A MCMC Estimation Steps for the Random Coefficients Models

The previously presented model (for the cumulative imitation specification) is based on the following equations:

$$
\begin{aligned}
c_{kt} &= \left(p_k + \sum_{k'=1}^{K} q_{kk'} \frac{C_{k',t-1}}{M_{k'}}\right)(M_k - C_{k,t-1}) + e_{kt} \qquad \text{(A.1)} \\
q_{kk'} &= \lambda_k \text{ if } k' = k \\
q_{kk'} &= \gamma_k \text{ if } k,\, k' \text{ are neighbors} \\
q_{kk'} &= 0 \text{ otherwise.} \\
e_{kt} &\sim N\left(0, \sigma_e^2\right) \cdot c_{k,t-1}
\end{aligned}
$$

with the parameters having the following distributions:

$$
\begin{aligned}
M_k &\backsim N\left(\mathbf{x}_{Mk}\boldsymbol{\beta}_M, \sigma_M^2\right) && \text{(A.2)} \\
p_k &\backsim N\left(\mathbf{x}_{pk}\boldsymbol{\beta}_p, \sigma_p^2\right) && \text{(A.3)} \\
\lambda_k &\backsim N\left(\mathbf{x}_{\lambda k}\boldsymbol{\beta}_p, \sigma_\lambda^2\right) && \text{(A.4)} \\
\gamma_k &\backsim N\left(\mathbf{x}_{\gamma k}\boldsymbol{\beta}_\gamma, \sigma_\gamma^2\right) && \text{(A.5)}
\end{aligned}
$$

In order for the model to be a meaningful diffusion model, all parameters are subject to non-negativity constaints. In addition, the parameter M_k is subject to the constraint that $M_k \geq \max_t(C_{kt})$. The parameters are estimated by drawing from their full conditional distributions. This process is illustrated below for M_k (the potential number of certifications in country k). M_k has a full conditional likelihood that combines the distribution of $c_{kt}|M_k$ with the empirical distribution of M_k itself.

$$
L\left(M_k \mid \text{rest}\right) \propto \prod_k \prod_t \left[\hat{c}_{kt} \mid M_k\right] \cdot \left[M_k\right] = \prod_k \prod_t N\left(\hat{c}_{kt}\left(M_k|\text{rest}\right), \sigma_{ek}^2\right) \cdot N\left(\mathbf{x}_{Mk}\boldsymbol{\beta}_M, \sigma_M^2\right), \quad \text{(A.6)}
$$

where

$$
\hat{c}_{kt}\left(M_k|\text{rest}\right) = \left(p_k + \sum_{k'=1}^{K} q_{kk'} \frac{C_{k',t-1}}{M_{k'}}\right)\left(M_k - C_{k,t-1}\right). \qquad \text{(A.7)}
$$

Because these distributions are a non-conjugate pair for M_k and we need to implement the logical constraints on the parameters, we use the Metropolis-Hastings algorithm to draw from the posterior distribution of M_k. We obtain a new value for M_k, M_k^{new}, by drawing δ_{Mk} from the appropriate Normal distribution, with mean zero and a variance obtained in a burn-in period that guarantees an acceptance rate of M_k^{new} between 30% and 40% (Gelman et al., 1995). M_k^{new} is defined by:

$$
M_k^{\text{new}} = M_k + \delta_{Mk}, \qquad \text{(A.8)}
$$

In each sweep of the sampler the new value M_k^{new} is accepted with probability θ:

$$
\theta = \min\left[\frac{L\left(M_k^{\text{new}} \mid \text{rest}\right)}{L\left(M_k \mid \text{rest}\right)}, 1\right] \qquad \text{(A.9)}
$$

Because it is impossible that $M_k < \max_t(C_{kt})$, we set the $L\left(M_k^{\text{new}} \mid \text{rest}\right) = 0$ when $M_k^{\text{new}} < \max_t(C_{kt})$. For the innovation parameter p_k and contagion parameters λ_k and γ_k, the estimation

steps are similar to the ones above for M_k. For the initial values of these parameters, we choose values that are larger than zero and similar to values presented in literature presented in the review section (again, we tested for different initial values and found insignificant differences in the final results).

To obtain the draws for the parameters from the lower level of the hierarchy, e.g., $\boldsymbol{\beta}_M$ and σ_M^2, we combine the higher level draws for M_k with the hierarchical model $M_k \backsim N\left(\mathbf{x}_{Mk}\boldsymbol{\beta}_M, \sigma_M^2\right)$. Denote the number of covariates in the matrix $\mathbf{x}_{Mk}$ by P_M. Further, define the priors $[\boldsymbol{\beta}_M] \sim N\left(\boldsymbol{\mu}_{0M}, \Sigma_{0M}\right)$ with $\boldsymbol{\mu}_{0M} = \mathbf{0}$ and $\Sigma_{0M} = 10000 \cdot \mathbf{I}_{P_M}$ and $\left[\sigma_M^2\right] \sim IG\left(1,1\right)$. Both prior distributions are diffuse for the parameter space of interest, having little impact on the final results. The posterior distributions are a result of a conjugation of the priors with the data distributions:

$$\begin{aligned}
\boldsymbol{\beta}_M &\sim N_{P_M}\left(\mathbf{V}_M\left(\Sigma_{0M}^{-1}\cdot\boldsymbol{\mu}_{0M} + \Sigma_M^{-1}\cdot\boldsymbol{\mu}_M\right), \mathbf{V}_M\right) && \text{(A.10)}\\
&= N_{P_M}\left(\mathbf{V}_M\left(\Sigma_{0M}^{-1}\cdot\boldsymbol{\mu}_{0M} + \frac{\mathbf{x}_M'\mathbf{M}}{\sigma_M^2})\right), \mathbf{V}_M\right), \text{ with} && \text{(A.11)}\\
\mathbf{V}_M &= \left(\Sigma_{0M}^{-1} + \Sigma_M^{-1}\right)^{-1} = \left(\Sigma_{0M}^{-1} + \frac{(\mathbf{x}_M'\mathbf{x}_M)}{\sigma_M^2}\right)^{-1} && \text{(A.12)}
\end{aligned}$$

and

$$\sigma_M^2 \sim IG\left(1 + 0.5K, 1 + 0.5\left((M - \mathbf{x}_M\boldsymbol{\beta}_M)'\left(M - \mathbf{x}_M\boldsymbol{\beta}_M\right)\right)\right) \tag{A.13}$$

For the error variance σ_e^2, the errors between estimated and actual number of certifications are calculated in each draw and standardized using the chosen scale, in this case, the last observed number of certifications:

$$\tilde{e}_{kt} = \frac{c_{kt} - \hat{c}_{kt}}{\sqrt{c_{k,t-1}}} \tag{A.14}$$

For the draws of the variance σ_e^2, we use the inverse gamma distribution. The diffuse prior is defined as $\sigma_{0e}^2 \sim IG\left(1,1\right)$. The draws are taken from the following posterior distribution:

$$\sigma_e^2 \sim IG\left(1 + 0.5K(T-1), 1 + 0.5\tilde{\mathbf{e}}'\tilde{\mathbf{e}}\right) \tag{A.15}$$

where $\tilde{\mathbf{e}}$ =vec($\tilde{e}_{kt}$).