TECHNICAL DETAILS

Model

There are 2017 model parameters, collectively denoted by the vector

$$\theta = \Big[\ \sigma_F,\ \underbrace{(\alpha_1,\delta_1,\varepsilon_1')}_{Fpre},\ \underbrace{(\alpha_2,\delta_2,\varepsilon_2')}_{Fpost},\ \underbrace{(\alpha_3,\delta_3,\varepsilon_3')}_{Start\,(t_{10})},\ \underbrace{(\alpha_4,\delta_4,\varepsilon_4')}_{End\,(t_{90})}\ \Big]'$$

The α and δ terms are scalars, the ε terms are 502x1 vectors of region-specific random effects, and σ_F is a non-negative scalar. The four sets of (α, δ, ε) terms correspond to the four logistic constants *Fpre*, *Fpost*, t_{10}, and t_{90} in the text, respectively. Given θ, the logistic constants for the trend in region *i*=1…502 are

$$Fpre_i = \alpha_1 + \delta_1\left(Ed_{i,1970} - \overline{Ed}_{1970}\right) + \varepsilon_{1i}$$

$$Fpost_i = \alpha_2 + \delta_2\left(Ed_{i,2000} - \overline{Ed}_{2000}\right) + \varepsilon_{2i}$$

$$t10_i = \alpha_3 + \delta_3\left(Ed_{i,1970} - \overline{Ed}_{1970}\right) + \varepsilon_{3i}$$

$$t90_i = \alpha_4 + \delta_4\left(Ed_{i,1970} - \overline{Ed}_{1970}\right) + \varepsilon_{4i}$$

where $\overline{Ed}_t = \frac{1}{502}\sum_i Ed_{it}$ is an average education level across locations at time *t*. The estimated logistic time trend for TFR in region *i* is

$$m_{it}(\theta) = Fpost_i + \frac{Fpre_i - Fpost_i}{1+\exp\left[-\beta_i\left(t-\tau_i\right)\right]}$$

where the speed (β) and midpoint (τ) of region *i*'s transition are

$$\beta_i = \ln 81/\left(t10_i - t90_i\right) \quad , \quad \tau_i = \left(t10_i + t90_i\right)/2$$

and time is scaled in decades since 1960 (thus $t10=0$ represents 1960, $t10=1.3$ represents 1973, and so on).

Likelihood

We model the likelihood of the census-estimated fertility rates given θ as

$$F_{it} \mid \theta \sim Normal\left[\; m_{it}(\theta)\,,\; \frac{\sigma_F^2}{n_{it}} \;\right]$$

and further assume that estimation errors $F_{it}-m_{it}(\theta)$ are independent across (i,t) cells. Using T_i to denote census dates for which estimates are available in region i and N=2405 as the total number of rate estimates, the likelihood of the fertility data is thus

$$L\big(F \mid \theta\big) \;\propto\; \sigma_F^{-N}\ \exp\left[\; \sum_i \sum_{t \in T_i} -\frac{1}{2\sigma_F^2}\, n_{it}\, \left[F_{it} - m_{it}(\theta)\right]^2 \right]$$

Priors

With no advance knowledge about the intercept terms for the four logistic constants, we use completely flat priors with (improper) densities

$$f\,(\alpha_k) \;\propto\; 1 \qquad k = 1...4$$

Knowing little in advance about the relationship between regional educational levels and regional logistic constants, we use prior distributions with high variance for the δ parameters. Specifically we assume $\delta_k \sim Normal(0,1000)$, so that

$$f\,(\delta_k) \;\propto\; \exp\left[\,-.0005\, \delta_k^2\,\right] \qquad k = 1...4$$

We also use an uninformative prior for the variance of F, modeling its inverse (known as the *precision*) as $(1/\sigma_F^2) \sim \Gamma(.01,.01)$. This specification is common in the Bayesian literature. It implies that we have little knowledge of σ_F, but before seeing the data we know that σ_F>0 and we believe high values to be more likely than low values.

The key priors in our analysis are those for 2008 region-specific effects ε. For each of the four logistic constants k=1…4, $\alpha_k + \delta_k\left(Ed_i - \overline{Ed}\right)$ gives the average value of the constant for regions with the same education level as region *i*. The additive ε_{ki} term represents an additional local deviation that makes the logistic constant higher or lower than expected. Our prior is that local deviations for each logistic constant are spatially patterned, so that when two regions *i* and *j* are adjacent on the map they are likely to have similar local deviations in *Fpre*, *Fpost*, t_{10}, and t_{90}. To model this we use the conditional normal autoregressive (*car.normal*) distribution in *WinBUGS*, with priors of the form

$$f\left(\varepsilon_k \mid \lambda_k\right) \propto \exp\left[-\frac{\lambda_k}{2}\sum_i\sum_j w_{ij}\left(\varepsilon_{ki}-\varepsilon_{kj}\right)^2\right] \qquad k=1...4$$

where w_{ij}=1 if regions *i* and *j* share a border and w_{ij} =0 otherwise. This prior states that parameter sets ε are more likely when neighboring regions have similar values. The higher the value of the precision λ_k, the more we 'insist' on spatially smooth parameter sets. Because we do not have advance knowledge about how smooth the map of regional effects should be, we make the spatial smoothness levels $\lambda_1 \ldots \lambda_4$ *hyperparameters* (parameters for the distribution of other parameters), and to complete the model we must add *hyperpriors* (priors for the hyperparameters). For each λ term we use the same diffuse gamma distribution that appeared earlier for the precision of *F* , $\lambda_k \sim \Gamma(.01,.01)$ for k=1…4, which gives us hyperpriors

$$f\left(\lambda_k\right) \propto \Gamma(.01,.01) \text{ density} \qquad k=1...4$$

Posterior

With this model specification the posterior distribution of a particular parameter set θ, given the set of fertility estimates $\{F_{\mathrm{it}}\}$ is

$$P(\theta \mid F) \propto \int_{\lambda} L(F \mid \theta)\ f(\theta \mid \lambda)\ f(\lambda)\ d\lambda$$

where priors can be decomposed as

$$\begin{aligned} f(\theta \mid \lambda)\ f(\lambda) = \ & f(\sigma_F) \\ & \text{x } f(\alpha_1)\cdots f(\alpha_4) \\ & \text{x } f(\delta_1)\cdots f(\delta_4) \\ & \text{x } f(\varepsilon_1 \mid \lambda_1)\cdots f(\varepsilon_4 \mid \lambda_4) \\ & \text{x } f(\lambda_1)\cdots f(\lambda_4) \end{aligned}$$

WinBUGS allows a simple syntax for this complex model, and draws Monte Carlo samples of θ values (recall that θ is 2017-dimensional) from the correct posterior distribution, given the data. Empirical summaries of these samples [especially mean values and percentiles for the logistic constants ($Fpre_1 \ldots Fpre_{502}$) …($t90_1 \ldots t90_{502}$)] provide point estimates and bands of likely values for parameters and other measures of interest.

Initialization

After extensive experimentation with the posterior distribution, we used initial parameter values of

$$\begin{aligned} &\sigma_F = 1, \\ &\alpha_1 = 6.5, \quad \alpha_2 = 1.7, \quad \alpha_3 = 1.0, \ \alpha_4 = 3.5, \\ &\delta_1 = -0.50, \ \delta_2 = -0.25, \ \delta_3 = -0.50, \ \delta_4 = -0.10, \\ &\varepsilon_{ki} = 0, \quad \lambda_k = 1 \quad \forall\ k = 1...4,\ i = 1...502 \end{aligned}$$

These initial values are near the mode of the posterior density, so that starting the Markov Chain at this point yields a thorough exploration of the parameter subspace that is most relevant for inference.

MCMC Sampling

Using *WinBUGS 1.4*, run from within *R*, we drew 52000 realizations from the posterior, discarded the first 2000 as a burn-in period, and saved every 25th sampled observation from the remaining 50000 draws to reduce serial correlation between successive sampled values. Standard checks indicated that the procedure had converged to a stable posterior distribution. This yielded a final sample of 2000 realizations of θ from the posterior, whose means and percentiles we report in the paper. *R* and *WinBUGS* programs for replicating the analysis are available from the web archive at http://schmert.net/BayesLogistic.