THE ALGEBRA OF THE SNLM

The model is

$$\underset{T\times 1}{Y} = X \underset{k\times 1}{\beta} + \varepsilon \tag{1}$$
$$\{\varepsilon \mid X, \beta\} \sim N(0, \sigma^2 I)\,.$$

This can be expressed equivalently as

$$\{Y \mid X, \beta\} \sim N(X\beta, \sigma^2 I)\,,\ \text{or} \tag{2}$$
$$p(Y \mid X, \beta) = (2\pi)^{-T/2}\sigma^{-T} \exp\left(-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right)\,. \tag{3}$$

1. $\{\beta \mid Y, \sigma^2\}$

Observe that the exponent is a quadratic form in Y and β. It is not hard to show that

$$\underset{\beta}{\operatorname{argmin}}\big((Y - X\beta)'(Y - X\beta)\big) = (X'X)^{-1}X'y = \hat{\beta}_{OLS}\,, \tag{4}$$

That is that the expression in the middle is the value of β that minimizes the sum of squared **residuals**, where the residuals are $u = Y - X\beta$. If we set $\hat{u} = Y - X\hat{\beta}_{OLS}$, then we can rewrite the exponent in (3) as

$$-\frac{1}{2\sigma^2}\hat{u}'\hat{u} - \frac{1}{2\sigma^2}(\beta - \hat{\beta}_{OLS})'X'X(\beta - \hat{\beta}_{OLS}) \tag{5}$$

We first derive the posterior distribution of the unknown parameters β, σ under a flat prior on β and $\log\sigma$. Note that this can be expressed as $d\sigma/\sigma$, $d\sigma^2/\sigma^2$, or $d\nu/\nu$, where $\nu = 1/\sigma$. All these improper priors are equivalent under the change of variables formula.

From (5) it is easy to see that the posterior pdf (prior times likelihood) as a function of β, with Y, X, σ^2 fixed, is proportional to a $N(\hat{\beta}_{OLS}, \sigma^2(X'X)^{-1}$. Thus $\{\beta \mid Y, X, \sigma^2\} \sim N(\hat{\beta}, \sigma^2(X'X)^{-1}$.

If instead we hold X, Y, β fixed and consider the likelihood as a function of σ^2, it is fairly easy to see that it is in the form of an inverse-gamma$(T, u'u)$. (Note that what appears here is u, the residual vector that depends on β, not $\hat{u}$, the OLS residual vector that does not depend on β.) With the prior in $d\sigma/\sigma$ form, this is slightly tricky, so we'll go through it. The likelihood multiplied by the prior is proportional to

$$\sigma^{-T-1} \exp\left(-\frac{u'u}{2\sigma^2}\right)\, d\sigma\, d\beta\,. \tag{6}$$

Making the change of variables to $\nu = 1/\sigma^2$ and recognizing that $|d\sigma| = \nu^{-3/2} d\nu$, this becomes

$$\nu^{T/2-1} \exp(-\tfrac{1}{2}\nu u'u)\, d\nu\, d\beta\,, \tag{7}$$

which is clearly proportional to a gamma$(T/2, u'u/2)$ pdf when treated as a function of ν. Note that if U is distributed as gamma$(T/2)$, then $2U$ is distributed as $\chi^2(T)$, so here $\nu u'u \sim \chi^2(T)$. The $\chi^2(T)$ distribution is the distribution of the sum of squares of T independent $N(0,1)$ random variables.

Since each of these conditional distributions is of standard form, it is easy to integrate w.r.t. either ν or β to arrive at a marginal pdf for the other parameter. Integrating (7) with respect to ν, we arrive at

$$(u'u)^{-T/2} = \frac{1}{\dfrac{\hat{u}'\hat{u}}{T-k}\left(1 + \dfrac{(\beta - \hat{\beta}_{OLS})'X'X(\beta - \hat{\beta}_{OLS})}{\hat{u}'\hat{u}/(T-k)}\right)^{T/2}}\,, \tag{8}$$

which, as a function of β, is proportional to a **multivariate** $\mathbf{t}_{T-k}(\beta, s^2(X'X)^{-1})$ pdf, where $s^2 = \hat{u}'\hat{u}/(T-k)$.

As we showed in lecture, if $\underset{k\times 1}{X} \sim N(0,\Sigma)$, then $X'\Sigma^{-1}X \sim \chi^2(k)$. Therefore, conditional on σ,

$$\frac{(\beta - \hat{\beta}_{OLS})X'X(\beta - \hat{\beta}_{OLS})}{\sigma^2} \sim \chi^2(k)\,. \tag{9}$$

Since its conditional distribution given σ turns out not to involve σ, this quantity is independent of σ and, therefore, of ν. But then

$$\frac{(\beta - \hat{\beta}_{OLS})X'X(\beta - \hat{\beta}_{OLS})/m}{\hat{u}'\hat{u}/(T-k)} \tag{10}$$

is the ratio of two independent χ^2 variables, the numerator with m and the denominator with $T-k$ degrees of freedom, multiplied by $(T-k)/m$. This is one definition of the $F(k, T-k)$ distribution. Thus we can find the amount of posterior probability inside the ellipse defined by the level curves of the posterior pdf (8) by looking up values in a table of the F distribution.

2. Non-Bayesian distributions: $\hat{\beta}, s^2 \mid \beta, \sigma^2$

Using the fact that if $X \sim N(a,b)$, then $c'x \sim N(c'a, c'bc)$ (which applies when c is any $m \times k$ matrix), it is easy to show that $\{\hat{\beta}_{OLS} \mid \beta, \sigma\} \sim N(\beta, \sigma^2(X'X)^{-1})$. Also $\hat{u} = (I - X(X'X)^{-1}X')\varepsilon$, which is distributed, conditional on X and σ^2, as $N(0, \sigma^2 M)$, where $M = I - X(X'X)^{-1}$. Using the fact that $X'M = 0$ and that uncorrelated jointly normal variables are independent, we can then say that the distribution of (10), considered as a random function of $\hat{\beta}_{OLS}$ and s^2 and conditioning on β, σ, is the ratio of two independent random variables. The numerator is clearly distributed as $\chi^2(k)$. It is easy to check that $M^2 = M$, which is what is meant by saying M is **idempotent**.

It can be shown that such a matrix can always be written as $M = W'DW$, where $W'W = I$ and D is a diagonal matrix with nothing but zeros and 1's on the diagonal. The **trace** (sum of diagonal elements) of D is the number of nonzero elements on the diagonal of D. It then follows that if $X \sim N(0, I)$, $MX \sim N(0, M)$, and further that $X'MX \sim \chi^2(\text{trace}(M))$.

The trace operator satifies $\text{trace}(AB) = \text{trace}(BA)$ and is linear. So $\text{trace}(I - X(X'X)^{-1}X') = T - k$. Thus the denominator of (10) is distributed as $\chi^2(T - k)$. So we conclude that (10) has an F distribution with k and $T - k$ degrees of freedom. Using this as a pivot will allow us to generate elliptical confidence regions for β that exactly coincide with the elliptical HPD regions whose posterior probabilities match the confidence levels.