IMPLEMENTING CONJUGATE PRIORS IN THE SNLM WITH DUMMY OBSERVATIONS

1. Starting from a normal-inverse-gamma

Supposse we want to implement a prior that has a marginal distribution on $\nu = 1/\sigma^2$ that is Gamma(p, α) and conditional distribution for $\beta \mid \sigma^2$ that is $N(\mu, \sigma^2 \Omega)$. Begin by finding W such that $W'W = \Omega^{-1}$. There is always such a W, indeed many of them. One is the inverse of the cholesky decomposition. In R, this would be `W<-t(solve(chol(Omega)))`, in Matlab `W=inv(chol(Omega)')`, in Rats `comp W = inv(decomp(Omega))`. Then choose as a first set of dummy observations

$$Y^* = W\mu$$
$$X^* = W$$

This will take care of the conditional distribution of $\beta \mid \sigma^2$.

To add a factor reflecting the proper prior on ν, add $2p$ additional dummy observations, each with $Y = \sqrt{\alpha/p}$ and $X = 0$. Note that since the number of dummy observations is discrete, this approach can only implement Gamma priors on ν that have p an integer multiple of $\frac{1}{2}$.

While this formally takes care of the problem, one more often works more directly with the dummy observations. One can think of reasonable X_t^* vector values (where X_t^* is a single hypothetical row of the X matrix), and for each one what a reasonable corresponding Y^* value would be. Then one can ask oneself what would a reasonable "standard error" on this correspondence between X_t^* and Y_t^* be, and multiply both Y_t^* and X_t^* by the ratio of what you think the equation residual standard error should be to your subjective standard error on this dummy observation. Proceeding this way, one need not be limited to a specific number of dummy observations. So long as there are at least $k+1$ dummy observations, with a full column rank X^* matrix, the implied prior pdf will be proper.

One reasonable standardized proper prior might set $\Omega = (X'X/T)^{-1}\kappa/k$, where k is the number of columns in X and $\kappa/(1+\kappa)$ is the fracation of pre-sample uncertainty about $Y'Y$ one wants to attribute to uncertainty about β, as opposed to uncertainty about the error terms. This of course still leaves open the choice of the mean of β and the values of p and α. Choosing $p = \frac{1}{2}$ or $p = 1$ is reasonable, as these priors do not imply a local peak in the prior pdf on ν anywhere except at zero.