{\theta}) p(y | \theta) p(\theta) dy d\theta$$

$$\hat{\theta} \in K_{\theta}$$

$$= \min_{\theta} \int \int L(\theta; \hat{\theta}) p(y | \theta) p(\theta) dy d\theta$$

$$= \min_{\theta} \int [ \int L(\theta; \hat{\theta}) p(\theta | y) d\theta ] p(y) d\theta$$

$$\min_{\theta} \int L(\theta; \hat{\theta}) p(\theta | y) d\theta \quad \text{i.e. pick Bayesian Est.}$$

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

If it diverges, then $\min_{\theta}$ for double doesn't exist, but $\min_{\theta}$ for single - This is a

Quasi-Bayesian estimator.

On an average risk criterion, the bayes estimator gives the best performance in repeated sampling.
Bayesian Intervals

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

\[
\Pr (\theta \in R | y) = \int_R p(\theta | y) \, d\theta
\]

This probability measures a degree of belief that \( \theta \in R \).

Conversely, we can fix the probability, \( \beta, \gamma \) and the region \( R \) (not necessarily unique).

Marginal Dist'n. of the Observations

Sometimes we want the marginal pdf of the observations. Then

\[
p(y) = \int_{\theta \in \Omega} p(y, \theta) \, d\theta
\]

\[
= \int_{\theta \in \Omega} p(y | \theta) p(\theta) \, d\theta
\]

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,

\[
p(\tilde{y} | y, \theta) = p(\tilde{y}, \theta | y) / p(y | \theta)
\]

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, \( p(\tilde{y} | y) = \int p(\tilde{y}, \theta | y) \, d\theta \)

\[
\therefore p(\tilde{y} | y) = \int p(\tilde{y} | \theta, y) p(\theta | y) \, d\theta.
\]
\[ p(\gamma) = \frac{(d_{\theta} p(d_{\theta} - e_{\theta}))}{(e_{\theta} p(e_{\theta} - e_{\theta}))} + \frac{(e_{\theta} p(e_{\theta} - e_{\theta}))}{(e_{\theta} p(e_{\theta} - e_{\theta}))} = p(\gamma) \]

To be more accurate:

\[ \text{as } e_{\theta} \to 0, \quad p(\gamma) \text{ becomes } N(\theta, 0) \]

\[ \text{as } e_{\theta} \to \infty, \quad p(\gamma) \text{ becomes } \text{gamma}(e_{\theta}^2) \]

\[ \text{as } e_{\theta} \to 0, \quad p(\gamma) \to \text{normal}(e_{\theta}) \]

\[ \text{as } e_{\theta} \to \infty, \quad p(\gamma) \to \text{gamma}(e_{\theta}) \]

For any \( \gamma(\theta) \in \gamma \) occurring at \( \theta \), we have:

\[ p(\theta) \propto p(\theta) \gamma(\theta) \]

\[ \text{Large-Sample Properties} \]

\[ p(\gamma) = \int_{\mathbb{R}} \gamma^2 p(\gamma) d\gamma \]

For a given \( \theta \), we can generally conclude:

\[ \text{Hence that we have the posterior pdf, } p(\gamma|\theta) \]

\[ \text{Recall the Lemma} \]

\[ \text{The large-family to consider:} \]

\[ \text{At the posterior pdf to be employed under} \]

\[ \text{the prior pdf, we can substitute a normal with } \]

\[ \text{and the resulting can be matched with } \gamma \text{ to get} \]

\[ \text{Closest} \]

\[ \text{Recall} \]

\[ \text{An} \]

\[ \text{Prior Parameters} \]

\[ \text{As can use the posterior pdf, } p(\gamma|\theta), \text{ to make} \]
Now, put \( y(\theta) = \frac{10}{\log \{ \lambda(\theta) \}} \).

\[
\exp \left\{ g(\theta) \right\} = \exp \left\{ g(\theta) + (\theta - \theta)g'(\theta) + \frac{1}{2}(\theta - \theta)^2g''(\theta) + \cdots \right\}
\]

But \( g'(\theta) = 0 \) (at maximum).

So, \( \exp \left\{ g(\theta) \right\} = \exp \left\{ g(\theta) \right\} \exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\} \exp \left\{ \frac{1}{6}(\theta - \theta)^3g'''(\theta) \right\} \cdots \)

\[
\lambda(\theta y) = \lambda(\theta y) \exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\} \exp \left\{ \frac{1}{6}(\theta - \theta)^3g'''(\theta) \right\} \cdots
\]

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

\[
\lambda(\theta y) \propto \exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\} \left[ 1 + \frac{1}{6}(\theta - \theta)^3g'''(\theta) + \cdots \right]
\]

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

\[
\exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\} \left[ 1 + \frac{1}{6}(\theta - \theta)^3g'''(\theta) + \cdots \right]
\]

And the leading term gives rise to:

\[
p(\theta y) \propto p(\theta) \exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\}
\]

\[
\propto \exp \left\{ \frac{1}{2}(\theta - \theta)^2g''(\theta) \right\}
\]

\[
E[\theta y] = \theta \quad \text{Var}(\theta y) = \left[ -g''(\theta) \right]^{-1}
\]

\[
= \left[ -\frac{d}{d\theta} \log \left\{ \lambda(\theta y) \right\} \right]^{-1}_{\theta = \theta}
\]

\[
\therefore \quad p(\theta y) \propto \frac{1}{\sqrt{2\pi}} \exp \left\{-\frac{1}{2}(\theta - \theta)^2\left[ -g''(\theta) \right]^{-1} \right\}
\]

\[
\lim_{\theta \to \infty} \frac{1}{\sqrt{2\pi}} \exp \left\{-\frac{1}{2}(\theta - \theta)^2\left[ -g''(\theta) \right]^{-1} \right\}
\]
Appendix: Diffuse Prior Distributions

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

1. If $\mu < 0$, then take a uniform distribution:
   $$p(\mu) \propto d\mu, \quad -\infty < \mu < 0.$$ 
   i.e., $p(\mu) \propto \text{constant}$.

   But $\int_{-\infty}^{0} p(\mu) d\mu = \infty$, so $\mu < 0$ represents "certainty".

2. If $0 \leq \mu \leq b$, then:
   $$\frac{\Pr(\mu \leq a)}{\Pr(\mu \leq b)} = \frac{0}{0}$$
   i.e., indeterminate
   So we can make no statement concerning the odds that $\mu$ 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 represent as little knowledge as possible, choose that PDF which minimizes information in some sense.

   Measure information by
   $$H = \int_{-\infty}^{\infty} p(\mu) \log p(\mu) d\mu.$$ 

   Then $-H = "entropy", \text{ or } "uncertainty."

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

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

   $\frac{\partial L}{\partial p} = \int_{-\infty}^{\infty} \frac{1}{p} \log p d\mu + \lambda \int_{-\infty}^{\infty} p d\mu + 0$ 

   $= \int_{-\infty}^{\infty} \log p + 1 \log p d\mu + \lambda \int_{-\infty}^{\infty} p d\mu$ 

   $= 2 \lambda \log p + 2 \lambda + \lambda \int_{-\infty}^{\infty} p d\mu$ 

   $= 2 \lambda \left( 1 + \lambda + \log p \right)$ 

   So, $1 + \lambda + \log p = 0$

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

   So set $1 + \lambda = 2M$

   $\therefore p(\mu) = \left(\frac{1}{2M}\right)^{1/2}$ 

   As the uniform distribution is the one to pick.
where \( \theta = \frac{\theta}{\sum_{i=1}^{n} \phi_i \theta_i} \) for the transformation.

From our result, we have \( \phi = \phi \), hence \( \phi = \phi \) in the limit.

Now, if we take \( \phi (x) \alpha \), we find:

$$f(x) = \int_{-\infty}^{\infty} \phi(x) \delta(x) \, dx$$

some information. The result

First, let's recall \( \phi \) and its properties.

To obtain a \( \phi \) in our system, we have:

$$\phi(x) = \int \phi(x) \delta(x) \, dx$$

According to the Dirac distribution, we have:

$$\int \phi(x) \delta(x) \, dx = \phi$$

Hence, the following:

$$\frac{\partial}{\partial x} \int \phi(x) \delta(x) \, dx$$

Now, we consider the following:

$$\phi = \phi$$

Note the following properties:

i. \( \phi (x) \alpha \)

ii. \( \phi (x) \alpha \)

iii. \( \phi (x) \alpha \)

iv. \( \phi (x) \alpha \)

v. \( \phi (x) \alpha \)

vi. \( \phi (x) \alpha \)

vii. \( \phi (x) \alpha \)

viii. \( \phi (x) \alpha \)

ix. \( \phi (x) \alpha \)

x. \( \phi (x) \alpha \)
where $A = F(\theta) \cdot \int F^{-1}(\theta) \, d\theta$.
\[ \text{Let } |\text{Inf}_{n}^{1}| = |\text{Inf}_{n}^{1} x. \\
\text{.. } |\text{Inf}_{n}^{1} x = (d\theta) |\text{Inf}_{n}^{1} x. \\
\text{.. } |\text{Inf}_{n}^{1} x d\theta = |\text{Inf}_{n}^{1} x d\theta. \text{ QED.} \]

Now, note the consistency regardless of parametrization, if the correct prior \( p(\theta) \) is chosen.

\[ p(\theta | y) \propto p(\theta) \cdot p(y | \theta) \]

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

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

Now, let \( \Omega = \text{parameter space; } \mathcal{S} = \text{sample space}. \)

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

Then if \( p(\theta | y) \propto p(\theta) \cdot l(\theta | y) \)

If \( p(\theta) \propto |\text{Inf}_{n}^{1} x \) the following

\[ \text{1) } \mathcal{S} - \text{labeling invariance:} \]

Let \( z = g(y) \) be a 1-1 & diff. func. for \( \mathcal{S} \rightarrow \mathcal{S}^* \).

Then \( p(\theta | z) \propto p(\theta | y) \).

\[ \text{2) } \mathcal{S} - \text{labeling invariance:} \]

If \( f = f(\theta) \) is 1-1 & diff. func., then

\[ (a) \exists \ p^* (y | \eta) = p(y | \theta). \]

\( (b) \ p(\eta | y) d\eta \propto p(\eta | y) d\eta. \)

\[ \text{3) } \mathcal{S} - \text{restriction invariance:} \]

Let \( \theta \in \Omega^* \subset \Omega \). Then the posterior based on \( p^* (y | \eta) \) with \( \theta \in \Omega^* \)

\[ \text{will be proportional to } p(\eta | y) \text{ based on } p(y | \eta) \text{ for } \theta \]
(4) **Sufficiency Invariance**

Let $t' = (t_1, \ldots, t_n)$ be sufficient for $\theta$.
Let $p^*(t' | \theta)$ be the corresponding pdf.

Then $p^*(\theta | t') \propto p(\theta | y)$,
where $p^*(\theta | t') \propto p^*(\theta) p^*(t' | \theta)$
and $p(\theta | y) \propto p(\theta) p(y | \theta)$
if both priors are $\propto |I_f|^{1/2}$.

5. **Direct Product Invariance**

Let $y_1, y_2$ be independent samples.
Take $p_i(\theta, y_i) = p_i(\theta | y_i) p_i(y_i)$ where $\theta \in \Omega_1, \theta \in \Omega_2$.

Then $p(\theta | y_1, y_2) \propto p_1(\theta) p_2(\theta) p(y_1, y_2)$
where $\theta \in \Omega = (\Omega_1 \times \Omega_2)$.

Then $p(\theta | y_1, y_2) \propto p_1(\theta, y_1) p_2(\theta, y_2)$
where $p_i(\theta, y_i) \propto p(\theta_i) p(y_i | \theta_i)$, $i = 1, 2$.

If all priors are $\propto |I_f|^{1/2}$.

6. **Repeated Product Invariance**

Take $y_1, y_2, \ldots, y_m$ independent samples.

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

Take $p^*(\theta) \propto |I_f|^{1/2}$.

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

This is "Repeated Product Invariance."

($\forall \theta \neq p(\theta | y_1) \propto p(\theta) p(y_1 | \theta)$. Hence other prior.)
Consider Jeffreys' prior: \( p(\theta) \propto 1/\sqrt{\pi} \). In what sense can this prior be considered as representing "knowing little"?

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

Define, \( I_y(\theta) = \int_{-\infty}^{\infty} p(y|\theta) \log p(y|\theta) dy \)

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

\[
\overline{I_y} = \int_{-\infty}^{\infty} I_y(\theta) p(\theta) d\theta
\]

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

\[ G = \overline{I_y} - \int_{-\infty}^{\infty} p(\theta) \log p(\theta) d\theta \]

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

**Example:** Suppose that \( p(y|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y-\theta)^2\right\} \).

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

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

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

\[ = -\frac{1}{2} \log 2\pi - \frac{1}{2} \theta^2 + \theta \cdot \mathbb{E}(y) - \frac{1}{2} \theta^2 \]

\[ = -\frac{1}{2} \log 2\pi - \frac{1}{2} + \theta^2 - \frac{1}{2} \theta^2 \]

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

So, \( \overline{I_y} = \int_{-\infty}^{\infty} I_y(\theta) p(\theta) d\theta = -\frac{1}{2} \left[ \log 2\pi + 1 \right] \int_{-\infty}^{\infty} p(\theta) d\theta \)

\[ = -\frac{1}{2} \left[ \log 2\pi + 1 \right]. \]
\[ G = \mathcal{I}_y - \int p(\theta) \log p(\theta) d\theta \]
\[ = -\frac{1}{2} [\log 2\pi + 1] - \int p(\theta) \log p(\theta) d\theta. \]

Q: Since the 1st term is independent of \( \theta \), we maximize \( G \) by minimizing \( \int p(\theta) \log p(\theta) d\theta \).

Q: This \( \Rightarrow \text{ uniform distribution} \).

i.e. \( p(\theta) \propto 1/\text{Inf}_0^H \propto \text{constant}. \)

So we have a justification for Jeffreys' prior.

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

\[ \mathcal{I}_y(\sigma) = \int p(y|\sigma) \log p(y|\sigma) dy \]
\[ = \int p(y|\sigma) \left[-\log \sigma - \frac{1}{2} \log 2\pi - \frac{y^2}{2\sigma^2}\right] dy \]
\[ = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{2\sigma^2} \int y^2 p(y|\sigma) dy \]
\[ = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{\sigma^2} [\text{var}(y) + (E(y))^2] \]
\[ = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{\sigma^2} [\sigma^2 + \sigma] \]
\[ = -\frac{1}{2} [\log 2\pi + 1] - \log \sigma. \]

\[ \Rightarrow \mathcal{I}_y = \int \mathcal{I}_y(\sigma) p(\sigma) d\sigma \]
\[ = -\frac{1}{2} [\log 2\pi + 1] - \int \log \sigma p(\sigma) d\sigma \]

\[ G = \mathcal{I}_y - \int p(\sigma) \log p(\sigma) d\sigma \]
\[ = -\frac{1}{2} [\log 2\pi + 1] - \int \log \sigma p(\sigma) d\sigma - \int p(\sigma) \log p(\sigma) d\sigma \]

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

\[ \mathcal{L} = -\frac{1}{2} [\log 2\pi + 1] - \int \log \sigma p(\sigma) d\sigma - \int p(\sigma) \log p(\sigma) d\sigma + \lambda \int \frac{1}{2} p(\sigma) d\sigma \]

\[ \frac{\partial \mathcal{L}}{\partial \sigma} = 0 - \int \frac{1}{\sigma p(\sigma)} [p \log p] d\sigma + \int \frac{1}{2} [p \log p] d\sigma + \lambda \int \frac{1}{2} p(\sigma) d\sigma \]
\[ = -\int \log \sigma d\sigma - \int [\log \sigma + 1] d\sigma + \lambda \int d\sigma \]
\[ = -\sigma \log \sigma - \sigma - \int [\log \sigma + 1] d\sigma + \lambda \int d\sigma \]
\[ = \sigma [1 - \log \sigma - \log p - 1 + \lambda] \]
\[ = 0. \]
\[ \lambda - \log \rho - \log \sigma = 0 \]
\[ \therefore \lambda = \log \rho + \log \sigma = \log (\rho \sigma) \]
\[ \therefore \rho \sigma = e^\lambda \]
\[ \therefore \rho = \frac{e^\lambda}{\sigma} \propto \frac{1}{\sigma} \]

So the result is also in accordance with
\[ p(\sigma) \propto 1/\sigma \propto \lambda. \]

**Example:**
\[ p(y|\theta, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y-\theta)^2 \right] ; -\infty < y < \infty. \]
\[ j(\theta, \sigma) = \int_\mathbb{R} p(y|\theta, \sigma) \log p(y|\theta, \sigma) \, dy \]
\[ = \int_\mathbb{R} p(y|\theta, \sigma) \left[ -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{2\sigma^2} (y-\theta)^2 \right] \, dy \]
\[ = \left( -\log \sigma - \frac{1}{2} \log 2\pi \right) - \frac{1}{2\sigma^2} \int_\mathbb{R} (y-\theta)^2 p(y|\theta, \sigma) \, dy \]
\[ = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{\sigma^2} \left( \int_\mathbb{R} y^2 \right) \]
\[ = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{\sigma^2} \left( \int_\mathbb{R} y \right) \]
\[ = -\frac{1}{2} \left( \log 2\pi + 1 \right) - \log \sigma. \]

Then
\[ \mathbb{E}_y = \iint p(\theta, \sigma) \, d\theta \, d\sigma \]
\[ = \iint \left\{ -\frac{1}{2} \left[ \log 2\pi + 1 \right] - \log \sigma \right\} p(\theta, \sigma) \, d\theta \, d\sigma \]
\[ = -\frac{1}{2} \left[ \log 2\pi + 1 \right] \iint p(\theta, \sigma) \, d\theta \, d\sigma - \iint \log \sigma p(\theta, \sigma) \, d\theta \, d\sigma \]
\[ = -\frac{1}{2} \left[ \log 2\pi + 1 \right] - \iint \log \sigma p(\theta, \sigma) \, d\theta \, d\sigma \]

\[ G = \mathbb{E}_y = \iint \log p(\theta, \sigma) \, d\theta \, d\sigma \]
\[ = -\frac{1}{2} \left[ \log 2\pi + 1 \right] - \iint \log \sigma p(\theta, \sigma) \, d\theta \, d\sigma - \iint p(\theta, \sigma) \log p(\theta, \sigma) \, d\theta \, d\sigma \]

Max. \( G \) s.t. \( \iint p(\theta, \sigma) \, d\theta \, d\sigma = 1. \)
\[ = -\frac{1}{2} \left[ \log 2\pi + 1 \right] - \iint \log \sigma p(\theta, \sigma) \, d\theta \, d\sigma - \iint p(\theta, \sigma) \log p(\theta, \sigma) \, d\theta \, d\sigma \]
\[ + 1 \left[ \iint p(\theta, \sigma) \, d\theta \, d\sigma - 1 \right] \]
\[
\frac{\partial \lambda}{\partial \rho} = 0 - \int \int \frac{2}{\rho} \left[ \rho \log \sigma \right] d\sigma d\theta - \int \frac{2}{\rho} \left[ \rho \log \rho \right] d\sigma d\theta \\
+ \lambda \int \int \frac{1}{\rho} d\sigma d\theta + \lambda \\
= - \int \log \sigma d\sigma d\theta - \int \left[ \log + 1 \right] d\sigma d\theta + \lambda \int d\sigma d\theta \\
= - \int \left[ \sigma \log \sigma - \sigma \right] d\sigma d\theta - \left[ \log + 1 \right] \int d\sigma d\theta + \lambda \int d\sigma d\theta \\
= \sigma \left[ 1 - \log \sigma - \log \rho + 1 \right] \\
= 0 \\
\]

\[
\lambda - \log \sigma - \log \rho = 0 \\
\lambda = \log \rho + \log \sigma = \log (\rho \sigma) \\
\rho = \left( \frac{e^\theta}{\sigma} \right) \propto \frac{1}{\sigma}, \text{ as required.} \\
i.e. \quad \rho (\theta, \sigma) \propto \frac{1}{\sigma}. \\
\]

Example: Consider the information matrix for 
\[
\rho(y|\theta, \sigma) = \frac{1}{\sigma 2\pi} \exp \left[ -\frac{1}{2\sigma^2} (y - \theta)^2 \right] \\
\]
\[
\log \rho(y|\theta, \sigma) = -\log \sigma - \frac{1}{2} \log 2\pi - \frac{1}{2\sigma^2} (y - \theta)^2 \\
\]
\[
\frac{\partial \ln \rho}{\partial \theta} = -\frac{n \log \sigma}{\sigma} + \frac{1}{2} \log 2\pi - \frac{1}{2\sigma^2} (y - \theta)^2 \\
\frac{\partial \ln \rho}{\partial \sigma} = -\frac{n}{\sigma} + C + \frac{1}{2\sigma^3} \sum (y_i - \theta)^2 \\
\frac{\partial^2 \ln \rho}{\partial \sigma^2} = \frac{1}{\sigma^3} \left[ \sum \delta \delta (y_i - \theta)^2 \right] \\
= \left( \frac{1}{\sigma^3} \right) \frac{\delta}{\delta \sigma} \left[ \sum (y_i^2 + \theta^2 - 2y_i \theta) \right] \\
= \left( \frac{1}{\sigma^3} \right) \sum (2\theta - 2y_i) \\
= \frac{2}{\sigma^3} \left[ n \theta - \bar{y} \right] \\
\bar{\theta} = -\frac{n}{\sigma} \left[ n \theta - \bar{y} \bar{\theta} \right] \\
= \frac{2}{\sigma^3} \left[ n \theta - n \bar{\theta} \right] = 0. = \bar{E} \left[ \frac{\partial \log \rho}{\partial \theta} \right]. \\
\]
\[
\frac{d}{d\sigma} \frac{20}{\sigma} = \frac{n}{\sigma} + 0 + \frac{1}{\sigma^2} \sum (y_i - \theta)^2
\]

\[
\left( \frac{d^2}{d\sigma^2} \right) \frac{20}{\sigma^2} = \frac{n}{\sigma^2} - \frac{3n}{\sigma^4} \sum (y_i - \theta)^2
\]

\[
\therefore E \left[ \frac{d^2}{d\sigma^2} \right] = \left( \frac{n}{\sigma^2} \right) - \left( \frac{3n}{\sigma^4} \right) \cdot n \sigma
\]

\[
= \frac{n}{\sigma^2} - \frac{3n}{\sigma^4}
\]

\[
= \frac{-2n}{\sigma^4}
\]

\[
\therefore E(\cdot) = \frac{2n}{\sigma^4}
\]

\[
\frac{d}{d\theta} \frac{20}{\theta} = 0 + 0 - \frac{1}{\sigma^2} \sum \frac{2}{\theta} (y_i - \theta)^2
\]

\[
= -\frac{1}{\sigma^2} \sum \frac{2}{\theta} (y_i^2 + \theta^2 - 2y_i \theta)
\]

\[
= -\frac{1}{\sigma^2} \sum (2\theta - 2y_i)
\]

\[
= -\frac{1}{\sigma^2} \sum (\theta - y_i)
\]

\[
\therefore \frac{d}{d\theta} \frac{20}{\theta} = \frac{-n}{\sigma^2}
\]

\[
\therefore E(\cdot) = \frac{n}{\sigma^2}
\]

\[
\therefore \text{Inf}_{\theta, \sigma} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{\sigma^2} \end{bmatrix}
\]

\[
\therefore \left| \text{Inf}_{\theta, \sigma} \right|^\frac{1}{2} = \left( \frac{n}{\sigma^2} \cdot \frac{n}{\sigma^2} \right)^\frac{1}{2} = \left( \frac{2n^2}{\sigma^4} \right)^\frac{1}{2} = \sqrt{2} \left( \frac{n}{\sigma} \right)
\]

\[
\alpha \left( \frac{1}{\sigma^2} \right) \neq \alpha(\cdot)
\]
Note also that
\[ p(\theta, \xi) \propto \frac{1}{\xi} \]

is invariant to transformations of the kind \( \eta = \theta + k \xi \).

If we consider an asymptotic form of \( G \) and write:
\[ G_a = \int p(\theta) \log \sqrt{\text{Info}} \, d\theta - \int p(\theta) \log p(\theta) \, d\theta \]
Maximize \( G_a \) s.t. \( \int p(\theta) \, d\theta = 1 \)
\[ \Rightarrow p(\theta) \propto |\text{Info}|^{\frac{1}{2}} \]

So, for the asymptotic form of \( G \), Jeffrey's prior is a minimal information prior. However, for the usual form of \( G \), Jeffrey's invariant priors are not always minimal information prior pdfs since they do not always maximize \( G \). [e.g. see third example above.]

If the Jeffrey's' prior is not 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 discrepancy may be important!
Let \( y_i = \beta_1 + \beta_2 x_i + u_i \); \( i = 1, 2, \ldots, n \).

Assume:
(1) \( u_i \sim \mathcal{N}(0, \sigma^2) \)
(2) \( x_i \) fixed & non-stochastic
(3) \( x_i \) random but distributed indep. of \( u_i \), \( x \) with a p.d.f. independent of \( \beta_1, \beta_2, \sigma \).

Then, \( p(y|x, \beta_1, \beta_2, \sigma) \propto \left( \frac{1}{\sigma^r} \right) \exp \left[ -\frac{1}{2\sigma^2} \sum (y_i - \beta_1 - \beta_2 x_i)^2 \right] \)

Take a diffuse prior:
\( p(\beta_1) \propto \text{const} ; \quad -\infty < \beta_1 < \infty \)
\( p(\beta_2) \propto \frac{1}{\beta_2} ; \quad 0 < \beta_2 < \infty \)
\( p(\sigma) \propto \frac{1}{\sigma} ; \quad 0 < \sigma < \infty \).

The prior is \( p(\beta_1, \beta_2, \sigma | x, y) \propto \left( \frac{1}{\sigma^{r-1}} \right) \exp \left[ -\frac{1}{2\sigma^2} \sum (y_i - \beta_1 - \beta_2 x_i)^2 \right] \)

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

\[ \gamma = \sum \{ (y_i - \beta_1 - \beta_2 x_i) - [\beta_1 - \beta_1] + (\beta_2 - \beta_2) x_i \}^2 \]
\[ = \sum (y_i - \beta_1 - \beta_2 x_i)^2 + \sum (\beta_1 - \beta_1)^2 x_i^2 + n(\beta_1 - \beta_1)^2 + 2(\beta_1 - \beta_1) (\beta_2 - \beta_2) \]
\[ = 2 \sum (y_i - \beta_1 - \beta_2 x_i) [ (\beta_1 - \beta_1) + (\beta_2 - \beta_2) x_i ] \]
\[ = \sum y_i^2 + (\beta_1 - \beta_1) \sum x_i^2 + \alpha \]
\[ = u^2 + (\beta_1 - \beta_1) \sum x_i^2 + \alpha \]

Now, \( \alpha = -2 \sum (y_i - \beta_1 - \beta_2 x_i) (\beta_1 - \beta_1) = -2 \sum (y_i - \beta_1 - \beta_2 x_i) (\beta_1 - \beta_1) \)
\[ = -2 (\sum y_i \beta_1 - \beta_1^2) = -2(\beta_1 \beta_2) \sum x_i \beta_1 \]
\[ = 0 \]
\[ = 0 \]

So, \( \gamma = u^2 + (\beta_1 - \beta_1) \sum x_i^2 + n(\beta_1 - \beta_1)^2 + 2(\beta_1 - \beta_1)(\beta_2 - \beta_2) \sum x_i \beta_1 \)
\[ p(y \mid \beta, \sigma \mid \mathbf{y}, \mathbf{X}) \propto \left(\frac{1}{\sigma^m} \right) \exp \left\{ -\frac{1}{\sigma} \left[ \frac{1}{2} \sum_{i=1}^{n} \left( \beta_i - \beta_{i0} \right)^2 \mathbf{X}_i \mathbf{X}_i^T + n (\beta - \beta_0)^2 \right] \right\} \]

Now, consider the case where \( \sigma \) is given

Then \[ p(y \mid \beta, \sigma, \sigma_0 \mid \mathbf{y}, \mathbf{X}) \propto \left(\frac{1}{\sigma_0^{m+1}} \right) \exp \left\{ -\frac{1}{\sigma_0} \left[ (\beta - \beta_0)^2 \mathbf{X}_i \mathbf{X}_i^T + n (\beta - \beta_0)^2 \right] \right\} \]

And this is bivariate normal.

Recall the MVN model:

\[ p(x \mid \Theta, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \Theta)^T \Sigma^{-1} (x - \Theta) \right\} \]

Now, let \( x = (\beta_1, \beta_2) \)
\( \Theta = (\beta_0, \beta_0) \)
\( \Sigma = \sigma^2 \begin{bmatrix} \mathbf{X}_i \mathbf{X}_i^T & \mathbf{X}_i \mathbf{X}_i^T \end{bmatrix} \)

\[ (x - \Theta)^T \Sigma^{-1} (x - \Theta) \]

\[ = \frac{1}{\sigma} \left[ \frac{1}{2} \sum_{i=1}^{n} \left( \beta_i - \beta_{i0} \right)^2 \mathbf{X}_i \mathbf{X}_i^T + n (\beta - \beta_0)^2 \right] \]

\[ = \frac{1}{\sigma} \left[ \frac{1}{2} \left( n (\beta - \beta_0)^2 + (\beta - \beta_0)^2 \mathbf{X}_i \mathbf{X}_i^T + 2 (\beta - \beta_0)(\beta_0 - \beta_0) \mathbf{X}_i \mathbf{X}_i^T \right) \right] \]

As required.

But \( \sigma^2 \) is rarely known, so this is of little help.
To look at the marginal pdf for $\beta_1$ only,

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

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

where $c = \left[ \frac{v}{s^2} + n(\beta - \beta_0)^2 + (\beta - \beta_2)^2 \Sigma_i x_i^2 + 2(\beta - \beta_1)(\beta - \beta_2) \Sigma_i x_i \right]$.

Let $t = \left( \frac{\beta - \beta_2}{\sigma} \right)$

so $\frac{1}{\sigma^{n+1}} = \frac{1}{(\frac{t}{v})^{n+1}}$

so $\int_{-\infty}^{\infty} \frac{1}{(\frac{t}{v})^{n+1}} e^{-\frac{ct}{v}} dt$

$\propto c^{-n} \int_{0}^{\infty} z^{-n-1} e^{-cz} dz$

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

So, $p(\beta_1 | y, x) \propto c^{\frac{n}{2}} (\Sigma_i x_i - \beta_1)^2 \prod_{i=1}^{n} \left( \frac{(x_i - \beta_1)^2}{s_i^2 \Sigma x_i} \right)^{\frac{1}{2}}$

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

$$p(x | \theta, v, u, m) \propto \left[ v + (x - \theta)^2 v(x - \theta) \right]^{-\frac{(m+1)/2}{2}}$$

So,

$$p(\beta_1 | y, x) \propto \left[ v + \frac{\Sigma_i (x_i - \beta_1)^2}{s_i^2 \Sigma x_i} (\beta_1 - \beta_2)^2 \right]^{-\frac{(m+1)/2}{2}}$$

and

$$p(\beta_2 | y, x) \propto \left[ v + \frac{\Sigma_i (x_i - \beta_1)^2}{s_i^2 \Sigma x_i} (\beta_2 - \beta_1)^2 \right]^{-\frac{(m+1)/2}{2}}$$

Then

$$\left( \frac{\Sigma_i (x_i - \beta_1)^2}{s_i^2 \Sigma x_i} \right)^{\frac{1}{2}} (\beta_1 - \beta_0) = t_v$$

and

$$\left( \frac{\Sigma_i (x_i - \beta_2)^2}{s_i^2 \Sigma x_i} \right)^{\frac{1}{2}} (\beta_2 - \beta_0) = t_v$$
\[ p(\sigma | y, x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mu, \beta, \sigma | y, x) \, d\mu, \, d\beta \]
\[ \propto \left( \frac{1}{\sigma^{\alpha n}} \right) \exp \left[ -\frac{\sum x_i^2}{2\sigma^2} \right], \quad 0 < \sigma < \infty. \]

And this is now the form of an inverted \( \Gamma \) density.

\[ p(\sigma | u, v) = \frac{2}{\Gamma(u/2)} \left( \frac{uv}{2} \right)^{u/2} \frac{1}{\sigma^{uv}} \exp \left[ -\frac{uv}{2\sigma^2} \right] \]
\[ \sigma \, E(\sigma) = u\left(\frac{v}{2}\right)^{1/2} \frac{\Gamma(\frac{u-1}{2})}{\Gamma(u/2)} \]
\[ \text{var}(\sigma) = \frac{v\sigma^2}{u-2} - \left[ E(\sigma) \right]^2. \]

Transform \( \sigma \) to \( \sigma^2 \):

\[ p(\sigma^2 | y, x) \propto \left( \frac{uv}{2} \right)^{u/2} \frac{1}{\sigma^{uv}} \exp \left[ -\frac{uv}{2\sigma^2} \right] \]
\[ 0 < \sigma < \infty \]

Simply by re-writing the expression obtained already.

Or, \( p(h | y, x) \propto (h^{(u+1)}) \exp \left( \frac{-uv^2}{2} \right) \); \( 0 < h < \infty \)

where \( h = \frac{1}{\sigma^2} \).

So since \( \chi^2 \) density is just \( \Gamma \) with \( \alpha = u/2, \, \beta = v/2 \),

Then \( p(x | \alpha, \beta) \propto x^{(\alpha-1)} e^{-x \beta} \)

\( \therefore \) \( p(h | y, x) \) is just \( \chi^2 \).

As we can make inferences about \( \alpha, \beta \), \( \frac{1}{\sigma^2} \) individually, what about joint inference on \( (\alpha, \beta) \)?

Consider the expression

\[ \psi = \frac{1}{25} \left[ \alpha (\beta - \beta_0)^2 + (\beta - \beta_0)2x_i^2 + 2(\beta - \beta_0)(x_i - \bar{x}) \right] \]

Then \( \psi \sim F_{2, 0} \).
To show this, we must convert $p(\rho_1, \rho_2 | y, x)$ to $p(y, x)$. Do this in several steps:

(a) Let $\delta' = (\rho_1 - \bar{\rho}_1, \rho_2 - \bar{\rho}_2) = (\delta_1, \delta_2)$ say.

Then if $A = \frac{1}{2} \int \frac{2 \delta_1 z_1}{z_1^2} \int \frac{2 \delta_2 z_2}{z_2^2}$

$\Rightarrow \delta' A \delta = (\rho_1 - \bar{\rho}_1, \rho_2 - \bar{\rho}_2) \frac{1}{2} \begin{bmatrix} 2 \delta_1 z_1 & 0 \\ 0 & 2 \delta_2 z_2 \end{bmatrix} \begin{bmatrix} \rho_1 - \bar{\rho}_1 \\ \rho_2 - \bar{\rho}_2 \end{bmatrix}$

$\Rightarrow \delta' A \delta = \psi, \quad \text{(add from $\rho = \delta$} \rightarrow \text{in region $\cdots$)}$

Now, $p(\delta | y, x) = p(\delta | y, x) \frac{1}{|J|}$

$\Rightarrow |J|^{-1} = \begin{vmatrix} \frac{\partial \delta_1}{\partial \rho_1} & \frac{\partial \delta_1}{\partial \rho_2} \\ \frac{\partial \delta_2}{\partial \rho_1} & \frac{\partial \delta_2}{\partial \rho_2} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 0 & 1 \end{vmatrix} = 1.$

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

Now, $\Rightarrow p(\delta | y, x) \propto \left[ \frac{\nu s^2 + (\rho - \bar{\rho})^2 + \delta_1^2 (\rho_1 - \bar{\rho}_1)^2 + \delta_2^2 (\rho_2 - \bar{\rho}_2)^2}{2kV} \right]^{-\nu/2}$

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

$\propto \left[ 1 + \frac{(\delta')^2 A \delta}{\nu} \right]^{-\nu/2}$

Now, $A = K^T K, \Rightarrow \delta' A \delta = (K \delta)'(K \delta) = V^T V, \text{ s.t.}$

$V = K \delta = \left( \begin{array}{c} \psi_1 \\ \psi_2 \end{array} \right), \Rightarrow K \delta = \left( \begin{array}{c} \psi_1 \\ \psi_2 \end{array} \right)$

So, $\delta = (\psi K) = \frac{1}{k} (\begin{array}{c} \psi_1 \\ \psi_2 \end{array}) = (\begin{array}{c} \delta_1 \\ \delta_2 \end{array})$

$\Rightarrow |J| = \begin{vmatrix} \frac{\partial \psi_1}{\partial \psi_1} & \frac{\partial \psi_1}{\partial \psi_2} \\ \frac{\partial \psi_2}{\partial \psi_1} & \frac{\partial \psi_2}{\partial \psi_2} \end{vmatrix} = \begin{vmatrix} k & 0 \\ 0 & k \end{vmatrix} = k^2$

$p(V | y, x) = \frac{1}{k^2} p(\delta | y, x)$

$\propto \left( \frac{1}{k^2} \right) \left[ 1 + \frac{(\delta')^2 A \delta}{\nu} \right]^{-\nu/2}$

$\propto \left[ 1 + \frac{(\delta')^2 A \delta}{\nu} \right]^{-\nu/2}$

$\Rightarrow V = \left( \begin{array}{c} \psi_1 \\ \psi_2 \end{array} \right) = \left( \psi \cos \theta \\ \psi \sin \theta \right)$
\[ p(\psi | y, x) = 1 \psi \cdot p(\nu | y, x) \]
\[ \text{and } \begin{vmatrix} \frac{\partial \psi}{\partial \nu} & \frac{\partial \psi}{\partial \psi} \\ \frac{\partial \nu}{\partial \psi} & \frac{\partial \nu}{\partial \psi} \end{vmatrix} = \begin{vmatrix} \frac{\partial \psi}{\partial \nu} \cos \theta & -\psi \sin \theta \\ \frac{\partial \nu}{\partial \psi} \cos \theta & \psi \cos \theta \end{vmatrix} = \frac{\psi (\sin^2 \theta + \cos^2 \theta)}{2} = \frac{\psi}{2} \]

\[ p(\psi | y, x) \propto \frac{1}{2} \left[ 1 + \left( \frac{3}{\nu^2} \right) \psi \right]^{-\frac{\nu + 2}{2}} \]

But, \( \nu' = \psi (\cos \theta + \sin \theta) \)

\[ p(\psi | y, x) \propto \left[ 1 + \frac{\nu}{2} \psi \right]^{-\frac{\nu + 2}{2}} \]

\[ p(F_{u, v}) \propto F^{\frac{m+2}{2}} \left[ 1 + F(\frac{m}{2}) \right]^{-(m+2) \psi} \]

So put \( m = 2 \).

Then \( p \propto \left[ 1 + \left( \frac{3}{\psi} \right) \psi \right]^{-\frac{3}{2}} \)

So, \( \psi \sim F_{2, u} \), as required.

Note that if 2 variables are distributed as a Bivariate \( t \)-distribution, then a linear combination of these variables is distributed as a univariate \( t \).

Let \( Z(y | x = x_0) = \beta_1 + \beta_2 x_0 \). (Use for predicting)

Then \( \hat{x}_0 = \hat{\beta}_1 + x_0 \hat{\beta}_2 \), i.e., predict with estimated regression.

Change variables \( \beta_1, \beta_2 \) to \( \beta_0, \beta_1 \) in the joint posterior pdf for \( \beta_1, \beta_2 \).

Viz., let \( \beta_0 = \beta_1 - \beta_2 = (\beta_1 - \beta_2) + (\beta_2 - \beta_2) x_0 \)

\[ \beta_2 = \beta_2 - \beta_2 \]

Then \( \text{det} = \begin{vmatrix} \frac{\partial \beta_0}{\partial \beta_1} & \frac{\partial \beta_0}{\partial \beta_2} \\ \frac{\partial \beta_1}{\partial \beta_1} & \frac{\partial \beta_1}{\partial \beta_2} \end{vmatrix} = \begin{vmatrix} 1 & x_0 \end{vmatrix} = 1 \)
\[ p(\eta_0, \beta_1 | y, x) = \int p(\eta_1, \beta_1 | y, x) \]
\[ \alpha \left[ \frac{\nu_0 + \nu (\beta_1 - \beta_1)^2 + \xi (\eta_0 - \eta_0)^2 + \xi (\eta_1 - \eta_1)^2}{\nu_0} \right]^{-\frac{\nu_0 + \nu}{2}} \]

Get
\[ p(\eta_0 | y, x) = \int p(\eta_0, \beta_1 | y, x) \, d\beta_1 \]
\[ \alpha \left[ \frac{\nu_0 + \nu (\eta_0 - \eta_0)^2 + \xi (\eta_0 - \eta_0)^2}{\nu_0 + \nu} \right]^{-\frac{\nu_0 + \nu}{2}} \]

Then,
\[ \frac{\eta_0 - \bar{x}}{s / \sqrt{\sum (x_i - \bar{x})^2} \sqrt{\nu_0 + \nu}} \sim t_n. \]

So use this for forecast intervals.

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

We know that
\[ \frac{(\beta_1 - \beta_1)}{s / \sqrt{\sum (x_i - \bar{x})^2}} \sim t_n. \]

Now,
\[ dt_n = k_0 \, d\beta_1, \quad \text{where} \quad k_0 = \frac{\sqrt{\sum (x_i - \bar{x})^2}}{s}. \]

Now, \( p(t) \) is given by the tables of the \( t \)-distribution.

So
\[ p(t) \, dt = p(t) \, k_0 \, d\beta_1. \]

So \( p(\beta_1) \) is got by taking \( k_0 \, p(t) \).
Take the multiple regression model.

\[ y = x\beta + \epsilon \]

\[ \therefore p(y \mid x, \beta, \sigma) \propto \left( \frac{1}{\sigma} \right) \exp \left[ -\frac{1}{2\sigma^2} (y - x\beta)'(y - x\beta) \right] \]

\[ \therefore \{ y - x\beta \}'(y - x\beta) = \{ y - x\beta - x(\beta - \bar{\beta}) \}'\{ y - x\beta - x(\beta - \bar{\beta}) \} \]

\[ = (y - x\beta)'(y - x\beta) + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}) \]

\[ = \sigma^2 + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}) \]

So,

\[ p(y \mid x, \sigma, \beta) \propto \left( \frac{1}{\sigma^2} \right) \exp \left[ -\frac{1}{2\sigma^2} \{ \sigma^2 + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}) \} \right] \]

Let \( p(\beta) \propto \text{const} \) ; \(-\infty < \beta < \infty \)

\[ p(\sigma) \propto \frac{1}{\sigma} \] \( \sigma > 0 \)

So independence \( \Rightarrow p(\beta, \sigma) \propto \frac{1}{\sigma} \).

\[ p(\rho, \sigma, y, x) \propto \left( \frac{1}{\sigma^2} \right) \exp \left[ -\frac{1}{2\sigma^2} \{ \sigma^2 + (\rho - \beta)'x'x(\rho - \beta) \} \right] \]

\[ \therefore p(\rho \mid \sigma, y, x) \sim \text{M.V.N.} \left( \rho, (xx')^{-1}\sigma \right) \]

\[ \therefore \]

\[ p(\beta \mid x, y) = \int_{\infty}^{\infty} p(\beta, \sigma \mid y, x) d\sigma \]

Let \( c = \{ \sigma^2 + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}) \} \)

Let \( z = \frac{\sigma^2}{\sigma^2} \)

\[ \therefore d\sigma = \frac{z}{\sigma^2} dz \]

\[ \therefore \sigma^2 = \left( \frac{c}{z} \right)^2 \therefore \sigma^{-2} = \left( \frac{c}{z} \right)^{-2} \]

\[ I = \int_{0}^{\infty} \left( \frac{c}{z} \right)^{-2} e^{-2c \left( \frac{z - \sigma^2}{\sigma^2} \right)} dz \]

\[ = \int_{0}^{\infty} \left( \frac{c}{z} \right)^{-2} e^{-2c \left( \frac{z}{\sigma^2} \right)} \left( \frac{1}{z} \right)^{3/2} e \left( \frac{1}{z} \right)^{3/2} \left( \frac{1}{\sigma^2} \right)^{1/2} \]

\[ = \int_{0}^{\infty} \left( \frac{c}{z} \right)^{-2} e^{-2c \left( \frac{z}{\sigma^2} \right)} \left( \frac{1}{z} \right)^{3/2} e \left( \frac{1}{z} \right)^{3/2} \left( \frac{1}{\sigma^2} \right)^{1/2} \]

\[ = \int \left( \frac{c}{z} \right)^{-2} e^{-2c \left( \frac{z}{\sigma^2} \right)} \left( \frac{1}{z} \right)^{3/2} e \left( \frac{1}{z} \right)^{3/2} \left( \frac{1}{\sigma^2} \right)^{1/2} \]
\[
I = \int_0^\infty x^{n-1} e^{-x} \, dx
\]
\[
\propto c^{-n} \int_0^\infty z^{n-1} e^{-cz} \, dz
\]
\[
\propto c^{-n} \left( \text{Ei}(c\cdot) \right) \quad (\text{c.dist.})
\]

\[
p(\beta \mid y, x) \propto \left[ \frac{1}{\sigma^2} + (\beta - \beta_0)^T x^T x (\beta - \beta_0) \right]^{-n/2}
\]

\[
q(\sigma \mid y, x) = \int p(\beta, \sigma \mid y, x) \, d\beta
\]
\[
\propto \int \frac{1}{(\sigma^{n+1})} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \frac{1}{\sigma^2} + (\beta - \beta_0)^T x^T x (\beta - \beta_0) \right] \right\} \, d\beta
\]
\[
\propto \left( \frac{1}{\sigma^{n+1}} \right) \exp \left[ -\frac{1}{2\sigma^2} \sum x_i^2 \right]
\]

Now, let \( z_i = \frac{1}{\sigma} x_i (\beta_i - \beta_0) \)

\[
dz_i = \frac{x_i}{\sigma} \, d\beta_i
\]
\[
d\beta_i \propto \sigma \, dz_i
\]

\[
q(\sigma \mid y, x) \propto \left( \frac{1}{\sigma^{n+1}} \right) \exp \left[ -\frac{1}{2\sigma^2} \sum x_i^2 \right]
\]
\[
\propto \left( \frac{1}{\sigma^{n+1}} \right) \exp \left[ -\frac{1}{2\sigma^2} \sum z_i^2 \right]
\]

But this is just an inverted \( \text{I-dist.} \)
Consider references concerning a single element $j$ say $j_i$.

We want $p(j_i \mid y, x) = \int \cdots \int p(j_i \mid y, x) \, dj_1 \cdots \, dj_n.$

where $p(j_i \mid y, x) \propto [uv + (j_i - \bar{j_i})^T x (j_i - \bar{j_i})]^{-n/2}$.

Now let $A = (j_i - \bar{j_i})$; let $H = \frac{1}{2}x^T x$; $\delta = (\bar{j_i} - \bar{j_i})$.

Let $A = (j_1 - \bar{j_1}, j_2 - \bar{j_2}, \ldots, j_n - \bar{j_n})$.

\[ p(j_1 \mid y, x) \propto [s^T]^{-n/2} \left[ u + \frac{1}{2} (j_i - \bar{j_i})^T x (j_i - \bar{j_i}) \right]^{-n/2} \]

\[ \propto [u + \delta^T H \delta]^{-n/2}. \]

\[ p(\delta \mid y, x) = p(j_1 \mid y, x) \mid_{\delta = 0} \]

\[ |J_1| = \left| \frac{\delta}{\delta j_1} \right| = 1 \]

\[ \delta^T H \delta = \delta^T (H_{11} - H_{12} H_{22}^{-1} H_{21}) \delta + (\delta^T + \delta^T H_{22}^{-1} H_{12}) H_{12} (\delta^T + \delta^T H_{22}^{-1} H_{12}) \]

[Completing the square on $\delta^T$]

\[ \alpha x^2 + bx + c = \alpha (x + \frac{b}{2\alpha})^2 + (c - \frac{b^2}{4\alpha}) \]

\[ \therefore p(\delta \mid y, x) \propto [u + \delta^2 (H_{11} - H_{12} H_{22}^{-1} H_{21}) + (\delta^T + \delta^T H_{22}^{-1} H_{12}) H_{12} (\delta^T + \delta^T H_{22}^{-1} H_{12})]^{-n/2} \]

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

\[ \delta^T H \delta = \delta^T (H_{11} - H_{12} H_{22}^{-1} H_{21}) \delta + (\delta^T + \delta^T H_{22}^{-1} H_{12}) H_{12} (\delta^T + \delta^T H_{22}^{-1} H_{12}) \]

\[ \left[ H_{11} - H_{12} H_{22}^{-1} H_{21} \right]^{-1} \]

\[ C = H_{22} / (u + \delta^T H_{22}^{-1} \delta) \]

\[ p(\delta \mid y, x) \propto [u + \delta^T C H_{11}^{-1} \delta]^{-n/2}. \left[ H_{11} H_{22}^{-1} H_{12} \right] \left( \delta^T + \delta^T H_{22}^{-1} H_{12} \right) C \left( \delta^T + \delta^T H_{22}^{-1} H_{12} \right) \left[ H_{11} H_{22}^{-1} H_{12} \right]^{-1} \]

\[ p(\delta \mid y, x) \propto [u + \delta^T H_{11}^{-1} \delta]^{-n/2} \int [1 + (\delta^T + \delta^T H_{22}^{-1} H_{12}) C (\delta^T + \delta^T H_{22}^{-1} H_{12})]^{-n/2} \delta \]
Now, MVT is in the general form:

\[ p(x|\theta, \nu, \psi, \phi) \propto |\nu|^{1/2} \left[ \nu + (x-\theta)^T \nu (x-\theta) \right]^{-\nu/2}. \]

\[ \int_{\nu} = |\nu|^{-1/2} \int_{\nu} |\nu|^{1/2} \left[ 1 + (\nu + \delta_i H_i \nu H_i^T)^{-1} \right]^{-\nu/2}. \]

\[ \propto |\nu|^{-1/2} \]

\[ p(\delta_i|y, x) \propto \left[ \nu + \frac{\delta_i^2}{\nu} \right]^{-\nu/2} |\nu|^{-1/2}. \]

Now, let \( \nu = H_{11} \left( \nu + \frac{\delta_i^2}{\nu} \right)^{-1} \)

\[ \nu = H_{11} \left( \nu + \frac{\delta_i^2}{\nu} \right)^{-1} \]

\[ \propto (\nu + \frac{\delta_i^2}{\nu})^{-\nu/2}. \]

\[ p(\delta_i|y, x) \propto \left[ \nu + \frac{\delta_i^2}{\nu} \right]^{-\nu/2} \propto \left[ \nu + \frac{\delta_i^2}{\nu} \right]^{-\nu/2}. \]

This is the form of a t-distribution.

So, \( \frac{\delta_i}{\sqrt{\nu}} \sim t_\nu. \)

\[ \frac{\delta_i}{\sqrt{\nu}} \sim t_\nu. \]

Where \( \nu \) is the (1,1)th element of \( (X'X)^{-1} \)

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

\[ h_{11} = \text{vec}(X'X) / \sigma^2. \]

\[ (h'')^2 \sim \left[ \frac{S}{(X'X)^{1/2}} \right]. \]
Now let \( F = \frac{(\beta - \beta_0)}{\hat{v}} \) = \( \frac{(\beta_1 - \beta_0)}{\hat{v}} \) = \( \frac{(\beta_2 - \beta_0)}{\hat{v}} \)

Then \( dF = \frac{1}{\hat{v}} \delta \; d\delta \; \propto \; 2F \frac{d\beta}{d\delta} F \frac{d\beta}{d\delta} \)

\[ \therefore \; d\delta \propto F^{-1} dF \]

So, \( p(\delta_i | y, x) \propto \left[ v + \frac{\delta_i^2}{\hat{v}} \right]^{-(\nu + 1)/2} \)

\[ \Rightarrow \; p(F | y, x) \propto F^{-\frac{1}{2}} \left[ v + F \right]^{-(\nu + 1)/2} \]

which is \( F_{1, \nu} \).

So,
\[ \frac{(\beta_i - \hat{\beta})^2}{\hat{v}_i} \sim F_{1, \nu}. \]

Now consider inferences on a subset \( g(\beta) \) i.e. we want a joint test on some function of \( \beta \).

Let \( \delta' = (\beta - \beta_0)' = (\delta'_1 : \delta'_2) \)

\[ \therefore \delta' H \delta = (\delta'_1 : \delta'_2)' \left( \begin{array}{c} H_{11} & H_{12} \\ H_{21} & H_{22} \end{array} \right) \left( \begin{array}{c} \delta'_1 \\ \delta'_2 \end{array} \right) \]

\[ = (\delta'_1 H_{11} + \delta'_2 H_{12} + \delta'_1 H_{21} + \delta'_2 H_{22}) (\delta'_1 + \delta'_2) \]

\[ = (\delta'_1 H_{11} \delta_1 + \delta'_2 H_{12} \delta_2 + \delta'_1 H_{21} \delta_1 + \delta'_2 H_{22} \delta_2) \]

\[ = \delta'_1 H_{11} \delta_1 + 2 \delta'_2 H_{21} \delta_1 + \delta'_1 H_{22} \delta_2. \]

We want to find \( p(\delta_i | y, x) \) i.e. we want to get rid of \( \delta_2 \), so \textbf{complete the square} on \( \delta_2 \).

\[ \therefore \delta' H \delta = \left[ \delta_2 + \delta'_1 H_{11} \delta_2 \right]' H_{22} \left[ \delta_2 + \delta'_1 H_{11} \delta_2 \right] \]

\[ + \left[ \delta'_1 H_{11} \delta_1 - \delta'_1 H_{12} \delta_1 H_{22} \delta_2 \right] \]

So, \( p(\delta_1, \delta_2 | y, x) \propto \left[ v + \left[ \delta_2 + \delta'_1 H_{11} \delta_2 \right]' H_{22} \left[ \delta_2 + \delta'_1 H_{11} \delta_2 \right] \right]^{-\nu/2} \]

\[ + \left[ \delta'_1 \delta_1 \delta_2 - \delta'_1 \delta_1 H_{12} \delta_2 \delta_1 \delta_2 \right] \]

Now, for given \( \delta_1 \), we observe that

\[ p(\delta_2 | \delta_1, y, x) \] is \textbf{MUS-t} with condition mean

\[ -H_{22}^{-1} H_{21} \delta_1 \]

and \( \delta_2 = \beta_2 - \hat{\beta}_2 \).

: \text{Conditional mean of } \delta_2 \text{ is } \hat{\beta}_2. \]
\[ E(\beta_2 | \beta_1) = E[\delta_2 + \hat{\beta}_0 | \beta_1] \]
\[ = E[\delta_2 | \beta_1] + \hat{\beta}_0 \]
\[ = \delta_2 - H_{H_21} H_{H_21} \delta_1 \]
\[ \therefore E(\beta_2 | \beta_1) = \delta_2 - H_{H_21} H_{H_21} (\beta_1 - \hat{\beta}_1). \]

Now, more important, continue to find \( p(\delta_1 | x, y) \):
\[ p(\delta_1 | x, y) = \int \cdots \int p(\delta_1, \delta_2 | x, y) d\delta_2 \quad ; \quad k \in \mathbb{R}, \text{integrals.} \]
\[ \propto \int \cdots \int [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1 + (\delta_2 + H_{H_2} H_{H_21} \delta_1)' H_{H_21} (\delta_2 + H_{H_2} H_{H_21} \delta_1)]^{-\frac{\eta}{2}} \quad d\delta_2. \]

Let \( C = H_{H_2} / [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1] \)
\[ \Rightarrow p(\delta_1, \delta_2 | x, y) \propto [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} \]
\[ \times \left[ 1 + \frac{(\delta_2 + H_{H_2} H_{H_21} \delta_1)' H_{H_21} (\delta_2 + H_{H_2} H_{H_21} \delta_1)}{[v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]} \right]^{-\frac{\eta}{2}} . \]

\[ \propto \left\{ \begin{array}{l} [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} |C|^{-\frac{1}{2}} \times \end{array} \right\} \]
\[ \left\{ \begin{array}{l} [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} |C|^{-\frac{1}{2}} \times \end{array} \right\} \]
\[ \int \cdots \int [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} |C|^{-\frac{1}{2}} \times \]
\[ \left\{ \begin{array}{l} [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} |C|^{-\frac{1}{2}} \times \end{array} \right\} \]

But the integrand is just \( \text{M.I.S.} \).
\[ \therefore p(\delta_1 | x, y) \propto [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{\eta}{2}} |C|^{-\frac{1}{2}} \]
\[ \text{And} \quad |C|^{-\frac{1}{2}} = |H_{H_2} / [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]|^{-\frac{1}{2}} \]
\[ \propto [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{k_2/2} \]
\[ \propto p(\delta_1 | x, y) \propto [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-\frac{(n-k_2)}{2}} \]
\[ \text{But,} \quad (n-k_2) = (n-k) + (k-k_2) = (n-k) + k_1, \]
\[ = n + k. \]
\[ \therefore p(\delta_1 | x, y) \propto [v + \delta_1 (H_{H_1} - H_{H_2} H_{H_21} H_{H_21}) \delta_1]^{-(n+k)/2}. \]
And this is just a $M.V.S-t$ distribution.

**Note:**

$E(\delta_i) = 0$

$E(\beta_1 - \beta_1) = 0$

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

And,

$E(\delta, \delta') = E[\frac{(\beta - \beta')(\beta - \beta')}]{\left(H_0 - H_1, H_1^T H_1\right)^{-1}}$

(_see p. 386_.)

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

Let $\alpha = 1'\beta$, where $1'$ is a vector of constant (not necessarily unit) elements.

**Note that**

$p(\beta, \sigma | y, x) = p(\beta | \sigma, y, x)p(\sigma | y, x)$

Now, we have that $p(\beta | \sigma, y, x)$ has $M.V.N$, while $p(\sigma | y, x)$ is $I.G.$

Now $\Rightarrow p(\alpha | \sigma, y, x)$ is also $M.V.N$

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

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

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

$= 1' (x'x) 1 \sigma^{-2}$.

Now,

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

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

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

Now,

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

Where $c = 1' (x'x) 1 \sigma^{-2}$

$\Rightarrow p(\sigma | y, x) \propto \frac{1}{\sigma^{\frac{N+1}{2}}} \left(\frac{1}{\sigma^{\frac{1}{2}}} \exp\left[-\frac{1}{2 \sigma} \frac{(\alpha - \hat{\alpha})^2}{2} \right] \right)$

$\Rightarrow p(\alpha, \sigma | y, x) \propto \frac{1}{\sigma^{\frac{N+1}{2}}} \exp\left[-\frac{1}{2 \sigma} \left(\frac{1}{\sigma^{\frac{1}{2}}} \exp\left[-\frac{1}{2 \sigma} \frac{(\alpha - \hat{\alpha})^2}{2} \right] \right) \right]$
\[ p(\alpha | x, y) \propto \int \left( \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} \left( U y^2 + (\frac{\alpha - \overline{x}}{\sigma})^2 \right) \right] \right) \, d\sigma \]
\[ \propto \left[ \frac{U y^2 + (\alpha - \overline{x})^2}{\sigma^2} \right]^{-\frac{1}{2} + \frac{1}{2}} \]
\[ \propto \left[ U + \frac{(\alpha - \overline{x})^2}{\sigma^2} \right]^{-\frac{1}{2} + \frac{1}{2}} \]

which is Student's t.

So \( \left( \frac{\alpha - \overline{x}}{\sigma^2} \right) \) is \( t_U \), as expected.

Now, suppose we have our posterior on an informative 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:

\[ p(\beta, \sigma | y_1, x_1) \propto \left( \frac{1}{\sigma^a} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ U y_1^2 + \left( \beta - \overline{x}_1 \right) x_1 \right] \right\} \left( \frac{1}{\sigma^b} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left( \beta - \overline{x}_1 \right)^2 \right\} \]

And \( p(\beta, \sigma | y_1, x_1) = p(\sigma | y_1, x_1) \cdot p(\beta | y_1, x_1) \).

So we have the 2 marginal pdf's.

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

\[ l(\beta, \sigma | y_2, x_2) \propto \left( \frac{1}{\sigma^a} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left( y_2 - \overline{x}_2 \beta \right)^2 \right\} \]

\[ p(\beta, \sigma | y_1, y_2, x_1, x_2) \propto \left( \frac{1}{\sigma^a} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ (y_1 - \overline{x}_1 \beta)^2 + (y_2 - \overline{x}_2 \beta)^2 \right] \right\} \]

Now,

\[ [y_1 - \overline{x}_1 \beta] (y_1 - \overline{x}_1 \beta) + (y_2 - \overline{x}_2 \beta) (y_2 - \overline{x}_2 \beta) \]

\[ = [(y_1 - \overline{x}_1 \beta) - (x_1 \beta - x_1 \overline{\beta})] [(y_1 - \overline{x}_1 \beta) - (x_1 \beta - x_1 \overline{\beta})] \]

\[ + [(y_2 - \overline{x}_2 \beta) - (x_2 \beta - x_2 \overline{\beta})] [(y_2 - \overline{x}_2 \beta) - (x_2 \beta - x_2 \overline{\beta})] \]

\[ = [(y_1 - \overline{x}_1 \beta) (y_1 - \overline{x}_1 \beta) + (y_2 - \overline{x}_2 \beta) (y_2 - \overline{x}_2 \beta) + (x_1 \beta - x_1 \overline{\beta}) (x_1 \beta - x_1 \overline{\beta}) + (x_2 \beta - x_2 \overline{\beta}) (x_2 \beta - x_2 \overline{\beta})] \]

\[ = U y^2 + \left[ (\beta - \overline{x}_1 \beta) x_1 (\beta - \overline{x}_1 \beta) + (\beta - \overline{x}_2 \beta) x_2 (\beta - \overline{x}_2 \beta) \right] \]

\[ = U y^2 + \left( \beta - \overline{x}_1 \beta \right)' M (\beta - \overline{x}_1 \beta) \]

say.

\[ p(\beta, \sigma | y_2, x_2, x_1) \propto \left( \frac{1}{\sigma^a} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ U y^2 + (\beta - \overline{x}_1 \beta)' M (\beta - \overline{x}_1 \beta) \right] \right\} \]
And this is the same form as was the case with Sample 1. So analyze it in the same way if we wish.

Now, if we had posted the 2 samples, got a joint $2/2$, & used a different prior, we would have got the same end result.

Predictive pdf—

\[ q \text{ new observations } \sim q' = (y_{n+1}, \ldots, y_{n+n}) \]

\[ q = \overline{X}\beta + \overline{u}. \]

We want to obtain $p(\overline{y} | y, x, \overline{x})$.
We could take $p(\overline{y} | \beta, \sigma, y, x)$ and integrate w.r.t. $\sigma, \beta$.

Now, $p(\overline{y}, \beta, \sigma | y, x, \overline{x}) = p(\overline{y} | \beta, \sigma, x) p(\beta, \sigma | y, x)$
And, $p(\beta, \sigma | y, x) \propto \left(\frac{1}{\sigma^2}\right)^{n_1/2} \exp\left[-\frac{1}{2\sigma^2} \left\{ y + (\beta-x) x^T \right\}\right]$

\[ p(\overline{y} | \beta, \sigma, x) \propto \left(\frac{1}{\sigma^2}\right)^{(n+1)/2} \exp\left[-\frac{1}{2\sigma^2} \left(\overline{y} - \overline{x}\beta\right) \right]\]

\[ = \frac{1}{\sigma^{(n+1)/2}} \exp\left[-\frac{1}{2\sigma^2} \left(\overline{y} - \overline{x}\beta\right) \right]. \]

Integrate w.r.t. $\sigma$:

\[ p(\overline{y} | \beta, x, \overline{x}) \propto \int \left(\frac{1}{\sigma^2}\right)^{(n+1)/2} \exp\left[-\frac{1}{2\sigma^2} \left(\overline{y} - \overline{x}\beta\right) \right] d\sigma \]

\[ \propto \left[ (\overline{y} - \overline{x}\beta)^T (\overline{y} - \overline{x}\beta) \right]^{-\alpha/2}. \]

\[ p(\overline{y} | x, \overline{x}) \propto \int \left[ (\overline{y} - \overline{x}\beta)^T (\overline{y} - \overline{x}\beta) \right]^{-\alpha/2} \right) d\beta. \]

Now,

\[ (\overline{y} - \overline{x}\beta)^T (\overline{y} - \overline{x}\beta) \]

\[ = y'y - y'x\beta - \beta'x'y + \beta'x'x\beta + \overline{y}'\overline{y} - \overline{y}'\overline{x}\beta - \beta'\overline{x}'\overline{y} + \beta'\overline{x}'\overline{x}\beta \]

\[ = (y'y + \overline{y}'\overline{y}) + \beta' \left[ (x'y + x'y) \right] \beta - 2\beta' \left[ x'y + x'y' \right] \]

\[ = (y'y + \overline{y}'\overline{y}) + \beta' M \beta - 2\beta' (x'y + x'y') \]

Complete the square on $\beta$—

\[ \Rightarrow y'y + \overline{y}'\overline{y} - (y'x + \overline{y}'\overline{x}) M^{-1} \left( x'y + x'y' \right) \]

\[ + [\beta - M^{-1} \left( x'y + x'y' \right)]' M \left[ \beta - M^{-1} (x'y + x'y') \right]. \]
\[ p(y, x, \bar{x}) \propto (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \]
\[ + \left[ \beta_{-M^{-1}(x'y + x'\beta)} \right] \left[ M_{-M^{-1}(x'y + x'\beta)} \right]^{-(n+2)/2} \]
\[ \alpha \left[ (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \right] \]
\[ \left\{ 1 + \left[ \beta_{-M^{-1}(x'y + x'\beta)} \right] \left[ M_{-M^{-1}(x'y + x'\beta)} \right]^{-(n+2)/2} \right\}^{-n/2} \]

\[ p(y | x, \bar{x}) \propto \left[ (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \right]^{-(n+2)/2} \]
\[ \int \left\{ 1 + \left[ \beta_{-M^{-1}(x'y + x'\beta)} \right] \left[ M_{-M^{-1}(x'y + x'\beta)} \right]^{-(n+2)/2} \right\}^{-n/2} \]
\[ d\beta. \]

Let \( C = M / \{ y'y + \beta^2 - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \} \).

Do integrand = \( \left\{ 1 + \left[ \beta_{-M^{-1}(x'y + x'\beta)} \right] \left[ M_{-M^{-1}(x'y + x'\beta)} \right]^{-(n+2)/2} \right\}^{-n/2} \)
\[ \int \int |C|^{-1/2} \left\{ 1 + \left[ \beta_{-M^{-1}(x'y + x'\beta)} \right] \left[ M_{-M^{-1}(x'y + x'\beta)} \right]^{-(n+2)/2} \right\}^{-n/2} \]

But integrand is MVN-\( t \).

\[ p(y | x, \bar{x}) \propto \left[ (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \right]^{-(n+2)/2} |C|^{-1/2} \]

\[ \text{Now,} \quad |C|^{-1/2} = \left[ \det \left[ M / \{ y'y + \beta^2 - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \} \right] \right]^{-1/2} \]
\[ \alpha \left[ (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \right]^{-(n+2)/2} \]

So, \[ p(y | x, \bar{x}) \propto \left[ (y'y + \beta^2) - (x'y + x'\beta)^{M}(y' + \bar{x})^\top \right]^{-(n+2)/2} \]

where \( v = n-k \).

Now, simplify the expression in brackets on the RHS to get it into a more recognisable form.
\[
\begin{align*}
[y'y + \tilde{g}'\tilde{g}] &= (x'y + \tilde{x}'\tilde{x})M^{-1}(x'y + \tilde{x}'\tilde{x}) \\
= y'[I - \tilde{x}'M^{-1}x'y + \tilde{y}'[I - \tilde{x}'M^{-1}x'y]]^{-1}\tilde{x}M^{-1}x'y \\
= y'[I - \tilde{x}'M^{-1}x'y - y'M^{-1}\tilde{x}'[I - \tilde{x}'M^{-1}x'y]^{-1}\tilde{x}M^{-1}x'y] \\
&\quad + \tilde{y}'[\tilde{g}' - (I - \tilde{x}'M^{-1}x'y)^{-1}\tilde{x}M^{-1}x'y]'(I - \tilde{x}'M^{-1}x'y)[\tilde{g}' - (I - \tilde{x}'M^{-1}x'y)^{-1}\tilde{x}M^{-1}x'y] \\
\text{But,} \\
(I - \tilde{x}'M^{-1}x'y)^{-1} &= (I + \tilde{x}M^{-1}x'y)' \text{ OVER->} \text{ \footnote{\textit{See over}}}
\end{align*}
\]

So,
\[

tilde{g}'M^{-1}x'y = [I + \tilde{x}(x'y)^{-1}(x'y)]
tilde{x}M^{-1}
\]
\[
= (I + \tilde{x}M^{-1}x'y)(x'y + x'y)^{-1} \\
= (I + \tilde{x}(x'y)^{-1}(x'y))\tilde{x}(x'y + x'y)^{-1} \\
= (I + \tilde{x}(x'y)^{-1}(x'y))\tilde{x}(x'y + x'y)^{-1} \\
= \tilde{x}(x'y)^{-1} \\
\]
As the bracketed expression becomes
\[
y'[I - \tilde{x}'M^{-1}x'y - y'M^{-1}\tilde{x}'(x'y)^{-1}x'y] \\
\quad + \tilde{y}'[\tilde{g}' - x'(x'y)^{-1}x'y]'(I - x'M^{-1}x')\tilde{g}' - x'(x'y)^{-1}x'y] \\
\text{Let } \hat{\beta} = (x'y)^{-1}x'y
\]
\[
\Rightarrow \tilde{y}'[I - \tilde{x}'M^{-1}x'y] - y'M^{-1}\tilde{x}'(x'y)^{-1}x'y \\
\quad + \tilde{y}'[\tilde{g}' - x'(x'y)^{-1}x'y]'(I - x'M^{-1}x')\tilde{g}' - x'(x'y)^{-1}x'y] \\
\]
\[
\begin{align*}
\text{Note, } \tilde{y}'(I - x'(x'y)^{-1}x'y)y &= y'y' - y'(x'y)^{-1}x'y' \\
p(y'y, x, \tilde{x}) &\propto [x'y + (\tilde{y} - x'y)^'(I - x'M^{-1}x')\tilde{g}' - x'y]^{-1/2} \\
\text{Let } H &= (I - x'M^{-1}x') \\
p(y'y, x, \tilde{x}) &\propto [x'y + (\tilde{y} - x'y)^'H\tilde{g}' - x'y]^{-1/2} \\
\text{which is } MUS - t_y \\
\text{So, } E(y'y) &= x'y, \quad \text{var}(y) = (x'y)^{-1}\text{H}^{-1} \\
&= (x'y)^{-1}(I + x'(x'y)^{-1}x'y)' \\
\end{align*}
\]
\[
= [I - \bar{x}(x'x)^{-1}\bar{x}'] [I + \bar{x}(x'x)^{-1}\bar{x}'] \\
= [I - \bar{x}(x'x)^{-1}\bar{x}'] + \bar{x}(x'x)^{-1}\bar{x}' - \bar{x}(x'x)^{-1}x'(x'x)^{-1}\bar{x}'
\]

Now, 
\[
(x'x)(x'x)^{-1} - [M - (x'x)](x'x)^{-1}
\]
\[
= M(x'x)^{-1} - I
\]

\[
\bar{x} = [I - \bar{x}(x'x)^{-1}\bar{x}'] + \bar{x}(x'x)^{-1}\bar{x}' - \bar{x}(x'x)^{-1}x'(x'x)^{-1}\bar{x}'
\]
\[
= I - \bar{x}(x'x)^{-1}\bar{x}' + \bar{x}(x'x)^{-1}\bar{x}' - \bar{x}(x'x)^{-1}x'(x'x)^{-1}\bar{x}' + \bar{x}M^{-1}\bar{x}'
\]
\[
= I, \quad \text{as required}
\]
So, for a single element, \( \bar{y}_i \), if \( i \neq j \), we get:

\[
\left( \frac{\bar{y}_i - \bar{x}_i \beta}{\sqrt{\Sigma_{ii}}} \right) \sim t_\nu.
\]

where \( \Sigma_{ii} \) is the \((i,i)\) element of \( \Sigma^{-1} \).

**What if \( (X'X)^{-1} \) is singular?**

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

i.e. use an *informative* prior pdf. To choose a N.I. prior pdf for the parameters.

\[
\begin{align*}
N(\mu,\Sigma) &= N(\mu,\Sigma) = N(\mu,\Sigma) = \mathcal{N}(\mu,\Sigma) \\
\mu &\sim t_\nu \\
\Sigma &\sim \mathcal{W}(\nu, \Sigma_0)
\end{align*}
\]

and take the L.F.:

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

\[
\begin{align*}
p(\beta,\sigma^2|y,x) &\propto \left( \frac{1}{\sigma^2} \right) \exp \left[ -\frac{1}{2\sigma^2} \left( \frac{1}{\sigma^2} \right) \right] \\
&\propto \left( \frac{1}{\sigma^2} \right) \exp \left[ -\frac{1}{2\sigma^2} \left( \frac{1}{\sigma^2} \right) \right]
\end{align*}
\]

Now, look at:

\[
\begin{align*}
\gamma &= \beta' y + \bar{y}' (\bar{y}', \bar{x}')^{-1} (\bar{y} + \bar{x}' \beta) \\
&= \beta' A y = \beta' \bar{y} + \beta' \bar{x}' + y' \beta - \beta' x' = \beta' (A + x' x) \beta + \beta' \bar{y} + y' \beta - \beta' x' = \beta' A \beta + \beta' \bar{y} + y' \beta - \beta' x' = \beta' [A + x' x] \beta - \beta' [A \bar{y} + x' y + x' \beta] + y' \beta + \bar{y}' A \bar{y} \\
&= \beta' [A + x' x] \beta - \beta' [A \bar{y} + x' y + x' \beta] + y' \beta + \bar{y}' A \bar{y} + \bar{y}' A \bar{y} - \beta' x' = \beta' A \beta + \beta' \bar{y} + y' \beta - \beta' x' = \beta' [A + x' x] \beta - \beta' [A \bar{y} + x' y + x' \beta] + y' \beta + \bar{y}' A \bar{y}.
\end{align*}
\]

Complete the square on \( \beta - \bar{x}' \gamma = y' y + \bar{y}' A \bar{y} - (A \bar{y} + x' y)' (A + x' x)^{-1} (A \bar{y} + x' y) \\
+ [\beta - (A + x' x)^{-1} (A \bar{y} + x' y)]' (A + x' x) [\beta - (A + x' x)^{-1} (A \bar{y} + x' y)]
Let \( \hat{\beta} = (A + x'x)^{-1}(A\beta + x'y) \)

\[
\hat{\gamma} = y'y + \frac{1}{n} A\hat{\beta} - (A\hat{\beta} + y') (\hat{\beta} + (\gamma - \hat{\gamma}) (A + x'x) (\gamma - \hat{\gamma}) - y') (A + x'x)^{-1}\hat{\gamma}
\]

\[
p(\beta, \sigma^2 | y) \propto \left( \frac{1}{\sigma^{n/2 + 1}} \right) \exp \left\{ -\frac{1}{2\sigma} \left[ \frac{y'y + \frac{1}{n} A\hat{\beta} - (A\hat{\beta} + y') (\hat{\beta} + (\gamma - \hat{\gamma}) (A + x'x) (\gamma - \hat{\gamma}) - y') (A + x'x)^{-1}\hat{\gamma} \right] \right\}
\]

where \( n' = n + n_0 \)

Let \( n' \sigma^2 = \left[ \frac{y'y + \frac{1}{n} A\hat{\beta} - (A\hat{\beta} + y') (\hat{\beta} + (\gamma - \hat{\gamma}) (A + x'x) (\gamma - \hat{\gamma}) - y') (A + x'x)^{-1}\hat{\gamma} \right] \)

\[
p(\beta, \sigma^2 | y) \propto \left( \frac{1}{\sigma^{n'/2 + 1}} \right) \exp \left\{ -\frac{1}{2\sigma} \left[ n' \sigma^2 + (\gamma - \hat{\gamma}) (A + x'x) (\gamma - \hat{\gamma}) \right] \right\}
\]

Which is MVS-\( t_0 \). Post-Mean is \( \hat{\beta} \).

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

\[\text{[NB]}\] Why not look for the least informative prior that will enable you to still handle the estimation problem.

\[\text{[NB]}\] Why not use an information-theoretic approach to getting such a "minimal-information" prior.

\[
n' = T + n_0
\]

\[
n' + k = T + n_0 + k
\]
$y = X\delta + u = X\hat{\delta} + u$

where $P$ is orthogonal, so $P^TP = I$.

$\therefore \delta = P\hat{\delta}$

Let $X^2$ have rank $q$; then $\exists y \text{ non-zero}$

C.R. $\delta$ of $(X^tX)$.

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

Set up the "normal" equations:

$P^tX^ty = P^tX^tX\delta$

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

$\therefore (P^tX^ty) = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}(\delta_1)$

$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = (P^tX^ty)$

i.e. $P^tX^ty = 0$.

Now, solve the system $\ast$ (use $g$-inverse).

We have $(P^tX^tX)\delta = P^tX^ty$

i.e. $A\delta = b$.

Then $g$-inverse $\Rightarrow$

$\delta = A^g b + (I - A^gA)z$; arbitrary $z$.

So $\delta = (P^tX^tX)^gP^tX^ty + [I - (P^tX^tX)^g(P^tX^tX)]z$

Now, there are a variety of ways $\delta$ choosing the $g$-inverse and choosing $z$.\]
We have that:
\[
\bar{Y} = (P'X'XP)^* P'X'y + \left[ I - (P'X'XP)^* P'X'XP \right] z
\]

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

Then this is a GI for \((P'X'XP)\):
\[
\begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{D} \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix}
\]
\[
= \begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix}
\]
\[
\begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} P'X'y + \left[ I - \begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} \right] z_2
\]
\[
\begin{pmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{F} \end{pmatrix} P'X'y + \begin{pmatrix} \mathbf{0} \\ \mathbf{c} \end{pmatrix}
\]
\[
\begin{pmatrix} \mathbf{D}^{-1} P'X'y \\ \mathbf{c} \end{pmatrix}
\]
\[
\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{c} \end{pmatrix}
\]
\[
\begin{pmatrix} \mathbf{D}^{-1} P'X'y \\ \mathbf{c} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{z}_2 - \mathbf{cD} \mathbf{z}_1 \end{pmatrix}
\]
\[
\begin{pmatrix} \mathbf{D}^{-1} P'X'y \\ \mathbf{z}_2 - \mathbf{cD} \mathbf{z}_1 \end{pmatrix}
\]

So \(\bar{Y}_1\) is independent of the choice of GI or \(z\), but \(\bar{Y}_2\) depends on the choice of both GI and \(z\).

Primarily we are interested in the \(\bar{Y}_1\) part, because that is the part of the coefficient vector not affected directly by the multicollinearity.

Examples: Pick the Moore-Penrose GI.
\[
(P'X'XP)^* = \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}
\]

Then for this GI or for any \(\exists \mathbf{c} = 0\), we have:
\[
\begin{align*}
\bar{Y}_1 &= \mathbf{D}^{-1} P'X'y \\
\bar{Y}_2 &= \mathbf{z}_2
\end{align*}
\]

So \(\bar{Y}_2\) still depends on the choice of \(z\), but no longer on the choice of the GI.
\[ \beta = P \gamma \]

So \( \tilde{\beta} = P \tilde{\gamma} \)

\[ \begin{align*}
\tilde{\beta} &= P \left( (P'X'XP)^* P'X'Y + (I - (P'X'XP)^* P'X'XP)Z \right) \\
&= P \left( (P'X'XP)^* P' \right) X'Y + \left[ I - P (P'X'XP)^* P' \right] X'X \right) P Z
\end{align*} \]

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 orthonormal.

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

\(\therefore (X'X)[P(P'X'XP)^*P'](X'X)^* - (X'X) \]

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

\[ \begin{align*}
\tilde{\beta} &= \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \tilde{\beta}_4 \\ \tilde{\beta}_5 \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \\ \tilde{\gamma}_3 \\ \tilde{\gamma}_4 \\ \tilde{\gamma}_5 \end{pmatrix} \\
&= \begin{pmatrix} P_{11} & P_{12} \end{pmatrix} \begin{pmatrix} D^{-1}P'X'Y \\ P_{21} \end{pmatrix} \\
&= \begin{pmatrix} P_{11} D^{-1} \ast P'X'Y + P_{12}(Z_2 + CD(\tilde{\gamma}_1, -Z_1)) \\ P_{21} D^{-1} \ast P'X'Y + P_{22}(Z_2 + CD(\tilde{\gamma}_1, -Z_1)) \end{pmatrix}
\end{align*} \]

So, although \(\tilde{\gamma}_1\) is independent f the choice of the GI \(\tilde{\gamma}_2\), it turns out that \(\tilde{\beta}_1\) (\(\tilde{\beta}_2\)) depends on both the GI \(\tilde{\gamma}_1\) and the choice \(Z_2\).

Even if we pick the Moore-Penrose GI, or any other \(E = 0\),

then both still depend on \(Z_2\) only.

\[ \text{viz.} \quad \tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \tilde{\beta}_4 \\ \tilde{\beta}_5 \end{pmatrix} = \begin{pmatrix} P_{11} D^{-1} \ast P'X'Y + P_{12}Z_2 \\ P_{21} D^{-1} \ast P'X'Y + P_{22}Z_2 \end{pmatrix} \]
Bayesian interpretation

Now, give a Bayesian interpretation to this sampling theory approach to handling the singularity of \((x'x)^{-1}\).

We have the \(LIF\)
\[
L(\beta, \sigma^2 | y) \propto \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ (y-x\beta)'(y-x\beta) \right] \right\}
\]

But \(\beta = P\gamma\)
\[
\Rightarrow |P| \propto \text{const.}
\]

So,
\[
L(\gamma, \sigma, \sigma' | y) \propto \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ (y-xP)'(y-xP) \right] \right\}
\]

Let \(\tilde{\gamma} = \tilde{\gamma}'P'x'y\)

9. Let \(a = (y-xP)'(y-xP)\)

Now,
\[
y = x\beta + u = xP\gamma + u
\]
\[
\Rightarrow (y-x\beta) = u = (y-xP\gamma)
\]
\[
y - x(P_{1,1}P_{1,1})\tilde{\gamma}
\]
\[
y - xP_{1,1}\tilde{\gamma} \quad ; \, \text{since} \, xP_{2} = 0
\]
\[
\Rightarrow (y-xP\gamma) = (y-xP_{1,1}\tilde{\gamma})
\]
\[
\Rightarrow L(\gamma, \sigma, \sigma' | y) \propto \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ (y-xP_{1,1}\tilde{\gamma})'(y-xP_{1,1}\tilde{\gamma}) \right] \right\}
\]

But,
\[
(y-xP_{1,1}\tilde{\gamma})'(y-xP_{1,1}\tilde{\gamma})
\]
\[
= \left[ (y-xP_{1,1}\tilde{\gamma}) - (xP_{1,1}\tilde{\gamma} - xP_{1,1}\tilde{\gamma}) \right]' \left[ (y-xP_{1,1}\tilde{\gamma}) - (xP_{1,1}\tilde{\gamma} - xP_{1,1}\tilde{\gamma}) \right]
\]
\[
= (y-xP_{1,1}\tilde{\gamma})'(y-xP_{1,1}\tilde{\gamma}) + (xP_{1,1}\tilde{\gamma} - xP_{1,1}\tilde{\gamma})'(xP_{1,1}\tilde{\gamma} - xP_{1,1}\tilde{\gamma})
\]
\[
= a + (\tilde{\gamma}' - \tilde{\gamma}' P'x'P_{1,1}x'P_{1,1}\tilde{\gamma})
\]

But \(P'x'xP_{1,1} = D\)
\[
\Rightarrow (P'x'P_{1,1}x'P_{1,1}) = D
\]
\[
L(\gamma, \sigma, \sigma' | y) \propto \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ a + (\tilde{\gamma}' - \tilde{\gamma}' D)D(\tilde{\gamma}' - \tilde{\gamma}) \right] \right\}
\]

\text{NB} The LIF is quite independent of \(x_2\).

That is, it is now on a form where none of the (true values of the) coefficients in the equation are omitted from the denominator — we are concentrating only on those not affected directly by the singularity of \((x'x)^{-1}\).
Now, take the general case, where
\[(P_1^*P_1) = \begin{pmatrix} c & \mathbf{f} \\ \mathbf{0} & F \end{pmatrix} \quad ; \quad c \neq 0.\]

Let \(P(\tau_1, y) \propto \frac{1}{\sigma} ; 0 < \sigma < \infty ; -\infty < \tau_1 < \infty \); \(i = 1, \ldots, q\).

Then we have:
\[
\tau_2 = \tau_1 + cD(\tau_1 - z_i).
\]

Then by assigning values to \(c\) and to \(z_1, z_2\), we are able to introduce our given knowledge.
\[
p(\tau_1, y) \propto \left(\frac{1}{\sigma} \right) \exp \left[ -\frac{1}{2\sigma} \left( \mathbf{a} + (\tau_1 - \bar{y})' \mathbf{D} (\tau_1 - \bar{y}) \right) \right]
\]

And this pdf is proper for \(\tau_1 \) (ie, the d.f.).
And \(p(\tau_1, y) \propto \left[ \mathbf{a} + (\tau_1 - \bar{y})' \mathbf{D} (\tau_1 - \bar{y}) \right]^{-\frac{n}{2}}
\]

So, \(E(\tau_1) = \bar{y}
\]

Then \(\tau_2 = \tau_1 + cD(\tau_1 - z_i)
\]

\[
\Rightarrow \quad E(\tau_2) = \tau_1 + cD(\bar{y}_1 - z_i)
\]

\[
= \bar{y}_1 \quad \text{as before}
\]

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

Take the special case where \(c = 0\) (eg the

\text{Pearson} - \text{Lorenz} \quad \text{situation} \quad -
\]

\[\tau_2 = \tau_1.
\]

So if we assign a value to \(z_1\), we assign a specific value to \(\tau_2\) - we are using "degmanics". The prior pdf

\(x_1 = \text{then degmanics} \quad \text{with \(y\)} \) its mass is concentrated

\(\text{at the point} \ z_1. \quad \text{Further,} \quad c = 0 \implies \tau_2 \text{is independent of} \ y_1.
\]

Clearly, to assume that \(c = 0\) so that \(\tau_2 = \tau_1\),

is very restrictive in practice. Can we loose this

assumption a little?
One way of doing this is to put

\[ p(\gamma_2 \mid \sigma) \propto \left( \frac{1}{\Delta \gamma} \right) \exp \left\{ -\frac{1}{2\Delta \gamma} (\gamma_2 - z_2)'^T (\gamma_2 - z_2) \right\} \]

**Note:** 3 (k-2) obs. for \( \gamma_2 \).

Then, since \( p(\sigma, \sigma) \propto \sigma \)

\[ \Rightarrow p(\sigma, \gamma_2, \sigma) = p(\sigma, \sigma) p(\gamma_2 \mid \sigma) \]

\[ \propto (\sigma^{k+1} \gamma_2) \exp \left\{ -\frac{1}{2\Delta \gamma} (\gamma_2 - z_2)' Q(\gamma_2 - z_2) \right\} \]

i.e., \( p(\sigma, \gamma_2, \sigma \mid y) \propto (\sigma^{k+1} \gamma_2) \exp \left\{ -\frac{1}{2\Delta \gamma} (\gamma_2 - z_2)' Q(\gamma_2 - z_2) \right\} + \lambda \gamma_2 \mathcal{Q} (\gamma_2 - \bar{\gamma}_2) \right\} \]

let \( \beta = \beta \), \( \gamma_2 = (\gamma_2', \gamma_2) \), \( \Delta = (\Delta \gamma) \)

Then,

\[ p(\beta, \sigma \mid y) \propto (\sigma^{k+1} \gamma_2) \exp \left\{ -\frac{1}{2\Delta \gamma} \left[ \lambda + (\beta - \beta_0)' P_{FP}' (\beta - \beta_0) \right] \right\} \]

Note that \( P_{FP}' = (P_1, D_1)' + P_2 \mathcal{Q} P_2 \) is non-singular, even though both \( P_1, D_1 \) and \( P_2 \mathcal{Q} P_2 \) are singular.

So the posterior is proper for \( \lambda > 0 \), and can be used for posterior inferences on \( \beta \).

Finally consider the situation if a diffuse prior is used for \( \beta \), then \( (X'X) \) is singular.

Then \( p(\beta) \propto \text{const.} \quad -\infty < \beta < \infty \) \quad \forall \gamma.

So, for given \( \sigma \) we get:

\[ p(\beta \mid \sigma, y) \propto \exp \left\{ -\frac{1}{2\Delta \gamma} (\gamma - \beta)' (\gamma - \beta) \right\} \]

\[ \alpha \exp \left\{ -\frac{1}{2\Delta \gamma} \beta^T X' X (\beta - \bar{\beta}) \right\} \]

For any \( \lambda > 0 \), so \( \beta \) is the normal in the \( \beta \) is singular. \( \beta \) is not unique.

Now, \( \beta = \bar{\beta} \)

So, \( p(\sigma, \gamma_2, \sigma \mid y) \propto (\sigma^{k+1} \gamma_2) \exp \left\{ -\frac{1}{2\Delta \gamma} (\gamma_2 - \bar{\gamma}_2)' Q(\gamma_2 - \bar{\gamma}_2) \right\} \]

Which is independent of \( \gamma_2 \). So the posterior for \( \gamma_2 \) is diffuse, as is its prior. Even if we introduce no prior information, it is possible to make inferences concerning the linearly independent combinations of the elements of \( \beta \).
A) Autocorrelated Errors:

\[
\begin{align*}
\begin{cases}
    y_t &= \beta x_t + u_t \\
    u_t &= \rho u_{t-1} + \epsilon_t
\end{cases}
\end{align*}
\]

\(\epsilon_t \sim \text{IN}(0, \sigma^2)\).

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 unknown, then,

\[
y_0 = \rho y_{-1} + \beta (x_0 - \rho x_{-1}) + \epsilon_0
= \rho (y_{-1} - \beta x_{-1}) + \beta x_0 + \epsilon_0
= (y_0 - \beta x_0) = M + \epsilon_0, \text{ say.}
\]

i.e. \(M\) is the initial level of the process.

Now, \(\epsilon_0 \sim \text{N}(0, \sigma^2)\).

\[
E(y_0) = \beta x_0 + M
\]

\[
\text{var}(y_0) = \sigma^2.
\]

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,

\[
p(y_0, y | \beta, \sigma, \rho, M) = p(y_0 | \beta, \sigma, \rho, M) p(y | y_0, \beta, \sigma, \rho, M)
\]

\[
\propto \left(\frac{1}{\sigma^m}\right) \exp \left\{ - \frac{1}{2\sigma^2} \left[ (y_0 - \beta x_0 - M)^2 + \sum_{t=1}^{T} \left[ y_t - \rho y_{t-1} - \beta (x_t - \rho x_{t-1}) \right]^2 \right] \right\}
\]

\(\sigma > 0; -\infty < \rho < 1; -\infty < \sigma < \infty; -\infty < M < \infty\).
And let \( p(\beta, \sigma, \rho, \mu) \propto (1/\sigma) \).

\[
p(\beta, \rho, \sigma, \mu | y) \propto \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\}
\]

And, usually, \( \mu \) is of little interest to us. So:

\[
p(\beta, \rho, \sigma | y) = \int_{\infty}^{\rho_{max}} p(\beta, \rho, \sigma, \mu | y) d\mu = \int_{\infty}^{\rho_{max}} \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\} d\mu
\]

\[
\alpha \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\} \int_{\infty}^{\rho_{max}} \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\} d\mu
\]

Which is the same posterior we would have got by ignoring complications regarding \( y_0 \), etc.

Now, \( p(\beta, \rho | y) = \int_{\sigma} p(\beta, \rho, \sigma | y) d\sigma \)

\[
\alpha \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\} d\sigma
\]

\[
\alpha \left\{ \frac{1}{\sigma} \sum_{i=1}^{n} \left( y_i - \beta \mu - \rho (x_i - \rho x_{t-1}) \right)^2 \right\} \cdot \left( \frac{\sigma}{\Gamma(n/2)} \right)
\]

Complete the square on \( \beta \) to get:

\[
p(\beta | y) = \int_{-\infty}^{\infty} p(\beta, \rho | y) d\beta
\]

\[
p(\beta | y) \propto \left[ \sum (x_t - \rho x_{t-1}) \right]^{-\frac{n}{2}} \cdot \left[ \left\{ \frac{\sum (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum (x_t - \rho x_{t-1})} \right\} \right]^{-\frac{(n-1)}{2}}
\]

Complete square on \( \rho \) and integrate to get:

\[
p(\rho | y) \propto \left[ \sum (y_t - \beta x_t) \right]^{-\frac{n}{2}} \cdot \left[ \left\{ \frac{\sum (y_t - \beta x_t)(y_t - \beta x_t)}{\sum (y_t - \beta x_t) \left( y_t - \beta x_t \right)} \right\} \right]^{-\frac{(n-1)}{2}}
\]
Note that for $p(y | y)$ to be a proper pdf, we require that $\sum (x_t - \beta x_{t-1})^2 > 0$.

i.e. For a given $\beta$, $\exists$ some $t \neq$

$\forall t \neq \beta x_{t-1}$. (This isn't too restrictive.)

The posterior pdf for $\beta$ enables us to make inferences about this parameter which allows for departures from serial independence. Recall that if we assume that $\beta = 0$, then

$$\frac{(y - \beta)^2}{S/(2 \tilde{y}^2)} \sim t_v.$$

And this will give a completely different posterior distribution.

Recall that

$$p(\beta | y) = \int_0^\infty p(\beta | y, \rho) \rho d\rho \int_0^\infty p(\beta | \rho, y) p(\rho | y) d\rho.$$

is a weighted average of $p(\beta | y, \rho)$, with $p(\rho | y)$ as the weighting factor. Note that $p(\beta | y, \rho)$ provides inferences about $\beta$ for a given $\rho$, while $p(\rho | y)$ reflects the plausibility of assuming $\rho$ in the light of our prior notion of the data evidence.

Note that

$$p(\beta | y, \rho) \propto [S(\rho)]^{-\frac{1}{2}} \left[ \sum (x_t - \beta x_{t-1})^2 \right]^{-\frac{1}{2}} \left[ \frac{(y - \beta x_t)^2}{S(\rho)} \right]^{\frac{1}{2}}$$

where $V = T - 1$

$$S(\rho) = \sum (x_t - \beta x_{t-1})^2$$

$$\hat{S}(\rho) = \sum \frac{(y_t - \beta y_{t-1} - \beta'(\rho)(x_t - \beta x_{t-1}))^2}{S(\rho)}$$

$$S(\rho) = \left[ \sum (x_t - \beta x_{t-1})^2 \right]^{\frac{1}{2}}$$

$$V = T - 1$$

$$\frac{(\beta - \hat{\beta}(\rho))}{S(\rho)} \sim t_v.$$
Let \( y = X \beta + \epsilon \)
\[ U = \rho U_{-1} + \epsilon \]
\[ y - \rho y_{-1} = (X - \rho X_{-1}) \beta + \epsilon \]

Let \( y_0 = \rho (y_{-1} - \beta X_{-1}) + X \beta + \epsilon \)
\[ = U + X \beta + \epsilon \]
\[ M = \rho (y_{-1} - \beta X_{-1}) \]

\[ \therefore \quad p(y, y_0 | \beta, \sigma, \rho, M) \propto \left( \frac{1}{\sigma} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ y_0 - X_0 \beta - M \right]^2 \right\} \]
\[ \propto \rho y_{-1} \exp \left\{ -\frac{1}{2\sigma^2} \left[ y_0 - X_0 \beta - M \right]^2 \right\} \]
\[ = \rho(y_0, \beta, \sigma, \rho, M) \]

\[ \alpha \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ y_0 - X_0 \beta - M \right]^2 \right\} \]
\[ \quad - \frac{1}{2\sigma^2} \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right]^2 \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right] \]

\[ \alpha \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ y_0 - X_0 \beta - M \right]^2 \right\} \]
\[ \quad \times \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right] \]

\[ p(y, \beta, \sigma, \rho, M) \propto \left( \frac{1}{\sigma^{n+1}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ y_0 - X_0 \beta - M \right]^2 \right\} \]
\[ \quad - \frac{1}{2\sigma^2} \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right]^2 \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right] \]

\[ p(y, \beta, \rho) \propto \left\{ \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right]^2 \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right] \right\}^{-(n+1)} \]

Now, the OLS estimator for
\[ (y - \rho y_{-1}) = (X - \rho X_{-1}) \beta + \epsilon \]
so the Cochrane–Orcutt estimator
\[ \hat{\beta}^{(CO)} = H^{-1}(X - \rho X_{-1})' (y - \rho y_{-1}) \]

Where \( H = (X - \rho X_{-1})' (X - \rho X_{-1}) \)

\[ \hat{u} s^2 = \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right]^2 \left[ y - \rho y_{-1} - (X - \rho X_{-1}) \beta \right] \]
And, for the OLS squares case we had

\[ p(\beta | y) \propto \left\{ \frac{1}{y^2} + \left[ \beta - \beta(\bar{y}) \right] \right\}^{-\frac{n}{2}} \]

\[ \propto \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}}. \]

So here -

\[ p(\beta | y) \propto \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}}. \]

And this is a MV- \( t \) distribution, with mean \( \beta(\bar{y}) \).

We have assumed that \( H \) is non-singular \( \text{if} \ k = 1 \)

This requires that \( x_t = p \cdot x_{t-1} \) for some \( t \). \( \text{If} \ k > 1 \)

This requires that any linear combination of the columns \( y_j \), \( \bar{y}_j (x - p x_{t-1}) \) must not be exactly 1st-order autoregressive.

Then,

\[ p(\beta | y) = \int p(\beta | y_0) \, d\beta \]

\[ = \int \left\{ \frac{1}{y^2} + [\beta - \beta(\bar{y})]^2 (x'x)^{-1} \right\}^{-\frac{n}{2}} \, d\beta \]

and

\[ p(\beta | y) = \int p(\beta | y_0) \, d\beta \]

Consider the 2nd and one-

\[ p(\beta | y) \propto \left\{ \frac{1}{y^2} + [\beta - \beta(\bar{y})]^2 (x'x)^{-1} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]

\[ \propto \left[ \frac{1}{y^2(\beta - \beta(\bar{y}))} \right]^{-\frac{n}{2}} \int \left\{ 1 + \frac{(\beta - \beta(\bar{y}))^2 (x'x)^{-1}}{y^2} \right\}^{-\frac{n}{2}} \, d\beta \]
Because:

\[(y - \rho y_1)' [I - (x - \rho x_1)H^{-1}(x - \rho x_1)'y] (y - \rho y_1)\]

\[= (y - \rho y_1)'(y - \rho y_1) - (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

And:

\[\[y - \rho y_1 - (x - \rho x_1)y\] [y - \rho y_1 - (x - \rho x_1)y] \]

\[= (y - \rho y_1)'(y - \rho y_1) - \beta(\rho)'(x - \rho x_1)'(y - \rho y_1)\]

\[- (y - \rho y_1)'(x - \rho x_1)'\beta(\rho) + \beta(\rho)'(x - \rho x_1)'(x - \rho x_1)'\beta(\rho)\]

\[= (y - \rho y_1)'(y - \rho y_1) - \left[H^{-1}(x - \rho x_1)'(y - \rho y_1)\right]'(x - \rho x_1)'(y - \rho y_1)\]

\[- (y - \rho y_1)'(x - \rho x_1)[H^{-1}(x - \rho x_1)'(y - \rho y_1)]\]

\[+ \left[H^{-1}(x - \rho x_1)'(y - \rho y_1)\right]'(x - \rho x_1)'(x - \rho x_1)'\beta(\rho)\]

\[= (y - \rho y_1)'(y - \rho y_1) - (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

\[- (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

\[+ (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

\[= (y - \rho y_1)'(y - \rho y_1) - 2(y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

\[+ (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

\[= (y - \rho y_1)'(y - \rho y_1) - (y - \rho y_1)'(x - \rho x_1)H^{-1}(x - \rho x_1)'(y - \rho y_1)\]

So the two are equivalent.

Now, consider:

\[p(\beta|y) = \int p(\beta, \rho|y)\,d\rho = \int \left[\left[y - \chi \beta - \rho(y_1 - x_1)\beta\right] \left[y - \chi \beta - \rho(y_1 - x_1)\beta\right]\right]^{-\frac{1}{2}}\,d\rho\]
\[ \alpha \int \left\{ (y-x_\beta)'(y-x_\beta) + \sigma^2 (y_i-x_\beta)'(y_i-x_\beta) - \tau (y_i-x_\beta)(y-x_\beta) \right\} \frac{1}{\phi} \, d\phi \]

\[ \int \left\{ (y_i-x_\beta)'(y_i-x_\beta) \left[ \phi - \left[ (y_i-x_\beta)'(y_i-x_\beta) \right]^{-1} (y_i-x_\beta)(y-x_\beta) \right] \right\} \frac{1}{\phi} \, d\phi \]

\[ \int \left\{ H \left[ \phi - \left[ (y_i-x_\beta)'(y_i-x_\beta) \right]^{-1} (y_i-x_\beta)(y-x_\beta) \right] \right\} \frac{1}{\phi} \, d\phi \]

\[ \frac{\beta}{\delta} \int\left\{ \frac{y_i-x_\beta}{\phi} \left[ \frac{y_i-x_\beta}{\phi} \right] \right\} \frac{1}{\phi} \, d\phi \]

\[ \gamma \left( \frac{\tau}{\delta} \right)^{\frac{1}{2}} \int \frac{1}{\phi} \, d\phi \]

\[ \left[ (y_i-x_\beta)'(y_i-x_\beta) \right]^{-\frac{1}{2}} \left\{ (y-x_\beta)'(y-x_\beta) - \frac{\left[ (y-x_\beta)'(y-x_\beta) \right]^{2}}{(y_i-x_\beta)'(y_i-x_\beta)} \right\} \]

---

What if we are interested in a single element \( \frac{\beta}{\tau} \)? Integrate out all of the unwanted elements? Messy — especially if \( k \) is large. So, alternatively,

\[ p(\beta_1, p(y)) = p(p(y), \beta_1, p, y) \]

And we have just derived \( p(\beta_1, p(y)) \), and we can obtain \( p(\beta_1, p, y) \) by integrating \( p(\beta_1, p, y) \) w.r.t. the remaining \((k-1)\) elements of \( \beta_1 \).

Then,

\[ \frac{\beta_i - \beta_i(p)}{s(\phi) \sqrt{n''}} \sim t \nu \]

where \( n'' \) is the \((1,1)\)th element of \( H^{-1} \).
Then use bivariate integration to get:

\[ p(\beta, \rho y) = \int p(\beta, \rho y') d\rho. \]

We can do this because result * tells us the form of the pdf \( p(\beta, \rho y). \) We already know the form of \( p(\rho y), \) so now we know the form of \( p(\beta, \rho y) \)

then integrate out \( \rho, \) as above.

Alternatively, partition \( \beta \to \beta' = (\beta_1 : \beta') \)

Let \( X = (x_1 : x) \)

Let \( \omega = y - \rho y_{-1} - (x_1 - \rho x_{-1}) \beta_1 \)

Then

\[ p(\beta, \rho y) \propto \left[ y - \rho y_{-1} - (x_1 - \rho x_{-1}) \beta_1 \right] \left[ y - \rho y_{-1} - (x - \rho x_{-1}) \beta_1 \right]^{-1/2} \]

\[ \times \left[ y - \rho y_{-1} - (x_1 - \rho x_{-1}) \beta_1 - (x - \rho x_{-1}) \beta_1 \right] \left[ y - \rho y_{-1} - (x - \rho x_{-1}) \beta_1 \right]^{-1/2} \]

\[ \times \left[ \omega - (x - \rho x_{-1}) \beta_1 \right] \left[ \omega - (x - \rho x_{-1}) \beta_1 \right]^{-1/2} \]

\[ \therefore p(\beta_1, \rho y) = \int p(\beta_1, \rho y) d\rho \]

\[ \times \int \left\{ \omega' \omega + \beta' \left( \beta - \beta_{-1} \right) \left( \beta - \beta_{-1} \right) \omega - 2 \beta' \left( \beta - \beta_{-1} \right) \omega \right\}^{-1} d\beta \]

Let \( \mathbb{H} = \left( \beta - \beta_{-1} \right) \left( \beta - \beta_{-1} \right) \)

\[ \times \int \left\{ \mathbb{H} \left[ \mathbb{H}^{-1} \left( \beta - \beta_{-1} \right) \omega \right]^2 + (\omega' \omega) \right\}^{-1} \left\{ \mathbb{H} \left[ \mathbb{H}^{-1} \left( \beta - \beta_{-1} \right) \omega \right]^2 + (\omega' \omega) \right\}^{-1} d\beta \]

\[ \times \left\{ \omega' \omega - (\omega') \left( \beta - \beta_{-1} \right) \mathbb{H}^{-1} \left( \beta - \beta_{-1} \right) \omega \right\} \left\{ \omega' \omega - (\omega') \left( \beta - \beta_{-1} \right) \mathbb{H}^{-1} \left( \beta - \beta_{-1} \right) \omega \right\}^{-1/2} \]

\[ \times \left\{ \left( \mathbb{H}^{-1} \right) \left( \beta - \beta_{-1} \right) \omega \right\} \left\{ \left( \mathbb{H}^{-1} \right) \left( \beta - \beta_{-1} \right) \omega \right\}^{-1} \]

\[ \times \left\{ \left( \mathbb{H}^{-1} \right)^{-1} \left( \beta - \beta_{-1} \right) \omega \right\} \left\{ \left( \mathbb{H}^{-1} \right)^{-1} \left( \beta - \beta_{-1} \right) \omega \right\}^{-1} \]

\[ \times \left\{ \left( \mathbb{H}^{-1} \right)^{-1} \left( \beta - \beta_{-1} \right) \omega \right\} \left\{ \left( \mathbb{H}^{-1} \right)^{-1} \left( \beta - \beta_{-1} \right) \omega \right\}^{-1} \]
\( \alpha \{ (w^\top w) - w^\top (X - \rho X_1) H^{-1} (X - \rho X_1)^\top w \} \sum_{i=1}^{q} (\tau-k+i)^{-\frac{1}{2}} \propto |H|^{-\frac{1}{2}} \)

Then integrate w.r.t. \( \beta \) to get \( p(\beta | y) \). The advantage is that \( H^{-1} \) is inversion of a \((k-1) \times (k-1)\) matrix, c.f. \( H^{-1} \) is inversion of a \((k \times k)\) matrix.

**B) Regressions with Unequal Variances:**

Consider 2 regressions based on the same model, but 2 different samples.

\[
\begin{align*}
    y_1 &= X_1 \beta + u_1, & u_1 &\sim \mathcal{N}(0, \sigma_1^2) \\
    y_2 &= X_2 \beta + u_2, & u_2 &\sim \mathcal{N}(0, \sigma_2^2)
\end{align*}
\]

If \( \sigma_1^2 = \sigma_2^2 \), or if \( \sigma_1^2 = \sigma_2^2 \) (for known \( \sigma \) then pool the data for \( y = X \), e compute:

\[
y = X \beta + u.
\]

But what if \( \sigma_1^2 \neq \sigma_2^2 \)? Then consider 2 cases:

(a) \( \sigma^2 \) known, but \( \sigma_1^2, \sigma_2^2 \) unknown.

(b) Both \( \sigma_1^2, \sigma_2^2 \) unknown.

\( a) \) \( \sigma^2 \) known, \( \sigma_1^2, \sigma_2^2 \) unknown:

\[
L(\beta, \sigma^2 | y, \sigma) \propto \left( \frac{1}{\sigma^2} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y_1 - X_1 \beta)'(y_1 - X_1 \beta) \\
-\frac{1}{2\sigma^2} (y_2 - X_2 \beta)'(y_2 - X_2 \beta) \right\}
\]

\[
p(\beta, \sigma^2) \propto \left( \frac{1}{\sigma^2} \right) \prod_{i=1}^{k} \bigg\{ -\infty < \beta_i < \infty \bigg\}.
\]

\[
p(\beta | \sigma^2, y) \propto \left( \frac{1}{\sigma^2} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y_1 - X_1 \beta)'(y_1 - X_1 \beta) \\
-\frac{1}{2\sigma^2} (y_2 - X_2 \beta)'(y_2 - X_2 \beta) \right\}.
\]

\[
p(\beta | y, \sigma) \propto \exp \left\{ -\frac{1}{2\sigma^2} (y_1 - X_1 \beta)'(y_1 - X_1 \beta) \right\} \cdot (y_2 - X_1 \beta)'(y_2 - X_1 \beta)^{-\frac{1}{2}}.
\]
Let \( z_i = (x_i', x_i) \); \( \hat{z}_i = (x_i', x_i') \); \( \overline{z}_i = z_i^{-1} (x_i', y_i) \);
\( \overline{z}_i = z_i^{-1} (y_i, x_i') \); \( u_i = n_i - k \); \( u_i = n_i - k \);
\( u_i = n_i - k \);
\( \overline{z}_i = u_i + \beta \hat{z}_i \).

\[
\begin{align*}
\frac{1}{2} \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( y_i - x_i' \beta \right)' \left( y_i - x_i' \beta \right) \right\} \\
&= \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( y_i, y_i + \beta' x_i' x_i' \beta - 2y_i' x_i' \beta \right) \right\} \\
&= \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( y_i', y_i + \beta' z_i (\beta - 2y_i' x_i' \beta) \right) \right\}
\end{align*}
\]

And \( \beta_i = z_i^{-1} (x_i', y_i) \).

\[
\begin{align*}
\beta_i \hat{z}_i = x_i' y_i \\
\beta_i \hat{z}_i = z_i^{-1} (x_i', y_i) \\
\beta_i \hat{z}_i = z_i^{-1} (x_i', y_i) \\
\beta_i \hat{z}_i = z_i^{-1} (x_i', y_i)
\end{align*}
\]

\[
\begin{align*}
\beta_i \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( y_i' y_i + \beta' x_i' x_i' \beta - 2y_i' x_i' \beta \right) \right\} \\
= \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( \beta' \beta_i' \right)' \left( \beta_i \beta_i' \right) \right\}
\end{align*}
\]

Then,
\[
\begin{align*}
\left( y_i' - x_i' \beta \right)' \left( y_i' - x_i' \beta \right)^{-tn} \\
= \left[ y_i' y_i + \beta' x_i' x_i' \beta - 2y_i' x_i' \beta \right]^{-tn} \\
= \left[ \left( y_i' - x_i' \beta_i \right)' - (x_i \beta' - x_i \beta_i) \right]^{-tn} \left[ \left( y_i' - x_i' \beta_i \right)' - (x_i \beta' - x_i \beta_i) \right]^{-tn} \\
= \left[ y_i' - x_i' \beta_i \right]' \left( y_i' - x_i' \beta_i \right) + (x_i \beta' - x_i \beta_i)' (x_i \beta' - x_i \beta_i) \right]^{-tn} \\
= \left[ z_i s_i + \left( \beta' \beta_i \right)' \left( x_i' x_i \left( \beta - \beta_i \right) \right) \right]^{-tn} \\
= \left[ z_i s_i + \left( \beta' \beta_i \right)' \left( z_i (\beta - \beta_i) \right) \right]^{-tn} \\
\alpha \left\{ 1 + \frac{(\beta - \beta_i)' z_i (\beta - \beta_i)}{z_i s_i} \right\}^{-tn} \\
\]

But \( n_i = u_i + k \).

\[
\rho(\beta | \sigma_i, y) \propto \exp \left\{ - \frac{1}{2 \sigma_i^2} \left( \beta - \beta_i \right)' z_i' (\beta - \beta_i) \right\}\left\{ 1 + \frac{(\beta - \beta_i)' z_i (\beta - \beta_i)}{z_i s_i} \right\}^{-tn}.
\]
So the posterior density for \( \theta \) is the product of a \( \mathcal{MVN} \) density and a \( \mathcal{MV} \) density. We call the product a "normal - t" density.

Now, expand the second factor (the \( \mathcal{MV} \) factor) of this density asymptotically, and approximate by taking only the first term:

Let \( Q_0 = \frac{1}{2 \pi} (\beta - \hat{\beta}_1)' \Sigma_1 (\beta - \hat{\beta}_1) \)

Then the \( \mathcal{MV} \) term is of the form \( (1 + \frac{Q_0}{\nu_k})^{-\frac{1}{2}} (\nu_k + k)\)

\[ (1 + \frac{Q_0}{\nu_k})^{-\frac{1}{2}} = \exp \left[ \log (1 + \frac{Q_0}{\nu_k}) - \frac{1}{2} (\nu_k + k) \right] \]

\[ = \exp \left\{ -\frac{(\nu_k + k)}{2} \log (1 + \frac{Q_0}{\nu_k}) \right\} \]

Now, \( \log (1 + \frac{Q_0}{\nu_k}) = (\frac{Q_0}{\nu_k}) - \frac{1}{2} (\frac{Q_0}{\nu_k})^2 + \frac{1}{4} (\frac{Q_0}{\nu_k})^3 \)

\[ = (\frac{Q_0}{\nu_k}) + R. \]

\[ (1 + \frac{Q_0}{\nu_k})^{-\frac{1}{2}} (\nu_k + k) = \exp \left\{ -\frac{(\nu_k + k)}{2} \left[ R + \frac{Q_0}{\nu_k} \right] \right\} \]

\[ = \exp \left\{ -\frac{1}{2} \left[ (\nu_k + k) R + k (\frac{Q_0}{\nu_k}) + Q_0 \right] \right\} \]

\[ = \exp \left\{ -\frac{Q_0}{\nu_k} \right\} \exp \left\{ -\frac{1}{2} \left[ k (\frac{Q_0}{\nu_k}) + (\nu_k + k) R \right] \right\} \]

But \( R = (1 + x + 2x^2 + \cdots) \)

So, \( \exp \left\{ -\frac{1}{2} \left[ k (\frac{Q_0}{\nu_k}) + (\nu_k + k) R \right] \right\} \)

\[ = 1 - \frac{1}{2} \left[ k (\frac{Q_0}{\nu_k}) + (\nu_k + k) R \right] + \frac{1}{4} \left[ k (\frac{Q_0}{\nu_k}) + (\nu_k + k) R \right]^2 + \cdots \]

\[ (1 + \frac{Q_0}{\nu_k})^{-\frac{1}{2}} (\nu_k + k) = \exp \left\{ -\frac{Q_0}{\nu_k} \right\} \sum_{i=0}^{\infty} \frac{1}{2^i} \left[ k (\frac{Q_0}{\nu_k}) + (\nu_k + k) R \right]^i \]

\[ = \exp \left\{ -\frac{Q_0}{\nu_k} \right\} \frac{\nu_k}{Q_0} \hat{\beta}_1 \Sigma_1^{-\frac{1}{2}} \]

\[ p(\sigma_1) \propto \exp \left\{ -\frac{1}{2\sigma_1^2} (\beta - \hat{\beta}_1)' \Sigma_1 (\beta - \hat{\beta}_1) \right\} \]

\[ \propto \exp \left\{ -\frac{1}{2\hat{\sigma}_1^2} (\beta - \hat{\beta}_1)' \Sigma_1 (\beta - \hat{\beta}_1) \right\} \]

Now, complete the square on \( \beta \):
\[
-\frac{1}{\sigma^2} \left\{ \beta^2 \hat{z}_1 \beta + \hat{z}_1 \hat{z}_2 \hat{\beta}_1 - 2\beta' \hat{z}_1 \beta \right\} + \frac{1}{\sigma^2} \left[ \beta' \hat{z}_2 \beta + \hat{z}_2 \hat{z}_2 \hat{\beta}_2 - 2\beta' \hat{z}_2 \hat{\beta}_2 \right]
\]

\[
= -\frac{1}{\sigma^2} \left\{ \beta' \left[ \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right] \beta - 2\beta' \left[ \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right] \right\}
+ \left[ \frac{1}{\sigma^2} \hat{z}_1 \hat{z}_2 \hat{\beta}_1 + \frac{1}{\sigma^2} \hat{z}_2 \hat{z}_2 \hat{\beta}_2 \right]
\]

\[
= -\frac{1}{\sigma^2} \left\{ \beta - \left( \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right) \right\} A \left\{ \beta - A^{-1} \left( \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right) \right\}
\]

\[
= -\frac{1}{\sigma^2} \left\{ \beta' A \beta - 2\beta' \left[ \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right] \right\} + \frac{1}{\sigma^2} \hat{z}_1 \hat{z}_2 \hat{\beta}_1 + \frac{1}{\sigma^2} \hat{z}_2 \hat{z}_2 \hat{\beta}_2
\]

Let \( \hat{\beta} = A^{-1} \left( \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right) \)

\[
A \hat{\beta} = \left( \frac{\hat{z}_1}{\hat{\sigma}_1^2} + \frac{\hat{z}_2}{\hat{\sigma}_2^2} \right)
\]

So the exponent becomes:

\[
-\frac{1}{\sigma^2} \left\{ \beta' A \beta - 2\beta' A \hat{\beta} + \sigma^2 \hat{z}_1 \hat{z}_2 \hat{\beta}_1 + \sigma^2 \hat{z}_2 \hat{z}_2 \hat{\beta}_2 \right\}
\]

\[
= -\frac{1}{\sigma^2} \left\{ \beta' A \beta - 2\beta' A \hat{\beta} + \hat{\beta}' A \hat{\beta} \right\}, \text{ say.}
\]

\[
Y = \beta' A \hat{\beta}.
\]

So,

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

\[
\propto \exp \left\{ -\frac{1}{\sigma^2} \left( \beta - \hat{\beta} \right)' A \left( \beta - \hat{\beta} \right) \right\}.
\]
Also, \( \hat{\beta} = \beta^{-1} \left( \frac{d}{d^2} x_2 y_1 + \frac{d}{d^2} x_1 y_2 \right) \)

\( \hat{\beta} = \left( \frac{d}{d^2} x_2 + \frac{d}{d^2} x_1 \right)^{-1} \left( \frac{d}{d^2} x_2 y_1 + \frac{d}{d^2} x_1 y_2 \right) \)

And this happens to coincide with the estimator recommended by the theory. It gives a large-sample justification for this estimator. This is in keeping with the fact that we got our final solution by taking an asymptotic expansion.

(b) **Both \( \sigma_i \) and \( \sigma_i^2 \) Unknown**

Clearly, this is the more realistic case.

Then, \( L(\beta, \sigma_i, \sigma_i^2 | y) \propto \left( \frac{1}{\sigma_i^m \sigma_i^{m-1}} \right) \exp \left\{ -\frac{1}{\sigma_i^2} \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\hat{\beta}} \right) \right\} \)

Let \( p(\beta, \sigma_i, \sigma_i^2) \propto \frac{1}{\sigma_i^m} \).

\( p(\beta, \sigma_i, \sigma_i^2 | y) \propto \left( \frac{1}{\sigma_i^m \sigma_i^{m-1}} \right) \exp \left\{ -\frac{1}{\sigma_i^2} \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\hat{\beta}} \right) \right\} \)

\[ p(\beta, \sigma_i^2 | y) \propto \left( \frac{1}{\sigma_i^{m-1}} \right) \exp \left\{ -\frac{1}{2 \sigma_i^2} \left( y_i - \frac{1}{\hat{\beta}} \right)' \left( y_i - \hat{\beta} \right) \right\} \int_{\frac{1}{\sigma_i^m}} \left( \frac{\sigma_i}{\sigma_i^2} \right)^\frac{m-1}{2} \sigma_i^{m-1} \exp \left\{ -\frac{1}{2 \sigma_i} \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\hat{\beta}} \right) \right\} d\sigma_i \]

\[ \propto \left( \frac{1}{\sigma_i^m} \right) \exp \left\{ -\frac{1}{2 \sigma_i^2} \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\hat{\beta}} \right) \right\} \left[ \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \hat{\beta} \right) \right]^{-\frac{n}{2}} \]

\[ p(\beta | y) \propto \left[ \left( y_i - \hat{\beta} \right)' \left( y_i - \hat{\beta} \right) \right]^{-\frac{n}{2}} \left[ \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\beta} \right) \right]^{-\frac{n}{2}} \]

Consider \( \left( y_i - \hat{\beta} \right)' \left( y_i - \hat{\beta} \right) \):

\[ \left\{ \left( y_i - \hat{\beta} \right)' \left( y_i - \hat{\beta} \right) \right\} = \left\{ \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\beta} \right) \right\} - \frac{2}{\beta} \]

\[ \left\{ \left( y_i - \hat{\beta} \right)' \left( y_i - \hat{\beta} \right) \right\} = \left\{ \left( y_i - \frac{1}{\beta} \right)' \left( y_i - \frac{1}{\beta} \right) \right\} + \left( \frac{1}{\beta} - \hat{\beta} \right)' \left( \frac{1}{\beta} - \hat{\beta} \right) \]

\[ \left\{ \left( y_i - \hat{\beta} \right)' \left( y_i - \hat{\beta} \right) \right\} = \left\{ \left( \beta - \hat{\beta} \right)' \left( \beta - \hat{\beta} \right) \right\} \]

\[ p(\beta | y) \propto \left[ \left( \beta - \hat{\beta} \right)' \left( \beta - \hat{\beta} \right) \right]^{-\frac{n}{2}} \left[ \left( \beta - \frac{1}{\beta} \right)' \left( \beta - \frac{1}{\beta} \right) \right]^{-\frac{n}{2}} \]

\[ p(\beta | y) \propto \left[ \left( \beta - \hat{\beta} \right)' \left( \beta - \hat{\beta} \right) \right]^{-\frac{n}{2}} \left[ \left( \beta - \frac{1}{\beta} \right)' \left( \beta - \frac{1}{\beta} \right) \right]^{-\frac{n}{2}} \]
\[ p(\beta | y) \propto \left\{ 1 + \frac{(z_1 - \beta)^2}{\nu_i z_i} \right\}^{-\frac{(\nu_i + k)}{2}} \times \left\{ 1 + \frac{(z_2 - \beta)^2}{\nu_2 z_2} \right\}^{-\frac{(\nu_2 + k)}{2}} \]

And this pdf is the product of 2 MVT pdfs, so we call it a "multivariate double-t" pdf.

Now, just as we were able to take an asymptotic expansion to get analogy to Blals estimator in the case where one variance was unknown, here with both variances unknown we can expand each 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.

Using the earlier result, we have:

\[
1 + \frac{a_i}{u_i} \exp \left\{ -\frac{a_i}{2} \right\} \frac{\nu_i}{\nu_2} p_i u_i^{-1}
\]

So,

\[
p(\beta | y) \propto \exp \left\{ -\frac{1}{2} (Q_1 + Q_2) \right\} \frac{\nu_1}{\nu_2} q_i p_i u_i^{-1} u_i^{-1}
\]

\[
= \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\hat{s}_1} (z_1 - \hat{\beta}_1) z_1 (z_1 - \hat{\beta}_1) + \frac{1}{\hat{s}_2} (z_2 - \hat{\beta}_2) z_2 (z_2 - \hat{\beta}_2) \right] \right\}
\]

Let

\[
M_1 = \frac{z_1}{\hat{s}_1}
\]

\[
D = \frac{Z_1'Z_1}{\hat{\beta}_1} + \frac{Z_2'Z_2}{\hat{\beta}_2}
\]

\[
\beta = D^{-1} (M_1 \hat{\beta}_1 + M_2 \hat{\beta}_2)
\]

\[
0(\beta | y) \propto \exp \left\{ -\frac{1}{2} \left[ (\beta - \hat{\beta}_1)'M_1 (\beta - \hat{\beta}_1) + (\beta - \hat{\beta}_2)'M_2 (\beta - \hat{\beta}_2) \right] \right\}
\]

\[
\propto \exp \left\{ -\frac{1}{2} (\beta - \hat{\beta})' D (\beta - \hat{\beta}) \right\}.
\]

Which gives the leading normal term

\[
N(\beta - \hat{\beta}, D^{-1} (M_1 \hat{\beta}_1 + M_2 \hat{\beta}_2))
\]

\[
\beta = D^{-1} (M_1 \hat{\beta}_1 + M_2 \hat{\beta}_2) = \left( \frac{1}{\hat{s}_1} z_1' z_1 + \frac{1}{\hat{s}_2} z_2' z_2 \right)^{-1} \left( \frac{1}{\hat{s}_1} z_1' z_1 \right)^{-1} z_1' y_1
\]

\[
+ \frac{1}{\hat{s}_2} z_2' z_2 \left( \frac{1}{\hat{s}_2} z_2' z_2 \right)^{-1} z_2' y_2
\]

\[
= \left( \frac{1}{\hat{s}_1} z_1' z_1 + \frac{1}{\hat{s}_2} z_2' z_2 \right)^{-1} \left( \frac{1}{\hat{s}_1} z_1' y_1 + \frac{1}{\hat{s}_2} z_2' y_2 \right)
\]

And \( \hat{\beta} \) is the mean of the leading normal term in the asymptotic expansion of \( p(\beta | y) \).

And \( \hat{\beta} \) is just the GLS estimator with \( \hat{s}_i^2 \) replacing \( s_i^2 \), \( i = 1, 2 \).
via: the GLS estimator is

\[(X'X)^{-1}(X'y)\]

\[= (\frac{\sigma_1'^2}{\sigma_1'^2}x'_1x'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2x'_2)^{-1}(\frac{\sigma_1'^2}{\sigma_1'^2}x'_1y'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2y'_2)\]

since \(X' = (x'_1' \ x'_2')\); \(y' = (y'_1 \ y'_2)\); \(Z = (\frac{\sigma_1'^2}{\sigma_1'^2} \ 0)\).

If we set \(\sigma_2'^2 = \sigma_2^2\), we get an approx. to the GLS estimator which is generally given asymptotic justification.

Now we can also get close to the GLS estimator from the Bayesian approach by looking at conditional posterior pdf for \(\beta\) -- \(p(\beta | \sigma_1, \sigma_2, y)\)

\[p(\beta | \sigma_1, \sigma_2, y) \propto (\sigma_1, \sigma_2, \sigma, \sigma, \sigma)^{-1} \exp \left\{ -\frac{1}{\sigma_1} (y'_1 - x'_1 \beta)'(y'_1 - x'_1 \beta) \right\}
\]

\[-\frac{1}{\sigma_2} (y'_2 - x'_2 \beta)'(y'_2 - x'_2 \beta) \right]\]
\[\alpha \exp \left\{ -\frac{1}{\sigma_1} (y'_1 - x'_1 \beta)'(y'_1 - x'_1 \beta) - \frac{1}{\sigma_2} (y'_2 - x'_2 \beta)'(y'_2 - x'_2 \beta) \right\}.
\]

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

\[= (\frac{\sigma_1'^2}{\sigma_1'^2}x'_1x'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2x'_2)^{-1}(\frac{\sigma_1'^2}{\sigma_1'^2}x'_1y'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2y'_2)\]

So,

\[(\beta - \beta')(X'X)^{-1}(X'y)\]

\[= (\beta - (X'X)^{-1}(X'y)) (X'X)^{-1}(X'y)\]

\[= \beta' (X'X)^{-1}(X'y)\]

\[= \beta' (\frac{\sigma_1'^2}{\sigma_1'^2}x'_1x'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2x'_2)^{-1}(\frac{\sigma_1'^2}{\sigma_1'^2}x'_1y'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2y'_2)\]

\[= \beta' (\frac{\sigma_1'^2}{\sigma_1'^2}x'_1x'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2x'_2)^{-1}(\frac{\sigma_1'^2}{\sigma_1'^2}x'_1y'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2y'_2)\]

\[\alpha \exp \left\{ -\frac{1}{\sigma_1} (y'_1 - x'_1 \beta)'(y'_1 - x'_1 \beta) - \frac{1}{\sigma_2} (y'_2 - x'_2 \beta)'(y'_2 - x'_2 \beta) \right\}.
\]

\[p(\beta | \sigma_1, \sigma_2, y) \propto \exp \left\{ -\frac{1}{\sigma_1} (X'X)^{-1}(\beta - \beta') \right\}\]

Which is normal with mean \(\beta\) \& c.v. matrix \((X'X)^{-1}\)

\[c.v. = (\frac{\sigma_1'^2}{\sigma_1'^2}x'_1x'_1 + \frac{\sigma_2'^2}{\sigma_2'^2}x'_2x'_2)^{-1} \]
As the conditional posterior mean for $\beta$ is $\hat{\beta}$, and

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

which is just the G.L.S. estimator.

As long as we have a large sample, there will be little loss involved in replacing $\sigma_1^2$ by $\hat{\sigma}_1^2$ in our estimator. So the use of the conditional pdf may be satisfactory enough for large samples. However, in general we do not work with large samples. Then it is best to integrate out $\sigma_1, \sigma_2$ to have a marginal posterior pdf for $\beta$. Use this for making inferences. That is, use the double-$t$ 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 variances, i.e., in $\sigma_1^2$ & $\sigma_2^2$.

Then let $A = \left( \frac{\sigma_1^2}{\sigma_2^2} \right)$

$$\frac{\partial A}{\partial \sigma_2} = -2 \sigma_1^2 \sigma_2^{-3} \times \sigma_1^2 \times \sigma_1 \lambda^{-3/2}$$

$$\therefore J \propto \sigma_1 \lambda^{-3/2}$$

Now, $p(\lambda) d\lambda \propto p(\sigma_2) \sigma_1 \lambda^{-3/2} d\sigma_2$

$$p(\lambda, \sigma_1, \beta | y) \propto p(\sigma_1, \sigma_2 | \beta | y) \cdot \sigma_1 \lambda^{-3/2}$$

And so,

$$p(\lambda, \sigma_1, \beta | y) \propto \left( \frac{\lambda^{(n-2)/2}}{\sigma_1^{n+1}} \right) \exp \left\{ -\frac{1}{\sigma_1^2} \left[ (y_1 - X_1 \beta)^T (y_1 - X_1 \beta) + \lambda (y_2 - X_2 \beta)^T (y_2 - X_2 \beta) \right] \right\}.$$ 

Now, before integrating this pdf to get $p(\sigma_1 | A | y)$, complete the square on the exponent:

$$(y_1 - X_1 \beta)^T (y_1 - X_1 \beta) + \lambda (y_2 - X_2 \beta)^T (y_2 - X_2 \beta)$$

$$= y_1 y_1^T + \beta' X_1 X_1 \beta - 2 \beta' X_1 y_1 + \lambda y_2 y_2 + \lambda \beta' X_2 X_2 \beta - 2 \lambda \beta' X_2 y_2$$

$$= \beta' (X_1' X_1 + \lambda X_2' X_2) \beta - 2 \beta' (X_1' y_1 + \lambda X_2' y_2) + (y_1 y_1^T + \lambda y_2 y_2)$$

Let: $C_1 = (X_1' X_1 + \lambda X_2' X_2)$; $C_2 = (X_1' y_1 + \lambda X_2' y_2)$. 

$$\therefore$$

$$p(\lambda, \sigma_1, \beta | y) \propto \left( \frac{\lambda^{(n-2)/2}}{\sigma_1^{n+1}} \right) \exp \left\{ -\frac{1}{2 \sigma_1^2} C_1 \beta^T C_1^{-1} \beta - \frac{C_2}{\sigma_1^2} \right\}.$$
Then, the exponent becomes:
\[
\beta' c_1 \beta - 2 \beta' c_2 + y_1 y_1 + \lambda y_2 y_2
\]
\[
= (\beta - c_1^{-1} c_2)' c_1 (\beta - c_1^{-1} c_2) + y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2
\]
\[
p(\beta, \sigma, \lambda | y) \propto \left( \frac{(n+2)\lambda}{\sigma_1 \sigma_2} \right) \exp \left\{ -\frac{1}{\sigma_1^2} \left[ (\beta - c_1^{-1} c_2)' c_1 (\beta - c_1^{-1} c_2) + y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2 \right] \right\}
\]
\[
\text{Now the exponent is:}
\exp \left\{ -\frac{1}{\sigma_1^2} (y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2) \right\} \exp \left\{ -\frac{1}{\sigma_2^2} (\beta - c_1^{-1} c_2)' c_1 (\beta - c_1^{-1} c_2) \right\}
\]
\[
p(\sigma, \lambda | y) \propto \left( \frac{(n+2)\lambda}{\sigma_1 \sigma_2} \right) \exp \left\{ -\frac{1}{\sigma_1^2} (y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2) \right\} \int \exp \left\{ -\frac{1}{\sigma_2^2} (\beta - c_1^{-1} c_2)' c_1 (\beta - c_1^{-1} c_2) \right\} d\beta
\]

And the integral is
\[
\int \exp \left\{ -\frac{1}{\sigma_1^2} (\beta - c_1^{-1} c_2)' c_1 (\beta - c_1^{-1} c_2) \right\}
\]

where \( \Sigma^{-1} = \frac{c_1 c_2}{\sigma_1} \) \( \therefore \Sigma = \frac{c_1 c_2}{\sigma_1} \)

\[
\therefore I \propto \frac{1 c_{1,1-t}}{(\sigma_1 - k)} \int \text{MVN}
\]

\[
\propto \frac{1 c_{1,1-t}}{(\sigma_1 - k)}
\]

\[
p(\sigma_1, \lambda | y) \propto 1 c_{1,1-t} \left( \frac{1}{\sigma_1 \sigma_2} \right)^{\frac{(n+2)\lambda}{2}} \exp \left\{ -\frac{1}{\sigma_1^2} (y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2) \right\}
\]

And, \( p(\lambda | y) = \int p(\sigma, \lambda | y) d\sigma \)

\[
\propto 1 c_{1,1-t} \left( \frac{1}{\sigma_1 \sigma_2} \right)^{\frac{(n+2)\lambda}{2}} \int \exp \left\{ -\frac{1}{\sigma_1^2} (y_1 y_1 + \lambda y_2 y_2 - c_1' c_1^{-1} c_2) \right\}
\]

\[
\propto 1 c_{1,1-t} \left( \frac{1}{\sigma_1 \sigma_2} \right)^{\frac{(n+2)\lambda}{2}} \left\{ \frac{\Sigma_{1+n_2-k}}{(\sigma_1 \sigma_2)^{\frac{1}{2}(n+2)\lambda/2}} \right\}
\]

\[
= \left( \frac{1}{\sigma_1 \sigma_2} \right)^{\frac{(n+2)\lambda}{2}} \left\{ \frac{\Sigma_{1+n_2-k}}{(\sigma_1 \sigma_2)^{\frac{1}{2}(n+2)\lambda/2}} \right\}
\]

\[
\propto \left( \frac{1}{\sigma_1 \sigma_2} \right)^{\frac{(n+2)\lambda}{2}} \left\{ \frac{\Sigma_{1+n_2-k}}{(\sigma_1 \sigma_2)^{\frac{1}{2}(n+2)\lambda/2}} \right\}
\]
And this joint pdf on \( \lambda \) (i.e. on \( \lambda^2 \)) is a little different from an \( F \)-distrib. The difference arising because we assumed in our original model that the same \( \beta \) appears in each model, i.e. we have imposed a constraint to get away from the general situation where a different \( \beta \) appears in each model. Then an \( F \)-distrib 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 subs vector is common to each.

\[
\begin{align*}
\eta_1 &= W_1 \beta_1 + W_2 \beta_2 + u_1 \\
\eta_2 &= Z_1 \beta_1 + Z_2 \beta_2 + u_2
\end{align*}
\]

\( \text{NB} \) \( u_1, u_2 \sim \text{IN}(0, \Omega) \) : i.e. same variance.

\[
\begin{align*}
\eta_1 &\sim (n_1 \times 1) \\
\eta_2 &\sim (n_2 \times 1) \\
(W_1 : W_2) &\sim (n_1 + n_2 \times k_1) \\
(Z_1 : Z_2) &\sim (n_2 + n_2 \times k_2) \\
\beta_1 &\sim (m_1 \times 1) \\
\beta_2 &\sim (m_2 \times 1) \\
\beta &\sim (m_1 + m_2 \times 1)
\end{align*}
\]

\[
\begin{align*}
W_1 &\sim (n_1 \times m_1) \text{, so } W_2 \sim (n_1 \times m_2) \\
Z_1 &\sim (n_2 \times m_2) \text{, so } Z_2 \sim (n_2 \times m_2)
\end{align*}
\]

\[
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix}
= 
\begin{bmatrix}
W_1 & W_2 \\
Z_1 & Z_2
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
+ 
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix}
\]

\[
\Rightarrow \quad y = X\beta + u \sim (n_1 + m_2) \text{ obs. in all.}
\]

Assume a diffuse prior, then the basic Bayesian result is that

\[
\begin{align*}
\rho(\beta | y) &\propto \left\{ \nu s^2 + (\beta - \bar{\beta})' X' X (\beta - \bar{\beta}) \right\}^{-\frac{(n_1 + m_2)\nu}{2}} \\
\text{where } &\begin{align*}
\nu &= n_1 + n_2 - m_1 - m_2 \\
\bar{\beta} &= (X' X)^{-1} X' y \\
\nu s^2 &= (y - \bar{\beta})' (y - \bar{\beta}).
\end{align*}
\end{align*}
\]
What is the marginal posterior for $\beta_1$?

Partition: $\beta' = (\beta_1', \beta_2')$

where $\gamma' = (\beta_2', \gamma_2')$

Let $(X'X) = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$

Now: $(X'X) = \begin{pmatrix} \omega_1 & \omega_2 & 0 \\ \omega_1 & \omega_2 & 0 \\ \omega_1 & \omega_2 & 0 \\ \omega_1' & \omega_2' & 0 \\ 0 & \omega_2' & 0 \end{pmatrix}$

$= \begin{pmatrix} \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \end{pmatrix}$

$\begin{pmatrix} \omega_2' \omega_2 \\ \omega_2' \omega_2 \\ \omega_2' \omega_2 \end{pmatrix}$

$= \begin{pmatrix} \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \end{pmatrix}$

$= \begin{pmatrix} \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \end{pmatrix}$

$= \begin{pmatrix} \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \end{pmatrix}$

$= \begin{pmatrix} \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \\ \omega_1 \omega_1' + \omega_2' \omega_2 \end{pmatrix}$

So: $M_{11} = (\omega_1 \omega_1' + \omega_2' \omega_2)$

$M_{12} = (\omega_1 \omega_2', \omega_2' \omega_2)$

$M_{21} = (\omega_2' \omega_1)$

$M_{22} = (\omega_2' \omega_2)$

Then, making use of the usual results derived for $\mathbf{Ch} - p(\beta_1 | y) \propto \left\{ \psi^{\lambda} + (\beta_1 - \bar{\beta}_1)' \mathbf{H} (\beta_1 - \bar{\beta}_1) \right\}^{-\lambda} (n_1 + m_2 - m_1 - m_2)/2$

where $\mathbf{H} = M_{11} - M_{12} M_{22}^{-1} M_{21}$

$\mathbf{H} = \omega_1 \omega_1 + \omega_2' \omega_2 - (\omega_1 \omega_2', \omega_2' \omega_2) \begin{pmatrix} (\omega_2' \omega_2)'^{-1} & 0 \\ 0 & (\omega_2' \omega_2)'^{-1} \end{pmatrix} \begin{pmatrix} \omega_1 \omega_1' \\ \omega_2' \omega_2 \end{pmatrix}$

$= \omega_1 \omega_1 + \omega_2' \omega_2 - (\omega_1 \omega_2' (\omega_2' \omega_2)'^{-1} \omega_2' \omega_2 (\omega_2' \omega_2)'^{-1} \omega_1 \omega_1' + \omega_2' \omega_2 (\omega_2' \omega_2)'^{-1} \omega_2' \omega_2)$

$= \omega_1 \omega_1 + \omega_2' \omega_2 - \omega_1 \omega_2' (\omega_2' \omega_2)'^{-1} \omega_2' \omega_2 (\omega_2' \omega_2)'^{-1} \omega_1 \omega_1'$

$\mathbf{H} = n_1 + m_2 - m_1 - m_2$

So, $(\mathbf{H} + m) = (n_1 + m_2 - m_1 - m_2)$. 
\[ p(\beta, y) \propto \left[ w^2 + (\beta - \beta_1)' H (\beta - \beta_1) \right]^{-\frac{p+n}{2}} \]

Note that \( \beta_i \) is a sub-vector of \( \beta \), where

\[ \beta = (X'X)^{-1} X'y. \]

\[ \beta_i = H^{-1} (V_1 \beta_1 + V_2 \beta_2) \]

where \[ V_1 = w_1 w_1 - w_2 w_2 (w_2' w_2)^{-1} w_2' w_1 \]

\[ V_2 = z_1' z_1 - z_2' z_2 (z_2' z_2)^{-1} z_2' z_1 \]

\[ H = V_1 + V_2. \]

So \( \beta \) is O.L.S. estimate from lot again —

So \( \hat{\beta}_1 = \frac{1}{2} \left[ (w_1 : w_2)' (w_1 : w_2) \right]^{-1} \left[ (w_1 : w_2)' y_1 \right] \] (full data)

So \( \hat{\beta}_2 = \frac{1}{2} \left[ (z_1 : z_2)' (z_1 : z_2) \right]^{-1} \left[ (z_1 : z_2)' y_2 \right] \] (full data)

[So \( \hat{\beta}_1 \) is part of the full OLS est. Same for \( \hat{\beta}_2 \)].

So \( \beta \) is a weighted average of \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) that would be obtained if the 2 equations were handled separately.
Zellner: Chapter 6.
"Non-Linear Models."

(A) The Box-Cox Analysis:

Let $y_{x} > 0$; $x = 1, 2, \ldots, n$.

Then consider the following transformation of the dependent variable:

$$y_{x} \rightarrow \left(\frac{y_{x}^{a} - 1}{a}\right) ; \quad a > 1$$

$$\rightarrow \log y_{x} ; \quad a = 0.$$

Then the regression is:

$$\left(\frac{y_{x}^{a} - 1}{a}\right) = \beta_{1} + \beta_{2} x_{2x} + \beta_{3} x_{3x} + \ldots + \beta_{k} x_{kx} + u_{x}.$$

Then, assume that for some unknown $\lambda$, the transformed dependent variables satisfy the conditions:

1. Normally distributed.
2. Same variance, say $\sigma_{x}^{2}$.

So the 3 basic properties of the transformed dependent variable that interest us are:

1. Normality.
2. Stable covariance.
3. Simplicity of structure — in the sense that
   $$E \left[\frac{y_{x}^{a} - 1}{a}\right] = \beta_{1} + \beta_{2} x_{2x} + \cdots + \beta_{k} x_{kx}$$
   is a simple function of the $\beta$'s and $x$'s.

Now, $\lambda$ is unknown and so it will have to be estimated together with $\beta$'s and $\sigma$. 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: $$\left(\frac{y_{x}^{a} - 1}{a}\right) = \beta_{1} + \beta_{2} x_{2x} + \cdots + \beta_{k} x_{kx} + u_{x}$$

as: $y^{(a)} = X \beta + u$. 
Now, let's consider the distribution $N(0, \sigma^2)$. But now the dependent variable $(y^{(u)})$ has a different distribution from that of the errors $w$:

\[ p(y) = \mathcal{N}(0, \sigma^2) \]

\[ \sigma = \prod_{x=1}^{n} \left| \frac{1}{3} y_{x} \right| = \prod_{x=1}^{n} y_{x}^{\lambda - 1} = (\prod_{x=1}^{n} y_{x})^{\lambda - 1} = g^{(\lambda - 1)} \]

where $g = \text{geometric mean} = (\prod_{x=1}^{n} y_{x})^{\frac{1}{n}}$.

So, $p(y|\lambda, \beta, \sigma^2) \propto \frac{1}{\sigma} \exp \left\{ \frac{1}{\sigma^2} (y^{(u)} - \lambda \beta)^\prime (y^{(u)} - \lambda \beta) \right\}$

and $l(\lambda, \beta, \sigma^2|y) \propto \frac{1}{\sigma^2} \exp \left\{ -\frac{1}{2} (y^{(u)} - \lambda \beta)^\prime (y^{(u)} - \lambda \beta) \right\}$

As

\[ L = \log l = \text{const.} + \log \sigma - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y^{(u)} - \lambda \beta)^\prime (y^{(u)} - \lambda \beta) \]

\[ \frac{dL}{d\beta} = -\frac{1}{2\sigma^2} \left[ (y^{(u)} - \lambda \beta)^\prime - 2\lambda \beta^\prime x^{(u)} \right] \]

\[ = -\frac{1}{2\sigma^2} \left[ -2x^\prime y^{(u)} + 2x^\prime \lambda \beta \right] \]

\[ = 0 \quad \text{for MLE} \]

\[ \hat{\beta} = (x^\prime x)^{-1} x^\prime y^{(u)} \]

And, \( \frac{d^2 L}{d\beta^2} = \frac{1}{2\sigma^4} \left[ (y^{(u)} - \lambda \beta)^\prime (y^{(u)} - \lambda \beta) \right] - \frac{n}{2\sigma^2} \]

\[ = 0 \quad \text{for MLE} \]

\[ \frac{d^2 L}{d\sigma^2} = \frac{1}{\sigma^2} (y^{(u)} - \lambda \hat{\beta})^\prime (y^{(u)} - \lambda \hat{\beta}) \]

Now, given $\lambda$, we can compute $\hat{\beta}$ and $\hat{\sigma^2}$, but at present $\lambda$ is unknown, so we will have to estimate its value.

\[ L = \text{const.} + \log \sigma - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y^{(u)} - \lambda \beta)^\prime (y^{(u)} - \lambda \beta) \]

\[ L_{\max} = \text{const.} + \log \sigma - \frac{n}{2} \log (\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \left[ (y^{(u)} - \lambda \hat{\beta}(u))^\prime (y^{(u)} - \lambda \hat{\beta}(u)) \right] \]

\[ = \text{const.} + \log \sigma - \frac{n}{2} \log (\hat{\sigma}^2) - \frac{n}{2} \lambda^2 \]

\[ = \text{const.} + \log \sigma - \frac{n}{2} \log (\hat{\sigma}^2) \]

\[ = \text{const.} + (\lambda - 1) \sum_{x=1}^{n} \log y_{x} - \frac{n}{2} \log (\hat{\sigma}^2) \].
\[ z \]

As, plot \( L(\beta) \) for different values of \( \lambda \), and pick \( \lambda = \hat{\lambda} \) to maximize this function. Then evaluate \( \hat{\sigma}^2(\lambda) \) and \( \hat{\beta}(\lambda) \).

Zarembska looked at a model where the regressors are also subjected to the basic Box-Cox transform.

\[
\left( \frac{\lambda X_i - 1}{\lambda} \right) = \beta_1 + \beta_2 \left( \frac{Y_i - 1}{\lambda} \right) + \beta_3 \left( \frac{X_i^{\lambda} - 1}{\lambda} \right) + u_i
\]

So, if \( \lambda = 1 \), then we have a linear relationship.

\[
\text{if } \lambda = 0, \text{ then we have a log-log relationship.}
\]

Let:

\[
\begin{align*}
(m_i - 1)/\lambda &= y_{i1}^{(u)} \\
(y_i - 1)/\lambda &= x_{i1}^{(u)} \\
(x_i^{\lambda} - 1)/\lambda &= x_{i2}^{(u)}
\end{align*}
\]

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

where \( \beta' = (\beta_1, \beta_2, \beta_3) \) and \( X^{(u)} = (1, x_1^{(u)}, x_2^{(u)}) \).

\[
\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
\]

\[
L = \log L = -n \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} (y^{(u)} - X^{(u)} \beta)'(y^{(u)} - X^{(u)} \beta)
\]

\[
\frac{\partial L}{\partial \beta} = -\frac{2}{\sigma^2} [2 X^{(u)'}(y^{(u)} - X^{(u)} \beta)]
\]

\[
= 0 \text{ for MLE.}
\]

So, \( X^{(u)'y^{(u)} = X^{(u)'x^{(u)} \beta} \)

\[
\hat{\beta}^{(u)} = (X^{(u)'X^{(u)}})^{-1} X^{(u)'y^{(u)}}
\]

\[
\frac{\partial L}{\partial \sigma^2} = -n \frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (y^{(u)} - X^{(u)} \beta)'(y^{(u)} - X^{(u)} \beta)
\]

\[
= 0 \text{ for MLE.}
\]

\[
\hat{\sigma}^2 = \frac{1}{n} (y^{(u)} - X^{(u)} \hat{\beta}^{(u)})'(y^{(u)} - X^{(u)} \hat{\beta}^{(u)})
\]

\[
\log L_{null}(u) = -n \log (\hat{\sigma}^2) + (\lambda - 1) \frac{1}{2} \sum_{i=1}^{n} \log y_i
\]

Plot \( L_{null}(u) \) for values \( \lambda \). And Zarembska 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 \( \lambda \) because
\[
2 \left[ \ln \hat{\lambda} - \ln \hat{\lambda} \right] \sim \chi^2(0)
\]

Now, for Zarzadeh's model,

the approximate C.I. around \( \hat{\lambda} = 0.19 \) is given by
\[
\ln \hat{\lambda} \pm \frac{1}{2} \chi^2(0.05) = 1.92
\]

So, take \( \hat{\lambda} = 0.19 \), evaluate \( \ln \hat{\lambda} \). Take
\[
\ln \hat{\lambda} \pm \frac{1}{2} \chi^2(0.05) = 1.92
\]

Use the values of \( \ln \hat{\lambda} \) to find
\[
\left( \hat{\lambda}^2, \hat{\lambda}^2 \right)
\]

as the C.I.

Now introduce the possibility of autocorrelated disturbances.

Let \( y_x = \beta_1 x_{x-1} + e_x ; x = 2, 3, \ldots, n. \)

So,
\[
y_x = \beta_1 x_{x-1} + x_3 x_{x-1} + \beta_2 y_{x-1} - \beta_1 \rho \beta_2 x_{x-1} + e_x.
\]

Then, if we are given \( \hat{\lambda} \), we can evaluate conditional MLE for \( \beta \) and for \( \sigma^2 \).

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 \( p(\beta, \sigma^2; \lambda) = p_1(\beta, \sigma^2; \lambda) p_2(\lambda) \)

Take \( p_2(\lambda) \) a const.

We have to take a conditional prior for \( \beta \) and \( \sigma^2 \); take \( p_1(\beta, \sigma^2; \lambda) \), since the general size is range of the transformed observations, \( y_x \), may depend strongly on \( \lambda \).
So, put
\[ p(\beta, \log \sigma | \lambda) \propto \sigma^{\beta} \exp(\log \sigma) \propto \sigma(\beta) \ pro d(\log \sigma) \ \propto \ g(\lambda) \ d\beta, \ d(\log \sigma) \]

i.e. \( \beta \) or \( \log \sigma \) are uniform, but depend on a function of \( \lambda \).
\[ p(\beta, \sigma | \lambda) \propto \frac{g(\lambda)}{\sigma} \]

Now, Box & Cox take \( g(\alpha) = \frac{1}{\alpha} \), whence
\[ f = \frac{1}{(x_1^\alpha, x_2^\alpha)^{1-n}}. \]

So,
\[ p(\beta, \sigma | \lambda) \propto \left( \frac{1}{\sigma} \right)^{1/n} \]
\[ \propto p(\beta, \sigma, \lambda) \propto \left( \frac{1}{\sigma} \right)^{1/n} \]

Now, \( p(y | \lambda, \sigma^2, \beta) = \sigma^2 \exp \left( -\frac{1}{\sigma^2} (y - X\beta)'(y - X\beta) \right) \)
\[ p(\lambda, \sigma, \beta | y) \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i' \beta)}{\sigma^m} \right) \exp \left\{ -\frac{1}{\sigma^2} (y - X\beta)'(y - X\beta) \right\} \]
And,
\[ (y - x\beta)'(y - x\beta) = \nu \sigma^2(\lambda) + (\beta - \beta(\lambda))'X'X(\beta - \beta(\lambda)) \]
where
\[ \frac{\beta}{\sigma}(\lambda) = (x'x)^{-1} x'y. \]
\[ \nu = (n-1) \]
\[ \nu \sigma^2(\lambda) = (y - x\beta(\lambda))'(y - x\beta(\lambda)). \]

So,
\[ p(\lambda, \beta | y) \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \exp \left\{ -\frac{1}{\sigma^2} \left[ \nu \sigma^2(\lambda) + (\beta - \beta(\lambda))'X'X(\beta - \beta(\lambda)) \right] \right\}^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left[ \nu \sigma^2(\lambda) + (\beta - \beta(\lambda))'X'X(\beta - \beta(\lambda)) \right]^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto p(\lambda | y), \ (\sigma^2(\lambda))^{-\frac{n}{2}} | \frac{X'X}{\sigma^2(\lambda)} |^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]
\[ \propto \left( \frac{\prod_{i=1}^{n} (y_i - x_i\beta)}{\sigma^m} \right) \left( \frac{X'X}{\sigma^2(\lambda)} \right)^{-\frac{n}{2}} \]

This pdf can be analyzed numerically to permit inference about \( \lambda \). Also, the marginal value of the posterior pdf coincides with the M.L.E.

Note that the conditional pdf \( p(y | \beta, \lambda) \) given \( \beta \) is an M.V.T. distribution, so we may use \( \beta \) to test the sensitivity of inference about \( \lambda \) to the assumptions regarding the value of \( \beta \).
The C.E.S. Production Function:

Let the elasticity of substitution be \( \sigma \in [0, \infty) \).

Let the function be of the general form

\[
y(x) = \delta \left[ \delta k - \gamma (1-\delta) L - \beta \right]^{-\gamma / \delta} e^{\alpha x} \quad ; \quad x \geq 1, 2, \ldots, \infty.
\]

For the 2-input case, and assume that \( u(x) \sim \text{IID}(0, \sigma^2) \).

\[
\log y(x) = \log \mu + v \log \left\{ \left[ \delta k^{-\gamma} + (1-\delta) L^{-\gamma} \right]^{-\gamma / \delta} \right\} + u(x).
\]

Then,

\[
(\sigma^2, \rho, v, \sigma | y, k, L) \propto \left( \delta \right)^{v} \exp \left\{ -\frac{v}{\sigma^2} \left[ \delta k^{-\gamma} + (1-\delta) L^{-\gamma} \right]^{-\gamma / \delta} \right\}
\]

where:

\[
\Theta = \begin{pmatrix} \log \mu \\ v \end{pmatrix} \quad ; \quad \theta = (x^1 x^{-1})^t \quad y \quad \gamma = \begin{pmatrix} \log y^1 \\ \log y^2 \end{pmatrix} \\
\quad \quad \gamma = \begin{pmatrix} \delta k^{-\gamma} + (1-\delta) L^{-\gamma} \end{pmatrix}^{-\gamma / \delta}
\]

\[
\hat{\Theta} = (\hat{y}^{-1} \hat{\gamma}) (\hat{y} - \hat{\gamma}).
\]

\[
\log l = L = \text{const.} - \frac{v}{2} \log \sigma^2 - \frac{v}{\sigma^2} \left[ \delta k^{-\gamma} + (1-\delta) L^{-\gamma} \right]^{-\gamma / \delta}
\]

So,

\[
\hat{\sigma}^2 = \frac{L}{V} \hat{\Theta} \hat{\gamma} \quad ; \quad \hat{\sigma} = \hat{\Theta} \quad \text{where a best denotes a MLE, for given } \mu \text{ and } \delta.
\]

\[
L_{\text{max}} = \text{const} - \frac{v}{2} \log (\hat{\Theta})
\]

So, \( L_{\text{max}} \) is a function of \( \rho \) and \( \delta \), since \( \hat{\Theta} = \Theta (\rho, \delta) \).

So search over the \((\rho, \delta)\) grid to find \((\hat{\rho}, \hat{\delta})\) such that \( L_{\text{max}} \) is maximal. Then compute \( \hat{\sigma}^2 (\hat{\rho}, \hat{\delta}) \) and \( \hat{\rho}(\hat{\rho}, \hat{\delta}) \) for MLE.

Now, \( 0 < \delta < 1 \), \( -1 < \rho < \infty \), since \( \rho = -1 + \frac{1}{\delta} \).

So to ease the grid search, we might transform \( \rho \).

Let \( \lambda = 1 / (1+\rho) \)

Then since \( \rho = 1 + 1/\lambda \),

\[
\Rightarrow \quad \lambda = 1 / (1 + \rho) = (1+\rho) / (2+\rho).
\]

So when \( \rho = -1 \), \( \lambda = 0 \).

But also \( 2+\rho \lambda = 1+\rho \),

\[
\Rightarrow \rho = (2 \lambda - 1) / (1-\lambda)
\]
\[ \text{So, if } \lambda = 1, \rho = \infty. \]

So \( 0 < \lambda < 1 \) and we just search for \((\hat{\lambda}, \hat{\rho})\) in the unit square.

Having obtained \( \hat{\lambda} \), we get \( \hat{\rho} \) hence \( \hat{\theta} \).

Now the mean and variance of \( \hat{\theta} \) do not exist in finite sample size; but they do exist for the asymptotic normal distribution of \( \hat{\theta} \).

Thorntor investigates by means of a Monte Carlo study, the risk functions associated with alternative estimations for \( \theta \).

The loss function that he employs is

\[ L(\theta, \hat{\theta}) = \frac{(\theta - \hat{\theta})^2}{(1 + \theta)^2 (1 + \hat{\theta})^2}. \]

Now, this loss function implies relatively greater loss for an overestimate than for an underestimate.

To show this:

Consider an underestimate \( \Delta \sim \mathcal{N}(\theta - \Delta) = \Delta \)

Then

\[ L(\theta, \hat{\theta}) = \frac{\Delta^2}{(1 + \theta + \Delta)^2 (1 + \hat{\theta})^2}. \]

Consider an overestimate \( \Delta \sim \mathcal{N}(\theta - \Delta) = \Delta \)

\[ L(\theta, \hat{\theta}) = \frac{\Delta^2}{(1 + \theta - \Delta)^2 (1 + \hat{\theta})^2}. \]

\[ \frac{L_1(\theta, \hat{\theta})}{L_2(\theta, \hat{\theta})} = \frac{(1 + \theta - \Delta)^2}{(1 + \theta + \Delta)^2} < 1. \]

So \( L_1 < L_2 \), or \( L_2 > L_1 \).

Now, consider the prior pdf's used by Thorntor:

\[ p_1(\gamma, \nu, \sigma, \delta, \theta) \propto \frac{(1 + \theta)^2 e^{-\theta}}{\gamma \sigma}; \theta > 0. \]

\[ p_2(\gamma, \nu, \sigma, \delta, \theta) \propto \frac{\theta (1 + \theta)^2 e^{-\theta}}{\gamma \sigma}. \]

That is, the priors for \( \nu, \delta, \log \gamma \) and \( \log \sigma \) are taken as uniform, and

\[ \begin{cases} p_1(\theta) \propto (1 + \theta)^2 e^{-\theta} \\
 p_2(\theta) \propto \theta (1 + \theta)^2 e^{-\theta}. \end{cases} \]
Consider the normalizing factors for $p_1$ & $p_2$. the marginal prior pdf's for $\xi$.

(a) $p_1(\xi) \propto (1+\xi)^2 e^{-\xi}$.

So, the normalizing constant is the inverse of

$I_1 = \int_0^\infty (1+\xi)^2 e^{-\xi} d\xi$.

Let $u = (1+\xi)^2$; $du = 2(1+\xi) d\xi$; $v = -e^{-\xi}$

$\therefore I_1 = uv - \int_0^\infty u du$

$= \left[ (1+\xi)^2 e^{-\xi} \right]_0^\infty + 2 \int_0^\infty e^{-\xi} (1+\xi) d\xi$

$= 1 + 2 \int_0^\infty e^{-\xi} d\xi + 2 \int_0^\infty e^{-\xi} \xi d\xi$

$= 1 + 2 \left[ -e^{-\xi} \right]_0^\infty + 2 \int_0^\infty e^{-\xi} \xi d\xi$

$= 3 + 2 \int_0^\infty e^{-\xi} \xi d\xi$

Let $I_2 = \int_0^\infty e^{-\xi} \xi d\xi$

$\therefore u = \xi$; $du = e^{-\xi} d\xi$;

$\therefore I_2 = uv - \int_0^\infty u du$

$= \left[ -\xi e^{-\xi} \right]_0^\infty + \int_0^\infty \xi e^{-\xi} d\xi$

$= 0 - \left[ e^{-\xi} \right]_0^\infty$

$= 1$

$\therefore 2I_2 = 2$.

$I = 3 + 2 = 5$

So, $p_1(\xi) = \frac{1}{5} (1+\xi)^2 e^{-\xi}$.
\( p_2(\varepsilon) \propto \varepsilon (1+\varepsilon)^2 \varepsilon^{-\varepsilon} \).

So, the normalizing constant \( \phi \) is the inverse of:

\[
\int_0^\infty \varepsilon (1+\varepsilon)^2 \varepsilon^{-\varepsilon} d\varepsilon
\]

\[
= \int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon + 2 \int_0^\infty \varepsilon^2 e^{-\varepsilon} d\varepsilon + \int_0^\infty \varepsilon^3 e^{-\varepsilon} d\varepsilon
\]

\[
= 1 + 2 I_2 + I_3
\]

\( I_2 = \int_0^\infty \varepsilon^2 e^{-\varepsilon} d\varepsilon \).

Let \( u = \varepsilon^2 \); \( dv = e^{-\varepsilon} d\varepsilon \)\n
\[ du = 2\varepsilon d\varepsilon \quad v = -e^{-\varepsilon} \]

\[
I_2 = uv - \int_0^\infty vdu
\]

\[
= \left[ -\varepsilon^2 e^{-\varepsilon} \right]_0^\infty + 2 \int_0^\infty \varepsilon e^{-\varepsilon} d\varepsilon
\]

\[
= 0 + 2
\]

\[
= 2
\]

\( 2 I_2 = 4 \)

\( I_3 = \int_0^\infty \varepsilon^3 e^{-\varepsilon} d\varepsilon \).

Let \( u = \varepsilon^3 \); \( dv = e^{-\varepsilon} d\varepsilon \)\n
\[ du = 3\varepsilon^2 d\varepsilon \quad v = -e^{-\varepsilon} \]

\[
I_3 = uv - \int_0^\infty vdu
\]

\[
= \left[ -\varepsilon^3 e^{-\varepsilon} \right]_0^\infty + 3 \int_0^\infty \varepsilon^2 e^{-\varepsilon} d\varepsilon
\]

\[
= 0 + 3 I_2
\]

\[
= 6
\]

\( I_1 = 1 + 4 + 6 \)

\[
= 11
\]

\( \phi_0, p_2(\varepsilon) = \frac{1}{11} \varepsilon (1+\varepsilon)^2 \varepsilon^{-\varepsilon} \).
Now, consider an alternative approach to the analysis of the C.E.S. production function. Consider the deterministic form of the function, with inputs \( x \) and \( \epsilon \) CRPS:

\[
V_{\alpha} = Y \left[ \delta_1 x_{1w}^\theta + (1 - \delta) x_{2w}^\theta \right]^{1/\theta}
\]

and \( g = -\rho = (\epsilon - 1) / \epsilon \).

Then,

\[
V_{\alpha} = \gamma \left[ \delta_1 x_{1w}^\theta + (1 - \delta) x_{2w}^\theta \right]
\]

and

\[
\left( \frac{V_{\alpha}^{\theta - 1}}{\gamma} \right) = \left[ \delta_1 x_{1w}^\theta + \gamma (1 - \delta) x_{2w}^\theta \right] - 1
\]

\[
V_{\alpha}^{(3)} = \left[ \gamma \delta_1 x_{1w}^\theta - \gamma \delta_1 x_{2w}^\theta + \gamma x_{2w}^\theta - 1 \right] / \gamma
\]

\[
= \left[ \gamma \delta_1 (x_{1w}^\theta - 1) + \gamma \delta_1 - \gamma \delta_1 (x_{2w}^\theta - 1) - \gamma \delta_1
\right.
\]
\[
+ \gamma (x_{2w}^\theta - 1) + \gamma - 1 \right] / \gamma
\]

\[
= \gamma \delta_1 x_{1w}^{(3)} - \gamma \delta_1 x_{2w}^{(3)} + \gamma^{(3)} + \gamma (x_{2w}^{(3)})
\]

\[
V_{\alpha}^{(3)} = \gamma x_{2w}^{(3)} + \gamma \delta_1 \left[ x_{1w}^{(3)} - x_{2w}^{(3)} \right] + \gamma^{(3)} \left[ 1 + \gamma x_{2w}^{(3)} \right]
\]

And assume that the observed output, \( y_x \), satisfies

\[
y_x = \gamma x_{2w}^{(3)} + \gamma \delta_1 \left[ x_{1w}^{(3)} - x_{2w}^{(3)} \right] + \gamma^{(3)} \left[ 1 + \gamma x_{2w}^{(3)} \right]
\]

\[
y_x = V_{\alpha}^{(3)} + \gamma x_{2w}^{(3)}
\]

where \( y_x = (y_x - 1) / \lambda \) and \( \lambda = \phi_3 \), for \( \phi_3 \).

Note that this is more general than if we just took \( y_x^{(3)} \) instead of \( y_x \).

\[
y_x^{(3)} = \gamma x_{2w}^{(3)} + \beta_1 (x_{1w}^{(3)} - x_{2w}^{(3)}) + \beta_2 (1 + \gamma x_{2w}^{(3)}) + u_x
\]

where \( \beta_1 = \gamma \delta_1 \) and \( \beta_2 = \gamma^{(3)} \).

Then, if \( \gamma = 0 \), the form becomes C-D.

If \( \gamma = 1 \), it is linear in the variables.

And, by using the more general parameter, \( \lambda \), instead of \( \gamma \) for the dependent variable, we are able to cover a far wider range of possible functional forms.
Now, let $W = y^{(a)} - x^{(3)}_2 \ ; \ X = (x_1^{(3)} - x^{(a)}_2 ; u + x_1^{(3)})$ \\
$\Rightarrow \beta' = (\beta_1, \beta_2)$.

Then, if we assume that $\alpha \sim NID(0, \sigma^2), \forall \xi_2$,

$$\Rightarrow \ell(\beta, \sigma, \lambda, g(y)) \propto \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} (\omega - \lambda \beta)^T (\omega - \lambda \beta) \right\}$$

$$\times \left(\frac{1}{\sigma^2}\right) \exp \left\{ -\frac{1}{2\sigma^2} [n \hat{\sigma}^2 + (\beta - \beta')^T X'X(\beta - \beta')] \right\}$$

where:

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

$$\hat{\sigma}^2 = \frac{1}{n}(\omega - \hat{\lambda} \hat{\beta})^T (\omega - \hat{\lambda} \hat{\beta})$$

$$\theta = \frac{1}{\sigma^2}.$$

Now, $\hat{\beta}$ and $\hat{\sigma}^2$ are functions of $Y_g$ and $\lambda$, for given $X_1, \beta$ and $\theta$. Given $X_1, \beta$ and $\theta$, we have MLE. Now, to get global maximum estimates, look at:

$$L_{max}(g, \lambda) = \text{const.} + \left(\frac{1}{\theta} - \frac{1}{\theta} \right) \log y_{g \lambda} - \frac{1}{2} \log \hat{\sigma}^2.$$

Search over the $(g, \lambda)$ grid to find a maximal value for $L_{max}$. This then yields $(\hat{g}, \hat{\lambda})$, which yields $\hat{\beta}(\hat{g}, \hat{\lambda})$ and $\hat{\sigma}^2(\hat{g}, \hat{\lambda})$. These are the desired MLE. Given $\beta$, we then have $\hat{\beta}$ and $\hat{\sigma}^2$. We already have $\hat{\beta}$ and $\hat{\sigma}^2$, so we can obtain $\hat{g}, \hat{\lambda}$.

Now, the advantage of using this type of approach to handling the CES form is that it 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.

$$V = \gamma \left[ s_1 x_1^3 + s_2 x_2^3 + \cdots + s_{K-1} x_{K-1}^3 + (1 - s_1 - \cdots - s_{K-1}) x_K^3 \right]^{g/\lambda}$$

$$Q = \gamma = \rho = (E - 1)/E$$

$$V_{g/\lambda} = \gamma^{g/\lambda} \left[ s_1 (x_1^3 - x_2^3) + s_2 (x_2^3 - x_K^3) + \cdots + s_{K-1} (x_{K-1}^3 - x_K^3) + x_K^3 \right]^{g/\lambda}$$

$$s_{g/\lambda} = s^{g/\lambda}$$

$$V = \gamma x_K^3 + \beta_1 (x_1^3 - x_2^3) + \beta_2 (x_2^3 - x_K^3) + \cdots + \beta_{K-1} (x_{K-1}^3 - x_K^3) + \beta_K (1 + x_K^3)$$

where $\beta_i = \gamma^{g/\lambda}$, $i = 1, \ldots, K-1.$
Assume, for generality, that
\[ y^{(u)} = \beta \mathbf{1} + w. \]

Here, \( \beta = \phi(\mathbf{1}^T) \), with free parameter \( \phi \).

Then, \( w = X\beta + u \).

Where,
\[ u = y^{(u)} \quad \beta = (\beta_1, \ldots, \beta_k) \]
\[ u = (u_1, \ldots, u_n) \]
\[ X = (x_k^{(u)}, x_{ni}^{(u)} - x_k^{(u)}, \ldots, x_{ki}^{(u)} - x_k^{(u)}, u + g(x_{ki}^{(u)})) \]

\[ l(\beta, \lambda, \sigma, \gamma, u | y) \propto \left( \frac{1}{\sigma n} \right)^{-1} \exp \left\{ -\frac{1}{2\sigma} (w - x\beta)' (w - x\beta) \right\} \]

Then, for given \( \lambda \) and \( \gamma \), the conditional maximum likelihood values are:
\[ \hat{\beta} = (X'X)^{-1} X'w = (X'X)^{-1} X'y^{(u)}. \]
\[ \hat{\sigma}^2 = \frac{1}{n} (w - x\hat{\beta})'(w - x\hat{\beta}) \]
\[ = \frac{1}{n} (y^{(u)} - x\hat{\beta})'(y^{(u)} - x\hat{\beta}). \]

\[ l_{\text{max}} = \text{const.} + (\lambda - 1) \frac{2}{\alpha} \log \gamma + \frac{n}{2} \log \hat{\sigma}^2. \]

Search over the \((\lambda, \gamma)\) grid to obtain \((\hat{\lambda}, \hat{\gamma})\) so that
\[ l_{\text{max}}(\hat{\lambda}, \hat{\gamma}) \approx \text{max}. \]

Then we get MLE = \( \hat{\beta}(\hat{\lambda}, \hat{\gamma}); \hat{\sigma}^2(\hat{\lambda}, \hat{\gamma}). \)
Generalized production functions:

GP's are just another broad class of functions which are usually non-linear in both parameters and variables. They permit generalization in two directions—

(a) we want production forms with a preassigned but unknown elasticity \( g \) 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:

\[
\frac{dV}{df} = \frac{V \alpha(V)}{f^k}
\]

with solution \( V = g(f) \)

where \( \alpha(V) = r.t.s. \) as a fcn. of output, \( V \).
\( f = f(K, L) \) is a neo-classical prod. fcn.
\( \alpha_f = r.t.s. \) associated with \( f \).
and \( \alpha(V) \) is chosen so that \((dV/df) \to 0 \) if or \( f \to \infty \).

Then this leads to a production function of the form

\[
Ve^\theta = K^{a(1-s)} L^{ds}
\]

\( \log V_i + \theta V_i = \beta_0 + \beta_2 \log K_i + \beta_3 \log L_i + u_i \)

where \( \beta_0 = \log s; \beta_2 = a(1-s); \beta_3 = ds. \)

Let \( \text{ui} \sim NID(0, \sigma^2) \), \( u_i. \)

\( L(\beta, \theta, \sigma | \text{data}) \propto (\frac{1}{2\pi\sigma^n}) \exp\left[-\frac{1}{2\sigma^2}(Z_0 - xp_1)'(Z_0 - xp_1)\right] \)

where \( Z_0 = (\log V + \theta V) \)
\( \beta' = (\beta_0, \beta_2, \beta_3) \)
\( X = (1) \log K, \log L) \)
\( F = \frac{n}{x_i^2} (1 + \theta V + u_i) . \)

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, \( \beta_1 + \beta_2 = \alpha \).

So, initialize in terms \( \beta, \beta_1, \alpha, \sigma \).

\[
\log \frac{y_i}{\theta} = \beta_1 + \beta_2 \log k_i + \beta_2 \log l_i + u_i
\]

\[
\Rightarrow \log y_i + \theta u_i = \beta_1 + \beta_2 \log (k_i/l_i) + \alpha \log l_i + u_i
\]

Take, for given \( \theta \):

\[
p(\beta_1, \beta_2, \sigma, \alpha | \Theta) \propto g(\Theta)p_1(\beta_2 | \alpha)p_2(\alpha)p_3(\sigma)
\]

where: \( 0 < \Theta, \sigma < \infty \); \( 0 < \beta_2, \alpha < \infty \); \( 0 < \alpha < \infty \).

\[
g(\Theta) \propto \frac{1}{\Theta^{3/2}}
\]

\[
p_1(\beta_1 | \alpha) \propto \left( \frac{\beta_1}{\alpha} \right)^{1/2} \left( 1 - \frac{\beta_1}{\alpha} \right)^{1/2} \; \text{is \( \beta \)-dist.}
\]

\[
p_2(\alpha) \propto \text{const.}
\]

\[
p_3(\sigma) \propto \frac{1}{\sigma}
\]

Note that \( g(\Theta) \) is just a proportionality factor — it is not a pdf at all.

Do we need some marginal prior for \( \Theta \),

say \( p_4(\Theta) \).

Then if \( p_4(\Theta) \propto \text{const.} \), \( g \propto g_1 \propto g_2 \),

\[
\Rightarrow p(\beta_1, \beta_2, \sigma, \alpha, \Theta) \propto \left( \frac{1}{\Theta^{3/2}} \right)
\]

Now, transform this pdf to be in terms \( \beta_3 \) —

\[
\text{Let } \begin{align*}
\beta_1 + \beta_2 & = \alpha \\
\frac{\partial \alpha}{\partial \beta_3} & = 1 \\
\Rightarrow & \; \Theta = 1
\end{align*}
\]

\[
\Rightarrow p(\beta_1, \beta_2, \beta_3, \sigma, \Theta) \propto \left( \frac{1}{\Theta^{3/2}} \right)
\]

And we had:

\[
L(\beta, \sigma | \text{data}) \propto \left( \frac{1}{\sigma^{3/2}} \right)\exp \left\{ \frac{1}{2\sigma^2} (\theta - \beta)^t (\theta - \beta) \right\}
\]

\[
p(\beta, \sigma | \text{data}) \propto \left( \frac{1}{\sigma^{3/2}} \right)\exp \left\{ -\frac{1}{2\sigma^2} (\theta - \beta)^t (\theta - \beta) \right\}
\]

\[
\propto \left( \frac{1}{\sigma^{3/2}} \right)\exp \left\{ -\frac{1}{2\sigma^2} \left[ \beta^2 + (\beta - \beta_0)^t X (\beta - \beta_0) \right] \right\}
\]
2. 

\[ U = \eta - 3 ; \quad US_0^2 = (2_0 - X\beta_0)'(2_0 - X\beta_0) \]

\[ \beta_0 = (X'X)^{-1}X'2_0. \]

Then \( \beta | \theta, \sigma, y \) is M.V.N., with conditional mean \( \beta_0 \) and conditional C.V. \((X'X)^{-1} \sigma^2.\)

Now, how do we obtain the marginal posterior pdf's?

\[
\begin{aligned}
\rho(\sigma, \theta \mid \text{data}) &= \int \rho(\beta, \sigma, \theta \mid \text{data}) \, d\beta \\
&\propto \int \left( \frac{\Gamma(n+1)}{(\sigma^{n+1})} \exp \left\{ -\frac{US_0^2}{2\sigma^2} \right\} \right) \exp \left\{ -\frac{1}{2\sigma^2} \left( \beta - \beta_0 \right)'X'X(\beta - \beta_0) \right\} \, d\beta \\
&\propto \left( \frac{\Gamma(n+1)}{(\sigma^{n+1})} \right) \exp \left\{ -\frac{US_0^2}{2\sigma^2} \right\} \left( \frac{1}{\sigma^2} \right)^k \frac{1}{\nu^k} \\
&\propto \left( \frac{\Gamma(n+1)}{(\sigma^{n+1})} \right) \exp \left\{ -\frac{US_0^2}{2\sigma^2} \right\}.
\end{aligned}
\]

\[
\rho(\theta \mid \text{data}) = \int \rho(\theta, \sigma \mid \text{data}) \, d\sigma \\
\propto \left( \frac{\Gamma(n+1)}{(\sigma^{n+1})} \right) \int \exp \left\{ -\frac{1}{2\sigma^2} (US_0^2) \right\} \, d\sigma \\
\propto \left( \frac{\Gamma(n+1)}{(\sigma^{n+1})} \right) (US_0^2)^{-\nu/2} \\
\propto \frac{\Gamma(n+1)}{(\nu^k)^{\nu/2}}.
\]

And the mode of this last marginal posterior pdf coincides with the MLE.
A) 1st-Order Autoregressive Model:

\[ y_t = \beta_1 + \beta_2 y_{t-1} + u_t \]

\[ u_t \sim \mathcal{N}(0, \sigma^2) \; \forall t. \]

Suppose we are given \( y_0 \) at time two. See Ch. 4 for other alternatives. Then,

\[ \mathcal{L}(\beta_1, \beta_2, \sigma | y_0) \propto \frac{1}{\sigma^T}, \exp \left\{ \frac{1}{2 \sigma^2} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 \right\} \]

Assume \( \rho(\beta, \sigma) \propto \left( \frac{1}{\sigma} \right) \).

i.e., we are not constraining the model to be stable, we are allowing \(-\infty < \beta_2 < \infty\) so it may be that \( |\beta_2| > 1 \), so the system will explode.

\[ \rho(\beta, \sigma | y_0, y_0) \propto \left( \frac{1}{\sigma^T} \right) \exp \left[ -\frac{1}{2 \sigma^2} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 \right] \]

\[ \rho(\beta | y_0) \propto \left\{ \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 \right\}^{-\frac{1}{2}} \]

Let \( v = t - 2 \); \( v s^2 = \sum (y_v - \hat{\beta}_1 - \hat{\beta}_2 y_{v-1})^2 \).

Now, \( x = \begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \)

So, \( (x'x) = \begin{pmatrix} \sum y_{t-1} & \sum y_t \\ \sum y_{t-1} & \sum y_{t-1} y_t \end{pmatrix} = H; \) say,

\[ \sigma (x'y) = \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix} \]

\[ \hat{\beta} = (x'x)^{-1} (x'y) = \left( \begin{pmatrix} \sum y_{t-1} & \sum y_t \\ \sum y_{t-1} & \sum y_{t-1} y_t \end{pmatrix} \right)^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix} \]

So,

\[ \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 = \sum \left[ (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1}) - \left\{ (\hat{\beta}_1, \hat{\beta}_2) + (\hat{\beta}_1, \hat{\beta}_2) y_{t-1} \right\} \right]^2 \]

\[ = \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1})^2 + \sum \left\{ (\hat{\beta}_1, \hat{\beta}_2) + (\hat{\beta}_1, \hat{\beta}_2) y_{t-1} \right\}^2 \]

\[ - 2 \sum \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 y_{t-1}) \left\{ (\hat{\beta}_1, \hat{\beta}_2) + (\hat{\beta}_1, \hat{\beta}_2) y_{t-1} \right\} \]

\[ = vs^2 + \sum \left\{ (\hat{\beta}_1, \hat{\beta}_2), (\hat{\beta}_1, \hat{\beta}_2) \right\} \begin{pmatrix} \sum y_{t-1} & \sum y_t \\ \sum y_{t-1} & \sum y_{t-1} y_t \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 - \hat{\beta}_1 \\ \hat{\beta}_2 - \hat{\beta}_2 \end{pmatrix} + \gamma \]

\[ = vs^2 + \sum \left\{ (\hat{\beta}_1, \hat{\beta}_2), (\hat{\beta}_1, \hat{\beta}_2) \right\} \begin{pmatrix} \sum y_{t-1} & \sum y_t \\ \sum y_{t-1} & \sum y_{t-1} y_t \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 - \hat{\beta}_1 \\ \hat{\beta}_2 - \hat{\beta}_2 \end{pmatrix} + \gamma \]
\[ y = \sum (\hat{\beta}_1 + \hat{\beta}_2) \sum (y_t - \hat{\beta}_1) \sum (y_t - \hat{\beta}_2) \]

But,

\[ [\hat{\beta}_1 + \hat{\beta}_2 (y_t - \hat{\beta}_1) + \hat{\beta}_2 (y_t - \hat{\beta}_2)] = - \sum (y_t - \hat{\beta}_1) - \sum (y_t - \hat{\beta}_2) \]

\[ = 0 \]

So,

\[ p(\beta | y, y_0) \propto \left\{ y^2 + (\beta - \bar{\beta})' H (\beta - \bar{\beta}) \right\}^{-(v+2)/2}. \]

And this is a bivariate t-distribution, with mean \( \bar{\beta} \).

No the marginal posteriors for \( \beta_1 \) and \( \beta_2 \) will be univariate t-

\[ \frac{(\beta_1 - \hat{\beta}_1)}{s (h^2)^{1/2}} \sim t_v. \]

\[ \frac{(\beta_2 - \hat{\beta}_2)}{s (h^2)^{1/2}} \sim t_v. \]

IV: \( h^2 = \sum y_t^2 \); \( h^2 = T \).

Now suppose that we are interested in \( p(\sigma, y_0) \).

Then,

\[ p(\sigma | y, y_0) = \int p(\beta, \sigma | y, y_0) d\beta \]

Let the Ch. 2 result

\[ p(\mu, \sigma | y) \propto (2\pi\sigma^2)^{-1} \exp \left\{ -\frac{1}{2\sigma^2} \left\{ y - \mu \right\}^2 \right\} \]

Then,

\[ p(\sigma | y) \propto \sigma^{-v/2} \exp \left\{ -\frac{\sum y_t^2}{2\sigma^2} \right\} \]

So here:

\[ p(\sigma | y, y_0) \propto \sigma^{-v/2} \exp \left\{ -\frac{\sum y_t^2}{2\sigma^2} \right\} \]

So, the results for the marginal posterior distributions of \( \beta \) or \( \sigma \) are the same as in the basic theoretical model.

But this just because we have paid no attention to limitations on \( \beta_2 \) — such limitations are now important if the model is not to explode.

In particular, impose the condition:

\[ |\beta_2| < 1. \]
Now, if $|\beta_1| < 1$, this does in fact imply stability in a time series sense:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

$$= \beta_0 + \beta_1 (\beta_0 + \beta_1 y_{t-1} + \epsilon_{t-1}) + \epsilon_t$$

$$= \beta_0 + \beta_1 \beta_0 + \beta_1^2 y_{t-1} + \beta_1 \epsilon_{t-1} + \epsilon_t$$

So,

$$y_t = \frac{\beta_0}{1 - \beta_1} + \sum_{i=0}^{\infty} \beta_1^i \epsilon_{t-i}.$$  

So,

$$y_0 = \frac{\beta_0}{1 - \beta_1} + \sum_{i=0}^{\infty} \beta_1^i \epsilon_{-i}.$$  

$$= \left( \frac{\beta_0}{1 - \beta_1} \right) + \sum_{i=0}^{\infty} \beta_1^i \epsilon_{-i}.$$  

$$\therefore E(y_0) = \left( \frac{\beta_0}{1 - \beta_1} \right) + \sum_{i=0}^{\infty} \beta_1^i E(\epsilon_{-i})$$  

$$\therefore E(y_0) = \left( \frac{\beta_0}{1 - \beta_1} \right)$$

And,

$$\text{if } x = f(x_1, x_2, \ldots, x_n)$$

$$\Rightarrow \text{var}(x) = \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i} \right)^2 \text{var}(x_i)$$

where a zero subscript ⇒ a mean value.

$$\therefore \text{var}(y_0) = \sigma^2 + \beta_1^2 \sigma^2 + \beta_1^4 \sigma^2 + \ldots$$

$$= \sigma^2 \left[ 1 + \beta_1^2 + (\beta_1^2)^2 + \ldots \right]$$

$$\therefore \text{var}(y_0) = \left( \frac{\sigma^2}{1 - \beta_1^2} \right)$$

And normality is preserved.

$$\therefore l(\beta_1, \sigma, \sigma(y_0)) \propto \frac{1 - \beta_1^2}{\sigma} \exp \left\{ -\frac{(1 - \beta_1^2)}{2\sigma^2} (y_0 - \frac{\beta_1}{1 - \beta_1})^2 \right\}.$$  

$$l(\beta_1, \sigma, \sigma(y_0)) \propto \sigma^{-(\sigma^2+1)}(1-\beta_1^2)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum \left( 1 - \beta_1^2 \left( y_t - \frac{\beta_1}{1 - \beta_1} \right) \right)^2 \right\}$$

Let $\sigma, \beta_1, \beta_2$ be independently distributed. Let $\beta_1$ and $\sigma$ be uniformly distributed, and let prior of $\beta_1$ be $p(\beta_1)$

$$. p(\beta_1, \beta_2, \sigma) \propto \left( \frac{p(\beta_1)}{\sigma} \right) \quad \left\{ \begin{array}{l} -\infty < \beta_1 < \infty \ \ 0 < \sigma < \infty \ \ |\beta_2| < 1 \end{array} \right.$$
\[ p(\beta_1, \beta_2, \sigma | y, y_0) \propto \left[ \frac{\nu}{\sigma^{\nu+1}} \right] \exp \left\{ -\frac{\nu}{2\sigma^2} \left[ \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2 + (1 - \beta_2)(y_0 - \frac{\beta_1}{1 - \beta_2})^2 \right] \right\} \]

Look at the exponent:

\[
(1 - \beta_2)(y_0 - \frac{\beta_1}{1 - \beta_2})^2 + \sum (y_t - \beta_1 - \beta_2 y_{t-1})^2
\]

\[
= (1 - \beta_2) y_0^2 + (1 - \beta_2) \left( \frac{\beta_1}{1 - \beta_2} \right)^2 - 2 y_0 \beta_1 \left( \frac{1 - \beta_2}{1 - \beta_2} \right) - \sum (y_t - \beta_1)^2 - \sum (\beta_2 y_{t-1})^2 - 2 \sum (y_t - \beta_1) \beta_2 y_{t-1}
\]

\[
= y_0^2 (1 - \beta_2) + \frac{\beta_1^2}{(1 - \beta_2)} - 2 y_0 \beta_1 + \sum y_t^2 + \sum y_{t-1}^2 - 2 \beta_2 \sum y_t y_{t-1} + 2 \beta_1 \beta_2 \sum y_{t-1}^2
\]

\[
= \beta_1^2 \left[ \tau + \frac{(1 + \beta_2)}{(1 - \beta_2)} \right] - 2 \beta_1 \left[ \sum y_t + y_0 (1 + \beta_2) - \beta_2 \sum y_{t-1} \right]
\]

\[
+ \left[ y_0^2 (1 - \beta_2) + \sum y_t^2 + \beta_2 \sum y_{t-1}^2 - 2 \beta_2 \sum y_t y_{t-1} \right]
\]

\[
= \beta_1^2 \left[ \tau + \frac{(1 + \beta_2)}{(1 - \beta_2)} \right] - 2 \beta_1 \left[ \sum y_t + y_0 (1 + \beta_2) - \beta_2 \sum y_{t-1} \right]
\]

\[
+ \left[ y_0^2 (1 - \beta_2) + \sum y_t^2 + \beta_2 \sum y_{t-1}^2 - 2 \beta_2 \sum y_t y_{t-1} \right]
\]

\[
= c \beta_1^2 - \left( \beta_1 - \beta_2 \right) \left( y_0 (1 + \beta_2) + \sum (y_t - \beta_1 y_{t-1}) \right) + \left[ y_0^2 (1 - \beta_2) + \sum (y_t - \beta_1 y_{t-1}) \right]
\]

\[
= c \beta_1^2 - \left( \beta_1 - \beta_2 \right) \left( y_0 (1 + \beta_2) + \sum (y_t - \beta_1 y_{t-1}) \right) + \left[ y_0^2 (1 - \beta_2) + \sum (y_t - \beta_1 y_{t-1}) \right]
\]

\[
= \left( \beta_1 - \beta_2 \right) \left( y_0^2 + \sum (y_t - \beta_1 y_{t-1}) \right)
\]

\[
\theta = \frac{1}{c} \left( \tau \Xi_1 + \frac{(1 + \beta_2)}{(1 - \beta_2)} \sum y_t \right)
\]

Where

\[
\Xi_1 = \sum (y_t - \bar{y} - \beta_1 (y_{t-1} - \bar{y}_{t-1}))^2
\]

\[
\Xi_2 = \sum (y_t - \beta_1 y_{t-1} - (1 - \beta_2)(y_0))^2
\]

\[
\bar{y} = \frac{1}{T} \sum y_t \quad \bar{y}_{t-1} = \frac{1}{T} \sum y_{t-1}
\]

So

\[
p(\beta, \sigma | y, y_0) \propto \left( \frac{\nu}{\sigma^{\nu+1}} \right) \exp \left\{ -\frac{\nu}{2\sigma^2} \left[ \left( \beta_1 - \beta_2 \right) \left( y_0^2 + \sum (y_t - \beta_1 y_{t-1}) \right) \right] \right\}.
\]
Now, we have:
\[ \hat{\beta}_1 = \frac{(1+\bar{y}_0) + \sum (y_t - \bar{y}_0, y_{t-1})}{c} \]

But, let \( h_1 = (T/\sigma^2) \); \( h_2 = \frac{(1+\bar{y}_0)}{\sigma^2 (1-\beta_1)} \)

So \( h_1 + h_2 = \frac{T}{\sigma^2} + \frac{1}{\sigma^2 (1-\beta_1)} \)
\[ = \frac{1}{\sigma^2} (T + \frac{1}{1-\beta_1}) \]
\[ = \frac{1}{\sigma^2} \left( \frac{T}{1-\beta_1} \right) \]

And, \( h_1 \sum (y_t - \beta_1, y_{t-1}) / T \)
\[ = \frac{1}{\sigma^2} \sum (y_t - \beta_1, y_{t-1}) \]
And, \( h_2 (1-\beta_1) y_0 \)
\[ = \frac{(1+\bar{y}_0) (1-\beta_1)}{(1-\beta_1) \sigma^2} \]
\[ = \frac{1}{\sigma^2} (1+\bar{y}_0) \]

\[ \hat{\beta}_1 = \frac{h_1 \sum (y_t - \beta_1, y_{t-1}) / T + h_2 (1-\beta_1) y_0}{h_1 + h_2} \]

Now, suppose that we are given \( \beta_1 = \alpha \). Then the conditional posterior for \( \beta_1 \) is
\[ p(\beta_1 | \beta_2, \sigma, y, y_0) \propto \frac{p(\beta_1, 1-\beta_1) \exp \left\{ - \frac{\alpha}{2 \sigma^2} (\beta_1 - \bar{\beta}_1)^2 \right\}}{\sigma^2 (1-\beta_1)} \exp \left\{ - \frac{1}{2 \sigma^2} (T \sum y_t + \frac{1+\bar{y}_0}{1-\beta_1} \sum t) \right\} \]
\[ \propto \exp \left\{ - \frac{\alpha}{2 \sigma^2} (\beta_1 - \bar{\beta}_1)^2 \right\} \]

So the conditional posterior \( \beta_1 \) has mean \( \bar{\beta}_1 \), and variance \( \sigma^2 / c \).

Now \( c = \frac{(1+\bar{y}_0)}{(1-\beta_1)} + T \)
\[ \Rightarrow T \to \infty \implies \frac{c}{\sigma^2} \to 0 \]
\[ \Rightarrow T \to \infty \implies \hat{\beta}_1 \to \frac{1}{T} \sum (y_t - \beta_1, y_{t-1}) / T. \]

To get the marginal posterior for \( \beta_1 \), integrate
\[ p(\beta_1, \beta_2, \sigma, y, y_0) \] w.r.t. \( \sigma \) and use Bivariate numerical integration w.r.t. \( \beta_2 \).
What about the marginal posterior pdf for $\mathbf{\beta}_2$?

\[
\begin{align*}
 p(\mathbf{\beta}_1, \mathbf{\beta}_2, \sigma | y, y_0) & \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ c(\mathbf{\beta}_1 - \bar{\mathbf{\beta}}_1)^t \right] \right. \\
& \left. \quad \left(\frac{1}{2\Sigma_1} + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right) \right\} \\
\end{align*}
\]

\[
\begin{align*}
p(\mathbf{\beta}_2 | y, y_0) & \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right] \right\} \\
& \left. \quad \int \exp \left\{ -\frac{1}{2\sigma^2} \left[ c(\mathbf{\beta}_1 - \bar{\mathbf{\beta}}_1)^t \right] \right\} d\mathbf{\beta}_1 \\
& \int \exp \left\{ -\frac{1}{2\sigma^2} \left[ c(\mathbf{\beta}_1 - \bar{\mathbf{\beta}}_1)^t \right] \right\} d\mathbf{\beta}_1 = 1. \\
\end{align*}
\]

Now, \( \int \exp \left\{ -\frac{1}{2\sigma^2} \left[ c(\mathbf{\beta}_1 - \bar{\mathbf{\beta}}_1)^t \right] \right\} d\mathbf{\beta}_1 = (2\pi)^{1/2} \).

\[
\begin{align*}
p(\mathbf{\beta}_2 | y, y_0) & \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right] \right\} \\
& \left. \quad \alpha \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ \frac{h_1 \Sigma_1 + h_2 \Sigma_2}{h_1 + h_2} \right] \right\} \right.
\end{align*}
\]

Now, \( h_1 + h_2 = (c/\sigma^2) \)

\[
\begin{align*}
h_1 \Sigma_1 + h_2 \Sigma_2 \\
& \frac{(c/\sigma^2)}{h_1 + h_2} = \frac{(T\Sigma_1)}{(T\Sigma_1) + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n} \left(\frac{1}{c/\sigma^2} \right) \\
& = \left(\frac{1}{c} \right) \left( T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right) \\
& \propto \left(\frac{1}{\sigma^2} \right) \left[ T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right] \exp \left\{ -\frac{1}{2\sigma^2} \left[ \frac{h_1 \Sigma_1 + h_2 \Sigma_2}{h_1 + h_2} \right] \right\}. \\
\end{align*}
\]

Now, consider \( p(\mathbf{\beta}_1, \sigma | y, y_0) \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right] \right\} \)

\[
\begin{align*}
p(\mathbf{\beta}_2 | y, y_0) & \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \int \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \exp \left\{ -\frac{1}{2\sigma^2} \left[ T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n \right] \right\} d\mathbf{\beta}_2 \\
& \left. \quad \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \left[ \frac{1}{T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n} \right]^{-1/2} \right. \\
& \left. \quad \alpha \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \left[ \frac{1}{T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n} \right]^{-1/2} \right. \\
& \left. \quad \propto \left(\frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}}\right) \left[ \frac{1}{T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n} \right]^{-1/2} \right. \\
& \left. \quad \propto \frac{p(\mathbf{\beta}_2)(1-\mathbf{\beta}_2)^t}{\sigma^{1t}} \left[ \frac{1}{T \Sigma_1 + \frac{1}{N\bar{\mathbf{e}}_n} \mathbf{e}_n} \right]^{-1/2} \right. \\
& \left. \quad \propto 1. \right.
\end{align*}
\]
Now, consider the following expression

\[
\frac{p(\beta_2 | \Sigma)}{(\Sigma_1)^{TV_2}} \times \left(1 - \beta_2^2 \right)^{\frac{1}{2}} \left[ \frac{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(1 - \beta_2^2)}{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(\Sigma_2)} \right]^{-TV_2}
\]

\[
= p(\beta_2)(1 - \beta_2)^{\frac{1}{2}} \left[ \frac{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(1 - \beta_2^2)}{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(\Sigma_2)} \right]^{-TV_2}
\]

\[
= p(\beta_2)(1 - \beta_2)^{\frac{1}{2}} C(T-1)^{-TV_2} \left[ \frac{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(1 - \beta_2^2)}{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(\Sigma_2)} \right]^{-TV_2}
\]

\[
= p(\beta_2)(1 - \beta_2)^{\frac{1}{2}} C(T-1)^{-TV_2} \left( 1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2}) \right) \Sigma_2^{-TV_2}
\]

So,

\[
p(\beta_2 | y, y_0) \propto \frac{p(\beta_2)}{(\Sigma_1)^{TV_2}} \left[ \frac{1 - \beta_2^2}{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(1 - \beta_2^2)} \right]^{\frac{1}{2}} \times \left[ \frac{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(1 - \beta_2^2)}{1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/(\Sigma_2)} \right]^{-TV_2}
\]

Note, note that the second log term:

If we are to use this then we need information concerning \( p(\beta_2) \) as well as concerning the pdf of \( y_0 \).

But, consider the asymptotic properties of this marginal posterior pdf

\[
\log p(\beta_2 | y, y_0) \propto \log p(\beta_2) - TV_2 \log \Sigma_1 + \frac{1}{2} \log (1 - \beta_2^2)
\]

\[
- \frac{1}{2} \log \left( 1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2})/\Sigma_2 \right) + \frac{1}{2} \log \left( 1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2}) \right)
\]

\[
- TV_2 \log \left( 1 + \left( \frac{1}{2} \right)(1 + \frac{1}{2}) \right)
\]

And as \( T \to \infty \), the dominant term is \( -TV_2 \log \Sigma_1 \).

So, as \( T \to \infty \),

\[
p(\beta_2 | y, y_0) \propto \Sigma_1^{-TV_2}
\]

ie.,

\[
p(\beta_2 | y, y_0) \propto \left\{ \Sigma(y_k - \overline{y} - \beta_2(y_k - \overline{y})) \right\}^{TV_2}
\]

And this is independent of \( p(\beta_2) \) as assumptions concerning \( y_0 \). So, in large samples we have a neat and simple univariate approximation to the marginal posterior pdf for \( \beta_2 \).
However, if the sample is small, then we are going to require information concerning both $y_0$ and $p(\beta)$. Consider the latter —
we want to restrict $|\beta| < 1$.

Now, consider the Beta pdf

$$p(z) \propto (z)^{\alpha-1} (1-z)^{\beta-1}; \quad 0 < z < 1.$$ 

Now let

$$\begin{align*}
\alpha &= k_2 \\
\beta &= k_1 \\
\bar{z} &= \frac{1}{2} (1 + \beta_2) \\
\bar{1-z} &= \frac{1}{2} (1 - \beta_2)
\end{align*}$$

so

$$d\bar{z} \propto d\beta_2.$$ 

Thus

$$p(\beta_2) \propto (1 + \beta_2)^{k_2-1} (1 - \beta_2)^{k_1-1}.$$ 

And $\bar{z} < 1 \Rightarrow 0 < \frac{1}{2} (1 + \beta_2) < 1$ gives $-1 < \beta_2 < 1$, 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 weight $k$, i.e., a prior.

If we put $k_1 = k_2 = \frac{1}{2}$ then

$$p(\beta_2) \propto (1 - \beta_2)^{-\frac{1}{2}} (1 + \beta_2)^{-\frac{1}{2}}.$$ 

So

$$p(\beta_2) \propto \frac{1}{2} (1 - \beta_2)^{-\frac{1}{2}}.$$ 

And it can be shown that this is approximately the same as Jeffrey's invariant prior

$$\frac{1}{2} \left( \frac{1}{\sqrt{1 - \beta_2}} \right)^{\frac{1}{2}}.$$ 

So put

$$p(\beta_2, \sigma) \propto \frac{1}{\sigma (1 - \beta_2)^{\frac{1}{2}}}. $$
(C) - Stability of a 2nd-Order Autoregressive Structure

Bayesian techniques may be used to make inference 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:

\[ y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t \]

\[ u_t \sim NID(0, \sigma^2) \quad \forall t. \]

Assume that the initial values \( y_0 \) and \( y_{-1} \) are given. Then,

\[ \ell(\alpha, \sigma^2 | y) \propto \left( \frac{1}{\sigma} \right)^T \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{t=1}^{T} (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2 \right] \right\}. \]

Assume that \( \mathcal{P}(\alpha, \sigma^2) \propto \frac{1}{\sigma^2} ; \quad -\infty < \alpha_1, \alpha_2 < \infty ; \ \sigma > 0. \)

Then \( \mathcal{P}(\alpha_1, \alpha_2 | y) \propto \left[ \sum_{t=1}^{T} (y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2})^2 \right]^{-T/2} \]

\[ \times \left[ \sigma^2 + (\alpha - \hat{\alpha})' \Sigma (\alpha - \hat{\alpha}) \right]^{-T/2}, \]

where \( \Sigma = T^{-1} \mathbf{V} \mathbf{V}^T \)

\[ \mathbf{V}^2 = \frac{T}{2} \left( y_{t-2}, y_{t-1}, y_t - \hat{\alpha}_1 y_{t-1} - \hat{\alpha}_2 y_{t-2} \right)^2 \]

\[ \hat{\alpha}_1 = \left( \frac{\mathbf{V}_1}{\mathbf{V}_2} \right) = H^{-1} \left( \frac{\mathbf{E}_y y_{t-1}}{\mathbf{E}_y y_{t-2}} \right) \]

\[ \hat{\alpha}_2 = \left( \frac{\mathbf{V}_3}{\mathbf{V}_4} \right) = H^{-1} \left( \frac{\mathbf{E}_y y_{t-2}}{\mathbf{E}_y y_{t-1}} \right) \]

Then the joint posterior pdf for \( (\alpha_1, \alpha_2) \) is bivariate.
If we are given values for $y$, then we can make
inferences about $x_1$, $x_2$, $q$ and hence about the dynamic
properties of the model.

Once we have $p(x_1, x_2 | y)$, we can normalize
and integrate over various ranges for $x_1, x_2, q$.
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 or non-explosive is 0.9. etc.
Also, $p(NE) = p(NE-O) + p(NE-NO)$. etc.

The characteristic equation for the model is

$$x^2 - x(\alpha_1 + \alpha_2) - \alpha_2 = 0$$

$$x_1 = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2}; \quad x_2 = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}.$$ 

And the system will oscillate if $\alpha_1^2 + 4\alpha_2 < 0$.

So, it may be of interest to find $Pr.(\text{oscillates}).$

i.e. $Pr.(\alpha_1^2 + 4\alpha_2 < 0)$.

Let $V_1 = x_1$
$V_2 = \alpha_1^2 + 4\alpha_2$

$$J = \begin{vmatrix} \frac{\partial V_1}{\partial x_1} & \frac{\partial V_1}{\partial x_2} \\ \frac{\partial V_2}{\partial x_1} & \frac{\partial V_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 2x_1 & 4 \end{vmatrix} = 4$$

$$p(V_1, V_2 | y) = \frac{1}{4} p(x_1, x_2 | y)$$

$$\propto \left[V_1^2 + (\alpha_1 - x_1)^2 + (\alpha_2 - x_2)^2\right]^{-\frac{1}{2N}}.$$
\[ p(v_1, v_2 | y) \propto \left[ v_1^2 + (v_1 - z_1)^2 \right] h_{11} + \left( \frac{v_2 - v_1^2}{4} - z_2 \right)^2 h_{22} \\
+ 2 (v_1 - z_1) (\frac{v_2 - v_1^2}{4} - z_2) h_{21} \]

Now, we are interested in \( p(v_2 | y) \).
So, \( p(v_2 | y) = \int p(v_1, v_2 | y) \, dv_1 \)

To obtain this posterior marginal pdf, we normalize it. Then see when \( v_2 < 0 \).
For the given set of data, compute \( p(v_2 | y) \) for different values of \( v_2 \). Plot -

\[ p(v_2 | y) \]

So the majority of the area to the right of \( v_2 = 0 \)?
If so, it doesn't 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 \((1 \times t)\) row vector of independent variables.
Consider the usual "geometric" model:

\[ y_t = \alpha x_{t-1} + u_t \quad , \quad -\infty < \alpha < \infty \]

Then the Kiyokawa 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 \( \text{var} \left[ u_t - \lambda u_{t-1} \right] = \text{var}(u_t) + \lambda^2 \text{var}(u_{t-1}) = \sigma^2(1 + \lambda^2) \).

Further, \( E \left[ (u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2}) \right] = E \left( u_t u_{t-1} \right) + \lambda^2 E \left( u_{t-1} u_{t-2} \right) - \lambda E \left( u_{t-1} \right) - \lambda E \left( u_{t-2} \right) \)

\[ = 0 + 0 - \lambda \sigma^2 = 0 \quad \text{(by independence assumption).} \]

Further, \( E \left[ (u_t - \lambda u_{t-1})(u_{t-k} - \lambda u_{t-k-1}) \right] \quad ; \quad k \geq 1 \)

\[ = E \left[ u_t u_{t-k} \right] + \lambda^2 E \left[ u_{t-1} u_{t-k-1} \right] - \lambda E \left[ u_{t-1} u_{t-k} \right] - \lambda E \left[ u_{t-k} u_{t-k-1} \right] \]

\[ = 0 \cdot \]

So, \( E \left[ (u - \lambda u)(u - \lambda u)' \right] = \sigma^2 \mathbf{G} \)

\[ \mathbf{G} = \begin{bmatrix}
(1+\lambda^2) & -\lambda & 0 & \cdots & 0 \\
-\lambda & (1+\lambda^2) & -\lambda & \cdots & 0 \\
0 & -\lambda & (1+\lambda^2) & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & (1+\lambda^2)
\end{bmatrix} \]

Let \( y' = (y_1, \ldots, y_T) \) and assume \( y_0 \) given.

Then, \( \rho(y | \lambda, \alpha, \sigma, y_0) \propto \frac{161^{-\frac{T}{2}}}{\sigma^T} \exp \left\{ -\frac{1}{2\sigma^2} (y - \lambda y_{-1} - y')' \mathbf{G}^{-1} (y - \lambda y_{-1} - y') \right\} \)

Assume that \( \rho(\alpha, \lambda, \sigma) \propto \left( \frac{161}{T} \right) \)

\( \rho(x, \lambda, \sigma | y, y_0) \propto \left( \frac{161^{-\frac{T}{2}}}{\sigma^T} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y - \lambda y_{-1} - y')' \mathbf{G}^{-1} (y - \lambda y_{-1} - y') \right\} \)
so, \( p(\lambda, \alpha | y_0) \propto \left\{ \frac{(y - \lambda y_{-1} - \alpha x)'}{\sigma \sqrt{y}} \right\}^{n/2} \). \(161-2\).

The normalizing constant can be obtained from the properties of the bivariate normal bivariate numerical routine. Then joint inferences may be made about \( \alpha \) and \( \lambda \). The marginal posterior for \( \lambda \) can be obtained by completing the square on \( \alpha \) and integrating w.r.t. \( \alpha \), using the properties of the univariate \( t \)-dist’n.

This is feasible analytically.

To get \( p(\alpha | y_0, \lambda) \), however, we have to use numerical techniques (univariate) because we have restricted \( 0 \leq \lambda < 1 \), so the \( t \)-dist’n is not applicable.

Now consider a more elaborate model:

\[
y_t = \alpha \frac{\lambda}{\sigma_\lambda^2} x_{t-1} + \frac{\lambda}{\sigma_\lambda^2} u_{t-1} + \epsilon_t \quad ; \quad t = 2, 3, \ldots, T.
\]

\[
\log \lambda_{t-1} = \alpha \frac{\lambda}{\sigma_\lambda^2} x_{t-1} - \frac{\lambda}{\sigma_\lambda^2} u_{t-1} + \epsilon_t
\]

\[
= \alpha \frac{\lambda}{\sigma_\lambda^2} x_{t-1} + \frac{\lambda}{\sigma_\lambda^2} u_{t-1}
\]

\[
\therefore (y_t - \lambda_{t-1}) = \alpha x_t + u_t.
\]

\[
= y_t - \lambda y_t + \alpha x_t + u_t.
\]

And \( u_t \sim \text{NID}(0, \sigma^2) \); \( \forall t \).

Then, for given \( y_0 \)

\[
\ell(x, \lambda, \sigma^2, y_0) \propto \left( \frac{1}{\sigma^2} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y - \lambda y_{-1} - \alpha x)'(y - \lambda y_{-1} - \alpha x) \right\}
\]

And \( p(\alpha, \lambda, \sigma^2) \propto (\ell) \)

\[
= p(\alpha, \lambda | y_0) \propto \left( \frac{1}{\sigma^2} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y - \lambda y_{-1} - \alpha x)'(y - \lambda y_{-1} - \alpha x) \right\}
\]

\[
\therefore p(\alpha, \lambda | y_0) \propto \left\{ (y - \lambda y_{-1} - \alpha x)'(y - \lambda y_{-1} - \alpha x) \right\}^{-\nu/2}
\]

Now, if it were not for the fact that we have restricted \( 0 \leq \lambda < 1 \), the joint posterior for \( \alpha \) and \( \lambda \) would be a bivariate \( t \)-dist’n.

So to make joint inferences on \( \lambda \) and \( \alpha \), obtain the constant term for \( p(\alpha, \lambda | y_0) \) by bivariate numerical integration.

Further, to get the marginal posterior pdf’s:

(a) \( p(\lambda | y_0) \) can be obtained analytically by completing the square on \( \alpha \), and using the properties of the univariate \( t \)-dist’n.

(b) The restrictions on \( \lambda \) necessitate the use of numerical (univariate) integration to obtain \( p(\alpha | y_0) \).
Suppose that we have our model in the form
\[ y_t = \beta y_{t-1} + \alpha x_t + u_t \]

And we now wish to assume autocorrelated disturbances
\[ u_t = \rho u_{t-1} + \epsilon_t ; \quad \epsilon_t \sim NID(0, \tau^2) \]

Then
\[ \beta y_{t-1} = \beta (\beta y_{t-2} + \alpha x_{t-1} + \rho u_{t-1}) \]
\[ y_t - \beta y_{t-1} = \lambda (y_{t-1} - \beta y_{t-2}) + \alpha (x_t - \alpha x_{t-1}) + \epsilon_t \]

Now, assume that \( p(\alpha, \lambda, \rho, \tau) \propto (\tau^2)^{-\frac{1}{2}} \).\]

where \(-\infty < \rho, \alpha < \infty ; \quad 0 < \lambda < 1 ; \quad 0 < \tau < \infty \).

Now, assuming that we are given \( y_0, y_1, \ldots, y_T \), let
\[ y^T = (y_T, y_{T-1}, y_{T-2}, \ldots, y_1). \]

So,
\[ p(\lambda, \rho, \alpha, \tau | y) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=1}^{T-1} [y_t - \lambda y_{t-1} - \alpha x_t - \rho (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})]^2 \right\}. \]

So,
\[ p(\lambda, \rho, \alpha | y) \propto \left\{ \frac{1}{\tau^2} \sum_{t=1}^{T-1} [y_t - \lambda y_{t-1} - \alpha x_t - \rho (y_{t-1} - \lambda y_{t-2} - \alpha x_{t-1})]^2 \right\}^{-\frac{1}{2}}. \]

Look at the second line. If it hadn't been for the fact that we assumed that \( \rho < 1 \), then \( p(\alpha, \lambda | \rho, y) \) would be bivariate \( t \). Now, we can analyze the joint conditional posterior pdf numerically to make inferences for \( \alpha \) and \( \lambda \), given \( \rho \). But what if we choose the wrong value of \( \rho \) when we estimate it (since \( \rho \) is always unknown)? Then Zellner shows that the results obtained for \( p(\alpha, \lambda | \rho, y) \) and \( p(\lambda | \rho, y) \) are very sensitive to a misspecification of the value \( \rho \).

So, if we suspect that \( \rho \neq 0 \), then it is better to make use of \( p(\alpha, \lambda | y) \) and hence \( p(\alpha | y) \) and \( p(\lambda | y) \) when making inferences than to rely upon \( p(\alpha | \lambda, y) \) and \( p(\lambda | \alpha, y) \) for such situations.
The derivation of \( p(x, \lambda | y) \):

\[
p(x, \lambda | y) = \int_0^\infty p(x, \lambda | y, \rho) \, d\rho
\]

\[
x = \frac{1}{\sqrt{2\pi}} \int \left\{ \sum_{i} \left[ y_i - \lambda y_{i-1} - \alpha x_i - \rho(y_i - \lambda y_{i-1} - \alpha x_{i-1}) \right]^2 \right\}^{-\frac{1}{2}} \, d\rho.
\]

We have a quadratic in \( \rho \)

Complete the square \( a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a}) \).

\[
\frac{1}{\sqrt{2\pi}} \int \left\{ \sum_{i} \left[ y_i - \lambda y_{i-1} - \alpha x_i \right]^2 + \rho \frac{1}{2} \sum_{i} \left[ y_i - \lambda y_{i-1} - \alpha x_i \right] \left( y_{i-1} - \lambda y_{i-2} - \alpha x_{i-2} \right) - \rho \frac{1}{2} \sum_{i} \left[ y_i - \lambda y_{i-1} - \alpha x_i \right] \left( y_{i-1} - \lambda y_{i-2} - \alpha x_{i-2} \right) \right\} \, d\rho
\]

\[
= \frac{1}{\sqrt{2\pi}} \int \left[ \sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)^2 \right] \left[ \rho - \frac{\sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)(y_{i-1} - \lambda y_{i-2} - \alpha x_{i-2})}{\sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)^2} \right] \, d\rho
\]

Now,

\[
\left\{ \frac{a(x + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})}{2} \right\}^{-T/2} \left\{ 1 + \frac{a(x + \frac{b}{2a})^2}{(c - \frac{b^2}{4a})} \right\}^{-T/2}
\]

\[
p \propto (c - \frac{b^2}{4a})^{-T/2} \left\{ 1 + \frac{a(x + \frac{b}{2a})^2}{(c - \frac{b^2}{4a})} \right\}^{-T/2} \int \left( \frac{a}{c - \frac{b^2}{4a}} \right)^{x_2} \left\{ 1 + \frac{a(x + \frac{b}{2a})^2}{(c - \frac{b^2}{4a})} \right\}^{-T/2} \, dx
\]

\[
\propto a^{-T/2} (c - \frac{b^2}{4a})^{-T} \left( \frac{a}{c - \frac{b^2}{4a}} \right)^{x_2} \left\{ 1 + \frac{a(x + \frac{b}{2a})^2}{(c - \frac{b^2}{4a})} \right\}^{-T/2}
\]

So,

\[
p(x, \lambda | y) \propto \left\{ \sum_{i} (y_i - \lambda y_{i-1} - \alpha x_{i-1})^2 \right\}^{-\frac{T}{2}} \times
\]

\[
\left\{ \sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)^2 \right\} \left\{ \frac{\sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)(y_{i-1} - \lambda y_{i-2} - \alpha x_{i-2})}{\sum_{i} (y_i - \lambda y_{i-1} - \alpha x_i)^2} \right\}^{\frac{T}{2}}
\]

where \( x = \frac{T}{2} \).

We can now make joint inferences about \( x \) and \( \lambda \).

If we apply numerical integration we can derive \( p(x | y) \) and \( p(\lambda | y) \).
Also of interest may be functions of \( \alpha \) and \( \lambda \).

The "long-run" quantity:

\[ \eta = \frac{\alpha}{(1-\lambda)}. \]

To make posterior inferences about \( \eta \), change our variables from \((\alpha, \lambda)\) to \((\eta, \lambda)\).

Then

\[ \mathbf{J} = \begin{bmatrix} \frac{\partial^2 \eta}{\partial \alpha^2} & \frac{\partial^2 \eta}{\partial \lambda \partial \alpha} \\ \frac{\partial^2 \eta}{\partial \lambda \partial \alpha} & \frac{\partial^2 \eta}{\partial \lambda^2} \end{bmatrix} \]

\[ \therefore \mathbf{J} = \begin{bmatrix} (1-\lambda) & 0 \\ \frac{\alpha}{(\lambda^2)} & 1 \end{bmatrix} = (1-\lambda). \]

So,

\[ p(\eta, \lambda | y) \propto (1-\lambda) p(\alpha, \lambda | y) \propto (1-\lambda) p(\eta | \eta, \lambda) p(\eta, \lambda) \]

\[ \therefore p(\eta | y) \propto \int (1-\lambda) p(\eta | \eta, \lambda) p(\eta, \lambda) \, d\lambda. \]

And this can be evaluated exactly numerically.

\[ e) \quad \text{Applications to Consumption Functions:} \]

Suppose that we have the model

\[ C_t = kY_t^* + u_t \]

where \( Y_t^* \) is "expected" income (i.e. P.I.H.)

And

\[ (Y_t^* - Y_{t-1}^*) = (1-\lambda)(Y_t - Y_{t-1}) \]

\[ \therefore \]

\[ Y_t^* = Y_{t-1}^* + Y_t - \lambda Y_t - Y_{t-1}^* + \lambda Y_{t-1} \]

\[ = (1-\lambda)Y_t + \lambda (Y_{t-1}^*) \]

\[ - Y_t^* = (1-\lambda)Y_t + \lambda [k Y_{t-1}^* + (1-\lambda) Y_{t-1}] \]

\[ = (1-\lambda)[Y_t + \lambda Y_{t-1}^*] + \lambda^2 Y_{t-1}^* \quad \text{etc.} \]

So,

\[ Y_t^* = (1-\lambda)[Y_t + \lambda Y_{t-1}^* + \lambda^2 Y_{t-2}^* + \ldots + \lambda^k Y_{t-k}^*]. \]

And we assume that \( 0 < \lambda < 1 \).
\[ C_t = k(1-k)[\gamma_{t-1} + \gamma_{t-2} + \ldots + \gamma_{t-k}] + u_t \]

\[ \lambda C_{t+1} = k(1-k)[\gamma_{t} + \gamma_{t-1} + \ldots + \gamma_{t-k}] + \lambda u_{t+1} \]

\[ (C_t - \lambda C_{t+1}) = k(1-k)\gamma_t + (u_t - \lambda u_{t+1}) \]

\[ C_t = \lambda C_{t+1} + k(1-k)\gamma_t + (u_t - \lambda u_{t+1}). \]

Now, consider the various assumptions which could be made concerning the structure of the disturbances.

**II)** \[ (u_t - \lambda u_{t+1}) = \varepsilon_t \sim \text{NID}(0, \sigma^2). \]

**III)** \[ u_t \sim \text{NID}(0, \sigma^2) \]

**IV)** \[ u_t \text{ are 1st order auto.} \]
\[
\begin{align*}
\varepsilon_t & \sim \rho u_{t-1} + \varepsilon_t \sim \text{NID}(0, \sigma^2) \\
\varepsilon_t & \sim \text{NID}(0, \sigma^2)
\end{align*}
\]

**IV)** \[ (u_t - \lambda u_{t+1}) \text{ is 1st order auto} \]
\[
\begin{align*}
\varepsilon_t & \sim \rho u_{t-1} + \varepsilon_t + \varepsilon_t \sim \text{NID}(0, \sigma^2)
\end{align*}
\]

**Note:**

(a) \[ \rho = 1 \Rightarrow (\text{III}) \equiv (\text{II}) \]

(b) \[ \rho = 0 \Rightarrow (\text{III}) \equiv (\text{I}) \]

(c) \[ \gamma = 0 \Rightarrow (\text{IV}) \equiv (\text{I}) \]

Assume that the initial condition, \( C_0 \), is given.

**Case (I):**

\[ p(k, \lambda, \sigma^2) \propto \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2} k} \exp \left\{ -\frac{1}{2 \sigma^2} \sum_{t=1}^{k} [C_t - \lambda C_{t-1} - k(1-k)\gamma_t]^2 \right\} \]

\[ p(k, \lambda, \sigma^2) \propto \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2} k} \exp \left\{ -\frac{1}{2 \sigma^2} \sum_{t=1}^{k} [C_t - \lambda C_{t-1} - k(1-k)\gamma_t]^2 \right\} \]

\[ p(k, \lambda | C) \propto \left[ \frac{1}{\mathcal{N}} \sum_{t=1}^{k} [C_t - \lambda C_{t-1} - k(1-k)\gamma_t]^2 \right]^{-\frac{1}{2} k} \]

\[ p(k, \lambda | C) \propto \left[ \frac{1}{\mathcal{N}} \sum_{t=1}^{k} [C_t - \lambda C_{t-1} - k(1-k)\gamma_t]^2 \right]^{-\frac{1}{2} k} \]

Now, to get \( p(\lambda | C) \) we obtain \( \int p(k, \lambda | C) \, dk. \)

But, this is a truncated Student pdf, and we are integrating since the range of \( k \) is \( (0, 1) \). To handle this numerically. Similarly, we obtain \( p(k | C) \).
Case (II): \( U_t \sim \text{NID}(0, \sigma_t^2) \)

so, \( E[(U_t - \lambda U_{t-1})(U_t - \lambda U_{t-1})] = \sigma_t^2 G \), where \( G \) is as defined in the previous section of these notes.

\[
p(c(\lambda, \sigma_t) | \lambda) \alpha \left( \frac{1}{\sigma_t} \right) \exp \left\{ -\frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' - \frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' \right\}
\]

or \( p(\lambda, \sigma_t) \alpha \frac{1}{\sigma_t} \).

\[
p(c(\lambda, \sigma_t) | \lambda) \alpha \left( \frac{1}{\sigma_t} \right) \exp \left\{ -\frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' - \frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' \right\}
\]

\( p(c(\lambda, \sigma_t) | \lambda) \alpha \left( \frac{1}{\sigma_t} \right) \exp \left\{ -\frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' - \frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' \right\}
\]

\( \sigma_t \alpha \left( \frac{1}{\sigma_t} \right) \exp \left\{ -\frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' - \frac{1}{2 \sigma_t^2} [c-\lambda c_{t-1} - \lambda(1-\lambda)]G^{-1}c' \right\}
\]

for \( \sigma_t > 0 \), \( \lambda > 0 \).

By applying bivariate numerical integration, we can normalize this pdf to make joint posterior inferences for \( k \& \lambda \). The integrate (bivariate numerical) again to get the two marginal posterior pdf's.

Case (III): \( U_t = \rho U_{t-1} + e_{t} \) \( \quad E_{t} e_{t} \sim \text{NID}(0, \sigma_t^2) \)

And we have: \( C_t = \lambda C_{t-1} + k(1-\lambda)Y_t + (U_t - \lambda U_{t-1}) \)

Let \( \eta_t = C_t - U_t \)

\[
\eta_t = \lambda \eta_{t-1} + k(1-\lambda)Y_t + (U_t - \lambda U_{t-1})
\]

and \( \rho \eta_{t-1} = \rho \eta_{t-1} + \rho k(1-\lambda)Y_{t-1} \)

\[
(\eta_t - \rho \eta_{t-1}) = \lambda (\eta_{t-1} - \rho \eta_{t-1}) + k(1-\lambda)(Y_t - \rho Y_{t-1})
\]

\[
\eta_t(\rho) = \lambda \eta_{t-1}(\rho) + k(1-\lambda)Y_t(\rho)
\]

\[
= \lambda^2 \eta_{t-2}(\rho) + \lambda k(1-\lambda)Y_{t-1}(\rho) + k(1-\lambda)Y_t(\rho)
\]

\[
= \lambda^2 \eta_{t-2}(\rho) + \lambda^2 k(1-\lambda)Y_{t-2}(\rho) + \lambda k(1-\lambda)Y_{t-1}(\rho) + k(1-\lambda)Y_t(\rho)
\]

\[
\eta_t(\rho) = \lambda^t \eta_0(\rho) + k(1-\lambda) [Y_t(\rho) + \lambda Y_{t-1}(\rho) + \cdots + \lambda^{t-1}Y_1(\rho)]
\]

\[
= \lambda^t \eta_0 + k(1-\lambda) [Y_t(\rho) + \lambda Y_{t-1}(\rho) + \cdots + \lambda^{t-1}Y_1(\rho)].
\]
Let $C_t(\rho) = C_t - \rho C_{t-1}$

$\therefore \eta_t = C_t - U_t$

$\therefore \eta_t(\rho) = \eta_t - \rho \eta_t(1)$

$\therefore C_t(\rho) = \eta_t(\rho) + U_t - \rho U_{t-1}$

$\therefore C_t(\rho) + E_{3t}$

$\therefore C_t(\rho) = \lambda^t \eta_0 + \lambda(1-\lambda^t) \left[ Y_t(\rho) + \lambda Y_{t-1}(\rho) + \cdots + \lambda^{t-n} Y_{t-n}(\rho) \right] + E_{3t}$

$\therefore C_t(\rho) = \lambda^t \eta_0 + \lambda(1-\lambda) \sum_{t=1}^{\infty} \left[ C_t(\rho) - \lambda^t \eta_0 - (1-\lambda) k E_t(\rho, \lambda) \right]^2$

$\therefore C_t(\rho) + E_{3t}$

This joint distribution can be integrated numerically w.r.t. $\rho$ and $\eta_0$ to yield a bivariate posterior distribution in $\lambda$ and $k$. The latter can be normalized numerically and its desired marginal posterior pdf's for $\lambda$ and $k$ can be obtained.

\[ n_{IV}: \]

$U_t = U_{t-1} + \xi_t$.

$P(C | \lambda, k, \gamma, \sigma_4) \propto \left( \frac{1}{\sigma_4} \right) \exp \left\{ -\frac{1}{2\sigma_4^2} E_4^\gamma E_4 \right\}$

$\therefore E_4 = \left[ C - \lambda C_t - \lambda(1-\lambda) Y_t - \gamma \left[ C_{t-1} - \lambda C_{t-2} - k(1-\lambda) Y_{t-1} \right] \right] + p(\lambda, k, \gamma, \sigma_4) \propto \left( \frac{1}{\sigma_4} \right) \exp \left\{ -\frac{1}{2\sigma_4^2} (E_4 | E_4) \right\} \exp \left\{ -\frac{1}{2\sigma_4^2} (E_4 | E_4) \right\}$

$\therefore E_4 = \left[ C - \lambda C_t - k(1-\lambda) Y_t - \gamma \left[ C_{t-1} - \lambda C_{t-2} - k(1-\lambda) Y_{t-1} \right] \right] + p(\lambda, k, \gamma, \sigma_4) \propto \left( \frac{1}{\sigma_4} \right) \exp \left\{ -\frac{1}{2\sigma_4^2} (E_4 | E_4) \right\}$

$\therefore E_4 = \left[ C - \lambda C_t - k(1-\lambda) Y_t - \gamma \left[ C_{t-1} - \lambda C_{t-2} - k(1-\lambda) Y_{t-1} \right] \right] + p(\lambda, k, \gamma, \sigma_4) \propto \left( \frac{1}{\sigma_4} \right) \exp \left\{ -\frac{1}{2\sigma_4^2} (E_4 | E_4) \right\}$

$\therefore E_4 = \left[ C - \lambda C_t - k(1-\lambda) Y_t - \gamma \left[ C_{t-1} - \lambda C_{t-2} - k(1-\lambda) Y_{t-1} \right] \right] + p(\lambda, k, \gamma, \sigma_4) \propto \left( \frac{1}{\sigma_4} \right) \exp \left\{ -\frac{1}{2\sigma_4^2} (E_4 | E_4) \right\}$
\[ p(k, \lambda, \sigma | x) \propto \left\{ C - \lambda C_{-1} - k(1 - \lambda) Y - \gamma (C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1}) \right\}^{-\frac{1}{2}} \times \ \left\{ C - \lambda C_{-1} - k(1 - \lambda) Y - \gamma (C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1}) \right\}^{-\frac{1}{2}} \]

And to obtain the joint posterior pdf for \( k \) and \( \lambda \) complete the square on \( Y \) and use the properties of the univariate \( t \)-distribution, to integrate out the \( Y \) term.

\[ \propto \left\{ \frac{1}{\gamma} \left[ C - \lambda C_{-1} - k(1 - \lambda) Y_{-1} - \gamma (C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1}) \right] \right\}^{2} \]

\[ \propto \left\{ \frac{1}{\gamma} \left[ C - \lambda C_{-1} - k(1 - \lambda) Y_{-1} \right] \right\}^{2} + \frac{\sigma^{2}}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right]^{2} \]

\[ = -2 \gamma \left[ C - \lambda C_{-1} - k(1 - \lambda) Y_{-1} \right] \left( C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right) \]

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

\[ \Rightarrow \]

\[ a = \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right]^{2} \]

\[ b = -2 \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \left( C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right) \]

\[ c = \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right]^{2} \]

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

\[ p \propto a^{-\frac{1}{2}} (c - \frac{b^{2}}{4a})^{-\frac{1}{2}} \]

\[ p(\lambda, x | C) \propto \left\{ \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \right\}^{-\frac{1}{2}} \times \ \left\{ \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \right\}^{-\frac{1}{2}} \]

\[ \hat{\gamma} = \left\{ \frac{\sum \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \left( C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right) \}}{\sum \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right]^{2} \} \right\} \]

\[ p(\lambda, x | C) \propto \left\{ \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \right\}^{-\frac{1}{2}} \times \ \left\{ \frac{\gamma}{\gamma} \left[ C_{-1} - \lambda C_{-2} - k(1 - \lambda) Y_{-1} \right] \right\}^{-\frac{1}{2}} \]

(as is quoted in Zellner.)
Now this pdf is bivariate in $\mathbf{X}$ and $\mathbf{Y}$ but follows no particular distributional form. Further, it is a truncated distribution, since the ranges of $\mathbf{X}$ and $\mathbf{Y}$ are each in $(0,1)$.

As bivariate numerical integration may be used to normalize $p(\lambda, k|c)$, and to enable joint posterior inferences to be made; and also the same techniques may be used to obtain $p(\lambda|c)$ and $p(k|c)$, if desired.

2. Generalizations of Distributed Lag:

Let $y_t = \beta_0 + \sum \beta_i X(t-i) + \epsilon_t$.

Then:

$$y_t = \sum \beta_i X(t-i) + \epsilon_t$$

$$\lambda y_{t-1} = \sum \beta_i X(t-i-1) + \lambda \epsilon_{t-1}$$

$$y_t = \lambda y_{t-1} + X(t) + (\lambda \epsilon_{t-1} + \epsilon_t)$$

Now, suppose that $\epsilon_{t-1} = \epsilon_t = \epsilon_{t-1} - \epsilon_t$ where $\epsilon_t \sim N(0, \tau^2)$.

Then, $y_t = \lambda y_{t-1} + X(t) + \rho \left[ y_{t-1} - \lambda y_{t-1} - X(t-1) \right] + \epsilon_t$.

$$\lambda y_{t-1} - \rho (y_{t-1} - \lambda y_{t-1}) = \{ X(t) - \rho X(t-1) \} \epsilon_t + \epsilon_t$$

Suppose that our initial conditions, $y_0$ and $y_{-1}$, are given.

Then $p(\lambda, \rho, \tau, \beta, \beta) \propto \left( \frac{1}{\tau^m} \right)$.

Let $p(\lambda, \rho, \tau, \beta, y_0) \propto \left( \frac{1}{\tau^m} \right) \exp \left\{ -\frac{2t}{\tau} \left[ y - \lambda y_{-1} - \rho (y_{-1} - \lambda y_{-1}) - (x - \rho x_{-1}) \epsilon_t \right] \right\}$.

$$\exp \left\{ -\frac{2t}{\tau} \left[ W - (x - \rho x_{-1}) \epsilon_t \right] \right\}$$

Let $W = [ y - \lambda y_{-1} - \rho (y_{-1} - \lambda y_{-1}) ]$.
\[
p(\lambda, \rho, \beta | y, y_0) \propto \left[ \omega - (x - \rho x_{-1})\beta \right]' \left[ \omega - (x - \rho x_{-1})\beta \right]^{-\frac{1}{2}}.
\]

Then we use the properties of the MVT distr. to integrate out \( \beta \), having first expanded & completed the square on \( \beta \).

\[
p(\lambda, \rho, \beta | y, y_0) \propto \left[ \omega' \omega - \omega' (x - \rho x_{-1})\beta - \beta' (x - \rho x_{-1})' \omega + \beta' (x - \rho x_{-1})'(x - \rho x_{-1})\beta \right]^{-\frac{1}{2}}.
\]

Let \( H = (x - \rho x_{-1})'(x - \rho x_{-1}) \)

\[
p(\lambda, \rho, \beta | y, y_0) \propto \left[ \omega' \omega - 2\beta' (x - \rho x_{-1})' \omega + \beta' H \beta \right]^{-\frac{1}{2}}.
\]

Complete the square on \( \beta - \beta_0 \)

\[
\beta_0 = \left[ (x - \rho x_{-1})' \omega \right]' H^{-1} \left[ (x - \rho x_{-1})' \omega \right]^{-\frac{1}{2}}
\]

Now, see the earlier notes concerning the use of MVT when we have a function of this general form.

\[
p(\lambda, \rho, y, y_0) \propto |\lambda|^{-\frac{k}{2}} (c - \omega' \omega)^{-\frac{1}{2}}.
\]

Note the differences because we are now in a MVT distr. and not just a bivariate t distr.

\[
p(\lambda, \rho | y, y_0) \propto |H|^{-\frac{k}{2}} \left[ \omega' \omega - \omega' (x - \rho x_{-1}) H^{-1} (x - \rho x_{-1})' \omega \right]^{-\frac{k}{2}}
\]

\[
\propto |H|^{-\frac{k}{2}} \left[ \omega' \left[ I - (x - \rho x_{-1}) H^{-1} (x - \rho x_{-1})' \right] \omega \right]^{-\frac{k}{2}}.
\]

Then this pdf may be normalized by means of bivariate numerical integration. The same technique may be adapted to obtain the 2 marginal posterior pdf's:

\( p(\lambda | y, y_0) \) and \( p(\rho | y, y_0) \).
Now consider a second generalization of the basic distributed lag model:

\[ y_t = \alpha \frac{\theta^t}{\theta^t} x_{t-1} + \varepsilon(t) \gamma + u_t \]

\[ \Delta y_{t+1} = \alpha \frac{\theta^t}{\theta^t} x_{t-1} + \Delta z(t-1) \gamma + \mu_{t+1} \]

\[ y_t = \Delta y_{t+1} + \alpha x_t + \{ \Delta z(t) - \Delta z(t-1) \} \gamma + (\mu_t - \mu_{t+1}) \]

\[ (y - \Delta y_{t+1}) = \begin{bmatrix} x \mid (z - \Delta z(t-1)) \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + (u - \Delta u_{t+1}) \]

Now, assume that \( u_t \sim \text{IID}(0, \sigma^2) \).

\[ l(\beta, \lambda, \sigma \mid y_0) \propto \left( \frac{1}{\sigma^t} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y - \Delta y_{t+1} - W\beta) \right\}^t \]

where \( W = \begin{bmatrix} x \mid (z - \Delta z(t-1)) \end{bmatrix} \); \( \beta = [\alpha \gamma]. \)

Let \( p(\beta, \lambda, \sigma) \propto (\frac{1}{\sigma})^\lambda \); \( 0 < \sigma < \infty \); \( -\infty < \beta_i < \infty \); \( 0 < \lambda < 1 \).

\[ p(\beta, \lambda, \sigma \mid y_0) \propto \left( \frac{1}{\sigma^t} \right) \exp \left\{ -\frac{1}{2\sigma^2} (y - \Delta y_{t+1} - W\beta) \right\}^t \]

Let \( \omega = y - \Delta y_{t+1} \)

\[ p(\beta, \lambda, \sigma \mid y_0) \propto \left( \frac{1}{\sigma^t} \right) \exp \left\{ -\frac{1}{2\sigma^2} (\omega - W\beta) \right\}^t \]

\[ p(\beta, \lambda \mid y, y_0) \propto \frac{1}{\sigma^{t-1}} \left( \omega - W\beta \right)^{-T/2} \]

So, \( p(\beta \mid y, y_0) \) is the prior. So we could see the sensitivity of the conditional posterior pdf for \( \beta \) as we vary assumptions on \( \lambda \) if we wish.

Suppose that we want \( p(\lambda \mid y, y_0) \):

Complete the square on \( \beta \) in \( p(\beta, \lambda \mid y, y_0) \) and use the properties of the \( \text{NUTS} \) distribution to integrate out \( \beta \).
\[
\{ (\omega - \mathbf{W}\beta)'G^{-1}(\omega - \mathbf{W}\beta) \}^{-\frac{1}{2}} \\
= (\omega'G^{-1}\omega - \omega'G^{-1}\mathbf{W}\beta - \beta'\mathbf{W}'G^{-1}\omega + \beta'\mathbf{W}'G^{-1}\mathbf{W}\beta)^{-\frac{1}{2}}.
\]

\[
\alpha = \mathbf{W}'G^{-1}\mathbf{W} \\
b = -2\mathbf{W}'G^{-1}\omega \\
c = \omega'G^{-1}\omega.
\]

\[
p(\lambda|y, y_0) \propto |E|^{-\frac{1}{2}} |\mathbf{G}|^{-\frac{1}{2}} (c - \frac{b^2}{4a})^{-\frac{1}{2}} \\
\alpha \sim |\mathbf{G}|^{-\frac{1}{2}} |\mathbf{W}'G^{-1}\mathbf{W}|^{-\frac{1}{2}} \\
\{ \omega'G^{-1}\omega - (\mathbf{W}'G^{-1}\omega)'(\mathbf{W}'G^{-1}\omega)^{-1}(\mathbf{W}'G^{-1}\omega) \}^{-\frac{1}{2}}.
\]

\[
p(\lambda|y, y_0) \propto |\mathbf{G}|^{-\frac{1}{2}} |\mathbf{W}'G^{-1}\mathbf{W}|^{-\frac{1}{2}} \\
\{ \omega'G^{-1}\omega - \beta'(\mathbf{W}'G^{-1}\omega)'(\mathbf{W}'G^{-1}\omega) \}^{-\frac{1}{2}}.
\]

as is required.

And this is indeed a fact of \( \lambda \), since \( \omega = y - \lambda y_0 \).

And we could also get \( p(\beta|y, y_0) \) or \( p(\beta_0|y, y_0) \) if we so desire.
The Multivariate Regression Model:

\[ \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{11} \\ \vdots \\ \mathbf{y}_{1m} \\ \vdots \\ \mathbf{y}_{nm} \end{pmatrix} \]

12. \text{n observations on m equations.}

i.e. \[ \mathbf{Y} = \mathbf{XB} + \mathbf{U} \]

where \( \mathbf{X} \) is \((nxk)\) and \( \mathbf{Y} \) rank \( k \),

\[ \mathbf{B} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{1m} \\ \vdots \\ \beta_{km} \end{pmatrix} \]

13. \( k \) regressors available per equation — \( k \) fixed and the same \( k \) variables in each equation.

\[ \mathbf{U} = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{11} \\ \vdots \\ \mathbf{u}_{1m} \\ \vdots \\ \mathbf{u}_{nm} \end{pmatrix} \]

12. \( n \) observations or each \( j \) \( m \) disturbances.

Let the rows of \( \mathbf{U} \) be independently distributed.

Then there is no time dependence and no serial correlation.

Let each row be distributed \( \mathcal{N}(0, \mathbf{Z}) \), where \( \mathbf{Z} \) is an \((nm \times nm)\) covariance matrix.

Then,

\[ p(\mathbf{Y} | \mathbf{X}, \mathbf{B}, \mathbf{Z}) \propto \mathbf{Z}^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{U}'(\mathbf{Y}-\mathbf{XB})'\mathbf{U} \right\} \]

But, \((\mathbf{Y}-\mathbf{XB})' (\mathbf{Y}-\mathbf{XB})\)

\[ = (\mathbf{Y}-\mathbf{XB})' (\mathbf{Y}-\mathbf{XB}) + (\mathbf{B}-\mathbf{\hat{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B}-\mathbf{\hat{B}}) \]

\[ = \mathbf{S} + (\mathbf{B}-\mathbf{\hat{B}})' \mathbf{X}' \mathbf{X} (\mathbf{B}-\mathbf{\hat{B}}) \]

where \( \mathbf{S} = (\mathbf{Y}-\mathbf{XB})' (\mathbf{Y}-\mathbf{XB}) \)

and \( \mathbf{\hat{B}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \).
\[ K(\xi, \xi | \nu, \mu) \propto \left| zI^{-1} \right| \exp \left\{ -\frac{1}{2} tr \left[ S + (\nu - \mu)^{-1} xx' (\nu - \mu)^{-1} \right] \right\} \]

\[ \times \left| zI^{-1} \right| \exp \left\{ -\frac{1}{2} tr \left[ S - \frac{1}{2} \mu S - \frac{1}{2} \mu (\nu - \mu)^{-1} \mu S \right] \right\} \]

Now, \( z \) has \( n(m+1)/2 \) distinct elements:
\[ Z = \begin{pmatrix} 0_{nn} & \cdots & 0_{nm} \\ \vdots & \ddots & \vdots \\ 0_{mn} & \cdots & 0_{mm} \end{pmatrix} \quad \text{with} \quad Z_{ij} = (\nu \mu)_{ij} \quad \forall i \neq j. \]

And \( \delta \) is elements on the diagonal, defining \( \delta_j = \delta_{jj} \).

Any \( m \times m \) distinct elements \( \delta_j \) such that \( \delta_j = \delta_{ij} \quad \forall i \neq j \) and \( \sum_j \delta_j = m \).

Now, take a suitable prior pdf \( p(\delta) \) — then:
\[ p(\delta, \Sigma) = p(\delta) p(\Sigma) \]

9. \( p(\delta) \propto \text{const.} \)
\[ p(\delta) \propto \left| \Sigma \right| \left| \delta \right|^{-m(m+1)/2} \]

- \( p(\delta, \Sigma | \delta) \propto \left| \Sigma \right|^{-m(m+1)/2} \)

Note that if \( m=1 \), then \( p(\Sigma) \propto \left| \Sigma \right|^{-\frac{1}{2}} \) reduces to \( p(\delta) \propto \left( \delta_{jj} \right)^{-\frac{1}{2}} \), as is usual for the Jeffreys' prior.

Let \( \delta_{mm} = \left( \mu \nu \right)_{mm} \), element \( \delta_{mm} \).
Then transform \( \Sigma \rightarrow \Sigma^{-1} \):

\[ (\delta_{11}, \delta_{22}, \ldots, \delta_{mm}) \rightarrow (\delta_{11}, \delta_{22}, \ldots, \delta_{mm}), \]

which is a transformation involving \( m(m+1)/2 \) variables.

Then \[ J = \left| \frac{\gamma(\delta_{11}, \delta_{22}, \ldots, \delta_{mm})}{\gamma'(\delta_{11}, \delta_{22}, \ldots, \delta_{mm})} \right| = \left| \Sigma \right|^{m(m+1)/2} \]

\( K \) is the full log-likelihood ratio from Anderson's book.
\[ \begin{aligned}
&\mathcal{L}(e, \varepsilon) \propto |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \\
&\quad \times \frac{1}{|\varepsilon|^{-m/2}} \\
&\quad \times |\varepsilon|^{-m/2}.
\end{aligned} \]

As the prior is invarient to the transformation of \( \varepsilon \sim \varepsilon' \).

There is an alternative way of arriving at the prior probability density, \( p(\varepsilon^2) \propto |\varepsilon'-1|^{-2m+2} \).

Taking the (alternative) constant prior pdf \( \pi(x) \) for \( \varepsilon \), tend to zero.

Let \( p(\varepsilon^2) \propto |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| (e-\varepsilon)^2 \| \right\} \)

and let \( (n-1) \to 0 \)

\[ \varepsilon \to |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \]

Now, the likelihood function with which we are working is,

\[ \mathcal{L}(e, \varepsilon) \propto |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \]

\[ \times p(e, \varepsilon) \propto |\varepsilon|^{-m/2} \]

\[ \times \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \]

**Note:**

\[ p(e, \varepsilon) = p(e | \varepsilon, y, x) p(\varepsilon | y, x) \]

\( \mathcal{L}(e, \varepsilon) \propto |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \)

\[ \times p(e, \varepsilon) \propto |\varepsilon|^{-m/2} \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \]

\[ \times \exp \left\{ -\frac{1}{2} \| e \|^2 \right\} \]
The conditional posterior for $\beta$ is normal
in $\Sigma$, while that for $\Sigma$ (i.e., the marginal
posterior for $\Sigma$) is not the same $\mathcal{W}$ on inverted
 Wishart pdf. 

Thus, $p(\beta | \Sigma, y, x) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \text{tr} (\beta - \hat{\beta}) (x' x (\Sigma^{-1} \Sigma)^{-1}) \right\}$
\begin{align*}
&\times |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - \hat{\beta})^t \Sigma \Sigma^{-1} \Sigma (x' x) (\Sigma^{-1} \Sigma)^{-1} \right\} \\
&\times |\Sigma|^{-1} \exp \left\{ -\frac{1}{2} (\Sigma - \Sigma_0)^t (\Sigma^{-1} \Sigma)^{-1} (\Sigma - \Sigma_0) \right\}
\end{align*}

This normal conditional posterior pdf has
mean $\hat{\beta}$, and c.m. $\Sigma \propto (x' x)^{-1}$. 

$[\Sigma_0, \Sigma_0^{-1} (\Sigma^{-1} \Sigma)^{-1} = (\Sigma^{-1} \Sigma)^{-1}]$

Now, what if we are interested in just one
of the equations from the system — say the first.
Then we just look at $\beta_1$.

Then, $p(\beta_1 | \Sigma, y, x) \propto \left( \frac{1}{\sigma_1^2} \right) \exp \left\{ -\frac{1}{2\sigma_1^2} (x_1 \beta_1 - y_1)^2 \right\}$

$p(\beta_1) \sim N(\beta_{1i}, (x' x)^{-1} \Sigma_{11})$.

Now look at $p(\Sigma | y, x)$ again:

$p(\Sigma | y, x) \propto |\Sigma|^{-(n+k+m)} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma S \right\}$

Let $S = \left( \begin{array}{c} \Sigma_1 \\ \Sigma_{21} \\ \Sigma_{11} \end{array} \right)$

where $\Sigma_{11} \sim (\mu) \sim \text{p.m.}$

Then, $p(\Sigma_{11} | y, x) \propto |\Sigma_{11}|^{-\frac{v-2(m-p)+1/2}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma_{11}^{-1} S_{11} \right\}$

where $V = (n-k+m+1)$.

If $p = 1$,

$p(\Sigma_{11} | y, x) \propto \left[ \frac{1}{\sigma_1^2} \right] \exp \left\{ -\frac{S_{11}}{\sigma_1} \right\}$

And this is just of the "Inverted Gamma" form.
Now, suppose that we are interested in the marginal posterior pdf for $\beta_1$, not the conditional posterior pdf.

$p(\beta_1 | y, x) \propto \int p(\beta_1, \Sigma | y, x) \, d\Sigma$

$\propto \int \left( \frac{1}{(\sigma_n)^k} \right) \exp \left\{ -\frac{1}{2\pi \sigma_n} \left[ \sum (\beta_1 - \beta_1^*)' x' x (\beta_1 - \beta_1^*) \right] \right\} \sigma_n^{-2(n+m)/2} \, d\sigma_n$

$\propto \int \left( \frac{1}{(\sigma_n)^{k+m+1}} \right) \exp \left\{ -\frac{1}{2\pi \sigma_n} \left[ \sum + (\beta_1 - \beta_1^*)' x' x (\beta_1 - \beta_1^*) \right] \right\} \, d\sigma_n$

$\propto \left[ \sum + (\beta_1 - \beta_1^*)' x' x (\beta_1 - \beta_1^*) \right]^{-1/2(n+m)}$

by virtue of the properties of the $t$-distribution.

We now have a $t$-distribution.

If $m = 1$, then we have just the one again in the system.

So, $p(\beta_1 | y, x, m = 1) \propto \left[ \sum + (\beta_1 - \beta_1^*)' x' x (\beta_1 - \beta_1^*) \right]^{-1/2}$

This corresponds to the Chap. 3 result.

Now, what about the posterior pdf for the joint $\Sigma$? (i.e., the joint marginal posterior for $\Sigma$).

$\propto \int p(B, \Sigma | y, x) \, d\Sigma$

$= \int p(B, \Sigma^{-1} | y, x) \, d\Sigma$

And, $p(B, \Sigma^{-1} | y, x) \propto |\Sigma^{-1}|^{-(n-1)/2}$

$\exp \left\{ -\frac{1}{2} t \left[ \Sigma^{-1} + (B - \beta)' x' x (B - \beta) \right] \Sigma^{-1} \right\}$
\[ p(B \mid y, x) \propto \int \frac{1}{1 + \left( \frac{y - (x \cdot B)}{x \cdot (x - B)} \right)^{1/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left( 1 + \left( \frac{y - (x \cdot B)}{x \cdot (x - B)} \right)^{1/2} \right) \right\} \, d\Sigma^{-1} \]

\[ \propto \frac{1}{1 + \left( \frac{y - (x \cdot B)}{x \cdot (x - B)} \right)^{1/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left( 1 + \left( \frac{y - (x \cdot B)}{x \cdot (x - B)} \right)^{1/2} \right) \right\} \, d\Sigma^{-1} \]

since the integral is just the normalising constant for a Wishart pdf, this constant being kept up to \( \mathbb{R} \).

To demonstrate this, we cancel from \( \mathbb{R} \) the Wishart pdf as:

\[ p(A \mid y, x) = \frac{1}{\left| A \right|^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left| \frac{1}{A} \right| \right\} \]

\[ J = -\int \frac{1}{\left| A \right|^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left| \frac{1}{A} \right| \right\} \, d\Sigma^{-1} \]

\[ \Rightarrow \quad \frac{1}{\left| A \right|^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left| \frac{1}{A} \right| \right\} \]

Then let \( A = \left( S + (B - B)^{1/2} \right) \)

and let \( \nu = (n - m - 1) \)

\[ \nu - m - 1 = -n + m + 1 - m - 1 \]

\[ \Rightarrow \quad \int \frac{1}{\left| A \right|^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left| \frac{1}{A} \right| \right\} \, d\Sigma^{-1} \]

as required.

Further, the form of \( p(B \mid y, x) \) is a generalised Student pdf. This will be the product of

\[ \frac{1}{\left| A \right|^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \log \left| \frac{1}{A} \right| \right\} \]

which is a Student pdf.
Predictive p.d.f. for \( W \)

Traditional Mult. Regression Model:

Take the M.V. E. model:

\[
Y = XB + u.
\]

Now we want a predictive p.d.f. on further observations on \( Y \). Let these observations constitute \( W \), a \((pxm)\) matrix of \( p \) additional observations on the \( m \) endogenous variables. The same generating model is assumed as previously.

Also, \( W = ZB + V \)

where \( Z \) is a \((pxk)\) matrix of \( p \) future observations on the \( k \) exogenous variables common to each equation, and \( V \) is a \((pxm)\) matrix of \( p \) future observations on the \( m \) disturbances.

Each row of \( V \) is independent with mean zero and C.M. \( Z \), the same as that for \( U \).

Then the predictive p.d.f. for \( W \) is given by:

\[
p(W|Y, X, Z) = \int \cdots \int \left( B, \Sigma^{-1}, W | Y, X, Z \right) dB d\Sigma^{-1}
\]

\[
= \int \cdots \int p(W|B, \Sigma^{-1}, Z)p(B, \Sigma^{-1}|Y, X) dB d\Sigma^{-1}
\]

\[
\text{(W1)} \quad p(W|B, \Sigma^{-1}) \propto |\Sigma^{-1}|^{\frac{p}{2}} \exp \left\{ \frac{1}{2} \text{tr} \left[ \left( W - EB \right) \left( W - EB \right)^{T} \Sigma^{-1} \right] \right\}
\]

And, \( p(B, \Sigma^{-1}|Y, X) \propto |\Sigma^{-1}|^{\left(\frac{n-m-1}{2}\right)} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ S + (B-\tilde{B})^{T} X (X-\tilde{X}) (B-\tilde{B}) \right] \Sigma^{-1} \right\} \)

\[
\therefore \quad p(W|X, Y, Z) \propto \int \cdots \int \left| \Sigma^{-1} \right|^{\frac{p}{2}} \left| \Sigma^{-1} \right|^{\left(\frac{n-m-1}{2}\right)} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ (W - EB)^{T} (W - EB) \Sigma^{-1} \right] - \frac{1}{2} \text{tr} \left[ S + (B-\tilde{B})^{T} X (X-\tilde{X}) (B-\tilde{B}) \right] \right\} dB d\Sigma^{-1}
\]

\[
= \int \cdots \int \left| \Sigma^{-1} \right|^{\left(\frac{n+m}{2}\right)} \exp \left\{ -\frac{1}{2} \text{tr} \left[ A \Sigma^{-1} \right] \right\} dB d\Sigma^{-1}
\]
\[
\begin{aligned}
\text{when } & \quad A = \frac{1}{2} (\mathbf{w} - \mathbf{e})' (\mathbf{w} - \mathbf{e}) + (\mathbf{y} - \mathbf{B} \mathbf{e})' (\mathbf{y} - \mathbf{B} \mathbf{e}) \\
\text{since } & \quad (\mathbf{y} - \mathbf{B} \mathbf{e})' (\mathbf{y} - \mathbf{B} \mathbf{e}) = \mathbf{y}' \mathbf{B}^{-1} \mathbf{y} + (\mathbf{e}' \mathbf{B}^{-1} \mathbf{e})\, .
\end{aligned}
\]

As the integrand in \( p(\mathbf{w} | \mathbf{y}, \mathbf{x}, z) \) is proportional to:
\[
| \Sigma^{-1} |^{(p-m-1)/2} \exp \left[ -\frac{1}{2} (\mathbf{v} - \mathbf{A} \mathbf{z}^{-1})' \mathbf{A} \mathbf{z}^{-1} (\mathbf{v} - \mathbf{A} \mathbf{z}^{-1}) \right]
\]
and this is in the form \( g \) of a Wishart pdf.

With \( v = (n+p-m-1) \).

\[
\int \cdots \int | \Sigma^{-1} |^{(n+p-m-1)/2} \exp \left[ -\frac{1}{2} (\mathbf{v} - \mathbf{A} \mathbf{z}^{-1})' \mathbf{A} \mathbf{z}^{-1} (\mathbf{v} - \mathbf{A} \mathbf{z}^{-1}) \right] d \Sigma^{-1}
\]

\[= \left( \frac{1}{\mathbf{A}} \right)^{(n+p)/2} \]

Now we have to integrate \( | \mathbf{A} |^{-(n+p)/2} \) w.r.t.

the elements of \( \mathbf{B} : \\
A = (\mathbf{w} - \mathbf{e})' (\mathbf{w} - \mathbf{e}) + (\mathbf{y} - \mathbf{B} \mathbf{e})' (\mathbf{y} - \mathbf{B} \mathbf{e})
\]

\[
| \mathbf{A} |^{-(n+p)/2} = | (\mathbf{w} - \mathbf{e})' (\mathbf{w} - \mathbf{e}) + (\mathbf{y} - \mathbf{B} \mathbf{e})' (\mathbf{y} - \mathbf{B} \mathbf{e}) |^{-(n+p)/2}
\]

Let \( \mathbf{M} = (\mathbf{x}' \mathbf{x} + \mathbf{z}' \mathbf{z}) \)

Let \( \mathbf{B} = \mathbf{M}^{-1} (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w}) \)

Now, \((\mathbf{y} - \mathbf{B} \mathbf{e})' (\mathbf{y} - \mathbf{B} \mathbf{e}) + (\mathbf{w} - \mathbf{B} \mathbf{e})' (\mathbf{w} - \mathbf{B} \mathbf{e}) \)

\[= (\mathbf{y}' \mathbf{y} + \mathbf{B}' \mathbf{B}' \mathbf{y} - 2 \mathbf{B}' \mathbf{y} + \mathbf{w}' \mathbf{w} + \mathbf{B}' \mathbf{B}' \mathbf{w} - 2 \mathbf{B}' \mathbf{w} + \mathbf{w}' \mathbf{w})
\]

\[= (\mathbf{y}' \mathbf{y} + \mathbf{w}' \mathbf{w} + \mathbf{B}' \mathbf{B}' \mathbf{y} + \mathbf{w}' \mathbf{w} + \mathbf{M}' \mathbf{B}^{-1} \mathbf{B}^{-1} \mathbf{y} + \mathbf{M}' \mathbf{B}^{-1} \mathbf{B}^{-1} \mathbf{w}
\]

Completing the square on \( \mathbf{B} : \\
= \mathbf{B} - (\mathbf{x}' \mathbf{x} + \mathbf{z}' \mathbf{z})^{-1} (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w})' [\mathbf{x}' \mathbf{x} + \mathbf{z}' \mathbf{z}]^{-1} (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w})
\]

\[\quad + \mathbf{y}' \mathbf{y} + \mathbf{w}' \mathbf{w} - (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w})' [\mathbf{x}' \mathbf{x} + \mathbf{z}' \mathbf{z}]^{-1} (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w})
\]

\[= (\mathbf{B} - \mathbf{B})' \mathbf{M} (\mathbf{B} - \mathbf{B}) - (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w})' \mathbf{M}^{-1} (\mathbf{x}' \mathbf{y} + \mathbf{z}' \mathbf{w}) + \mathbf{y}' \mathbf{y} + \mathbf{w}' \mathbf{w}
\]

\[= (\mathbf{B} - \mathbf{B})' \mathbf{M} (\mathbf{B} - \mathbf{B}) - \mathbf{B}' \mathbf{M} \mathbf{M}^{-1} (\mathbf{B} - \mathbf{B}) + \mathbf{y}' \mathbf{y} + \mathbf{w}' \mathbf{w}
\]

\[= (\mathbf{B} - \mathbf{B})' \mathbf{M} (\mathbf{B} - \mathbf{B}) - \mathbf{B}' \mathbf{M} \mathbf{M}^{-1} (\mathbf{B} - \mathbf{B}) + \mathbf{y}' \mathbf{y} + \mathbf{w}' \mathbf{w}
\]
\[
\begin{align*}
\mathbf{A}^{-1} &= \begin{pmatrix} (E-B)')M(C-B) - C'MC + Y'Y + w'w \\ (E-B)'M(E-B) - C'MC \end{pmatrix}^{-1/2} \\
\rho(N|\theta, \eta, \omega) &= \mathcal{N}\left(\mu, \Sigma\right) \\
\alpha &= \begin{pmatrix} Y'Y + w'w - E'MC \\ -y' + \eta \end{pmatrix}^{-1/2}.
\end{align*}
\]

Now, simply multiply this expression:

\[
(Y'Y + w'w - E'MC)
\]

\[
= (Y'Y + w'w) - (Y'X + w'z)'M^{-1}(X'Y + z'w)
\]

\[
= Y'\left[ I - XM^{-1}X' \right] Y + w'\left[ I - ZeM^{-1}z' \right] w - Y'X M^{-1}z'w
\]

\[
- w'ZeM^{-1}X'Y
\]

Now complete the square on \(w'w\):

\[
\begin{align*}
\left[ w - (I - XM^{-1}X')^{-1} XM^{-1}X'Y \right]' \left[ w - (I - XM^{-1}X')^{-1} XM^{-1}X'Y \right] \\
+ (I - XM^{-1}X')^{-1} XM^{-1}X'Y w' \left[ I - XM^{-1}X' \right]^{-1} XM^{-1}X'Y
\end{align*}
\]

Let \(C = (I - XM^{-1}X')^{-1} XM^{-1}X'Y\)

\[
\Rightarrow \left[ w - C^{-1} XM^{-1}X'Y \right]' C \left[ w - C^{-1} XM^{-1}X'Y \right] \\
+ \left[ I - XM^{-1}X' \right]^{-1} XM^{-1}X'Y w' \left[ I - XM^{-1}X' \right]^{-1} XM^{-1}X'Y
\]

Now, \(C^{-1} = \left[ I - XM^{-1}X' \right]^{-1} = I + z(x'x)^{-1}z'\).

To show this:

\[
\begin{align*}
\left[ I - XM^{-1}X' \right] \left[ I + z(x'x)^{-1}z' \right]
\end{align*}
\]

\[
= \begin{pmatrix} M^{-1} - (x'x)^{-1}z'z \\ M^{-1} - (x'x)^{-1}z'z \end{pmatrix}
\]

\[
= \begin{pmatrix} M^{-1} - (x'x)^{-1}z'z \\ M^{-1} - (x'x)^{-1}z'z \end{pmatrix}
\]

\[
= \begin{pmatrix} (x'x)^{-1} - M^{-1}z'z \\ (x'x)^{-1} - M^{-1}z'z \end{pmatrix}
\]

\[
= \begin{pmatrix} (x'x)^{-1} - M^{-1}z'z + (z'z)(x'x)^{-1}z \end{pmatrix}
\]

\[
= \begin{pmatrix} (x'x)^{-1} - M^{-1}z'z \end{pmatrix}
\]

But, \(M = (x'x + z'z)

\[
(x'x)^{-1} - M = (x'x + z'z)
\]

\[
0 = \begin{pmatrix} 0 \end{pmatrix}
\]

\[
\therefore \left[ I \right] \cdot \left[ \begin{pmatrix} 0 \end{pmatrix} \right] = 0.
\]
So, returning to the main theme:

\[ C^{-1} \hat{z} M^{-1} = \left[ I - \hat{z} (x'x)^{-1} \hat{z} \right] \hat{z} M^{-1} \]

\[ = \hat{z} \left[ I - \hat{z} \left( (x' x)^{-1} \hat{z} \hat{z} \right) \right] M^{-1} \]

\[ = \hat{z} \left( (x' x)^{-1} (x' x) + (\hat{z} \hat{z}) \right) M^{-1} \]

\[ = \hat{z} \left( (x' x)^{-1} \right) M M^{-1} \]

\[ = \hat{z} \left( (x' x)^{-1} \right) \]

And, similarly:

\[ X M^{-1} x' + X M^{-1} \hat{z} \left( C^{-1} \hat{z} M^{-1} \right) x' \]

\[ = x \left[ \left( M^{-1} + M^{-1} \hat{z} \left( C^{-1} \hat{z} M^{-1} \right) \right) \right] x' \]

\[ = x M^{-1} \left[ I + \hat{z} C^{-1} \hat{z} \right] x' \]

\[ = x M^{-1} \left[ I + \hat{z} \right] (x' x)^{-1} x' \]

\[ = x M^{-1} \left[ x' x + \hat{z} x' \right] (x' x)^{-1} x' \]

\[ = x M^{-1} \left( (x' x)^{-1} x' \right) \]

\[ = x \left( x' x)^{-1} \right) \]

So:

\[ y'y + \omega' \omega - \hat{z} i M \hat{z} \]

\[ = y' \left[ I - x (x' x)^{-1} x' \right] y + \left( \omega - \hat{z} (x' x)^{-1} x' y \right) \]

\[ \left( \omega - \hat{z} (x' x)^{-1} x' y \right) \]

\[ \omega \]

Let \( \tilde{z} \) = \( (x' x)^{-1} x' y \)

\[ \Rightarrow y' \left[ I - x (x' x)^{-1} x' \right] y + \left( \omega - \tilde{z} \right) \left( \omega - \tilde{z} \right) \]

But, \( y' \left[ I - x (x' x)^{-1} x' \right] y = \left( y - x (x' x)^{-1} x' y \right) \)

\[ = \left( y - \tilde{z} \right) \left( y - \tilde{z} \right) \]

\[ = 0. \]

So:

\[ y'y + \omega' \omega - \tilde{z} i M \tilde{z} \]

\[ = 0 + \left( \omega - \tilde{z} \right) \left( \omega - \tilde{z} \right) \]
\[ p(x, y | Z) \propto \left( \sigma^2 \right)^{-1} \left( Z - Z' \beta \right)' \left( Z - Z' \beta \right) \]

And this is in the same form as the joint posterior pdf for \( \beta \) in the M.V.R. model:

\[ p(\beta | X, Y) \propto \left( \sigma^2 \right)^{-1} \left( X - X' \beta \right)' \left( X - X' \beta \right) \]

And this is in the "generalized" M.V.R. form. That we know that the marginal predictive pdf for any row or column of \( X \) will be M.V.R. forms.

Further, if we partition \( \alpha = (\alpha_1; \alpha_2) \), then the marginal predictive pdf for \( \alpha_1 \) will also be in the "generalized" M.V.R. form.

**The Track M.V.R. Model**

**With Exact Restrictions:**

Sometimes we will know that some elements of \( \beta \) are zero (e.g. some variables do not affect the dependent variable) or that \( 3 \) is an exact linear relation on \( \beta \). Then the joint pdf for \( \beta \) may be "conditioned" to reflect this fact into account. Then the conditioned posterior pdf may be used to make inferences about the remaining non-zero coefficients in \( \beta \).

If exact zero restrictions pertain only to one \( \beta \) vector in \( \beta \), say \( \beta_1 \), then this means that among the variables in \( x \) are not relevant to the 1st equation. If all \( \beta \) restrictions are confined to just \( \beta_1 \), then the M.V.R. posterior pdf for \( \beta_1 \) can be used subject to obtain a conditioned posterior pdf which incorporates the restrictive information.

There is no need to modify the posterior pdf's related to any other of the \( \beta \) vectors.

Let \( \beta = (\beta_1' \beta_2') \), where \( p_3 = 0 \).

Then the posterior pdf:

\[ p(\beta_1 | \beta_2 = 0) \] may be obtained. We can make use of the properties of the M.V.R. distribution to simplify these linear constraints on \( \beta_1 \). However, if the constraints relate to more than one column of \( \beta \),
Then the situation is more complex. In this case we have:

\[ Y_i = X_{i1} \beta_1 + X_{i2} \beta_2 + \ldots + X_{in} \beta_n + \epsilon_i \quad ; \quad i = 1, 2, \ldots, n. \]

where \( \beta_{i0} = 0 \). Then in this case the partitioning of \( Y \) is not the same for all equations, and this is the complicating factor.

To tackle this situation we look at the joint posterior pdf for \( \beta \), expand it, and then undifferentiate the leading term in this expansion by setting \( \beta_{i0} = 0 ; \forall i \).

Now,

\[ p(\beta | y, x) \propto | \Sigma + (B - \hat{B})^T X \hat{X} (B - \hat{B}) |^{-n/2} \]

Let \[
\begin{align*}
\Sigma &= \frac{1}{s} S \\
M &= \frac{1}{t} (X^T X)
\end{align*}
\]

\[ p(\beta | y, x) \propto | \Sigma + (B - \hat{B})^T M (B - \hat{B}) |^{-n/2} \]

Let \( H \) be \( \Sigma H^{-1} = I \)

and \( H (B - \hat{B})^T M (B - \hat{B}) H^{-1} = D \)

where \( D = \text{diag}(d_i) \) is a diagonal matrix with positive elements, and the e.c. (see above) \( B = (X^T X)^{-1} (X^T y) \).

Now, \( H^T H = I \)

\[ \Rightarrow H = (S H^{-1})^{-1} \]

\[ H^{-1} = (S H)^{-1} \]

\[ H^T H = (SH)^{-1} (S H^{-1})^{-1} \]

\[ = \left[ (SH) H^{-1} \right]^{-1} \]

\[ = S^{-1} \]

\[ \Rightarrow \Sigma + (B - \hat{B})^T M (B - \hat{B}) \]

\[ = (H^T H)^{-1} \left[ | I + H^T H (B - \hat{B})^T M (B - \hat{B}) | \right]^{n/2} \]

\[ = (H^T H)^{-n/2} | I + H^T H (B - \hat{B}) M (B - \hat{B}) |^{-n/2} \]
\[
\frac{\partial}{\partial x} f(x) = \frac{d}{dx} \left( x^2 \right) = 2x
\]

\[
\int f(x) dx = \int x^2 dx = \frac{x^3}{3} + C
\]

\[
\lim_{x \to 0} f(x) = \lim_{x \to 0} x^2 = 0
\]
\[ p(\theta_2 | y) \propto \exp \left\{ -\frac{1}{2} \left( \frac{1}{n} \text{tr} S^{-1} \theta_2 \right) \right\} \]
\[
\propto \exp \left\{ -\frac{1}{2} \text{tr} \left( S^{-1} \theta_2 \right) \right\} \exp \left\{ \frac{1}{4n} \text{tr} \left( S^{-1} \theta_2 S^{-1} \right) \right\}
\]
\[
\propto \exp \left\{ -\frac{1}{2} \text{tr} \left( S^{-1} \right) \right\}
\]
\[
\propto \exp \left\{ -\frac{1}{2} \text{tr} \left( S^{-1} \left( A - \hat{A} \right) A^{-1} \right) \right\}
\]
\[
\propto \exp \left\{ -\frac{1}{2} \left( A^{-1} \right)^{-1} \left( A - \hat{A} \right) A^{-1} \right\}
\]

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

So, $p(\theta_2 | y) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left( S^{-1} \right) \right\} \times \left\{ 1 + \frac{1}{4n} \text{tr} \left( S^{-1} \theta_2 S^{-1} \right) \right\}$

Let $\beta_0$ denote the non-zero coefficients

"$\beta_0$" "are always zero"

if $S^{-1}$ is the $(x, x')$th element of $\Sigma^{-1}$

and $\left( \hat{\beta}_{1}, \hat{\beta}_{2}, \ldots, \hat{\beta}_{n} \right) \neq \left( \beta_{1}, \beta_{2}, \ldots, \beta_{n} \right)$

i.e., this is the $(\beta_0 - \hat{\beta}_0)$, conditionalized on the zero restrictions are incorporated.

Then, $X = (x_1, x_0) \otimes (x_1, x_0)'$ as $x_1$ is $(x_1, x_2, \ldots, x_n)'$ (why?)

so $\left[ x' \otimes (x_1, x_0) \right] \otimes (x_1, x_0)'$ is a typical element of $R$ perturbed matrix.

\[ p(\theta_1 | y, \beta_0 = 0) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[ S^{-1} \left( x_1, x_0 \right)' \left( x_1, x_0 \right) \right] \right\}
\]

\[ \propto \exp \left\{ -\frac{1}{2} \left( \hat{\beta}_1 - \hat{\beta}_0 \right)' \left( x_1, x_0 \right)' \right\}
\]

\[ \exp \left\{ -\frac{1}{2} \left( \hat{\beta}_1 - \hat{\beta}_0 \right)' \left( x_1, x_0 \right)' \right\} \exp \left\{ -\frac{1}{2} \left( \hat{\beta}_1 - \hat{\beta}_0 \right)' \left( x_1, x_0 \right)' \right\}
\]

\[ \exp \left\{ -\frac{1}{2} \left( \hat{\beta}_1 - \hat{\beta}_0 \right)' \left( x_1, x_0 \right)' \right\} \propto \exp \left\{ -\frac{1}{2} \left( \hat{\beta}_1 - \hat{\beta}_0 \right)' \left( x_1, x_0 \right)' \right\}
\]

Let $V = \left( x_1, x_0 \right)' \otimes \left( x_1, x_0 \right)'$; $R = \left( x_1, x_0 \right)' \otimes \left( x_1, x_0 \right)'$

\[ p(\theta_1 | y, \beta_0) \propto \exp \left\{ -\frac{1}{2} \left( \theta_1 - \hat{\beta}_1 \right)' V^{-1} R V \right\}
\]

or

\[ V \left( \theta_1 - \hat{\beta}_1 \right)' V^{-1} R V \left( \theta_1 - \hat{\beta}_1 \right)
\]
And this is approximately a normal pdf, i.e., we have a conditioned posterior pdf with mean \((\hat{\beta}_1 + V^{-1}R'\beta_0)\) and c.m. \(\hat{\beta}\).

\[ V^{-1} = \left( X_{x_1}' X^L \bar{S} x_1 \right)^{-1}. \]
Now we wish to introduce informative prior information about \( \beta \). 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 \(X(X'X)^{-1}X'\). Then if we used a simple natural conjugate prior, we would find for example that the ratio of variances of corresponding coefficients in equations 1 and 2 would be equal. This problem is avoided if we use a general mv normal 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 readily with the LFL.

So, our prior pdf is

\[
p(\beta, \Sigma^{-1}) \propto \Sigma^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - c)(\Sigma^{-1})(\beta - c) \right\}
\]

where \( \bar{c} \) is an \((m \times 1)\) vector, the mean of the prior pdf, assigned by the analyst.

\( C \) is an \((m \times m)\) prior C.M., whose value is also assigned by the user. (All \( C \) is constant, so the leading term in \( p(\beta, \Sigma^{-1}) \) including \( C \) is absorbed into the proportionality sign.)

The usual diffuse prior is used for \( \Sigma^{-1} \) in the above expression.

Now,

\[
(\beta, \Sigma | y, x) \propto \mid \Sigma^{-1} \mid^{-N/2} \exp \left\{ -\frac{1}{2} x' \Sigma^{-1} x - \frac{1}{2} x' (\bar{c} - \bar{c})(\Sigma^{-1})x \right\}
\]

\[
\propto \mid \Sigma^{-1} \mid^{-N/2} \exp \left\{ -\frac{1}{2} x' \Sigma^{-1} x - \frac{1}{2} x' (\bar{c} - \bar{c})(\Sigma^{-1})x \right\}
\]
So,
\[ p(\beta, \Sigma | y, x) \propto |\Sigma|^{-(n-p-1)/2} \exp \left[ -\frac{1}{2} (\beta - \hat{\beta})' C^{-1} (\beta - \hat{\beta}) \right] \]
\[ \cdot \exp \left[ -\frac{1}{2} x'(\Sigma + (b - \hat{b})' x x (b - \hat{b})) x \right] \]
where \( \Sigma = (x' x)^{-1} x' y. \)

Now, integrate w.r.t. \( \Sigma \) to obtain
\[ p(\beta | y, x) \propto \exp \left[ -\frac{1}{2} (\beta - \hat{\beta})' C^{-1} (\beta - \hat{\beta}) \right] \cdot \left| \Sigma + (b - \hat{b})' x x (b - \hat{b}) \right|^{-\frac{1}{2}} \]

This is the product of a MVN pdf. and a generalized MVN pdf. This no rather messy to handle as it stands. One way of getting around the analytical problem is to expand the determinant as we did before, to normalize in the first term so obtained:
\[ \left| \Sigma + (b - \hat{b})' x x (b - \hat{b}) \right|^{-\frac{1}{2}} \]
\[ = |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \text{tr. } D - \frac{1}{2} \text{tr. } D + \frac{1}{2} \text{tr. } D^2 - \cdots \right] \]
where \( D = \frac{1}{2} H (b - \hat{b})' x x (b - \hat{b}) H' \)
\[ + \text{tr. } H' = I \quad , \quad \text{tr. } H = \frac{1}{2} S. \]
\[ \therefore \left| \Sigma + (b - \hat{b})' x x (b - \hat{b}) \right|^{-\frac{1}{2}} = \exp \left[ -\frac{1}{2} (\beta - \hat{\beta})' S^{-1} (\Sigma + (b - \hat{b})' x x (b - \hat{b})) (\beta - \hat{\beta}) \right] \]
\[ \cdot \exp \left[ -\frac{1}{2} (b - \hat{b})' (\Sigma + (b - \hat{b})' x x (b - \hat{b})) (b - \hat{b}) \right] \]
where \( F = \left( C^{-1} + S^{-1} \otimes (x' x) \right) \)
\[ \cdot \left( C^{-1} + S^{-1} \otimes (x' x) \right)^{-1} \cdot \left( C^{-1} \beta + S^{-1} \otimes x' x \beta \right) \]

Then \( b \) is the mean of the leading normal term of the expansion, \( F \) is the CM of the leading term appropriately.
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. \[ \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & \cdots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \]

$Y \times X$ is an $(n \times 1)$ vector of obs. on the $s$th dep. var.

$X_\alpha = (n \times k_\alpha)$ matrix of $k_\alpha$ indep. var.

$\beta_\alpha = (k_\alpha \times 1)$ vector of coeff.

$u_\alpha = (n \times 1)$ disturbance.

$Y = Z \theta + U$, say.

where $Z$ is the block-diagonal matrix shown above,

$E(u) = 0$

$E(u, u') = \Sigma \otimes I_n$.

And $\Sigma$ is $(m \times m)$ and pos. def. symm.

Then,

$\ell(\beta, \Sigma | y) \propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} (y - Z \beta)' \Sigma^{-1} \otimes I_n (y - Z \beta) \right]$

$\propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr} A \Sigma^{-1} \right]$

where $A = \begin{bmatrix} (y_1 - x_1 \beta_1)' (y_1 - x_1 \beta_1) & \cdots & (y_1 - x_1 \beta_1)' (y_m - x_m \beta_m) \\ \vdots & \ddots & \vdots \\ (y_m - x_m \beta_m)' (y_1 - x_1 \beta_1) & \cdots & (y_m - x_m \beta_m)' (y_m - x_m \beta_m) \end{bmatrix}$

And assume that $p(\beta, \Sigma^{-1}) = p(\beta) p(\Sigma^{-1})$. 
\[ p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} 
\]
\[ p(\beta|\Sigma^{-1}, y) \propto |\Sigma^{-1}|^{-(n-m-1)/2} \exp \left[ -\frac{1}{2} (y - \mu_0)^T \Sigma^{-1} (y - \mu_0) \right] \]

Then \( p(\beta|\Sigma^{-1}, y) \) is \( MVN \)
\[ p(\beta|\Sigma^{-1}, y) \propto |\Sigma^{-1}|^{-(n-m-1)/2} \exp \left[ -\frac{1}{2} (y - \mu_0)^T \Sigma^{-1} (y - \mu_0) \right] \]

\[ E[\beta|\Sigma^{-1}, y] = (\Sigma^{-1}(\Sigma^{-1} \otimes I_n)\Sigma^{-1})^{-1} \Sigma^{-1}(\Sigma^{-1} \otimes I_n) y = \hat{\beta} \]
\[ \text{Cov}(\beta|\Sigma^{-1}, y) = (\Sigma^{-1}(\Sigma^{-1} \otimes I_n)\Sigma^{-1})^{-1} \]

And the conditional posterior mean \( \hat{\beta} \), given \( \Sigma^{-1} \), is just the sampling theory GLS estimator, and the same applies to the CML 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 \( \beta \).

Now: 1. If all of the \( x_0 \)'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: 2. If \( \Sigma \) is diagonal, then again \( \hat{\beta} \) reduces to a vector of O.L.S. estimators.

Then \( \hat{\beta}_x = (X'X)^{-1}X'y_x \) ; \( x = 1, 2, \ldots, m \).

Now the sampling theory estimator \( \hat{\beta} \) is equivalent to the mean of the conditional posterior pdf for \( \beta \), given \( \Sigma^{-1} \). Now, in the sampling theory approach, \( \beta^* \) is independent \( \hat{\Sigma}^{-1} \), which is usually an estimate of \( \Sigma \) replaced by \( \hat{\Sigma}^{-1} \), formed from the residuals of the equations as estimated by O.L.S.

So, in the sampling theory case we get:
\[ \beta^* = (\hat{\Sigma}^{-1}(\hat{\Sigma}^{-1} \otimes I_n)\hat{\Sigma}^{-1})^{-1} \hat{\Sigma}^{-1}(\hat{\Sigma}^{-1} \otimes I_n)y \]

And note that \( \beta^* \) has the same large sample properties as \( \hat{\beta} \) in the Bayes case.

So \( \beta^* \) is what we would obtain in the Bayes case if we proceeded with \( \Sigma = \hat{\Sigma} \), in large samples.

If \( \Sigma \) is consistent, then \( \Sigma \approx \hat{\Sigma} \) are little different, so the assumption that \( \Sigma = \hat{\Sigma} \) will produce quite satisfactory results. However, in small samples,
things will not be so satisfactory. Then it is better to use the marginal posterior pdf for \( \beta \) and take its mean as the estimator of \( \beta \). Note that by adopting a Bayes technique, we have an advantage over the sampling theorist when it comes to small samples, in this particular case. We can get an exact small-sample result, whereas he is unable to do so.

Now,

\[
p(\beta, \Sigma^{-1}) \propto \frac{1}{\pi} \frac{1}{\| \Sigma^{-1} \|^{1/2}} \exp \left\{ -\frac{1}{2} \frac{1}{\Sigma} \right\}
\]

\[
p(\beta | y) \propto \int \frac{1}{\pi} \frac{1}{\| \Sigma^{-1} \|^{1/2}} \exp \left\{ -\frac{1}{2} \frac{1}{\Sigma} \right\} \Sigma^{-1} \, d \Sigma
\]

\[
\propto \frac{1}{\sqrt{\det \Sigma}}
\]

\[
p(\beta | y) \propto \left| \begin{array}{c}
(y_{1} - x_{1} \beta_{1})' \quad (y_{1} - x_{1} \beta_{1}) \\
(y_{2} - x_{2} \beta_{1})' \quad (y_{2} - x_{2} \beta_{1}) \\
\vdots \quad \vdots \\
(y_{m} - x_{m} \beta_{1})' \quad (y_{m} - x_{m} \beta_{1})
\end{array} \right|^{-1/2}
\]

Now, the marginal posterior pdf for \( \beta \) vaguely resembles a generalized MVS pdf, but in fact it cannot be brought exactly into that form, because not all of the \( X \)'s are the same. For \( p(\beta | y) \) in the form given we don't have any practical way of analyzing this pdf.

An alternative way of viewing the SIR model is to write it as a "restricted" traditional M.V.R. model, i.e.,

\[
(y_{1}, y_{2}, \ldots, y_{m}) = (X, X_{2} \ldots X_{m}) \begin{pmatrix} \beta_{1} \\ \beta_{m} \end{pmatrix} + (u_{1}, \ldots, u_{m})
\]

Then the (zero) restrictions on the coefficient matrix appear quite explicitly. If \((X, \ldots, X_{m})\) 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:

\[ Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mm} \end{pmatrix} \]

14. n observations on m equations.

\[ Y = XB + U \]

where \( X \) is \((nxk)\) and \( J \) rank \( k \),

\[ B = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1m} \\ \vdots & \ddots & \vdots \\ \beta_{m1} & \cdots & \beta_{mm} \end{pmatrix} \]

15. \( k \) regressors available per equation — \( k \) fixed and the same \( k \) variables in each equation.

\[ U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nn} \end{pmatrix} \]

16. \( n \) observations on each \( j \) in disturbance.

Let the rows of \( U \) be independently distributed. The rows are uncorrelated and so no serial correlation.

Let each row be distributed \( N(0, \Sigma) \), where \( \Sigma \) is an \((nmxm)\) covariance matrix.

Then,

\[ p(Y | X, B, \Sigma) \propto |2\pi \Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} tr [ (Y - XB)' (Y - XB) \Sigma^{-1}] \right\} \]

but,

\[ (Y - XB)' (Y - XB) \]

\[ = (Y - \hat{Y})' (Y - \hat{Y}) + (B - \hat{B})' X' X (B - \hat{B}) \]

\[ = S + (B - \hat{B})' X' X (B - \hat{B}) \]

where \( S = (Y - \hat{Y})' (Y - \hat{Y}) \)

and \( \hat{B} = (X' X)^{-1} X' Y \).
\[ L_1(\delta; \Sigma; Y, X) = | \Sigma |^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ \Sigma + (\delta - \bar{\delta}) \right] X' X (\delta - \bar{\delta}) \Sigma^{-1} \right\} \]

Now, \( \Sigma \) has \( n (n+1)/2 \) distinct elements -

\[ \Sigma = \begin{pmatrix} \sum_{ii} & \cdots & \sum_{in} \\ \vdots & \ddots & \vdots \\ \sum_{mi} & \cdots & \sum_{mm} \end{pmatrix} \quad \text{and} \quad \sum_{ij} = \sum_{ji} \quad \forall i \neq j. \]

And \( \delta \) in elements on the diagonal. Denoting

\[ \delta_i = \delta + m \cdot \eta_i \]

\[ \delta_i \text{ as distinct elements: } = \sum_i \sum_{i-1} \cdots (m-1) \cdot \eta_i \]

\[ = \sum_{i} \sum_{i} \cdot \eta_i = \sum_{i} \eta_i. \]

Now, replace a diagonal prior pdf - then

\[ p(\delta, \Sigma) = p(\delta) p(\Sigma) \]

9 \( p(\delta) \propto \text{const.} \)

\[ p(\Sigma) \propto | \Sigma |^{-(n+1)/2} \]

\[ \therefore p(\delta, \Sigma) \propto | \Sigma |^{-(n+1)/2}. \]

Note that if \( n = 1 \), then \( p(\Sigma) \propto | \Sigma |^{-(n+1)/2} \)

reduces to \( p(\delta_{11}) \propto (\delta_{11})^{-1} \), as is usual for the Jeffreys' diffuse prior.

Let \( \hat{\delta}^* = (\hat{\mu}, \hat{\sigma}) \), element \( \hat{\mu}^* \).

Then transform \( \Sigma \rightarrow \Sigma^{-1} \).

\[ (\Sigma_{11}, \Sigma_{12}, \ldots, \Sigma_{1n}) \rightarrow (\Sigma_{11}, \Sigma_{11}, \ldots, \Sigma_{1n}). \]

Which is a transformation involving \( \sum_{i} \eta_i \)

variables:

\[ \text{Then } J = \left| \frac{\chi(\Sigma_{11}, \Sigma_{12}, \ldots, \Sigma_{1n})}{\chi(\Sigma_{11}, \Sigma_{11}, \ldots, \Sigma_{1n})} \right| = | \Sigma |_{n+1} \]

\( \chi \) for this will be separate not clear from Anderson's text.
\[ p(\beta, \Sigma) \propto |\Sigma|^{-\frac{3}{2}} |\Sigma - \beta\beta^T|^{-\frac{1}{2}} \]

\[ \Rightarrow f = |\Sigma|^{m+1} \]

\[ \therefore p(\beta, \Sigma^{-1}) \propto |\Sigma|^{m+1} |\Sigma^{-1}|^{-\frac{1}{2}} \]

\[ \times |\Sigma|^{-(m+1)/2} \]

\[ \times |\Sigma^{-1}|^{-(m+1)/2} \]

So the prior is equivalent to the transformation of \( \Sigma \) to \( \Sigma^{-1} \).

There is an alternative way of arriving at the prior probability density, \( p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-\frac{1}{2}(m+1)} \).

Take our (improper) constant prior pdf \( \pi(\beta) \) and let the diff. tend to zero.

Let \( p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (m+1) \Lambda \Sigma^{-1} \right] \right\} \)

\[ \Rightarrow m \rightarrow m+1 \]

\[ \therefore p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (m+1) \Lambda \Sigma^{-1} \right] \right\} \]

Now, the likelihood function with which we are working is,

\[ L(\ell, \Sigma | y, x) \propto |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (\ell - \beta)^T \Sigma^{-1} (\ell - \beta) \right] \right\} \]

\[ + p(\ell, \Sigma) \propto |\Sigma|^{-(m+1)/2} \]

\[ \therefore p(\ell, \Sigma | y, x) \propto |\Sigma|^{-(m+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (\ell - \beta)^T \Sigma^{-1} (\ell - \beta) \right] \right\} \]

Now, write \( p(\ell, \Sigma | y, x) = p(\beta | \Sigma, y, x) p(\Sigma | y, x) \)

Let \( p(\ell, \Sigma | y, x) \propto |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (\ell - \beta)^T \Sigma^{-1} (\ell - \beta) \right] \right\} \]

\[ \times p(\ell, \Sigma | y, x) \propto |\Sigma|^{-(m+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (\ell - \beta)^T \Sigma^{-1} (\ell - \beta) \right] \right\} \]
The conditional distribution of $\beta$ is normal

Now, $p(\beta | \Sigma, y, x) \propto \Sigma^{-1/2} \exp \left\{ -\frac{1}{2} tr \left( (\beta - \beta_0)' (X'X)^{-1} (\beta - \beta_0) \right) \right\} \times \Sigma_{11}^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - \beta_0)' (X'X)^{-1} (\beta - \beta_0) \right\}$

This normal conditional posterior pdf has mean $\beta_{11}$ and c.o.v. $\Sigma_{11}$.

Also, $(A \otimes C)^{-1} = (A^{-1} \otimes C^{-1})$.

Now, what if we are interested in just one of the regression from the system — say the first. Then we just look at $\beta_{11}$.

Then, $p(\beta_{11} | \Sigma, y, x) \propto \left( \frac{1}{\sigma_{11}^{(n+1)}} \right) \exp \left\{ -\frac{1}{2} \sigma_{11}^{(n)} (\beta_{11} - \beta_{11})' (X'X)^{-1} (\beta_{11} - \beta_{11}) \right\}$

Here, $\beta_{11}$ is $N \left( \beta_{11}, (X'X)^{-1} \right)$.

Now look at $p(\Sigma | y, x)$ again:

$p(\Sigma | y, x) \propto \Sigma^{-\frac{n+m-1}{2}} \exp \left\{ -\frac{1}{2} tr \Sigma^{-1} S \right\}$

Let $\Sigma = \left( \begin{array}{c} \Sigma_{11} \\
\Sigma_{21} \\
\end{array} \right)$

Let $\Sigma_{11} = (p \exp) \otimes p^{-1}$.

Then, $p(\Sigma_{11} | y, x) \propto \left| \Sigma_{11} \right|^{-\frac{(n+m+1)}{2}} \exp \left\{ -\frac{1}{2} tr \Sigma_{11}^{-1} S_{11} \right\}$

where $S = (n-k+m+1)$.

If $p = 1$,

$p(\Sigma_{11} | y, x) \propto \left( \frac{1}{\sigma_{11}^{(n-k+m+1)}} \right) \exp \left\{ -\frac{S_{11}}{2\sigma_{11}^{(n-k+m+1)}} \right\}$

And this is just the "inverted Gamma" form.
Now suppose that we are interested in the marginal posterior pdf for $\beta_1$, not the
conditional posterior pdf.

\[ p(\beta_1 \mid y, x) \]

depends on $x$, but not on any of the other elements of $x$.

\[ p(\beta_1 \mid y, x) = \int p(\beta_1 \mid x, y, x) p(\sigma_2 \mid y, x) \, d\sigma_2 \]

\[ \alpha \int \left( \frac{1}{\sigma_1^{n-1}} \right) \exp \left\{ -\frac{1}{2\sigma_1} (\beta_1 - \hat{\beta}_1)' x' x (\beta_1 - \hat{\beta}_1) \right\} \]

\[ \left( \frac{1}{\sigma_1^{n-1}} \right) \exp \left\{ -\frac{\sigma_2}{2\sigma_2} \right\} \, d\sigma_2 \]

\[ \sigma_2 = \sigma_1 (n-1) \]

\[ p(\beta_1 \mid y, x) = \int \left( \frac{1}{\sigma_1^{n-m+1}} \right) \exp \left\{ -\frac{1}{2\sigma_1} (\sigma_1 + (\beta_1 - \hat{\beta}_1)' x' x (\beta_1 - \hat{\beta}_1)) \right\} \] \[ \alpha \left[ \sigma_1 + (\beta_1 - \hat{\beta}_1)' x' x (\beta_1 - \hat{\beta}_1) \right]^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (n-1) \right\} \]

by virtue of the properties of the $t$-density.

we now have a $M.V. \chi^2$ and $t$-density.

By $m = 1$, then we have just the one again in the system.

\[ p(\beta_1 \mid y, x, m = 1) \propto \left[ \sigma_1 + (\beta_1 - \hat{\beta}_1)' x' x (\beta_1 - \hat{\beta}_1) \right]^{-\frac{n}{2}} \]

This corresponds to the Chap. 8 result.

Now what about the posterior pdf for the joint $E$? (i.e., the joint marginal posterior for $a$).

\[ p(E \mid y, x) = \int p(E, \Sigma \mid y, x) \, d\Sigma \]

\[ = \int p(E, \Sigma \mid y, x) \, d\Sigma \]

\[ p(E, \Sigma \mid y, x) \propto |\Sigma|^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2} \Sigma + (E - \hat{E})' (E - \hat{E}) \right\} \Sigma^{-1} \]

\[ = \int p(E, \Sigma \mid y, x) \, d\Sigma \]

\[ p(E, \Sigma \mid y, x) \propto |\Sigma|^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2} \Sigma + (E - \hat{E})' (E - \hat{E}) \right\} \Sigma^{-1} \]
\[ p(B | y, x) \propto \frac{\Gamma \left( \frac{n+m}{2} \right)}{\Gamma \left( \frac{n}{2} \right) \Gamma \left( \frac{m}{2} \right) \left| B \right| \left( \frac{n+m}{2} \right)} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ S + (B-B')^\top X'X(B-B') \right] \right\} \]

\[ \times \frac{1}{\left| S + (B-B')^\top X'X(B-B') \right|^{n/2}} \]

\[ \times \int \frac{\exp \left\{ -\frac{1}{2} \text{tr} \left[ S + (B-B')^\top X'X(B-B') \right] \right\}}{\left| S + (B-B')^\top X'X(B-B') \right|^{n/2}} d\Sigma^{-1} \]

since the integral is just the normalizing constant for a Wishart pdf, this constant being denoted \( \mathbb{B} \).

To demonstrate this, we need to find the Wishart pdf of \( B \).

\[ p(B | x, y) = \frac{\mid A \mid^{(n+m)/2}}{\mid X \mid^{n/2} \mid Y \mid^{m/2}} \cdot \exp \left\{ -\frac{1}{2} \text{tr} \left[ A^{-1} B \right] \right\} \]

\[ f_0 = \frac{1}{\mid A \mid^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \text{tr} \left[ A^{-1} B \right] \right\} \]

\[ = \frac{1}{\mid A \mid^{(n+m)/2}} \cdot \exp \left\{ -\frac{1}{2} \text{tr} \left[ A^{-1} B \right] \right\} \]

Then let \( A = \left( S + (B-B')^\top X'X(B-B') \right) \)

and let \( U = -(n-m) \)

\[ V = (m-1) \]

\[ \therefore \quad V+m-1 = -(n+m-1) \]

\[ = \quad -\lambda \]

\[ \therefore \quad \int \frac{1}{\mid A \mid^{-(n+m)/2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ A^{-1} B \right] \right\} \]

as required.

Further, the form of \( p(B | y, x) \) is a generalized Wishart pdf. This will be the product of a Wishart pdf with a Student pdf.
Traditional Multi. Regression Model:

Take the M.V. R. model:

\[ Y = X\beta + \epsilon. \]

Now we want a predictive p.d.f. on further observations on \( Y \). Let these observations constitute \( W \), a (pxm) matrix of \( p \) additional observations on the \( m \) endogenous variables. The same generating model is assumed as previously.

Also, \[ W = Z\beta + v \]

where \( Z \) is a (pxm) matrix of \( p \) future observations on the \( m \) regressors variables common to each equation, and \( V \) is a (pxm) matrix of \( p \) future observations on the \( m \) disturbances.

Each row of \( V \) is independent, with mean zero and C.M. \( E \), the same as that for \( U \).

Then the predictive p.d.f. for \( W \) is given by:

\[ p(W | Y, X, Z) = \int \cdots \int p(B, \Sigma^{-1}, W | Y, X, Z) \, dB \, d\Sigma^{-1} \]

\[ = \int \cdots \int p(W | B, \Sigma^{-1}, Z) p(B, \Sigma^{-1} | Y, X) \, dB \, d\Sigma^{-1} \]

\[ \text{Hence, } p(W | B, \Sigma^{-1}) \propto |\Sigma^{-1}|^{p/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (W - Z\beta)(W - Z\beta)^\top \Sigma^{-1} \right] \right\} \]

And, \( p(B, \Sigma^{-1} | Y, X) \propto |\Sigma^{-1}|^{(m-1)/2} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ S + (B - \tilde{B})'X'(X - \tilde{X}) \Sigma^{-1} \right] \right\} \]

\[ \therefore p(W | X, Y, Z) \propto \int \cdots \int |\Sigma^{-1}|^{p/2} \times |\Sigma^{-1}|^{(m-1)/2} \times \]

\[ \int \cdots \int (S + (B - \tilde{B})'X'(X - \tilde{X}) \Sigma^{-1} \int dB \, d\Sigma^{-1} \]

\[ \times \int \cdots \int |\Sigma^{-1}|^{(m+p-n-1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[ A \Sigma^{-1} \right] \right\} \, dB \, d\Sigma^{-1} \]
\[
\begin{align*}
\text{where} \quad A &= \left( (w-z \mathbf{B})^\dagger (w-z \mathbf{B}) + (y-x \mathbf{B})^\dagger (y-x \mathbf{B}) \right) \\
\text{and} \quad (y-x \mathbf{B})^\dagger (y-x \mathbf{B}) &= \mathbf{A} + (\mathbf{B}-\mathbf{B})^\dagger \mathbf{X}' \mathbf{X} (\mathbf{B}-\mathbf{B}).
\end{align*}
\]

As, the integrand in \( p(x,y,z) \) is proportional to:
\[
\left| \Sigma \right|^{-\frac{p-m-1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{X} \Sigma^{-1} \mathbf{X}' \right\}
\]

And this is in the form \( \mathcal{J} \) in a Wishart pdf.
with \( \Sigma^{-1} = (n+p-m-1) \).
\[
\int \ldots \int \left| \Sigma \right|^{-\frac{p-m-1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{X} \Sigma^{-1} \mathbf{X}' \right\} \, d\Sigma
\]

\[\alpha \left| A \right|^{-\frac{(n+p)}{2}}\]

Now we have to integrate \( |A|^{-\frac{(n+p)}{2}} \) w.r.t.
the elements \( \mathcal{B} \):
\[
A = (w-z \mathbf{B})^\dagger (w-z \mathbf{B}) + (y-x \mathbf{B})^\dagger (y-x \mathbf{B})
\]
\[
|A|^{-\frac{(n+p)}{2}} = \left| (w-z \mathbf{B})^\dagger (w-z \mathbf{B}) + (y-x \mathbf{B})^\dagger (y-x \mathbf{B}) \right|^{-\frac{(n+p)}{2}}.
\]

Let \( M = (x'y + z'w) \)

Let \( \mathcal{B} = M^{-1} (x'y + z'w) \)

Now, \((y-x \mathbf{B})^\dagger (y-x \mathbf{B}) + (w-z \mathbf{B})^\dagger (w-z \mathbf{B})\)
\[= y'y + b'x'y - b'x'w + w'w + b'z'w - b'z'w\]
\[= y'y + w'w + b' \left[ x'y + z'w \right] \mathbf{M} - M^{-1} \mathbf{M} \left[ x'y + z'w \right]\]

Completing the square on \( \mathcal{B} \):
\[
\begin{align*}
&= \left[ b - (x'y + z'w)^{-1} (x'y + z'w) \right]^{-1} \left[ x'y + z'w \right] \left[ b - (x'y + z'w)^{-1} \right] \\
&= \left[ b - (x'y + z'w)^{-1} (x'y + z'w) \right]^{-1} \left[ x'y + z'w \right] \\
&= (b - \mathbf{M})^{-1} (b - \mathbf{M}) - (x'y + z'w)^{-1} (x'y + z'w) + y'y + w'w\]
\end{align*}
\]
\[ \begin{align*}
  |A|^{-\frac{\nu(\nu-1)}{2}} &= |(B-B')M(C-C') - C'MB + \gamma'Y + w\omega|^{-\frac{\nu(\nu-1)}{2}} \\
  \rho(A | X, Y, Z) &\propto \int \cdots \int Y'Y + w'w + (B-B')M(B-B') - C'MB \, dB \, dX \, dY \, dZ
\end{align*} \]

Now simplify this expression:

\[
(Y'Y + w'w - C'MB)
\]

\[
= (Y'Y + w'w) - (M(B-B')M)^{-1}(X'Y + w'w)
\]

\[
= Y'\left[ I - XM^{-1}X' \right] Y + w'[I - 2M^{-1}z']w - Y'XM^{-1}z'w - w'z'M^{-1}X'Y
\]

Now complete the square on \( w'w \):

\[
= [w - (X'XM^{-1}X')^{-1}2M^{-1}X'Y]'\left[ I - (X'XM^{-1}X')^{-1}2M^{-1}X'Y \right] Y
\]

\[
+ Y'[I - XM^{-1}X'] - X'M^{-1}z'c XM^{-1}X'Y
\]

Let \( C = (I - 2M^{-1}z') \)

\[
\Rightarrow \left[ w - c't \right] 2M^{-1}X'Y \right] C \left[ w - c't \right] 2M^{-1}X'Y
\]

\[
+ Y'[I - XM^{-1}X'] - XM^{-1}z'c XM^{-1}X'Y
\]

Now, \( C^{-1} = [I - 2M^{-1}z']^{-1} = I + 2(z'X)^{-1}z\).

To show this:

\[
[I - 2M^{-1}z'] \left[ I + 2(z'X)^{-1}z \right]
\]

\[
= I + 2M^{-1} - (z'X)^{-1}z = I - 2M^{-1}z' + 2(z'X)^{-1}z
\]

\[
= I - 2M^{-1} \left[ I + M(z'X)^{-1} + 2z'(z'X)^{-1}z \right]
\]

\[
= I - 2M^{-1} \left[ (z'X)^{-1}(I + Mz + 2z'z) \right]
\]

But, \( M = (z'X + 2z'z) \)

\[
(z'X)^{-1} - M + 2z'z = 0
\]

\[
\Rightarrow [I - 2M^{-1}z']^{-1} = I.
\]
So, returning to the main theme:

\[
C^{-1} z M^{-1} = \left[ I + Z (x'x)^{-1} z^{2} \right] z M^{-1}
\]
\[
= \hat{z} \left[ I + (x'x)^{-1} z^{2} \right] M^{-1}
\]
\[
= \hat{z} (x'x)^{-1} \left[ (x'x) + (z^{2} e) \right] M^{-1}
\]
\[
= \hat{z} (x'x)^{-1} M M^{-1}
\]
\[
= \hat{z} (x'x)^{-1}
\]

And, further:

\[
x M^{-1} x' + x M^{-1} \hat{z} (C^{-1} z M^{-1} x
\]
\[
= x \left[ M^{-1} + M^{-1} \hat{z} C^{-1} z M^{-1} \right] x
\]
\[
= x \left[ M^{-1} + M^{-1} \hat{z} C^{-1} \right] x
\]
\[
= x M^{-1} \left[ I + \hat{z} C^{-1} \right] x
\]
\[
= x M^{-1} \left[ I + \hat{z} \right] (x'x)^{-1} x
\]
\[
= x M^{-1} \left[ x'x + \hat{z} \right] (x'x)^{-1} x
\]
\[
= x (x'x)^{-1} M (x'x)^{-1} x
\]
\[
= x (x'x)^{-1} x
\]

So:

\[
y'y + \omega'\omega = \hat{z} M \hat{z}
\]
\[
= y' \left[ I - x (x'x)^{-1} x' \right] y + \left( \omega - \hat{z} (x'x)^{-1} x'y \right)' C \left( \omega - \hat{z} (x'x)^{-1} x'y \right).
\]

Let \( \hat{z} = (x'x)^{-1} x'y \)

\[
= y' \left[ I - x (x'x)^{-1} x' \right] y + \left( \omega - \hat{z} \right)' C \left( \omega - \hat{z} \right)
\]

But, \( y' \left[ I - x (x'x)^{-1} x' \right] y = (y - x (x'x)^{-1} x'y)' (y - x (x'x)^{-1} x'y) = (y - x \hat{z})' (y - x \hat{z}) = \hat{S} \).

So:

\[
y'y + \omega'\omega = \hat{z} M \hat{z}
\]
\[
= \hat{S} + \left( \omega - \hat{z} \right)' C \left( \omega - \hat{z} \right).\]
\[ p(z|x, y, z) \propto \frac{1}{S + (U - Standardize)x(x - Standardize)} \]

And this is in the same form as the joint posterior pdf for \( B \) in the M.V.R. model:

Recall:

\[ p(B|x, y) \propto \frac{1}{S + (C - \hat{C})' (x(X)(C - \hat{C}))^{-1} \hat{B}} \]

And this is the "generalized" M.V.S.-t distribution we termed, such that the marginal predictive pdf for any row or column of \( B \) will be M.V.S.-t.

Further, if we condition on \( z = (x^T, w^T) \), then the marginal predictive pdf for \( z \) will also be in the "generalized" M.V.S.-t. form.

(c). The Final M.V.R. Model

With Exact Restrictions:

Sometimes we want to handle some elements of \( B \) and zero (e.g., some variables do not affect one of the dependent variables) in such an exact linear restriction on \( B \). Then the joint pdf for \( B \) may be "conditioned" to take this fact into account. Thus the conditioned joint posterior pdf may be used to make inferences about the remaining non-zero coefficients in \( B \).

If exact zero restrictions pertain only to one \( \beta \) vector in \( B \), say \( \beta_1 \), then this means that some of the variables in \( X \) are not relevant to the 1st equation. If all of the restrictions are confined to just \( \beta_1 \), then the joint posterior pdf for \( \beta_1 \) can be undated to obtain a conditioned posterior pdf which incorporates the restrictive information. There is no need to modify the posterior pdf's related to any other of the \( \beta \)'s.

Let the \( \beta \) and \( \beta' = (\beta_1', \beta_1') \), where \( \beta_1 = 0 \).

Then the posterior pdf:

\[ p(\beta_1 | \beta_2, \beta_3 = 0) \] may be obtained. We can make use of the properties of the M.V.S. distribution to update these linear constraints on \( \beta_1 \). However, if the constraints relate to more than one column of \( \hat{B} \),
Then the situation is more complex. In this case we have:

\[ y_{\chi} = \tilde{X}_{\chi} \tilde{B}_{\chi} + x_{\chi} \tilde{B}_{\chi} + \tilde{u}_{\chi} \quad ; \chi = 1, 2, \ldots, m. \]

where \( x_{\chi} = 0 \). Then in this case the partitioning of \( x \) is not the same for all equations, and this is the complicating factor.

To tackle this situation we look at the joint posterior pdf for \( \beta \), expand it, and then conditionalize the leading term in this expansion by setting \( x_{\chi} = 0 \); \( y_{\chi} \).

Now,

\[ p(y, x) \propto \frac{1}{s + (E - \hat{\beta})'M(E - \hat{\beta})}^{\frac{1}{2}n} \]

Let \( S = \frac{1}{2} s \)

\[ M = \frac{1}{2} \tilde{X}' \tilde{X} \]

\[ p(y, x) \propto \frac{1}{S + (E - \hat{\beta})'M(E - \hat{\beta})}^{\frac{1}{2}n} \]

Let \( H \) such that \( H \tilde{S} H' = I \)

and \( H(E - \hat{\beta})'M(E - \hat{\beta})H' = \Sigma \)

where \( \Sigma = \Sigma(x) \) is a diagonal matrix with positive elements and the c.r. (i.e. Aickin) \( \hat{\beta}'M(E - \hat{\beta}) \).

Now,

\[ H \tilde{S} H' = I \]

\[ \Rightarrow H = (\tilde{S} H')^{-1} \]

\[ H' = (H \tilde{S})^{-1} \]

\[ H' H = (H \tilde{S})^{-1}(\tilde{S} H')^{-1} \]

\[ = \left[ (H \tilde{S})^{-1} H' + \tilde{S}^{-1} \right]^{-1} \]

\[ = \tilde{S}^{-1} \]

\[ s + (E - \hat{\beta})'M(E - \hat{\beta}) \]

\[ = (H' H)^{-1/2} \left[ \frac{1}{s} + H' (E - \hat{\beta})' M (E - \hat{\beta}) \right] \]

\[ = (H' H)^{-1/2} \frac{1}{s} + H' (E - \hat{\beta})' M (E - \hat{\beta}) \]

\[ = (H' H)^{-1/2} \frac{1}{s} + H' H (E - \hat{\beta})' M (E - \hat{\beta}) \]

\[ = (H' H)^{-1/2} \frac{1}{s} + H' H (E - \hat{\beta})' M (E - \hat{\beta}) \]
\[ \begin{aligned}
&= |\Sigma|^{-n/2} |I + H (B - \bar{B})' M (B - \bar{B}) H|^{-n/2} \\
&= |\Sigma|^{-n/2} \exp \left\{ \log |I + D|^{-n/2} \right\} \\
&= |\Sigma|^{-n/2} \exp \left\{ -\frac{n}{2} \sum \log (1 + z_i) \right\} \\
&= |\Sigma|^{-n/2} \exp \left\{ -\frac{n}{2} \sum \log (1 + \lambda_i) \right\}
\end{aligned} \]

(since \( D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \) and \( (I + D) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \).

\[ \Rightarrow \begin{aligned}
|\Sigma|^{-n/2} \exp \left\{ + \frac{n}{2} \left[ \sum (\lambda_1 - \frac{1}{2} \lambda_1^2 + \frac{1}{3} \lambda_1^3 - \cdots) \right] \right\} \\
&= |\Sigma|^{-n/2} \exp \left\{ + \frac{n}{2} \left[ \sum \lambda_i - \frac{1}{2} \sum \lambda_i^2 + \frac{1}{3} \sum \lambda_i^3 - \cdots \right] \right\} \\
&= |\Sigma|^{-n/2} \exp \left\{ + \frac{n}{2} \left[ \text{tr. } D - \frac{1}{2} \text{tr. } D^2 + \frac{1}{3} \text{tr. } D^3 - \cdots \right] \right\}
\end{aligned} \]

But \( (\text{tr. } D) = \text{tr. } H (B - \bar{B})' M (B - \bar{B}) H \)

Also \( \overline{\Sigma} = (B - \bar{B})' M (B - \bar{B}) \)

\[ \begin{aligned}
\text{tr. } D &= \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \\
\text{tr. } D^2 &= \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \\
\text{tr. } D^3 &= \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \text{tr. } \overline{\Sigma}^{-1} \overline{\Sigma} \\
\end{aligned} \]

\[ \rho(\overline{\Sigma}) \propto \left| \overline{\Sigma} + (B - \bar{B})' M (B - \bar{B}) \right|^{-n/2} \]

\[ \begin{aligned}
\text{But } D &= \overline{\Sigma}^{-1} \\
\overline{\Sigma} &= \overline{\Sigma}^{-1} \\
&= \overline{\Sigma}^{-1} \\
&\quad \text{by definition}.
\end{aligned} \]

\[ \begin{aligned}
\rho(\overline{\Sigma}) \propto \left| \overline{\Sigma} + (B - \bar{B})' M (B - \bar{B}) \right|^{-n/2} \\
&= \left| \overline{\Sigma} + (B - \bar{B})' M (B - \bar{B}) \right|^{-n/2} \\
&= \left| \overline{\Sigma} + (B - \bar{B})' \times (B - \bar{B}) \right|^{-n/2} \\
&= \left| \overline{\Sigma} + (B - \bar{B}) \right|^{-n/2} \\
&= \left| \overline{\Sigma} \right|^{-n/2} \\
&= \left| \overline{\Sigma} \right|^{-n/2} \text{, say.}
\end{aligned} \]
And this is approximately a normal pdf; i.e., we have a
conditionalized posterior pdf with
mean \( \beta_{\text{post}} + V^{-1}R' \beta_{\text{prior}} \) and c.o.v. \( \Sigma \). \[ V^{-1} = (X_{\text{new}}' X_{\text{new}} - s_{xx})^{-1}. \]
Now we wish to introduce informative prior information about \( \beta \). 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 \( Z = X(X'X)^{-1} \). 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 N.M.V. pdf as our prior for all of the coeff.

in the model. But the price that we pay then is a loss in easy analysis—the prior pdf will no longer combine readily with the L.I.F.

So, our prior pdf is:

\[
p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{- \frac{mn}{2}} \exp \left\{ -\frac{1}{2} (\beta - \bar{\beta})' C^{-1} (\beta - \bar{\beta}) \right\}
\]

where \( \bar{\beta} \) is an \((m \times 1)\) vector, the mean of the prior pdf, assumed by the analyst.

\( C \) is an \((m \times m)\) prior C.M., whose value is also assigned by the user. (And \( C \) is constant, so the leading term in \( p(\beta, \Sigma^{-1}) \) involving \( C \) is absorbed into the proportionality sign.)

The usual diffuse prior is used for \( \Sigma^{-1} \) in the above expression.

Now,

\[
J(\beta, \Sigma^{-1} | y, x) \propto |\Sigma^{-1}|^{-n \frac{p}{2}} \exp \left\{ -\frac{1}{2} (y - \beta)' \Sigma^{-1} (y - \beta) - \frac{n}{2} \Sigma^{-1} \right\}
\]

\[
\propto |\Sigma^{-1}|^{-n \frac{p}{2}} \exp \left\{ -\frac{1}{2} tr \Sigma^{-1} - \frac{n}{2} (\beta - \bar{\beta})' X' X (\beta - \bar{\beta}) \Sigma^{-1} \right\}
\]
\[ p(\beta, \Sigma^{-1} | y, x) \propto |\Sigma^{-1}|^{n+m-1/2} \times \exp \left[ -\frac{1}{2} (\beta - \bar{\beta})' \Sigma^{-1} (\beta - \bar{\beta}) \right] \times \exp \left[ -\frac{1}{2} \Sigma^{-1} \left( S + (\beta - \bar{\beta})' x' x (\beta - \bar{\beta}) \right) \right] \]

where \( \bar{x} = (x' x)^{-1} x' y \).

Now, integrate w.r.t. \( \Sigma^{-1} \) and obtain:
\[ p(\beta | y, x) \propto \exp \left[ -\frac{1}{2} (\beta - \bar{\beta})' \Sigma^{-1} (\beta - \bar{\beta}) \right] \times |S + (\beta - \bar{\beta})' x' x (\beta - \bar{\beta})|^{-1/2} \]

And this is the product of a MVN pdf and a generalized MVN 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, to normalize in the first term so obtained:
\[ |S + (\beta - \bar{\beta})' x' x (\beta - \bar{\beta})|^{-1/2} \]

\[ = |S|^{-1/2} \exp \left[ -\frac{1}{2} \left[ \text{tr.} D - \frac{1}{2} \text{tr.} D^2 + \frac{1}{4} \text{tr.} D^3 - \ldots \right] \right] \]

where \( D = \frac{1}{2} (B - \bar{B})' x' x (B - \bar{B}) H' \)

\[ + \frac{1}{2} H' = I, \quad \frac{1}{2} S = \frac{1}{2} S. \]

\[ = \exp \left[ -\frac{1}{2} (\beta - \bar{\beta})' \frac{1}{\gamma} - \frac{1}{2} \alpha(\chi x') (\beta - \bar{\beta}) \right] \]

where \( F = \left( \alpha^{-1} + \frac{1}{2} \alpha(\chi x') \right) \)

\[ \bar{x} = \left( \alpha^{-1} + \frac{1}{2} \alpha(\chi x') \right)^{-1} \left( \alpha^{-1} \bar{\beta} + \frac{1}{2} \alpha(\chi x') \right) \]

Then \( \bar{x} \) is the mean of the leading normal form of the expansion, \( F \) is the CM of the leading term approximately the posterior pdf of \( \beta \).
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. 

\[
\begin{pmatrix}
Y_1 \\
\vdots \\
Y_m
\end{pmatrix} = 
\begin{pmatrix}
X_1 & \cdots & X_m
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\vdots \\
\beta_m
\end{pmatrix} + 
\begin{pmatrix}
U_1 \\
\vdots \\
U_m
\end{pmatrix}
\]

$Y$ is an $(n \times 1)$ vector of $n$ obs. in the $n$th dep. var.

$X$ is an $(n \times k)$ matrix of $k$ indep. vars.

$\beta$ is a $(k \times 1)$ vector of coeff.

$U$ is an $(n \times 1)$ vector of disturbances.

$Y = Z\beta + U$, say.

where $Z$ is the block-diagonal matrix shown above.

\[E(U) = 0\]
\[E(UU') = \Sigma \otimes I_n\]

And $\Sigma$ is $(m \times m)$ and pos. def. symm.

Then,

\[
\ell(\beta, \Sigma | y) \propto \Sigma^{-1/2} \exp \left[ -\frac{1}{2} (y - Z\beta)' \Sigma^{-1} (y - Z\beta) \right]
\]

\[
\propto \Sigma^{-1/2} \exp \left[ -\frac{1}{2} \text{tr} A \Sigma^{-1} \right]
\]

where,

\[
A = \begin{bmatrix}
(y_1 - X_1\beta_1)'(y_1 - X_1\beta_1) & \cdots & (y_1 - X_1\beta_1)'(y_m - X_m\beta_m) \\
\vdots & \ddots & \vdots \\
(y_m - X_m\beta_m)'(y_1 - X_1\beta_1) & \cdots & (y_m - X_m\beta_m)'(y_m - X_m\beta_m)
\end{bmatrix}
\]

And assume that $\varphi(\beta, \Sigma^{-1}) = \varphi(\beta) \varphi(\Sigma^{-1})$. 

\[ p(\beta, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2} \]
\[ p(\beta, \Sigma^{-1} | y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp \left[ -\frac{1}{2} tr(\Sigma^{-1} \Sigma^{-1} | y) \right] \]

Then
\[ p(\beta | \Sigma^{-1}, y) \text{ is MVN -} \]
\[ p(\beta | \Sigma^{-1}, y) \propto |\Sigma^{-1}|^{(n-m-1)/2} \exp \left[ -\frac{1}{2} (y-\Sigma^{-1} \Sigma^{-1} | y) \right] \]

\[ E [\beta | \Sigma^{-1}, y] = (\Sigma^{-1} (\Sigma^{-1} \Sigma^{-1} | y) )^{-1} \Sigma^{-1} (\Sigma^{-1} \Sigma^{-1} | y) y = \hat{\beta} \]
\[ \text{Cov.} (\beta | \Sigma^{-1}, y) = (\Sigma^{-1} (\Sigma^{-1} \Sigma^{-1} | y) )^{-1} \]

And the conditional posterior mean \( \hat{\beta} \), given
\( \Sigma^{-1} \), is just the sampling theory ELS estimator.

This the same applies to the case 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 \( \beta \).

Now: (i) If all of the \( X_i \)'s are the same, or if they are proportional to one another, then \( \beta \) reduces to a vector of O.L.S. estimators.

and/or: (ii) If \( \Sigma \) is diagonal, then again \( \beta \) reduces to a vector of O.L.S. estimators.

Then \( \hat{\beta}_x = (X'X)^{-1}X'y \) ; \( x = 1, 2, \ldots, m \).

Now the sampling theory estimator of \( \beta \) is equivalent to the mean of the conditional posterior pdf for \( \beta \), given \( \Sigma^{-1} \). Now, in the sampling theory approach, \( \beta^* \) is independent of \( \Sigma^{-1} \), which is usually an unknown \( \Sigma \) is replaced by \( \hat{\Sigma}^{-1} \), formed from
the residuals of the equations as estimated by O.L.S.

So here, in the sampling theory case we get:
\[ \hat{\beta}^* = \left[ \Sigma^{-1} (\Sigma^{-1} \Sigma^{-1} | y) \right]^{-1} \left[ \Sigma^{-1} (\Sigma^{-1} \Sigma^{-1} | y) y \right] \]

And note that \( \hat{\beta}^* \) has the same large sample
properties as has \( \hat{\beta} \) in the Bayes case.

As \( \hat{\beta}^* \) is what we would obtain in the Bayes case if we proceeded with \( \Sigma = \hat{\Sigma} \), in large samples,
if \( \hat{\Sigma} \) is consistent, then \( \Sigma \approx \hat{\Sigma} \) are little different,
so the assumption that \( \Sigma = \hat{\Sigma} \) will produce quite satisfactory results. However, in small samples,
Things will not be so satisfactory, then it is better to use the marginal posterior pdf for $\beta$ and take it as the estimator of $\beta$. Note that by adopting a Bayesian technique, we have an advantage over the sampling theory when it comes to small samples, in this particular case. We can get an exact small-sample result, whereas we is unable to do so.

Now,

$$p(\beta, \Sigma^{-1} | y) \propto \left| \Sigma^{-1} \right|^{(n-m-1)/2} \exp \left\{ -\frac{1}{2} \Sigma^{-1} \right\}$$

$$\propto p(\beta | y) \propto \int \left| \Sigma^{-1} \right|^{(n-m-1)/2} \exp \left\{ -\frac{1}{2} \Sigma^{-1} \right\} d \Sigma^{-1}$$

$$= |A|^{-1/2}$$

$$\propto (y_1 - x_1 \beta)^T (y_1 - x_1 \beta) + \cdots + (y_m - x_m \beta)^T (y_m - x_m \beta)$$

Now, the marginal posterior pdf for $\beta$ vaguely resembles a generalized inverse pdf, but in fact it cannot be brought exactly into that form, because not all of the $X_i$'s are the same. For $p(\beta | y)$ in the form shown we don't have any practical way of analyzing this pdf.

An alternative way of viewing the S.R. model is to write it as a "restricted" traditional M.V.R. model, i.e.

$$(y_1, y_2, \ldots, y_m) = (X, X_2 \ldots X_m) \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + (u_1 \ldots u_m)$$

Then the (zero) restrictions on the coefficient matrix appear quite explicitly. If $(X, \ldots, X_m)$ has full rank, then we can see the restrictions in section (C) above to handle the zero restrictions, as for the traditional model.
"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 p.d.f. $p(y|\theta = \theta_1)$

Let $w$ be a dichotomous random variable:

$$w = \begin{cases} 0 & \text{if } H_0 \text{ is true;} \\ 1 & \text{if } H_1 \text{ is true.} \end{cases}$$

Then the prior probabilities are:

$$p(H_0) = p(w = 0)$$
$$p(H_1) = p(w = 1)$$

And, $p(H_0) + p(H_1) = p(w = 0) + p(w = 1) = 1$.

Now, consider the following:

$$p(y, w) = p(y|w) p(w)$$

$$= p(w|y) p(y)$$

$$\therefore p(w|y) = \frac{p(y|w) p(w)}{p(y)}$$

And

$$p(y) = \sum_w p(y|w_1) p(w_1) = p(y|w = 0) p(w = 0) + p(y|w = 1) p(w = 1);$$
$$\neq 0 \quad \text{by assumption.}$$
\[ p(H_0 | y) = \frac{p(y | H_0) p(H_0)}{p(y)} = \frac{p(y | \theta = \theta_0) p(H_0)}{p(y)} \]

And, \[ p(H_1 | y) = \frac{p(y | H_1) p(H_1)}{p(y)} = \frac{p(y | \phi = \phi_1) p(H_1)}{p(y)} \]

Consider then the posterior odds in favour of \( H_0 \):

\[ K_{01} = \frac{p(H_0 | y)}{p(H_1 | y)} = \frac{p(y | \theta = \theta_0) p(H_0) / p(y)}{p(y | \phi = \phi_1) p(H_1) / p(y)} \]

\[ = \left[ \frac{p(H_0)}{p(H_1)} \right] \times \left[ \frac{p(y | \theta = \theta_0)}{p(y | \phi = \phi_1)} \right] \]

\[ = \left( \text{posterior odds} \right) \times \left( \text{likelihood ratio} \right) \]

\[ = \text{total (posterior odds)} = (\text{prior odds}) \times (\text{likelihood ratio}) \]

So, if we are given the prior odds and the likelihood ratio, then we can arrive at the posterior odds quite simply.

But, having been given the posterior odds, what say that we want to make use of them in decision problems? How can we make use of the posterior odds to decide on our actions with regard to the acceptance or otherwise of a given hypothesis?

In the situation we face we have 2 states: \( H_0 \) and \( H_1 \), and we have 2 actions: accept or reject \( H_0 \) (or, alternatively, accept \( H_0 \) or reject \( H_1 \)). This set-up may be depicted:

<table>
<thead>
<tr>
<th>States</th>
<th>( H_0 ) true</th>
<th>( H_1 ) true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>( L(H_0, \hat{H}_0) )</td>
<td>( L(H_0, \hat{H}_0) )</td>
</tr>
<tr>
<td>Acc ( H_0 )</td>
<td>0</td>
<td>( L(H_1, \hat{H}_1) )</td>
</tr>
<tr>
<td>Acc ( H_1 )</td>
<td>( L(H_0, \hat{H}_1) )</td>
<td>0</td>
</tr>
</tbody>
</table>

Now, consider the expected losses involved in a posterior sense (after having seen the data).
Then,
\[ E(\Lambda | \hat{H}_0) = \Lambda (H_0, \hat{H}_0) p(H_0 | y) + \Lambda (H_1, \hat{H}_0) p(H_1 | y) \]
\[ = 0 + \Lambda (H_1, \hat{H}_0) p(H_1 | y). \]

And,
\[ E(\Lambda | \hat{H}_1) = \Lambda (H_1, \hat{H}_1) p(H_1 | y) + \Lambda (H_0, \hat{H}_1) p(H_0 | y) \]
\[ = 0 + \Lambda (H_0, \hat{H}_1) p(H_0 | y). \]

Now, we look at \( E(\Lambda | \hat{H}_0) \) and \( E(\Lambda | \hat{H}_1) \) and on the basis of these values we make a decision whether to accept \( H_0 \) or to accept \( H_1 \).

**Note:**
1. If \( E(\Lambda | \hat{H}_0) \leq E(\Lambda | \hat{H}_1) \), then accept \( H_0 \); i.e., \( \hat{H}_0 \).
2. If \( E(\Lambda | \hat{H}_1) < E(\Lambda | \hat{H}_0) \), then accept \( H_1 \); i.e., \( \hat{H}_1 \).

If the posterior expected losses are equal, then either action (\( \hat{H}_0 \) or \( \hat{H}_1 \)) will give rise to the same degree of risk.

Now, \[ \begin{cases} E(\Lambda | \hat{H}_0) = \Lambda (H_1, \hat{H}_0) p(H_1 | y) \\ E(\Lambda | \hat{H}_1) = \Lambda (H_0, \hat{H}_1) p(H_0 | y) \end{cases} \]

So, the decision rule may be expressed as:

1. Accept \( H_0 \) if:
\[ E(\Lambda | \hat{H}_0) < E(\Lambda | \hat{H}_1) \]

i.e., if \[ p(H_1 | y) \Lambda (H_1, \hat{H}_0) < p(H_0 | y) \Lambda (H_0, \hat{H}_1) \]

\[ \Rightarrow \left[ \frac{\Lambda (H_1, \hat{H}_0) p(y | \theta = \phi_1) p(H_1)}{p(y)} < \frac{\Lambda (H_0, \hat{H}_1) p(y | \theta = \phi_0) p(H_0)}{p(y)} \right] \]

\[ \Rightarrow \left[ \frac{p(H_1)}{p(H_0)} \right] \left[ \frac{\Lambda (H_1, \hat{H}_0)}{\Lambda (H_0, \hat{H}_1)} \right] < \left[ \frac{p(y | \theta = \phi_0)}{p(y | \theta = \phi_1)} \right] . \]
\[ \frac{P(y|H_0)}{P(y|H_1)} > \frac{P(H_0) \cdot L(H_1, \hat{\theta}_0)}{P(H_1) \cdot L(H_0, \hat{\theta}_1)} \]

So, we pick \( H_0 \) (i.e., select \( \hat{\theta}_0 \)) iff

\[ \text{Likelihood ratio} > \text{ratio of prior expected losses} \]

2. Similarly, accept \( H_1 \) iff

\[ \frac{P(y|H_1)}{P(y|H_0)} > \frac{P(H_1) \cdot L(H_0, \hat{\theta}_1)}{P(H_0) \cdot L(H_1, \hat{\theta}_0)} \]

Now, our usual decision theoretic approach to sampling theory gives us the well known Likelihood Ratio Test when testing between 2 hypotheses.

And in that particular case, the rule is:

Accept \( H_0 \) if \( \frac{P(y|H_0)}{P(y|H_1)} > \lambda \)

Accept \( H_1 \) 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, \( \lambda \) is determined by the significance level chosen for the test. This is an unavoidable arbitrary. The loss structure is introduced into the selection of \( \lambda \) only in a very implicit fashion. This seems rather unsatisfactory.

On the contrary, in the Bayesian approach the loss structure is used explicitly to find \( \lambda \). Namely, we put \( \lambda \) equal to ratio of the posterior expected losses. (Note recall we pick the lesser of the 2 posterior expected losses, which leads to a ratio test involving prior expected losses.)

That is, we put:

\[ \lambda = \left( \frac{P(H_1)}{P(H_0)} \right) \frac{L(H_1, \hat{\theta}_0)}{L(H_0, \hat{\theta}_1)} \]

Now, clearly the value of \( \lambda \), and hence the form of the ratio that will depend very much on the functional form taken by the loss structure.

A natural class of loss functions are those that are symmetric:

use \( \{ L(H_0, \hat{\theta}_1) = L(H_1, \hat{\theta}_0), L(H_0, \hat{\theta}_0) = L(H_1, \hat{\theta}_1) = 0 \} \).
In sampling theory terminology, this is the same as saying that a Type I and Type II error have the same levels associated with them.

\[
\text{Then, if } \frac{p(y|\bar{H}_0)}{p(y|\bar{H}_1)} > \frac{p(H_0)}{p(H_1)} \text{, then pick } H_0.
\]

So, in the case of a symmetric loss structure, the critical value, \( \lambda \), in the likelihood ratio test is just the ratio of the prior odds for \( H_1 \).

Let us select \( H_0 \) as true if the proportional data evidence overwhets the proportional 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 symmetry assumption \( \Rightarrow L(H_1, \hat{H}_0) = L(H_0, \hat{H}_1) = L \).

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 \( H_1 \):

i.e. \( E(L|\hat{H}_0) < E(L|\hat{H}_1) \).

Now, \( \Rightarrow L \cdot p(H_1|y) < L \cdot p(H_0|y) \)

\[
\Rightarrow \left( \frac{p(H_0|y)}{p(H_1|y)} \right) > 1.
\]

Similarly, accept \( H_1 \) if:

\[
\left( \frac{p(H_1|y)}{p(H_0|y)} \right) > 1.
\]

So, in the special case of a symmetric loss function, we just pick an hypothesis if the posterior odds...
Now, so far we have considered only simple hypotheses. This is generally quite unrealistic in practice, as we more usually encounter composite hypotheses. To extend the analysis developed so far to a more general framework.

Let \( H_0 \) and \( H_1 \) be composite mutually exclusive and
totally exhaustive hypotheses, i.e.,

\[
\begin{align*}
& \begin{cases} 
\Pr(H_0) = \Pr(w = 0) \\
\Pr(H_1) = \Pr(w = 1) 
\end{cases} \\
\text{are the prior pdf's.}
\end{align*}
\]

Let \( \Theta \) be the parameter vector associated with \( H_0 \), under which the pdf for \( y \) is \( p(y | \Theta) = p(y | w = 0, \Theta) \); and let \( \Phi \) be the parameter vector associated with \( H_1 \), under which the pdf for \( y \) is \( p(y | \Phi) = p(y | w = 1, \Phi) \). Then the joint

dist.

\[
p(y, w) = p(w, \Theta | y) p(y) = p(y | w, \Theta) p(w, \Theta)
\]

so,

\[
\Pr(w, \Theta | y) = \frac{p(y | w, \Theta) p(w, \Theta)}{p(y)} = \frac{p(y | w, \Theta) p(\Theta | w) p(w)}{p(y)}.
\]

Now, \( \Pr(\Theta | w = 0) = \Pr(\emptyset) \)

\( \Pr(\Theta | w = 1) = \Pr(\emptyset) \)

and \( p(w = 0) = p(\emptyset) \), \( p(w = 1) = p(\emptyset) \).

so,

\[
\Pr(H_0 | y) = \int p(w = 0, \Theta | y) d\Theta = \int p(y | w = 0, \Theta) p(\Theta | w) d\Theta = \int \frac{p(y | w = 0, \Theta) p(\Theta | w) p(w)}{p(y)} d\Theta
\]

\[
= \int \frac{p(H_0) p(y | \Theta) p(\Theta) d\Theta}{p(y)} = \left[ \frac{p(H_0) \int p(y | \Theta) p(\Theta) d\Theta}{p(y)} \right]
\]

and

\[
\Pr(H_1 | y) = \Pr(w = 1 | y) = \int p(w = 1, \Theta | y) d\Theta
\]

\[
= \int \frac{p(w = 1) p(y | w = 1, \Theta) p(\Theta) d\Theta}{p(y)} = \frac{p(H_1) \int p(y | \Phi) p(\Phi) d\Phi}{p(y)}.
\]


\[ P(H_0 | y) \] are given by:

\[ \frac{P(H_0 | y)}{P(H_1 | y)} = \frac{P(H_0)}{P(H_1)} \cdot \left\{ \frac{\int P(y | \theta) P(\theta) d\theta}{\int P(y | \theta) P(\theta) d\theta} \right\}. \]

That is, the posterior odds are the product of the prior odds and the ratio of the averaged likelihoods.

This result in which the maximized likelihoods are used — the MLE are used as if they were the true value of the parameter.

**B. Analysis of Hypotheses Using Diffuse Prior Information:**

Lindley's procedure to be outlined here is valid only if the prior information is fairly diffuse on the scale of the like likelihood function. For many, but not all problems, Lindley's 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.

Lindley's procedure involves the elimination of a test procedure based on a posterior pdf, starting from a diffuse prior pdf. Let the posterior pdf be \( p(\theta | y) \).

We want to construct a test of the hypothesis \( H_0 : \theta = \theta_0 \) at the \( \alpha \) significance level.

To construct an interval \((a, b) \) s.t. \( P(a < \theta < b | y) = 1 - \alpha \).

That is, minimize \((b - a)\) s.t. \( P(a < \theta < b | y) = 1 - \alpha \).

And \( a < \theta < b \) is accepted at the \( \alpha \) significance.

Lindley chooses to use the posterior pdf for \( \theta \) as a basis for making inferences about values of \( \theta \). If \( \theta = \theta_0 \) falls within a region where the posterior density is not high, then we have little belief in \( \theta = \theta_0 \), and the hypothesis will be rejected.

Now, the posterior distribution is based on the likelihood explicitly, so it contains all of the sample information. Compare this situation with that frequently
arriving with respect to the sampling theory approach. Here the tests are often based on test statistics which are not sufficient 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 posterior knowledge—that is, on all of the knowledge that we currently have. We are not basing our judgement on the fact that some test statistic takes an unusual value if \( \theta = \theta_0 \). In sampling theory tests should not 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 whole of the posterior distribution, whereas the sampling theory looks just at the significance level. This is only a partial expression of posterior beliefs.

Ex. Let: \( Y' = (y_1, y_2, \ldots, y_n) \sim \text{NID} (\mu, \sigma^2) \),
where \( \sigma^2 \) is known.

When will we now reject \( H_0 : \mu = \bar{\mu} \) at the 5% level?
Suppose the \( p(\mu) \) is constant.

\[
p(\mu | y, \sigma = \sigma_0) \propto \exp \left\{ -\frac{1}{2\sigma_0^2} (y_i - \mu)^2 \right\}
\times \exp \left\{ -\frac{1}{2\sigma^2} \sum (y_i - \hat{\mu})^2 \right\}
\times \exp \left\{ -\frac{n}{2\sigma^2} (\mu - \hat{\mu})^2 \right\}.
\]

So the posterior pdf for \( \mu \) is \( N(\hat{\mu}, \frac{\sigma^2}{n}) \).

\[
Z = \left( \frac{\mu - \bar{\mu}}{\frac{\sigma}{\sqrt{n}}} \right) \sim N(0, 1).
\]

\[
Pr \left\{ -1.96 < Z < 1.96 \right\} = 0.95
\]

\[
Pr \left\{ \hat{\mu} - \frac{1.96 \sigma}{\sqrt{n}} < \mu < \hat{\mu} + \frac{1.96 \sigma}{\sqrt{n}} \right\} = 0.95.
\]

\[
\text{If } \sigma \in \left( \hat{\mu} - \frac{1.96 \sigma}{\sqrt{n}}, \hat{\mu} + \frac{1.96 \sigma}{\sqrt{n}} \right), \text{ accept } H_0.
\]

The sampling theory interpretation of a computationally equivalent test \( Z' = \left( \frac{\hat{\mu} - \bar{\mu}}{\frac{\sigma}{\sqrt{n}}} \right) \) is considered random (not \( \mu \) itself), and \( H_0 \) is accepted if \( Z' \in (-1.96, 1.96) \).
An example is an example of a case where there is no sampling theory equivalent test, consider the case of the
autocorrelation coefficient, $p = \rho$.

$$p(\rho|y) \propto \left| I - (X-\rho X_a) Y^{-1}(X-\rho X_a)' \right| ^{-\frac{1}{2}} \{y-y_\cdot\}' \{I - (X-\rho X_a) Y^{-1}(X-\rho X_a)' \} (y-y_\cdot)$$

Then apply numerical integration techniques to find the shortest interval $(a, b)$, $\Pr \{ a < \rho < b \} = 1 - \alpha$.

To test $H_0: \rho = \rho_0$,
accept $H_0$ if $(a < \rho < b)$, otherwise reject.

And clearly, Lindley's general procedure may be extended to the case where we have a joint hypothesis about $\Theta$ or more parameters, say $\Theta$. We just obtain the Bayesian highest posterior density confidence region with a probability content $(1 - \alpha)$. If $H_0$ is $\Theta = \Theta_0$, then accept $H_0$ if $\Theta_0$ 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, a posteriori.

In some cases, Lindley's procedure is computationally equivalent to the sampling theory solution. In other cases, there is no sampling theory solution. In large samples
Lindley's procedure corresponds to M.L. procedure under the assumption of an asymptotically normal L.F.
3) Hypothesis & Non-Diffuse
Prior Information:

If prior information is not in fact diffuse, then it is most important that this fact be taken into account when testing hypotheses. For example, if we test one set of explanatory variables, and these seem to be insignificant, we may modify the set and test again, then we have some prior information when it comes to testing the second time, and we really should take account of this fact. Sampling theory techniques do not allow us to take account of this particular problem very satisfactorily.

There are a number of situations in which we really wish to introduce non-diffuse prior information. For example, consider the simple hypotheses

\[ H_0 : \theta = \theta_0. \]

Then if \( \theta = \theta_0 \) is the value suggested by the theory (as it is likely to be, otherwise we would not have tested \( H_0 \)) then our prior information is not diffuse at all, in fact it is tightly packed close to \( \theta_0 \). That is, \( \theta_0 \) is a priori as more likely than other values, is this posteriorly maintained a posteriori?

Can we introduce non-diffuse prior information successfully?

Consider the following:

\[ y' = (y_1, \ldots, y_n) \sim NID(\mu, 1) \]

\[ H_0 : \mu = 0. \]

\[ \theta \sim \begin{cases} p(\mu = 0) = \pi_1 \\ p(\mu \neq 0) = 1 - \pi_1. \end{cases} \]

\[ p(\mu = 0) \] is spread uniformly in \((-\Delta, +\Delta)\)

\[ \pi_1 = \frac{1}{2}. \]

\[ H_0 : \mu = 0 \quad H_1 : \mu \neq 0. \]
Now, the prior odds are \( \left( \frac{\pi_1}{1-\pi_1} \right) \).

Now, under \( H_0 \), the posterior pdf is obtained as
\[
P(y | \mu=0, \sigma=1) = \left( \frac{1}{(2\pi)^{\frac{n}{2}}} \right) \exp \left\{ -\frac{1}{2} \frac{y}{\sigma}^2 \right\}
= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (y \bar{y} + \bar{y} \bar{y}) \right\}
\]

Under \( H_1 \):
\[
P(y | \mu=\theta, \sigma=1) = (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{y}{\sigma}^2 (y \bar{y} - \mu \bar{y}) \right\}
= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (y \bar{y} + n(\mu - \bar{y})^2) \right\}
\]

And, \( K_{01} = \frac{P(H_0 | y)}{P(H_1 | y)} \)
\[
= \frac{P(H_0)}{P(H_1)} \cdot \frac{\int P(\theta) P(y | \theta) d\theta}{\int P(\phi) P(y | \phi) d\phi}
\]

\[
= \left( \frac{\pi_1}{1-\pi_1} \right)
\]

\[
\int P(y | \theta) P(\theta) d\theta = \int (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{y}{\sigma}^2 (y \bar{y} - \mu \bar{y}) \right\} \cdot d\mu
= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{y \bar{y}}{\sigma} \right\} \cdot \int \exp \left\{ -\frac{1}{2} (\mu - \bar{y})^2 \right\} d\mu
\]

\[
\int P(y | \phi) P(\phi) d\phi = \int (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{y}{\sigma}^2 (y \bar{y} + n(\mu - \bar{y})^2) \right\} d\mu
= (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{y \bar{y}}{\sigma} \right\} \cdot \int \exp \left\{ -\frac{1}{2} (\mu - \bar{y})^2 \right\} d\mu
\]

\[
K_{01} = \left( \frac{\pi_1}{1-\pi_1} \right) \frac{\exp \left\{ -\frac{1}{2} \frac{y \bar{y}}{\sigma} \right\} \cdot \int \exp \left\{ -\frac{1}{2} (\mu - \bar{y})^2 \right\} d\mu}{\int \exp \left\{ -\frac{1}{2} (\mu - \bar{y})^2 \right\} d\mu}
\]

Now note the following points:

1. We may have a basis for testing \( H_0 \) using non-differ prior information, as well as sample information.

2. Given the base 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 \( (-\frac{\sigma}{2}, +\frac{\sigma}{2}) \).

Then if we look only at the curve between \( -\frac{\sigma}{2} \) and \( +\frac{\sigma}{2} \), we can assume that the area of the shaded area above is small.

Now, because \( \sigma^2 = 1 \), \( \mu \sim N(\bar{y}, \sigma^2) \)

So, \( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left( \mu - \bar{y} \right)^2 \right\} d\mu = 1 \)

\( \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left( \mu - \bar{y} \right)^2 \right\} d\mu = (\frac{2\pi}{\sigma^2})^{1/2} \)

Now taking account of the admissibility of the range approximation just noted above —

\( \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left( \mu - \bar{y} \right)^2 \right\} d\mu \leq (\frac{2\pi}{\sigma^2})^{1/2} \)

\[ K_{01} = \left( \frac{\pi_{1}}{1_{-1}} \right) \frac{\exp\left\{-\frac{1}{2} \frac{x^2}{\sigma^2} \right\}}{(2\pi/\sigma)^{1/2}} \]

\[ = \left( \frac{\pi_{1}}{1_{-1}} \right) M. \left( \frac{n}{2\pi} \right)^{1/2} \exp\left\{-\frac{2x^2}{\sigma^2} \right\} \]

where \( 2x = \frac{s}{\sigma} \).

Now, if \( 2x > 1.96 \), a sampling theoretist would reject at the 5% level the hypothesis that \( \mu = 0 \).

Hence, put \( 2x = 1.96 \).

\[ K_{01} = \left( \frac{\pi_{1}}{1_{-1}} \right) M. \left( \frac{n}{2\pi} \right)^{1/2} \exp\left\{-\frac{1}{2} \right\} \]

\[ = f(P_{01}, M, n) \]

And it is in fact possible for \( K_{01} \) to be high (thus rejecting \( H_0 \)) even if \( 2x = 1.96 \).

E.g., put \( P_{11} = \frac{z}{2}, M = 1, \) \( K_{01} > 1 \) for \( n \geq 300 \).

This is Lindley's Paradox.
This fact 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

\[ H_0 : \mu = 0 \quad \Rightarrow \quad p(H_0) = \pi_1 \]

\[ H_1 : \mu \neq 0 \quad \Rightarrow \quad p(H_1) = 1 - \pi_1. \]

Now, if \( H_1 \) true, then \( \mu \neq 0 \), so let \( p(\mu) \) be a pdf for \( \mu \).

i.e. \( \int p(\mu) d\mu = 1 \) under \( H_1 \).

Let

\[ p(\omega = 0) = \text{prior } p(H_0), \]

\[ p(\omega = 1) = \text{prior } p(H_1). \]

Now,

\[ p(\omega, \mu | y) = \frac{p(y | \omega, \mu) p(\omega, \mu)}{p(y)} \]

\[ = \frac{p(y | \omega, \mu) p(\mu | \omega = 0, \mu = 0)}{p(y)} \]

Then the posterior probability for true \( H_0 \) is

\[ p(\omega = 0 | y) = \frac{p(y | \omega = 0, \mu = 0) p(\mu | \omega = 0, \mu = 0)}{p(y)} \]

\[ = \frac{\pi_1 p(y | \omega = 0, \mu = 0)}{p(y)} \]

\[ = \frac{\pi_1 p(y | \omega = 0, \mu = 0)}{p(y)} \cdot 1 \]

Since,

\[ p(\mu | \omega = 0) = \begin{cases} 1 & \text{if } \mu = 0 \\ 0 & \text{if } \mu \neq 0 \end{cases} \]

\[ \therefore p(\omega = 0 | y) = \frac{\pi_1 p(y | \omega = 0, \mu = 0)}{p(y)} \]

\[ p(\omega = 1 | y) = \frac{\int p(y | \mu, \omega = 1) p(\mu | \omega = 1) d\mu}{p(y)} \]

\[ = (1 - \pi_1) \frac{\int p(y | \mu, \omega = 1) p(\mu | \omega = 1) d\mu}{p(y)} \]
Now, \( p(\mu | y = 1) = p(\mu) \), since \( p(\mu) \) is defined only under \( H_1 \) true.

So,
\[
p(w = 1 | y) = \frac{(1 - \pi_1) \int p(y | \mu, w = 1) p(\mu) \, d\mu}{p(y)}
\]

\[
P_{01} = \left( \frac{\pi_1}{1 - \pi_1} \right) \left[ \frac{\int p(y | \mu, w = 0) p(\mu) \, d\mu}{\int p(y | \mu, w = 1) p(\mu) \, d\mu} \right].
\]

Now, take the case where \( y' = (y_1, \ldots, y_n) \in Y, \sim N_{n, D}(\mu, \sigma^2) \) known.

\[
P_{01} = \left( \frac{\pi_1}{1 - \pi_1} \right) \frac{(\sigma_0^2 / \alpha) \exp \left\{ -\frac{y_1^2}{2\sigma_0^2} \right\} \exp \left\{ -\frac{n \alpha^2}{2} \right\}}{(\alpha \sigma_0^2) \exp \left\{ -\frac{y_1^2}{2\sigma_0^2} \right\} \int p(\mu) \exp \left\{ -\frac{y_1^2}{2\sigma_0^2} (\mu - \mu')^2 \right\} \, d\mu}
\]

\[
= \left( \frac{\pi_1}{1 - \pi_1} \right) \left[ \frac{\exp \left\{ -\frac{n \alpha^2}{2} \right\}}{\int p(\mu) \exp \left\{ -\frac{y_1^2}{2\sigma_0^2} (\mu - \mu')^2 \right\} \, d\mu} \right]
\]

And if we are given \( p(\mu) \) and \( \pi_1 \), then we can evaluate \( P_{01} \).

Consider 2 extreme cases—

1. Assume \( f \subseteq (-a, a) \Rightarrow \int_a^a p(\mu) \, d\mu = 1 \).

If \( f \) lies well within \((a, a) \) or \( \kappa / \sigma_0^2 \) is so large, then \((\kappa / \sigma_0^2)\) is small for \(-a,a\)

Then \( P_{01} = 1 \), since \( e = 1 \Rightarrow \int_a^a p(\mu) \, d\mu = 1 \).

That is, when the standard deviation \( \sigma_0^2 \) is much larger than the range of \( \mu \), then we cannot distinguish between \( H_0 \) and \( H_1 \) — this seems fairly plausible.

2. Suppose that \( \kappa / \sigma_0^2 \) is fairly small. Then \( \kappa / \sigma_0^2 \) is fairly small.

Then \( \exp \left\{ -\frac{n \alpha^2}{2\sigma_0^2} (\mu - \mu')^2 \right\} \) can be very large on \((-a, a) \).

Then \( P_{01} \) increases with \( n \) — in the odds tend to favour \( H_0 \) as \( n \) gets larger.

Non-diffuse prior information affects the posterior
compared in both small and large samples, so care must be taken in specifying the prior part of this specification so non-difficult.

Compare this situation to that of Bayesian estimation. There is a large sample, the information in the data swamps any prior information, so the likelihood fits become 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 \( R^2 \) and the Bayesian approach of comparing \( \beta \) - 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 \( X_1 \) and \( X_2 \) are \((n \times k)\), so both \( \beta_1 \) and \( \beta_2 \) are \((k \times 1)\), i.e., the same number of exogenous variables appear in both \( M_1 \) and \( M_2 \). Further, \( \beta_1 \) and \( \beta_2 \) have no common elements.

If \( M_1 \) true, then \( u_{1t} \sim \text{NID}(0, \sigma_1^2) \); \( \forall t \).

If \( M_2 \) true, then \( u_{2t} \sim \text{NID}(0, \sigma_2^2) \); \( \forall t \).

Choose natural conjugate prior:

\[ p(\beta; \sigma) = p(\beta_1 | \sigma_1) p(\sigma) \quad ; \quad i = 1, 2, \]

where

\[ p(\beta_1 | \sigma_1) = \left( \frac{1}{(2\pi)^{k_1/2} \sigma_1^{k_1}} \right) \exp \left[ -\frac{1}{2\sigma_1^2} (\beta_1 - \bar{\beta}_1)^T C_1 (\beta_1 - \bar{\beta}_1) \right] \]

and

\[ p(\sigma) = \left( \frac{k_1}{\sigma^{k_1+1}} \right) \exp \left[ -\frac{k_1 \sigma^2}{2\sigma_1^2} \right]\]

where

\[ k_1 = \left[ 2 \left( \frac{\sigma_1 \gamma^{1/2}}{2} \right)^{2k_1/2} \right] / \Gamma(k_1/2). \]

That is, \( p(\sigma) \) is an **inverted Gamma**.
Note that in the normal pdf for \((\beta_i|\sigma_i)\) the mean \(\bar{\beta}_i = \bar{\beta}_i \times \text{c.m.} = (\beta_i|\sigma_i)\). Both \(C_i\) \& \(\beta_i\) must be assigned.

In the I.E. dist., both \(\beta_i \sim \text{G}\) have to be assigned

If we have prior odds on \(\omega_1, \omega_2\) then we get posterior odds

\[
K_{01} = \frac{p(\mu_1)}{p(\mu_2)} \int p(y | \beta_1, \sigma_1, M_1) p(\sigma_1) d\sigma_1 \int p(y | \beta_2, \sigma_2, M_2) p(\sigma_2) d\sigma_2
\]

\[+ p(y | \beta_1, \sigma_1, M_1) \text{ is the K/F, given } M_1 \text{ true.}
\]

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(\frac{\bar{y}}{\bar{x}_i} + (\beta_i - \bar{\beta}_i) \bar{x}_i \bar{x}_i \beta_i (\beta_i - \bar{\beta}_i)\right)\right] \]

where \(\bar{y} = \frac{y}{n} \); \(\bar{x}_i = (y_i x_i) / (y_i x_i)\); \(\bar{\beta}_i = (x_i x_i)^{-1} x_i y_i\).

Now, we have to evaluate:

\[
\int p(y | \beta_i, \sigma_i, M_i) p(\sigma_i|\sigma_i) p(\sigma_i) d\beta_i d\sigma_i, i = 2, 2.
\]

Consider, for example, \(i = 1\):

\[p(\sigma_1) = \left(\frac{k_1}{\sigma_1}\right) \exp\left[-\frac{k_1 \sigma_1^2}{2}\right]\]

\[p(y | \beta_1, \sigma_1, M_1) = \left(\frac{1}{\sqrt{2\pi} \sigma_1}\right)^n \exp\left[-\frac{1}{2\sigma_1^2} \left(\frac{\bar{y}}{\bar{x}_1} + (\beta_1 - \bar{\beta}_1) \bar{x}_1 \bar{x}_1 (\beta_1 - \bar{\beta}_1)\right)\right]\]

\[p(\beta_1 | \sigma_1) = \left(\frac{k_1 c_1 \kappa}{\sqrt{2\pi} \sigma_1^{n+1}}\right) \exp\left[-\frac{1}{2\sigma_1^2} \left(\frac{\beta_1 - \bar{\beta}_1}{\sigma_1}\right) c_1 (\beta_1 - \bar{\beta}_1)\right].\]

So, \(S = \frac{k_1 c_1 \kappa}{(2\pi)^{n+1}} \int \frac{1}{\sigma_1^{n+1}} \exp\left\{ -\frac{1}{2\sigma_1^2} \left(\frac{\bar{y}}{\bar{x}_1} + (\beta_1 - \bar{\beta}_1) \bar{x}_1 \bar{x}_1 \beta_1 (\beta_1 - \bar{\beta}_1)\right)\right\} \]

Integrate w.r.t. \(\beta_i\) — first complete the square in \(\beta_i\) in the exponent:

\[
\begin{align*}
\beta_i - \bar{\beta}_1, \bar{x}_1 (\beta_i - \bar{\beta}_1) + (\beta_i - \bar{\beta}_1) \bar{x}_1 \bar{x}_1 (\beta_i - \bar{\beta}_1) \\
= \beta_i \bar{c}_1 \bar{\beta}_1 + \bar{c}_1 (\beta_i - \beta_1) + \bar{\beta}_1 \bar{x}_1 (\beta_i - \beta_1) + \beta_1 \bar{x}_1 \beta_1 + \beta_1 \bar{x}_1 \beta_1 \\
- 2 \beta_1 \bar{x}_1 \beta_1 \\
\beta_1 \left[ \bar{c}_1 + (\bar{x}_1 x_1) \bar{c}_1 \right] + 2 \beta_1 \left[ \beta_1 \bar{c}_1 + (x_1 x_1) \bar{c}_1 \right] + \bar{\beta}_1 \bar{c}_1 + \bar{\beta}_1 \bar{x}_1 \beta_1 .
\end{align*}
\]
And let \( A_i = c_i + (x_i'x_i) \).

we obtain:

\[
\begin{align*}
[\beta_i - A_i^{-1} (c_i + x_i'x_i \beta_i)] & A_i^{-1} (c_i + x_i'x_i \beta_i) \\
+ \beta_i' (c_i + x_i'x_i \beta_i) & - (c_i + x_i'x_i \beta_i)' A_i^{-1} (c_i + x_i'x_i \beta_i)
\end{align*}
\]

Let \( \tilde{\beta}_i = A_i^{-1} (c_i + x_i'x_i \beta_i) \)

\[
\Rightarrow [\beta_i - \tilde{\beta}_i]' A_i [\beta_i - \tilde{\beta}_i] + \tilde{\beta}_i' (c_i + x_i'x_i \beta_i) - \tilde{\beta}_i' (c_i + x_i'x_i \beta_i) - \tilde{\beta}_i' A_i \tilde{\beta}_i.
\]

\[
= (\beta_i - \tilde{\beta}_i)' A_i (\beta_i - \tilde{\beta}_i) + (\tilde{\beta}_i' c_i + x_i'x_i \tilde{\beta}_i) - \tilde{\beta}_i' (c_i + x_i'x_i \beta_i) - \tilde{\beta}_i' (c_i + x_i'x_i \beta_i) - \tilde{\beta}_i' A_i \tilde{\beta}_i.
\]

\[
= (\beta_i - \tilde{\beta}_i)' A_i (\beta_i - \tilde{\beta}_i) + (\tilde{\beta}_i' c_i + x_i'x_i \tilde{\beta}_i) - (\tilde{\beta}_i' c_i + x_i'x_i \tilde{\beta}_i) - \tilde{\beta}_i' A_i \tilde{\beta}_i.
\]

\[
\int_0^\infty \int_0^{\beta_i - \tilde{\beta}_i} \frac{(k_i l_i c_i 1/n)}{(2\pi)^{h_i/2}} \exp \left\{ -\frac{1}{2\sigma_i} \left[ q_i \tilde{\beta}_i^2 + u s_i^2 + \omega a + \omega a_n \right] \right\} \, dp_i, ds_i,
\]

where \( q_i = \tilde{\beta}_i' c_i + x_i'x_i \tilde{\beta}_i \) ; \( u s_i = \tilde{\beta}_i' x_i'x_i (\beta_i - \tilde{\beta}_i) \)

\[
= \frac{k_i l_i c_i 1/n}{(2\pi)^{h_i/2}} \int_0^{1/(\alpha_i+1)} \exp \left\{ -\frac{1}{2\sigma_i} \left[ q_i \tilde{\beta}_i^2 + u s_i^2 + \omega a + \omega a_n \right] \right\} ds_i.
\]

Consider now the integration of this form w.r.t. \( \sigma_i \):

The integral itself is

\[
\int_0^{1/(\alpha_i+1)} \exp \left\{ -\frac{1}{2\sigma_i} \left[ q_i \tilde{\beta}_i^2 + u s_i^2 + \omega a + \omega a_n \right] \right\} ds_i
\]

which is \( \mathcal{D} \) the general form

\[
\int_0^{1/(\alpha_i+1)} \exp \left\{ -\frac{1}{2\sigma_i} \right\} d\sigma_i.
\]
Let \( x = (b/2\sigma^3) \)

\[
\begin{align*}
x \, dx &= -b\sigma^{-3} \, d\sigma \\
\sigma &= (b/2x)^{1/3} \\
\sigma^{-1} &= (b/2x)^{-1/3} \\
\exp \left( \frac{-b}{2b\sigma^3} \right) \, d\sigma &= \int_0^\infty \left( \frac{b}{2x} \right)^{-1/3} \sigma^{-1/3} \, e^{-\sigma^{-3/6}} \, dx \\
&= \int_0^\infty \left( \frac{b}{2x} \right)^{-1/3} \, e^{-\sigma^{-3/6}} \, dx
\end{align*}
\]

\[ \beta n, \quad (\sigma^{-3/6}) = (b^{1/3})/(2x)^{1/3} \]

\[
\begin{align*}
\int_0^\infty \left( \frac{b}{2x} \right)^{-1/3} \, e^{-\sigma^{-3/6}} \, dx
&= \int_0^\infty \frac{b^{-1/3}}{(2x)^{-1/3}} \, e^{-\sigma^{-3/6}} \, dx \\
&= \int_0^\infty \frac{b^{-1/3}}{(2x)^{-1/3}} \, (x^{-1/3}) \, e^{-x} \, dx \\
&= (1/2) (2\pi)^{-1/2} \Gamma(3/2)
\end{align*}
\]

\[ A = n + q_1, \quad b = \left[ q_1 s_1^2 + v s_1^2 + Q \right] \]

\[
\begin{align*}
\Gamma \left( \frac{n+q_1}{2} \right) \, \Gamma \left( \frac{n+q_1}{2} \right)
&= \left( \frac{k_1}{(2\pi)^{n/2}} \right) \left( \frac{1}{1A_1} \right)^{1/2} \frac{2^{(n+q_1)/2}}{\sqrt{q_1 s_1^2 + v s_1^2 + Q} \, \Gamma(1/2)} \\
K_{12} &= \frac{p(n_1)}{p(n_2)} \left( \frac{k_1}{k_2} \right) \left[ \frac{1}{1A_1} \right]^{1/2} \frac{2^{(n+q_1)/2}}{\sqrt{q_1 s_1^2 + v s_1^2 + Q} \, \Gamma(1/2)} \\
&\approx \frac{\left( q_1 s_1^2 + v s_1^2 + Q \right)^{(n+q_1)/2}}{\left( q_1 s_1^2 + v s_1^2 + Q \right)^{(n+q_1)/2}}
\end{align*}
\]
\[ K_i = 2 \left( \frac{q_{i+1}^2}{x_i} \right)^{q_{i+2}/2} / \Gamma \left( \frac{q_{i+2}}{2} \right) \]

\[ [K_i, 2^{(n+q_i)/2}] = \left[ \frac{1}{\Gamma \left( \frac{q_{i+1}}{2} \right)} \cdot \left( q_{i+1}^2 \right)^{q_{i+1}/2} \cdot 2^{q_{i+1}} \right] ; \quad i = 2, 3. \]

\[ K_{12} = \frac{p(M_1)}{p(M_2)} \cdot \left[ \frac{C_{11}/L_{A1}}{C_{12}/L_{A2}} \right]^{q_{11}^2} \cdot \frac{\Gamma \left( \frac{n+q_1^2}{2} \right)}{\Gamma \left( \frac{q_1^2}{2} \right)} \cdot \frac{\Gamma \left( \frac{n+q_2^2}{2} \right)}{\Gamma \left( \frac{q_2^2}{2} \right)} \cdot \left( \frac{q_{11}^2}{q_{22}^2} \right)^{q_{11}/2} \frac{\Gamma \left( \frac{n+q_1^2}{2} \right)}{\Gamma \left( \frac{q_1^2}{2} \right)} \frac{\Gamma \left( \frac{n+q_2^2}{2} \right)}{\Gamma \left( \frac{q_2^2}{2} \right)} \frac{(q_{11}^2 + q_{22}^2)}{(q_{11}^2 + q_{22}^2)} \]

Let \( \delta_i = \frac{q_{i+1}^2 + q_{i+2}^2 + q_{i+3}^2 + q_{i+4}^2}{n} \).

\[ \frac{(q_{11}^2 + q_{22}^2 + q_{33}^2 + q_{44}^2)}{(q_{11}^2 + q_{22}^2 + q_{33}^2 + q_{44}^2)} \]

\[ = \frac{(n \cdot \delta_i)^{-q_{11}/2}}{(n \cdot \delta_i)^{-q_{22}/2}} \]

\[ = \frac{(n \cdot \delta_i)^{-q_{11}/2}}{(n \cdot \delta_i)^{-q_{22}/2}} \]

\[ \delta_i, K_{12} = \frac{p(M_1)}{p(M_2)} \cdot \left[ \frac{C_{11}/L_{A1}}{C_{12}/L_{A2}} \right]^{q_{11}^2} \cdot \frac{\Gamma \left( \frac{n+q_1^2}{2} \right)}{\Gamma \left( \frac{q_1^2}{2} \right)} \cdot \frac{\Gamma \left( \frac{n+q_2^2}{2} \right)}{\Gamma \left( \frac{q_2^2}{2} \right)} \cdot \left( \frac{q_{11}^2}{q_{22}^2} \right)^{q_{11}/2} \]

Now, look at the four terms making up the posterior odds, \( K_{12} \):

(i) \( \frac{p(M_1)}{p(M_2)} = \) prior odds.

(ii) \( \frac{C_{11}/L_{A1}}{C_{12}/L_{A2}} \) is the relative precision for the prior pdf w.r.t. the precision for the posterior pdf.

Now, \( K_{12} \) increases if \( \frac{C_{11}/L_{A1}}{C_{12}/L_{A2}} \) increases relative to \( \frac{C_{11}/L_{A}}{C_{12}/L_{A2}} \). This implies more prior information in
M_1 relative to that of M_2.

(6) Look at \((s_j/d_j)^{-n_j}\). Note that \(s_j\) reflects what the sample has to say about the goodness of fit of the models:

\[ s_j = \left( \frac{\nu s_j^2 + Q_jc + Q_j a x + q_j s_j^2}{n} \right). \]

And \(\nu s_j^2\) is residual sum of squares, so high \(\nu s_j^2\) = goodness of fit.

\[ Q_jc = (\beta_j - \bar{\beta}_j')^T \Gamma_j (\beta_j - \bar{\beta}_j) \] is the closer together are \(\beta_j\) and \(\bar{\beta}_j\), the smaller is \(Q_jc\). So, if \(Q_jc \approx Q_ja\) are high, then the sample of prior information are incompatible.

So, now \(s_j\) = good fit.

If \(s_j\) high relative to \(s_2\), then \(M_1\) is poor rel to \(M_2\).

So \((s_j/d_j)^{-n_j}\) high \(\Rightarrow\) favour \(M_1\) (as required above).

\[ \left( \frac{s_j^2/d_j}{s_2^2/d_2} \right) \frac{\Gamma_j}{\Gamma_2} \left( \frac{(s_j^2/d_j)}{s_2^2} \right) \] reveals a dependence of \(K_{12}\) on the prior information regarding \(s_j\) or \(s_2\). If the prior of sample information regarding \(s_j\) 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

\[ K_{12} \to \left( \frac{s_1^2}{s_2^2} \right)^{-n_j} \exp(-\frac{1}{2} \frac{s_1^2}{s_2^2}) \exp(-\frac{1}{2} \frac{s_2^2}{s_2^2}) \]

where \(s_1^2 = s_2^2 = s^2\); \(q_1 = q_2 = q\).

Then:

\[ \text{If } s_1^2 = s_2^2, \quad \Rightarrow K_{12} = 1 \]

\[ \text{If } s_1^2 > s_2^2, \quad \Rightarrow K_{12} < 1 \]

\[ \text{If } s_1^2 < s_2^2, \quad \Rightarrow K_{12} > 1 \]

If the prior information becomes diffuse, then \(K_{12} \to 0\) if \(q_1 = q_2 = q \to 0\). Assume that \(p(M_1) = p(M_2)\) so \(s_1^2 = s_2^2\).
Then: \[ k_{12} = \left( \frac{s_1^2}{s_2^2} \right)^{-n/2} \]

- If \( s_1^2 = s_2^2 \) ; \( k_{12} = 1 \)
- If \( s_1^2 < s_2^2 \) ; \( k_{12} > 1 \)
- If \( s_1^2 > s_2^2 \) ; \( k_{12} < 1 \)

Now, we may choose the model with the 
lower expected loss \( \ell \) of the loss function 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 occurs when \( s_1^2 < s_2^2 \); i.e. when \( M_1 \) has a higher \( R^2 \) than \( M_2 \). This is true if either 
- we have a symmetric loss function \( \ell \) a diffusion prior; 
- we have large \( n \), so the other assumptions are not.

\[ E \) Comparing Distributed

\textit{Key Models:}

\[ C_t = \lambda C_{t-1} + (1 - \lambda) Y_t + \epsilon_t - \lambda \epsilon_{t-1} \]

- \( M_1 : \quad U_t - \lambda U_{t-1} = \epsilon_t \sim N ID(0, \sigma^2) \)
- \( M_2 : \quad U_t = \epsilon_t \sim N ID(0, \sigma^2) \)

If \( \lambda = 0 \), then \( M_1 \) is a deterministic.

Try 3 alternative prior pdfs:

1. \[ p(\lambda, k, \sigma_i) \propto (\sigma_i)^{3} \quad i = 1, 2 \quad 0 < \lambda, k < 1 \]
2. \[ p(\lambda, k, \sigma_i) \propto \left[ \lambda^{34} (1-\lambda)^{14} k^{59} (1-k)^{9} \right] / \sigma_i \]
3. \[ p(\lambda, k, \sigma_i) \propto \left[ \lambda (1-\lambda)^{4} k^{59} (1-k)^{9} \right] / \sigma_i \]

\( \sigma_i \) independent prior with \( p(k), p(\lambda), \beta\)-density.
\[ K_{12} = \frac{p(m_1)}{p(m_2)} \int \int \int p(x_1, k, \sigma) p(x_2, k, \sigma) \, dx_1 \, dx_2 \, dk \, d\sigma \]

\[ p_1(x_1, k, \sigma) \propto \frac{1}{(\sigma_1^{2\pi})} \exp \left\{ -\frac{1}{2\sigma_1^2} \left[ x_1 - \lambda C_{-1} - k(1-\lambda) \right] \right\} \]

\[ \lambda C_{-1} - k(1-\lambda) \]

And,

\[ p_2(x_2, k, \sigma) \propto \frac{1}{(\sigma_2^{2\pi})} \exp \left\{ -\frac{1}{2\sigma_2^2} \left[ x_2 - \lambda C_{-1} - k(1-\lambda) \right] \right\} \]

\[ \lambda C_{-1} - k(1-\lambda) \]

\[ \int \int \int p(x_1, x_2, k, \sigma) \, dx_1 \, dx_2 \, dk \, d\sigma \]

\[ \int \int p(x_1, k) \left\{ \left[ \lambda C_{-1} - k(1-\lambda) \right] \right\}^{-\frac{7}{2}} \, dk \, dx_1 \]

\[ = \int \int p(x_2, k) \left\{ \left[ \lambda C_{-1} - k(1-\lambda) \right] \right\}^{-\frac{7}{2}} \, dk \, dx_2 \]

\[ \int \int \int p(x_1, x_2, k, \sigma) \, dx_1 \, dx_2 \, dk \, d\sigma \]

\[ \int \int p(x_2, k, \sigma) \left\{ \left[ \lambda C_{-1} - k(1-\lambda) \right] \right\}^{-\frac{7}{2}} \, dk \, dx_2 \]

\[ K_{12} = \frac{p(m_1)}{p(m_2)} \int \int \int p(x_1, x_2, k, \sigma) \, dx_1 \, dx_2 \, dk \, d\sigma \]

\[ \int \int p(x_1, k) \left\{ \left[ \lambda C_{-1} - k(1-\lambda) \right] \right\}^{-\frac{7}{2}} \, dk \, dx_1 \]

And these integrals may be evaluated using bivariate numerical techniques, given \( p(x, k) \) as in (a), (b), (c) above

**Example:** Consider a binomial distributed log model -

\[ y_i = \sum_{i=1}^{n} x_i x_{i+1} + u_i \]

\[ \alpha_i = \frac{\lambda_i}{(1-\lambda_i)} (1-\lambda_i) \lambda_i \]

\[ \sigma_i = k \lambda_i \]

\[ u_i \text{ is just a probability from a Beta pdf.} \]

\[ \text{Mean} = \frac{\lambda}{(1-\lambda)} ; \text{Var} = \frac{\lambda}{(1-\lambda)^2} \]

If \( \lambda = 1 \), we have a geometric distribution.
Let \( L^i X_t = X_{t-i} \)

\[
\begin{align*}
\mathbb{E}_0 (r^{t-i-1}) L^i (1-L)^X_t &= \mathbb{E}_0 (1-L)^{t-i} \\
\mathbb{E}_0 (r^{t-i-1}) L^i &= (1-L)^{t-i}
\end{align*}
\]

\[
y_t = \sum_0^{N} c_k L^i X_t + \epsilon_t
\]

\[
= \sum_0^{N} k (r^{t-i-1}) (1-L)^r L^i X_t + \epsilon_t
\]

\[
= \sum_0^{N} k (r^{t-i-1}) (1-L)^r (1-L)^i X_t + \epsilon_t
\]

\[
= k (1-L)^r X_t + \sum_0^{N} (r^{t-i-1}) (1-L)^i X_t + \epsilon_t
\]

\[
= k (1-L)^r X_t + (1-\lambda L)^r \epsilon_t
\]

Now assume that \((1-\lambda L)\epsilon_t \sim \mathcal{N}(0, \sigma^2)\).

The problem then becomes: given our observations and prior assumptions about the parameters, what are the posterior probabilities of

\[
H_1 : r = 1; \quad H_2 : r = 2; \quad H_3 : r = 3; \quad H_4 : r = 4.
\]

\[
H_1 : r = 1 \Rightarrow y_t = \lambda y_{t-1} + (1-L) X_t + \epsilon_t
\]

\[
H_2 : r = 2 \Rightarrow y_t = 2 \lambda y_{t-1} - \lambda^2 y_{t-2} + (1-L)^2 X_t + \epsilon_t
\]

\[
H_3 : r = 3 \Rightarrow y_t = 3 \lambda y_{t-1} - 3 \lambda^2 y_{t-2} + 3 \lambda^3 y_{t-3} + (1-L)^3 X_t + \epsilon_t
\]

\[
H_4 : r = 4 \Rightarrow y_t = 4 \lambda y_{t-1} - 6 \lambda^2 y_{t-2} + 4 \lambda^3 y_{t-3} - 6 \lambda^4 y_{t-4} + (1-L)^4 X_t + \epsilon_t
\]

For \(r = i\), we have a prior pdf, \(p(\lambda, k, \sigma, \epsilon | r = i)\)

Now,

\[
K_{1i} = \frac{\text{Pr} \{r = 1 \mid X \}_{1i}}{\text{Pr} \{r = 1 \mid X \}_{1i}}
\]

Then we do the same for each pair of models, get the posterior odds, and hence the posterior probabilities.

Let the latter be \(T_i\), \(i = 1, \ldots, 4\). Then

\[
\sum_1^4 \frac{T_i}{1} = 1 = (T_1 + T_2 + T_3 + T_4)
\]
\[ 24. \]

\[ \therefore \quad 1 = \Pi_1 \left( 1 + \frac{\Pi_2}{\Pi_1} + \Pi_3 (\Pi_1 + \Pi_4 / \Pi_1) \right) \]

\[ 1 = \Pi_1 \left( 1 + k_{21} + k_{23} + k_{41} \right) \]

\[ \therefore \quad \Pi_1 = \frac{1}{1 + k_{21} + k_{23} + k_{41}} \]

\[ \text{i.e. in general,} \quad \Pi_i = \frac{1}{1 + \sum_{j=0}^{n} k_{ij}}. \]

\[ \text{So, we use numerical integration to get the} \ k_{ij}, \]

\[ \text{and then get the} \ \Pi_i \ \text{from these.} \]
"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 and 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 the set of 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 fact that information about parameter values as we move through time. This is an "adaptive control" problem. The Bayesian solution is a sequence of optimizing actions (i.e. settings of the control variables) which achieves the control goal while simultaneously playing has we learn from new data.

(A) One-Period Control Problem:

Let \( y_t = \beta x_t + u_t \)

Assume that the values of \( x_t \) can be controlled, and also assume that \( u_t \sim N(0, \sigma^2) \); \( t = 1, 2, \ldots, T \).

Now, consider period \( T+1 \).

Let \( z \equiv y_{T+1} \), let \( w \equiv x_{T+1} \).

\[ z = w + u_{T+1} \sim u_{T+1} \sim N(0, \sigma^2) \]

Suppose that we have not yet observed \( z \), nor determined a value for \( z \); let \( a \) denote the "forget" value for \( z \).

Let \( L(z, a) = (z - a)^2 \sim \) i.e. a quadratic loss function.
Now, we can show that minimizing expected loss when the loss function is quadratic, is equivalent to minimizing expected utility.

Let \( U = c_0 + x^2 - c_2 x^2 \quad ; \quad c_1, c_2 > 0 \)

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 \( a = \frac{c_1}{c_2} \); \( a_0 = c_0 + \frac{c_1^2}{c_2} \).

Now, \( U \) is an increasing function of \( (z-a)^2 \), so if we minimize \( E[(z-a)^2] \), we maximize \( EU \).

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, \( \min_{(a)} E[L(z, a)] = \min_{(a)} E[(z-a)^2] \) 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 \( y \) and the value \( w = x_{t+1} \).

Now, \( p(z|y, w) \) is in the form of a Student- \( t \) density if we have used the diffuse prior – \( p(\beta, \sigma) \propto 1/\sigma \).

\[p(z|y, w) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\nu^{\nu/2} \pi^{(\nu+1)/2}} \left( \frac{\nu}{\nu + \frac{1}{2} (z - \beta x_t)^2} \right)^{-(\nu+1)/2}. \]

where \( \nu = (t-1) \); \( \beta = (\bar{x}_t y_t)/(\bar{x}_t^2) \)

\[q = \frac{1}{\nu^2 \Sigma_x^2}; \nu x^2 = Z(y_t - \beta x_t)^2. \]

Then the mean \( \beta \) \( p(z|y, w) \) is \( \hat{\beta} \), which is clearly a function of \( w \). And \( \beta \) is also a function of \( w \), the control variable.

Now, \( E(Z-a)^2 = E \left[ (Z-E(Z) - (a-E(Z))^2 \right] \]

\[= Var(Z) + (a-E(Z))^2. \]
Now, \( t = \sqrt{\frac{u}{v}} (x - \omega \hat{\beta}) \) is \( t \)-shaped. \( \eta \) has mean zero, and variance \( \sqrt{\frac{u}{v-1}} \).

So, \( z = (\omega \hat{\beta} + \frac{t}{\sqrt{v}}) = (\omega \hat{\beta} + \frac{1}{\sqrt{v}} t) \)

\[ \text{Var}(z) = \left(\frac{1}{\sqrt{v}}\right)^2 \text{Var}(t) \]

\[ = \frac{v}{v(v-1)} \cdot \frac{u}{v} \cdot \frac{1}{(v-2)} \cdot \frac{1}{v} \cdot \frac{u}{v} \cdot \frac{1}{v} \cdot \frac{1}{v} \cdot \frac{1}{v} \]

\[ \therefore \text{E}(z-a)^2 = \frac{v}{v(v-2)} (a - E(z))^2 \]

\[ = \frac{1}{v(v-2)} + (a - \omega \hat{\beta})^2 \]

\[ = \frac{1}{v(v-2)} \left[ 1 + \frac{(\omega \hat{\beta} - a)^2}{\text{Var}(z)} \right] + (a - \omega \hat{\beta})^2 \]

Now, to find the value of \( \omega \) to minimize \( \text{E}(z-a)^2 \):

\[ \frac{\partial \text{E}(z-a)^2}{\partial \omega} = \left(\frac{u}{v-2}\right) (\frac{2a}{\text{Var}(z)}) - 2(a - \omega \hat{\beta}) \]

\[ = 0 \]

\[ \therefore \omega \left[ (\frac{2a}{\text{Var}(z)}) (\frac{u}{v-1}) + 2a \hat{\beta} \right] = 2a \hat{\beta} \]

\[ \therefore \omega = \frac{(2a \hat{\beta}) (1 + \frac{2u}{(v-1)\text{Var}(z)})^{-1}}{2a \hat{\beta}} \]

\[ = (\frac{1}{\text{Var}(z)}) \left[ 1 + \frac{(\omega \hat{\beta} - a)^2}{\text{Var}(z)} \right]^{-1} \]

Let \( \text{Var} = \text{Var}(z) \); \( \mu = \frac{u}{v-2} \); \( \hat{\beta} = \frac{\text{Var}}{\mu} \).

\[ \therefore \omega = (\frac{1}{\text{Var}}) (1 \pm \frac{1}{\sqrt{v}}) \]

So, \( \omega \) is the product of two factors. First, \( (\frac{1}{\text{Var}}) \) is the target value as a ratio of \( \omega \) to the mean of the posterior pdf. Second, we have a factor which is a function of \( \omega \) to \( \text{Var} \). Now, \( \text{Var} = \mu \text{Var} \); \( \hat{\beta} \); \( \therefore \text{Var} \) is the square of the coefficient of variance associated with the posterior pdf for \( \beta \).

Now, \( (\text{Var}/\mu) \) measures the precision of estimation. As precision increases, so too does \( \omega \).

So \( (1 + \frac{1}{\sqrt{v}}) \rightarrow 1 \).

So, if our estimate \( \hat{\beta} \) is precise, then \( \omega \rightarrow (\frac{1}{\text{Var}}) \). However, if we approximate \( \omega \) by \( (\frac{1}{\text{Var}}) \), we may have a sub-optimal setting of the control variable.
Now, if we were to adopt the "certainty equivalence" approach, then we would proceed conditional on the assumption that \( \beta = \hat{\beta} \).

Then, 
\[
L(z, a) \approx (w \hat{\beta} - a)^2
\]

Then, 
\[
\frac{\partial L(z, a)}{\partial \theta} = 2 \hat{\beta} (w \hat{\beta} - a) = 0
\]

i. \( \theta \text{ee} = \left( \frac{a}{\hat{\beta}} \right) \)

So the certainty equivalence solution to our minimization problem is just the first factor of our optimal solution, \( w^* \).

The second factor, which reflects the precision with which \( \hat{\beta} \) has been estimated, does not appear in \( \theta \text{ee} \), since there we have equated \( \hat{\beta} \) with \( \beta \). We are certain that the two are equivalent.

If we use the (sub-optimal) control value \( \theta \text{ee} \), then this gives rise to a higher expected loss than if \( w^* \) is employed. Compare the two results for \( L(z, a) = (z-a)^2 \).

\[
\begin{align*}
E[L \mid w=w^*] &= E \left[ (z-a)^2 \mid w=w^* \right] \\
&= \left( \frac{\sigma^2}{\theta^2} \right) \left[ 1 + \left( \frac{1}{\theta^2} \right) \left( \frac{\hat{a}^2}{\hat{\beta}^2} \left( \frac{\theta_0}{1+\theta_0} \right)^2 \right) \right] \\
&\quad + \left[ a - \hat{\beta} \left( \frac{\hat{a}}{\hat{\beta}} \right) \left( \frac{\theta_0}{1+\theta_0} \right) \right]^2 \\
&= \frac{\sigma^2}{\theta_0^2} + \left( \frac{\hat{a}}{\theta_0} \right)^2 \left[ \frac{\hat{a}^2}{\hat{\beta}^2} \left( \frac{\theta_0}{1+\theta_0} \right)^2 \right] + a^2 \left[ 1 - \left( \frac{\theta_0^2}{1+\theta_0} \right) \right]^2 \\
&= \frac{\sigma^2}{\theta_0^2} + \left( \frac{\hat{a}}{\theta_0} \right)^2 \left[ \frac{\hat{a}^2}{\hat{\beta}^2} \left( \frac{\theta_0}{1+\theta_0} \right)^2 \right] + a^2 \left[ 1 - \left( \frac{\theta_0^2}{1+\theta_0} \right) \right]^2 \\
&= \frac{\sigma^2}{\theta_0^2} + \left( \frac{\hat{a}}{\theta_0} \right)^2 \left( \frac{\theta_0}{1+\theta_0} \right)^2 + a^2 \left[ 1 + \left( \frac{\theta_0^2}{1+\theta_0} \right) - \frac{2 \theta_0^2}{1+\theta_0} \right] \\
&= \frac{\sigma^2}{\theta_0^2} + \left( \frac{\hat{a}}{\theta_0} \right)^2 \left( \frac{\theta_0}{1+\theta_0} \right)^2 + a^2 \left( \frac{\theta_0^2}{1+\theta_0} \right)^2 - \frac{2 \theta_0^2 a^2}{1+\theta_0} \\
&= \frac{\sigma^2}{\theta_0^2} + a^2 \left( \frac{\theta_0}{1+\theta_0} \right)^2 \left[ 1 + \frac{1}{\theta_0^2} \right] + a^2 - \frac{2 \theta_0^2 a^2}{1+\theta_0} \\
&= \frac{\sigma^2}{\theta_0^2} + a^2 \left( \frac{\theta_0}{1+\theta_0} \right)^2 \left[ \frac{\theta_0^2+1}{\theta_0^2} \right] + a^2 - \frac{2 \theta_0^2 a^2}{1+\theta_0} \\
&= \frac{\sigma^2}{\theta_0^2} + a^2 - \frac{a^2 \theta_0^2}{1+\theta_0} \\
&= \frac{\sigma^2}{\theta_0^2} + a^2 - \frac{a^2 \theta_0^2}{1+\theta_0} \\
&= \sigma^2 + \frac{a^2 + a^2 \theta_0^2 - a^2 \theta_0^2}{1+\theta_0} \\
&= \sigma^2 + \frac{a^2}{1+\theta_0}.
\end{align*}
\]
And, $E[L \mid w = w_{ec}] = E[(z - a)^2 \mid w = w_{ec}]$

\[
= \left(\frac{w_{ec}^2}{u^2}\right) \left[ 1 + \frac{1}{w_{ec}^2} \left(\frac{a^2}{\beta^2}\right) \right] + (a - \beta)^2
\]

\[
= \tilde{s}^2 + \frac{a^2}{to^2} (a^2 / \beta^2)
\]

\[
= \tilde{s}^2 + \frac{a^2}{to^2}
\]

\[\therefore E[L \mid w = w_{ec}] - E[L \mid w = w^*] = \tilde{s}^2 + \frac{a^2}{to^2} - \tilde{s}^2 - \frac{a^2}{1+to^2} = a^2 \left[ \frac{1+to^2-to^2}{to^2(1+to^2)} \right]
\]

\[
= \frac{a^2}{to^2(1+to^2)}.
\]

So, the increase in expected loss when $w_{ec}$ is used instead of $w^*$ depends not only on precision, as measured by $to$, but also on the value of the target, $\beta$. Only if $a = 0$ to have no increase in expected loss when $w_{ec}$ is used in place of $w^*$. And this is because we assumed that the regression goes through the origin i.e., there is a zero intercept.

For target values far from the origin, then the extra loss when $w_{ec}$ is used may well be substantial.

Write the relative expected loss as:

\[REL = \frac{E[L \mid w = w^*]}{E[L \mid w = w_{ec}]} = \left(\frac{\tilde{s}^2 + \frac{a^2}{(1+to^2)}}{\frac{\tilde{s}^2 + \frac{a^2}{to^2}}{to^2}}\right)\]

\[= \left[1 + \frac{\frac{a^2}{\beta^2}}{(1+to^2)}\right] / \left[1 + \frac{\frac{a^2}{(1/\beta)^2}}{(to^2)}\right] \leq 1.\]

Hence:

\[\text{In particular, the reduction in expected loss obtained by using $w^*$ (rather than $w_{ec}$) is greater the smaller is $\frac{to^2}{\beta^2}$ and the larger is $(\frac{1}{\beta})^2$.}\]

Consider the first statement -- as $to^2$ gets small, it becomes relatively better to use $w^*$. This makes sense, since $to^2 = \beta^2 / (3/4u^2)$. And $\beta^2 / u^2$ is a variance related to $\beta^2$.

as $to^2$ gets small, variance $(\beta)$ gets big.

So, the less accurate is your estimate of $\beta$, the more profitable it becomes to use $w^*$ instead of $w_{ec}$. 

\[
\]
Example: Let \( \begin{align*} 
\frac{d\hat{y}_t}{dt} &= \Delta GNP \\
\hat{x}_t &= \Delta M_t 
\end{align*} \)

Data for 1921-1929 given, use a simple prior pdf \( \pi \).

Find \( x_{1930} \Rightarrow \hat{y}_{1930} \) for 1930 = $10m.

Take \( L(2, a) = (2 - 10)^2 \); \( 2 = y_{1930} \).

The regression estimates are \( \hat{y}_t = 2.0676 \hat{x}_t \)
\( (0.4993) \)

\[ \begin{align*} 
W^* &= 2.693 \quad ; \quad E(L|W=W^*) = 35.29 \\
W_{cc} &= 4.837 \quad ; \quad E(L|W=W_{cc}) = 60.03.
\end{align*} \]

So there is an 8% gain in using \( W^* \) cf. \( W_{cc} \).

Now, let us generalize the analysis a little. In particular, modify the loss form so that it takes account of possible costs involved in changing the control variable.

Let \( L = (2-a)^2 + c(W - x_t)^2 \)
where \( c \) is known, \( a \) - negative, \( c \) constant. And \( W = x_{1930} \).

\[ \begin{align*} 
E(L) &= E(2-a)^2 + E[c(x-x_t)^2] \\
&= \text{Var}(2) + (a - E(2))^2 + c(W - x_t)^2 \\
&= \left( \frac{\sigma^2}{V-2} \right) (1 + \frac{\text{Var}}{E^2}) + (a - \bar{x})^2 + c(\bar{x} - x_t)^2 \\
&= \sigma^2 (1 + \frac{\text{Var}^2}{\text{Max}^2}) + \bar{x}^2 (\frac{\bar{x}}{\text{Max}^2} - \bar{x})^2 + c(\bar{x} - x_t)^2
\end{align*} \]

And we wish to minimize \( E(L) \).

Let min. 1st term \( \frac{\partial}{\partial a} = 0 \)

\[ \begin{align*} 
\frac{\partial}{\partial W} &= 0 \quad ; \quad W = (a/\beta) \\
\frac{\partial}{\partial x_t} &= 0 \quad ; \quad x_t = x_t
\end{align*} \]

Now, \( \frac{\partial E(L)}{\partial W} = \sigma^2 \left( \frac{2W}{\text{Max}^2} \right) + 2\beta^2 (-1)(\frac{a}{\beta} - W) \\
+ 2c(1)(W - x_t) \\
= 0 \\
W^* \left[ \frac{2\sigma^2}{\text{Max}^2} + 2\beta^2 + 2c \right] - 2a\beta - 2c x_t = 0\)
\[ w^{**} = \left( \frac{a^2 \beta + c X_T}{a^2 + c + \frac{1}{\hat{\gamma} \text{max}}} \right) \]

and this just a weighted average of \( y_{20}, (x/\beta), \) 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

Multiple Regression:

Let \( y' = (y_1, y_2, \ldots, y_T) \) be generated according to

\[ y = X \beta + u \]

Assume that each explanatory variable can be controlled. (This assumption may be relaxed. Zellner looks at this possibility later in the chapter.)

Let \( u_t \sim N(0, \sigma_u) \); \( t = 1, 2, \ldots, T. \)

Assume \( \rho(\beta, \sigma) \propto \frac{1}{\sigma}; \quad 0 < \sigma < \infty; \quad -\infty < \beta < \infty; \quad i = 1, \ldots, k \)

Let \( z = y_{t+1}, \omega = (x_{1t+1}, x_{2t+1}, \ldots, x_{kt+1}) \)

\[ z = w^1 \beta + u_{t+1}. \]

where \( u_{t+1} \sim N(0, \sigma_u) \)

Let \( Z(z, \omega) = (z - \omega) \]

\[ p(z | y, \omega) \propto [\frac{1}{u + (z - w^1 \beta)^2} H]\]^{-\frac{(u+1)/2}{2}}

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

\[ H = \frac{1}{\hat{\gamma}} (1 + \omega' (X'X)^{-1} \omega)^{-1} \]

\[ u = (T-k) \]

\[ u_{st} = (y - X \hat{\beta})' (y - X \hat{\beta}) \].
\[
\begin{align*}
\text{Let } E(z) &= \psi_0 \\
}\text{Var}(z) &= \left( \frac{H - \nu}{\nu - 2} \right) = \beta^2 \left[ 1 + \omega' (x'x)^{-1} \omega \right] \\
\text{where } \beta^2 &= \left( \frac{\nu s^2}{\nu - 2} \right).
\end{align*}
\]

Now, 
\[
E(z - a)^2 = \text{Var}(z) + (a - E(z))^2
\]
\[
= \beta^2 \left[ 1 + \omega' (x'x)^{-1} \omega \right] + (a - \omega \beta)^2
\]
\[
\therefore \frac{\partial E(z)}{\partial \omega} = 2 \beta^2 (x'x)^{-1} \omega - 2a \hat{\beta} + 2 \hat{\beta} \omega \omega = 0
\]
\[
i \left( \beta^2 (x'x)^{-1} + \hat{\beta} \hat{\beta}' \right) \omega^* = a \hat{\beta}
\]
\[
\therefore \omega^* = \frac{a \hat{\beta}}{\beta^2 (x'x)^{-1} + \hat{\beta} \hat{\beta}'}
\]
\[
\begin{align*}
\omega^* &= \frac{x' \beta a}{(\beta^2 + \hat{\beta} \beta') (x'x)^{-1}} \\
\omega^* &= \omega^* \hat{\beta}' \quad \text{say}
\end{align*}
\]

Then, 
\[
\hat{\beta}^* = \hat{\beta} \left(1 + \frac{1}{(\beta'x'x\beta')/\beta^2} \right)
\]

And this is just the implied estimator \( \hat{\beta}^* \).

Then, 
\[
E(L| \omega = \omega^*) = \beta^2 \left[ 1 + \omega^* \hat{\beta}' (x'x)^{-1} \omega^* \right] + (a - \omega^* \hat{\beta})^2
\]
\[
= \beta^2 \left[ 1 + \frac{a' \hat{\beta}' (x'x)^{-1} (x'x) \beta a}{\beta^2 + \hat{\beta} \beta'} \right] \\
+ \left( a - \frac{a' \hat{\beta}' (x'x)^{-1} (x'x) \beta a}{\beta^2 + \hat{\beta} \beta'} \right)^2
\]
\[
= \beta^2 \left[ 1 + \frac{a^2}{\beta^2 + \hat{\beta} \beta'} \right]
\]

This is the expected bias when \( \omega \) is set at its optimal value, \( \omega^* \). The first term, \( \beta^2 \), does not disappear as \( n \to \infty \), but the second term \( \to 0 \) as \( n \to \infty \).

Also, 
\[
\hat{\beta}^* \to \hat{\beta} \left(1 + \frac{1}{(\hat{\beta}' \hat{\beta})/\beta^2} \right) \to \hat{\beta} \quad \text{as } n \to \infty.
\]
Note: Section C on Multivariate Regression is omitted.

(1) Sensitivity & Control to Form of the Loss Function:

Consider the whole problem of robustness under changes in the loss function. Let's 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:
\[ u_t \sim N(0, \sigma^2) \]
\[ x_t \sim N(0, 0.64) \]

Let:
\[ L(z, a) = |z - a|^x \quad ; \quad x = 0.5, 2, 4. \]

For each \( J \) be 4 alternative loss forms, compute the value

\[ \int_{-\infty}^{\infty} L(z, a) p(z | y, \omega) \, dz \]

where:
\[ p(z | y, \omega) = \frac{\Gamma(\nu/2)}{(\pi \nu)^{1/2}} \left( \frac{3}{4} \right)^{1/2} \left[ 1 + \frac{3}{4} (z - \omega)^2 \right]^{-\nu/2} \]

where \( \nu = 14. \)

Zellner finds that for this problem the optimal solution for the required error loss function is remarkably robust under changes in the form of the loss function, as long as we restrict ourselves to symmetric loss functions.