APPENDIX G

This appendix describes the derivation of the full conditional distribution that are used for the Gibbs Sampler in Chapter 6. These distributions can be separated into three groups: group one is normaly distributed, group two gamma, and the distributions of group three do not belong to any known distribution. The derivation of the distributions within each group is very similar; to avoid tedious replications only the derivations of the distributions that are related to the parameters α, α_μ, and α_σ are shown.

For presentational ease it is convenient to define the following expressions:

$$\alpha \sim N(\alpha_\mu, \alpha_\sigma^{-1}), \tag{G.1}$$

$$\alpha_\mu \sim N(0,10000), \tag{G.2}$$

$$\alpha_\sigma \sim \Gamma(0.01,0.01), \tag{G.3}$$

$$p_{i,t} \sim P(\mu_{i,t}), \tag{G.4}$$

$$\mu_{i,t} = \alpha + Y_t + S_i + \tau\, TAX_{i,t} \quad . \tag{G.5}$$

<u>Full conditional distribution for α_μ:</u>

The distribution for α_μ conditional on $n_{-\alpha\mu}$ (all unknown nodes except α_μ) is proportional to the product of the normal prior of α_μ, given by G.2, and the normal likelihood, given by G.1. It is straightforward to show[1] that this conditional distribution is also normal, because

$$\begin{aligned} P(\alpha_\mu \mid n_{-\alpha_\mu}) &= \frac{100}{\sqrt{2\pi}}\, e^{-\frac{100}{2}(\alpha_\mu - 0)^2} \cdot \frac{\sqrt{\alpha_\sigma}}{\sqrt{2\pi}}\, e^{-\frac{\alpha_\sigma}{2}(\alpha - \alpha_\mu)^2} \\ &\propto e^{-\frac{1}{2}\left(100\,(\alpha_\mu - 0)^2 \,+\, \alpha_\sigma\,(\alpha - \alpha_\mu)^2\right)} \\ &\propto N\left(\frac{\alpha_\sigma \cdot \alpha}{.0001 + \alpha_\sigma^{-1}}\,,\, \frac{1}{.0001 + \alpha_\sigma^{-1}}\right) \end{aligned} \tag{G.6}$$

The full conditional distributions of τ_μ, $S_{\mu i}$ and $Y_{\mu t}$ can similarly be shown to be normal distributions, so that

$$P(\tau_\mu \mid n_{-\tau_\mu}) \propto N\left(\frac{\tau_\sigma \cdot \tau}{.0001 + \tau_\sigma^{-1}}\,,\, \frac{1}{.0001 + \tau_\sigma^{-1}}\right) , \tag{G.7}$$

---

[1] Box and Tiao (1973), pp. 74-75.

$$P(S_{\mu_i} \mid n_{-S_{\mu_i}}) \propto N\left(\frac{S_\sigma \cdot S_i}{.0001 + S_\sigma^{-1}} \; , \; \frac{1}{.0001 + S_\sigma^{-1}} \right) \; , \tag{G.8}$$

$$P(Y_{\mu_t} \mid n_{-Y_{\mu_t}}) \propto N\left(\frac{Y_\sigma \cdot Y_t}{.0001 + Y_\sigma^{-1}} \; , \; \frac{1}{.0001 + Y_\sigma^{-1}} \right) \; . \tag{G.9}$$

An efficient sampling algorithmus for the normal distribution is the algorithm by Box and Muller (1958).

<u>Full conditional distribution for α_σ:</u>

The distribution for α_σ conditional on $n_{-\alpha\sigma}$ is proportional to the product of the gamma prior of α_σ, given by G.3, and the normal likelihood, given by G.1. As

$$\begin{aligned} P(\alpha_\sigma \mid n_{-\alpha_\sigma}) &= \frac{.01^{.01} \cdot \alpha_\sigma^{.01-1} \cdot e^{-.01\ \alpha_\sigma}}{\Gamma(.01)} \cdot \frac{\sqrt{\alpha_\sigma}}{\sqrt{2\pi}}\ e^{-\frac{\alpha_\sigma}{2}(\alpha-\alpha_\mu)^2} \\ &\propto \alpha_\sigma^{.01-1+.5} \cdot e^{-\alpha_\sigma \left(.01 + \frac{1}{2}(\alpha-\alpha_\mu)^2 \right)} \\ &\propto \Gamma\left(.01 + \frac{1}{2} \; , \; .01 + \frac{1}{2}\ (\alpha-\alpha_\mu)^2 \right) , \end{aligned} \tag{G.10}$$

the full conditional distribution is proportional to a gamma distribution. The full conditional distributions of τ_σ, S_σ and Y_σ can similarly be shown to be gamma distributions, so that

$$P(\tau_\sigma \mid n_{-\tau_\sigma}) \propto \Gamma\left(.01 + \frac{1}{2} \; , \; .01 + \frac{1}{2}\ (\tau-\tau_\mu)^2 \right) , \tag{G.11}$$

$$P(S_\sigma \mid n_{-S_\sigma}) \propto \Gamma\left(.01 + \frac{M}{2} \; , \; .01 + \frac{1}{2} \sum_{i=0}^{M} (S_i - S_{\mu_i})^2 \right) \; , \tag{G.12}$$

$$P(Y_\sigma \mid n_{-Y_\sigma}) \propto \Gamma\left(.01 + \frac{T}{2} \; , \; .01 + \frac{1}{2} \sum_{i=0}^{T} (Y_i - Y_{\mu_t})^2 \right) \; , \tag{G.13}$$

where M is the total number of municipalities, and T is the total number of years examined. The full conditional distributions for S_σ and Y_σ are more complicated because the likelihood term does not only relate to one single expression as for α_σ and τ_σ, but to M or T different terms. To sample from the standardized gamma distribution with single parameter γ, it is necessary to use different algorithms that depend on the size of γ. For $\gamma<1$ I used Ahrens and Dieter's (1974) algorithm, and for $\gamma>1$ Cheng and Fast's (1979) algorithm. For $\gamma=1$ the gamma distribution becomes the exponential distribution, and I used the algorithm described by Press *et al* (1995).

Full conditional distributions for α:

The conditional distribution for α is proportional to the product of the normal prior of α, given by G.1, and $M{\cdot}T$ Poisson likelihood terms, given by G.4 and G.5, where $M{\cdot}T$ is the total number of observations:

$$
\begin{aligned}
P(\alpha \mid n_{-\alpha}) &\propto \sqrt{\frac{\alpha_\sigma}{2\pi}}\, e^{-\frac{\alpha_\sigma (\alpha-\alpha_\mu)^2}{2}} \cdot \prod_{i=1}^{M}\prod_{t=1}^{T} \frac{\mu_{i,t}^{p_{i,t}} e^{-\mu_{i,t}}}{x!} \\
&\propto e^{-\frac{\alpha_\sigma (\alpha-\alpha_\mu)^2}{2}} \cdot \prod_{i=1}^{M}\prod_{t=1}^{T} e^{p_{i,t}\cdot\alpha} \cdot \exp\left(-Mon_{i,t}\cdot Pop_{i,t}\cdot e^{\alpha}\cdot e^{Y_t+S_i+\tau\cdot TAX_{i,t}} \right) \\
&\propto \exp\Bigg(\frac{\alpha_\sigma}{2} \cdot \alpha^2 + \left(\alpha_\sigma \alpha_\mu + \sum_{i=1}^{M}\sum_{t=1}^{T} p_{i,t}\right) \cdot \alpha \; - \\
&\qquad\qquad e^{\alpha} \cdot \sum_{i=1}^{M}\sum_{t=1}^{T} \left(Mon_{i,t} \cdot Pop_{i,t} \cdot e^{Y_t+S_i+\tau\cdot TAX_{i,t}} \right) \Bigg) \; .
\end{aligned}
\tag{G.14}
$$

It is not possible to simplify the term any further, and to link it to a known distribution. However, it is straightforward to show that the distribution is log-concave,which makes it possible to use the method of adaptive-rejection sampling described in Gilks and Wild (1992). The full conditional distributions for S_i and Y_t can be derived similarly, so that

$$
\begin{aligned}
P(S_i \mid n_{-S_i}) &\propto \exp\Bigg(\frac{S_\sigma}{2} \cdot S_i^2 + \left(S_\sigma\; S_{\mu_i} + \sum_{i=1}^{M} p_{i,t}\right) \cdot S_i \; - \\
&\qquad\qquad e^{S_i} \cdot \sum_{i=1}^{M} \left(Mon_{i,t} \cdot Pop_{i,t} \cdot e^{\alpha+Y_t+\tau\cdot TAX_{i,t}} \right) \Bigg) \; ,
\end{aligned}
\tag{G.15}
$$

$$
\begin{aligned}
P(Y_t \mid n_{-Y_t}) &\propto \exp\Bigg(\frac{Y_\sigma}{2} \cdot Y_t^2 + \left(Y_\sigma\; Y_{\mu_t} + \sum_{t=1}^{T} p_{i,t}\right) \cdot Y_t \; - \\
&\qquad\qquad e^{Y_t} \cdot \sum_{t=1}^{T} \left(Mon_{i,t} \cdot Pop_{i,t} \cdot e^{\alpha+S_i+\tau\cdot TAX_{i,t}} \right) \Bigg) \; .
\end{aligned}
\tag{G.16}
$$

The sampling process from these three distributions does not take much time. However, sampling from the full conditional distribution for τ, which is proportional to

$$P(\tau \mid n_{-\tau}) \propto \exp\left(\frac{\tau_\sigma}{2} \cdot \tau^2 + (\tau_\sigma \ \tau_\mu + \sum_{i=1}^{M} \sum_{t=1}^{T} p_{i,t} \ TAX_{i,t}) \cdot \tau - \sum_{i=1}^{M} \sum_{t=1}^{T} (Mon_{i,t} \cdot Pop_{i,t} \cdot e^{\alpha + S_i + Y_t + \tau \cdot TAX_{i,t}}) \right) , \tag{G.17}$$

is rather time consuming, because τ is a part of the summation over all observations, which therefore needs to be recalculated whenever the value of τ changes. Given that the sampling routine requires about 50 evaluations of G.17 and its derivative with respect to τ for each sample, this expression accounts for 90 percent of the sampler's running time.