1 Hierarchical models of transformed quantiles

The microcredit data can be described simply with a few parameters: a Box-Cox (BC) parameter, a mean, and a standard deviation. However, these are not the parameters to which we want to apply a hierarchical model. For example, it is not clear at all what a common normal distribution on the Box-Cox parameter would mean, nor does it make sense to describe the a priori correlation between that and the pre-transform mean. What we really want is to put a group prior on observed quantities. Fortunately, if we are willing to restrict ourselves to putting a prior on a function of as many quantiles as there are free parameters in the transform, this can be done with a simple change of variables.

Let us simply model the control and treatment separately for the moment – we can always use simulation to infer any needed posterior quantities, such as quantile differences. Let's also assume for simplicity that the profit is all positive – we can either model the positive and negative profit separately, or we can choose from the zoo of BC extensions that allow negative numbers. The setup then is that we have profit data which is distributed normally after a BC transform:

$$\begin{aligned} y_{gn} &= \text{Profit of observation n in group g} \\ \gamma_{gn} &= \frac{y_{gn}^{\lambda_g} - 1}{\lambda_g} \\ \gamma_{gn} &\sim \mathcal{N}\left(\mu_g, \sigma_g^2\right) \end{aligned}$$

The $\alpha-$quantiles of γ_{gn}, which we'll call $\tilde{q}_g^\alpha$, have a closed form as a function of μ_g and σ_g^2:

$$\tilde{q}_g^\alpha = \sigma_g \Phi^{-1}(\alpha) + \mu_g$$

Because quantiles are transformed 1:1 under invertible transformations, this means that the quantiles of y_{gn}, q_g^α also have a closed form:

$$q_g^\alpha = \left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{1/\lambda_g} = \left(\lambda_g \left(\sigma_g \Phi^{-1}(\alpha) + \mu_g\right) + 1\right)^{1/\lambda_g}$$

Thus, we can put priors on the quantiles of q_g^α or smooth functions of it with a change of variables, since we can calculate the Jacobian of the mapping $(\mu_g, \sigma_g, \lambda_g) \rightarrow \left(q_g^{\alpha_1}, q_g^{\alpha_2}, q_g^{\alpha_3}\right)$ in closed form for three distinct values of α.

$$\begin{aligned} \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha} &= \frac{1}{\lambda_g}\left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{\frac{1}{\lambda_g}-1} \lambda_g = \left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{\frac{1}{\lambda_g}-1} \\ \frac{\partial q_g^\alpha}{\partial \mu_g} &= \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha}\frac{\partial \tilde{q}_g^\alpha}{\partial \mu_g} = \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha} \\ \frac{\partial q_g^\alpha}{\partial \sigma_g} &= \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha}\frac{\partial \tilde{q}_g^\alpha}{\partial \sigma_g} = \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha}\Phi^{-1}(\alpha) \\ \frac{\partial q_g^\alpha}{\partial \lambda_g} &= \frac{1}{\lambda_g}\left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{\frac{1}{\lambda_g}-1} \tilde{q}_g^\alpha = \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha}\tilde{q}_g^\alpha \lambda_g^{-1} \end{aligned}$$

So

$$\nabla q_g^\alpha = \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha} \left(\begin{array}{c} 1 \\ \Phi^{-1}(\alpha) \\ \tilde{q}_g^\alpha \lambda_g^{-1} \end{array} \right)$$

The second derivatives (which you don't need) are given by

$$\frac{\partial^2 q_g^\alpha}{\partial \left(\tilde{q}_g^\alpha\right)^2} = \left(\frac{1}{\lambda_g} - 1\right) \left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{\frac{1}{\lambda_g} - 2} \lambda_g = (1 - \lambda_g) \left(\lambda_g \tilde{q}_g^\alpha + 1\right)^{\frac{1}{\lambda_g} - 2}$$

$$\begin{aligned}
\nabla \nabla^T q_g^\alpha &= \frac{\partial^2 q_g^\alpha}{\partial \left(\tilde{q}_g^\alpha\right)^2} \left(\begin{array}{c} 1 \\ \Phi^{-1}(\alpha) \\ -\tilde{q}_g^\alpha \lambda_g^{-2} \end{array} \right) \frac{\partial \tilde{q}_g^\alpha}{\partial (\mu_g, \sigma_g, \lambda_g)} + \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha} \frac{\partial}{\partial (\mu_g, \sigma_g, \lambda_g)} \left(\begin{array}{c} 1 \\ \Phi^{-1}(\alpha) \\ \tilde{q}_g^\alpha \lambda_g^{-1} \end{array} \right) \\
&= \frac{\partial^2 q_g^\alpha}{\partial \left(\tilde{q}_g^\alpha\right)^2} \left(\begin{array}{c} 1 \\ \Phi^{-1}(\alpha) \\ \tilde{q}_g^\alpha \end{array} \right) \left(\begin{array}{ccc} 1 & \Phi^{-1}(\alpha) & 0 \end{array} \right) + \frac{\partial q_g^\alpha}{\partial \tilde{q}_g^\alpha} \left(\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & \Phi^{-1}(\alpha) & -\tilde{q}_g^\alpha \lambda_g^{-2} \end{array} \right)
\end{aligned}$$

Denote the Jacobian by

$$|J_g| = \left| \left[\begin{array}{ccc} \nabla q_g^{\alpha_1} & \nabla q_g^{\alpha_2} & \nabla q_g^{\alpha_3} \end{array} \right] \right|$$

A sensible thing to put an informative prior (shared across groups) on might be the ratio of the median to the inter-quartile range:

$$\eta_g := \frac{q_g^{0.5}}{q_g^{0.75} - q_g^{0.25}}$$

You could then put a non-shared, non-informative prior on the actual quartiles, $q_g^{0.75}$ and $q_g^{0.25}$. The model (in vector notation) would then be

$$\begin{aligned}
\log P(y_g, \mu_g, \sigma_g, \lambda_g) = &- \frac{1}{2\sigma_g} (\gamma_g - \mu_g)^T (\gamma_g - \mu_g) - \frac{1}{2} N \log \sigma_g^2 + \log |J_g| + \\
&- \frac{1}{2\sigma_\eta^2} (\eta_g - \eta)^2 - \frac{1}{2} \log \sigma_\eta^2 + \log P\left(q_g^{0.75}\right) + \log P\left(q_g^{0.25}\right) + \\
&\log \left| \frac{\partial}{\partial \left(q_g^{0.25}, q_g^{0.5}, q_g^{0.75}\right)} \left(\begin{array}{c} \eta_g \\ q_g^{0.25} \\ q_g^{0.75} \end{array} \right) \right|
\end{aligned}$$

(I often get the Jacobian the wrong way around, so someone should double check that this is the right way. Also, I often get the Jacobians themselves wrong, so someone should check that, too.)

2 Sufficient statistics for the normal model

For normal observations in groups where

$$
\begin{aligned}
y_n &\sim \mathcal{N}\left(x_n^T\beta, \sigma^2\right) \\
Y &\sim \mathcal{N}\left(X\beta, \sigma^2 I\right) \\
\sum_n \log p\left(y_n|\beta, \sigma^2\right) &= -\frac{1}{2\sigma^2}\left(Y - X\beta\right)^T\left(Y - X\beta\right) - \frac{n}{2}\log\sigma^2 \\
&= -\frac{1}{2\sigma^2}\left(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta\right) - \frac{n}{2}\log\sigma^2
\end{aligned}
$$

So to run the sampler all we need are the quantities Y^TY , Y^TX, and X^TX. These can be pre-computed as they are the same for every draw. For the spike we just need the counts:

$$
\begin{aligned}
\rho &= \alpha + \gamma T_i \\
p &= \frac{\exp(\rho)}{1 + \exp(\rho)} \\
P\left(n_{zero}\right) &\propto p^{n_{zero}}\left(1 - p\right)^{N - n_{zero}} \\
&= \left(\frac{p}{1-p}\right)^{n_{zero}}\left(1 - p\right)^N
\end{aligned}
$$