EXERCISE ON DUMMY OBSERVATION PRIORS AND ODDS RATIOS ON RESTRICTIONS

You have already estimated, without a prior (i.e. with a flat prior) a regression of log wages on years of education (**educ**) state of birth (**pob**), year of birth (**yob**), and an education/year-of-birth interaction (**educ** $\cap$ **pob**). We saw that F-tests of restrictions on blocks of variables in this regression tended to deliver extreme p-values, and that BIC in a few cases gave a different answer than an F-test.

In this exercise you formulate a proper prior on that model, and calculate posterior odds ratios, checking sensitivity of results to some parameters of the prior and comparing to the results from BIC and the F tests.

Part of the point of the exercise is to get you thinking about how to specify a prior in a high-dimensional space, so some parameters of the prior I suggest below are left to your judgment.

You will specify a prior using dummy observations. If Y^* and X^* are dummy observations, with X^* $n \times k$, adding them to the sample implies a $N(\hat{\beta}, \sigma^2(X'X)^{-1})$ prior for β conditional on σ^2, where $\hat{\beta}$ is the usual OLS estimator based on X^* and Y^*. Because it is conditional on σ^2, the variance of the full model's residual, it does not imply a proper prior where the regression of Y^* on X^* fits perfectly, even if $n > k$. For example, with dummy observations of the form $\beta_1 = 0$, $\beta_2 = 0$ and $\beta_1 + \beta_2 = 0$, The Y^* values would all be zero and the corresponding $\hat{\beta} = 0$ would produce a perfect fit. (The $\beta_1 + \beta_2$ dummy observation is not redundant, as it would generate negative correlation between β_1 and β_2 in the prior.)

The simplest way to arrive at a proper prior is usually to add a dummy observation in which Y^* is non-zero, say λ, and all right-hand-side variables (including the constant) are zero. If the rest of the dummy observations have a $k \times k$ non-singular X^*, this additional dummy observation implies a $\text{Gamma}(3/2, \lambda^2/2)$ proper prior on $1/\sigma^2$. In other words, it implies a prior with the modal value for $1/\sigma^2$ equal to $1/\lambda^2$.

For the **educ** and **yob** dummy variables the prior should reflect the fact that coefficients on dummies for adjacent years should be closer than coefficients on dummies for distant years. Dummy observations that implement this idea, without pre-judging whether the effects of education or year of birth should be positive or negative, or even monotone, could have the form

$$X^* = A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} , \quad Y^* = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} , \tag{1}$$

of course weighted appropriately. This is a matrix that is all zeros except for ones down the main diagonal and -1's down the first diagonal below the main one. The dummy observation implies the sequence of coefficients is a random walk.

Of course there is one block of dummy observations with A in the columns corresponding to **educ**, and another block where a matrix of the form A (but of a different size) appears in the columns corresponding to **yob**.

For **pob** it probably makes sense to have X^* the identity, assuming that there is a constant in the system so that **pob** coefficients are only deviations from the common mean.

For **educ**$\cap$**pob**, X^* of the form $A \otimes I$ make sense, where $\otimes$ indicates Kronecker product. This forms assumes that in the real data X matrix the interaction dummies are ordered with the **pob** varying first, then **educ**. If it's the opposite, the dummies should take the form $I \otimes A$.

Finally, you need a dummy for the constant term. This should have a nonzero Y^*, so your prior belief is that the constant coefficient mean is around the mean you expect for log wages of the base category — 0 years of schooling, born in 1930, state 1.

You should use your judgment in deciding how to scale these dummy observations. The residual standard deviation of log wages is on the order of .5 or 1.0. You might also bear in mind that dummy observations Y^*, X^* imply that the prior expected explained sum of squares (assuming zero prior mean for β) is

$$E[\beta'(X'X)\beta] = \operatorname{tr}(\sigma^2 (X^{*\prime} X^*)^{-1} X'X) , \tag{2}$$

where X is the real data X matrix and σ^2 is the residual covariance. This in turn implies that the expected R^2 of the equation is

$$\left(\frac{N}{\operatorname{tr}((X^{*\prime} X^*)^{-1} X'X)} + 1 \right)^{-1} , \tag{3}$$

where N is sample size. This implies that choosing high a priori uncertainty about β, i.e. a small $X^{*\prime}X^*$, entails asserting a high expected R^2. You should check the implied R^2 from your prior, and possibly scale it up or down if the implied R^2 is unreasonable.

The easy way to form posterior odds on a set of linear restrictions is just to use the marginal data densities of the restricted linear combinations of the coefficients implied by substituting the estimate $\hat{\sigma}^2$ for σ^2 in the normal posterior distribution for the coefficients conditional on σ^2. It is possible to find the unconditional marginal posterior densities analytically in this standard normal linear model case, but doing so is not required for the exercise.

With the prior set up, find the four posterior odds for these blocks of coefficients being zero: **educ**, **yob**, **pob**, and **educ∩pob**. Compare these results to BIC and the standard F-tests. Also check sensitivity to at least one or two aspects of the prior. For example, scale the dummy observations up or down by two, or tighten or loosen the prior on **educ∩pob** relative to the other blocks of coefficients.

You may find it useful, if you work in R, to use `model.matrix` or `sparse.model.matrix`. However, keep in mind that with all the dummies present the X matrix will be singular. I'm not sure whether model.matrix recognizes the singularity and deletes some columns, or instead hands `lm.fit` a singular X and lets lets `lm.fit` figure out which columns to drop. In any case, you will need to keep track of this in order to properly implement the dummy observations.