MCMC, LOGIT/PROBIT, GLS

1. Gibbs Sampling: Why it works
2. Data-Augmentation
3. When it won't work
4. Checking convergence and accuracy

5. GLS

- This is the framework for several standard extensions of the SNLM. Of course extensions of a model always imply specification tests (via looking at posterior odds between a model and the extended version).
- We replace the assumption $\operatorname{Var}(\varepsilon) = \sigma^2 I$ with $\operatorname{Var}(\varepsilon) = \sigma^2\Omega$.
- If Ω is known, the same sort of algebra that led us to the result that in the SNLM with $d\sigma/\sigma$ prior we get a normal-inverse-gamma posterior on σ^2 and β leads us in this model to a different normal-inverse-gamma posterior:

$$\hat\beta_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$
$$\hat u = y - X\hat\beta_{GLS}$$
$$\left\{\frac{1}{\sigma^2}\Big| Y, X\right\} \sim \text{Gamma}\left(\tfrac{1}{2}(T-k), \tfrac{1}{2}\hat u'\Omega^{-1}\hat u\right)$$
$$\left\{\beta \mid Y, X, \sigma^2\right\} \sim N\left(\hat\beta_{GLS}, \sigma^2(X'\Omega^{-1}X)^{-1}\right)$$

6. Examples of Ω's

- Conditional heteroskedasticity: $\omega_{ij} = 0$, $i \neq j$, $\omega_{ii} = \sigma^2 g(X_i, \alpha)$.
- General serial correlation: $\omega_{ij} = \sigma^2\rho(i-j)$. I.e., Ω is constant down all diagonals. ($\rho(0) = 1$ as a normalization, or else omit σ^2.)
- First-order autoregressive serial correlation: $\rho(s) = \rho(1)^s$. So all of Ω is a function of a single parameter.
- Spatial correlation: $\omega_{ij} = \sigma^2\rho(\delta(i,j))$, where δ measures "distance".
- The cluster model: y indexed by, say, state i and county j. $\omega(i,j,k,\ell) = 0$ if $i \neq k$ (i.e. for observations from different states). $\omega(i,j,i,\ell) = \sigma^2\rho(i,j,\ell)$. Ω is block diagonal, if observations are grouped by state. This is a difficult model, easily leading to complicated likelihood functions. There is only a small amount of literature, all technically demanding, on the properties of the estimators. It is available with a single "button-push" in STATA, so it is widely used, even treated as standard in some fields, despite weak understanding of its properties.

7. OLS ANYWAY?

- OLS is unbiased, no matter what Ω might be. Its presample distribution is
$$\hat{\beta}_{OLS} \sim N\big(\beta, \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}\big)$$
- The unbiasedness is "robust" against non-scalar Ω, but to make any claims about precision of our estimates, we need to estimate Ω anyway.
- If σ^2 is known, and we observe $Y'X$ and $X'X$ but not Y and X separately, the posterior distribution under a flat prior on β is of exactly the same form as the normal presample distribution for $\hat{\beta}_{OLS}$, with the mean at $\hat{\beta}_{OLS}$.
- However if σ^2 is unknown, information on $Y'X$ and $X'X$ alone does not tell us much about σ^2. If we add $Y'Y$ to the list of observed statistics, we can derive Bayesian inference, but it will imply a nasty distribution for σ^2 and thus not the usual s^2 estimate based on OLS residuals.
- Still, often in large samples the uncertainty about σ^2 is small, so that even though $\hat{u}'\hat{u}/T$ is not the best possible estimate of σ^2, it is so accurate that treating σ^2 as known is reasonable.
- So using OLS as a convenient shortcut when we have an estimate of Ω has a Bayesian interpretation, so long as there is not much uncertainty about σ^2, given the sample.

8. ESTIMATING Ω

- Ω contains T^2 elements, which reduce to $T(T+1)/2$ free parameters when we take account of symmetry restrictions. (If σ^2 is also a free parameter, at least one element or function of elements, of Ω has to be normalized to some fixed value, so the total number of free parameters stays the same.)
- As you might suspect, estimating Ω with a flat prior therefore doesn't work. The number of free parameters in the model exceeds the number of observations, and by a larger amount the larger is T.
- So we must always either use an informative prior on Ω, and accept that the prior will affect inference even in large samples, or else treat Ω as a function of a finite number of parameters that does not grow, or grows only slowly, with sample size.
- This reflects a more general point: improper priors are reasonable only when we can expect that the likelihood will dominate the prior. We will see that this is likely to be true in large samples under broad regularity conditions. But if the number of parameters grows linearly in T or faster, there is usually no hope that the likelihood dominates the prior.

9. FEASIBLE GLS

- This is GLS with Ω estimated, always by making it a function of a smaller number of parameters.

- Commonly in practice, one starts with OLS, forms $\hat{u}_{OLS}$, uses these estimated residuals as if they were the actual ε's to form an estimate $\hat{\Omega}$ of Ω, then applies the GLS formulas with $\hat{\Omega}$ replacing Ω, ignoring the fact that $\hat{\Omega}$ is only an estimator.
- How can this be justified? There are large-sample justifications that we will consider later. They consist of showing that when the sample is large there is little cross-dependence between Ω and β in the likelihood, so that inference that accounts for uncertainty in Ω is not much different from inference that treats estimates as exact.
- However this relies on the number of free parameters in Ω being small. We don't actually usually have solid a priori knowledge about Ω, so there is a tendency to push the limits, using fairly large numbers of free parameters. So in practice it is not uncommon for inference about β to be rather strongly affected by whether we account for uncertainty in Ω.

10. Gibbs for GLS

It is clear that a nice Gibbs sampling scheme is available for GLS:

(i) Fix Ω, apply GLS to get a standard posterior for $\{\beta, \sigma^2 \mid \Omega\}$ and make a draw from it.

(ii) Use the draw of β to form residuals, and use these and the draw of σ^2 to form a conditional pdf for Ω. This can of course ignore uncertainty about σ^2 and β, but the inference is sometimes not straightforward. Might need Metropolis-within Gibbs or the like.

11. Can it happen that GLS≡OLS even though $\Omega \neq I$?

- Yes.
- The condition is that $\Omega \underset{T\times k}{X} = X\Lambda$, for some $k \times k$ Λ. This means that the columns of X are in the space spanned by k eigenvectors of Ω. This may seem a knife-edge special case, but there is one situation where it is fairly often relevant in practice.
- If X contains a column of ones, as it usually does, and if $\Omega = I + \lambda \mathbf{1}\mathbf{1}'$, we are in the special case. And such an Ω matrix arises if we assume that correlation among residuals arises because $\varepsilon_i = \nu + \xi_i$, where ξ_i is i.i.d. and ν is a common error component.
- Note, though, that even though $\hat{\beta}_{OLS} = \hat{\beta}_{GLS}$, $X'\Omega^{-1}X \neq X'X$. And the usual two-step approach to feasible GLS won't work here. Since the model is implicitly
$$y_t = \beta_0 + X_{1t}\beta_1 + \nu + \varepsilon_t\,, \tag{1}$$
and there is only one ν in a given sample, we cannot hope to separate β_0 from ν by looking at the data.